Research versus Rhetoric: An
Examination of Standardized Testing

Lois A. Eaton
laetn@bsu.edu
Ball State University

Dr. Donald Gilman

Graduating in December of 2004
I would like to thank Dr. Donald Gilman for advising me with this honors thesis. He played a key role in helping me develop my thoughts and present them in an organized manner in order to write a strong thesis.
Abstract

A push for standards is occurring across the nation and is creating great changes in United States schools. For example, standardized tests are used more than they ever were before to assess students’ eligibility for grade promotion, graduation from high school, acceptance into institutions of higher learning, and many other important decisions. With such emphasis being placed on standardized test scores alone for the basis of many important decisions, some educators and educational theorists want to know if the standardized tests are reliable, valid, and unbiased. Many politicians and test-makers maintain that the tests do achieve their main goal of providing an accurate measure of students’ academic performance. However, many parents, educators, and educational theorists believe that current standardized tests are too structurally flawed to accurately reflect students’ academic abilities. Alfie Kohn, Peter Sacks, and James Popham are three educational theorists who oppose the current uses of standardized test scores and highlight cultural bias in test items and the inappropriateness of standardized test objectives, among other problems. In my honors thesis, I will summarize the viewpoints of Kohn, Sacks, and Popham and emphasize the importance of involving parents and educators in the standardized testing debate.


<table>
<thead>
<tr>
<th>Chapter One: Introduction</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter Two: Theories of Alfie Kohn</td>
<td>7</td>
</tr>
<tr>
<td>Chapter Three: Theories of Peter Sacks</td>
<td>19</td>
</tr>
<tr>
<td>Chapter Four: Theories of James Popham</td>
<td>27</td>
</tr>
<tr>
<td>Chapter Five: Conclusion</td>
<td>43</td>
</tr>
<tr>
<td>References</td>
<td>45</td>
</tr>
</tbody>
</table>
Chapter One

We live in a hierarchical society. Living in the United States presupposes that everything and everyone will be graded, rated, and categorized. We place values on everything in our society even though many things were not meant to be quantified. On January 7, 2001, President George W. Bush presented a radio address to the nation that changed the face of American education. He announced that from that date on, each and every child would be assessed yearly using standardized tests. The premise behind this new education initiative was loud and clear: “We must go back to the fundamentals of early reading and regular testing, local control, and accountability for results, clear incentives for excellence, and clear consequences for failure” (Bush 2001). Bush’s vision sounds like an aggressive, straightforward approach to raising the educational standards of our youth, but the consequences for not meeting those standards are severe.

Later in the address, Bush asserts that we, as a nation, cannot and will not continue to give tax dollars to failing schools that do not improve. Once again, the measurement of success relies on rigid procedures that ignore the individual characteristics of each student. For example, Bush places an ultimatum on failing schools and imposes a strict deadline to the teachers of these schools to comply. It is surprising that the President of our nation only finds one factor to which he attributes the low achievement of some schools. According to Bush (2001), schools are failing not because children are growing up in poverty, not because housing discrimination forces minority families to live in substandard housing, not because schools are locally funded and poor communities give rise to poor schools, but because of the incompetence of the teachers, the very people he calls on later in the address to help him raise the bar for
education. The students’ success or failure is determined solely upon quantitative measurements. It is safe to say that the United States has many important problems such as inequality and discrimination based on race, gender, sexual orientation, and a number of other factors that impede the progress of a community and its school. To attribute our failing schools to the lack of good teaching, based upon students’ test-taking abilities on a single test, is insensitive and flawed according to some educational theorists.

The testing movement is occurring in full force in the United States, and has led many, from parents to school psychologists, to question the validity and consequences of battering our children with standardized tests. Children, by nature, come into this world with open hearts and open minds. As adults, we tend to envy the innocence and curiosity of children but quickly suppress these interests as they enter public schools. Even though we agree that part of what makes children so charming is their never-ending curiosity and tendency to ask questions to the point of exhaustion, we encourage them to conform when they enter public schools. We, as a nation, do a wonderful job of stifling children’s imaginations at a young age by bombarding them with instructions such as “Sit up straight. Stop asking so many questions. You’re being disruptive. Learn these ten vocabulary words, and be able to repeat them back to me in a week.” With the overuse and abuse of high-stakes standardized testing to assess children’s abilities, we have succeeded in curbing their internal motivations to achieve.

Standardized test scores have been used as the sole basis of many important decisions that carry with them harsh consequences. Faculty members at many schools have been fired because their students received low test scores on standardized tests.
Students have been retained instead of progressing to the next grade level because their scores on criterion-based tests have not met the standard passing score by one point (Kohn 2000). Entire schools have even been closed because its students had not performed well on standardized tests. It is the opinion of many experienced educators and school psychologists that a single test score should not be the most important factor in deciding whether a student, teacher, or school is passing or failing. However, in recent years, standardized test scores have become an extremely popular way for the public and policymakers to label a school as being either a “good school” or a “bad school.” On a day that some believe started out as a slow news day, one newspaper journalist decided to rate a few local schools according to what each received on a recent standardized test. From one newspaper article arose a powerful movement of publicizing test scores that allowed the readers to feel as if they could fully grasp what one school had to offer as opposed to another based on the scores. Parents felt as if they could now wisely choose the “right” school for their children based on standardized test scores (Popham 2001).

Now schools’ test scores are published regularly in newspapers that quantify the success of the school. People who read these “report cards” on different schools in a certain area can clearly see which schools are “successful” and which schools are “failing.” However, a single test score can never capture the complexities of the inner-workings of a school. The published test score has not actually measured the effectiveness of the teachers’ instruction alone, but has also measured the socioeconomic status of the student, the particular mood of the student that day, the innate cognitive capacities of the student, and many other factors that should not be used to punish and reward a school.
Theorists of educational psychology as well as a number of educators and parents are becoming more and more concerned about the use of test scores for high-stakes decisions. With more standardized testing being used in schools at all levels and more high-stakes being placed on those test scores, the negative effects of high-stakes testing have become more apparent to parents and school faculty. For example, many parents around test-taking time are becoming more concerned about the levels of stress their children experience that are related to the fears these students have regarding the high stakes placed on standardized test scores (Popham 2001). However, educators and parents feel that they cannot change the use of high-stakes testing in their states. Some feel that the mechanics of standardized tests are too complicated for them to understand. Others naively believe that the individuals who create the standardized tests are experienced former teachers or others who have had first-hand familiarity with a primary or secondary school setting (Kohn 2000). Alfie Kohn, Peter Sacks, and James Popham are all educational theorists who oppose high-stakes testing and are trying to eliminate the mystery that surrounds the standardized testing movement. They encourage parents and educators to speak out against the use of a single test score on a standardized test as the sole basis for high-stakes decisions.
Alfie Kohn is an activist and theorist on education who believes that high-stakes standardized testing undermines children’s natural drives to learn and instead teaches them how to perform for extrinsic rewards. In particular, he maintains that standardized testing produces standardized minds (Kohn 1992). For instance, a child who reaches school-age full of curiosity will quickly learn that curiosity and questions are an annoyance to teachers, and that rewards can be obtained through obedience and rote memorization of facts. Likewise, a teacher who enters the profession ready to improve the education and higher-order thinking skills of children is likely to feel pressure from superiors to “teach to test” or to teach only what is necessary to pass a standardized test. According to Kohn (2000), teachers are being turned into test-prep technicians, and standardized testing is actually hurting children’s education because it is taking away from actual, in-depth learning that requires original thought.

According to Kohn, the overuses and misuses of standardized tests to judge a school’s effectiveness originated from the cry of “back to basics” that many policymakers have used to create the image that they were “cracking down” on public schools, and that they were bringing every student and school back up to a strict standard of education. One useful tool that policymakers have used to gain support for their standardized testing campaigns involves giving the impression that public schools have been moving toward more innovative and frivolous ways of teaching instead of giving students the basic facts they need to know to survive and to be successful (Kohn 1992). In order for the standardized testing movement to gain support, policymakers claimed that our nation’s
schools have turned into extended playtime instead of imparting the most important knowledge to the students in a more traditional manner. Kohn (1992) states that these policymakers want to be viewed as "a brave minority under siege, fighting an uphill battle for old-fashioned methods that have been driven out of the schools by an educational establishment united in its desire for radical change" (p. 5).

From the politicians' perspective, public schooling has succumbed to newfangled imitations of teaching in which the student never really learns the basics such as math, English, or science. However, Kohn (1992) and most educators would agree that schools have never really left the basics, and that classroom thinking retains a traditional, teacher-centered approach. Teacher-centered classrooms usually involve the students obediently waiting until the teacher asks them straightforward questions, which are inserted throughout a lengthy lecture. Educator John Goodlad visited more than a thousand classrooms in 1979 and found that such traditional classroom thinking dominated United States schools. In only a few of the classrooms did teachers display novel approaches and strategies. In Goodlad's assessment, he found that the majority of the classrooms were "almost entirely teacher dominated with respect to seating, grouping, content, materials, use of space, time utilization, and learning activities" (Kohn, 1992, p. 7).

Kohn remarks that very little has changed since 1979. John Dewey, one of the most influential educational theorists of the twentieth century, commented with great regret and sorrow shortly before his death on the seemingly minimal influence of his democratic theories of education, which at one time shook the foundations of American schools. He stated that the schools that had actually adopted segments of his innovative
educational reforms used them in a way that was “atmospheric” (Kohn, 1992, p. 7). According to Dewey, the indestructible authoritarianism of education had remained intact. Many other educational theorists have found that our schools have never really deviated from the more traditional type of teaching. Therefore, the politicians’ cries of faddish teaching techniques ruining our schools may be nothing more than a political strategy designed to create an image of deep concern for our nation’s future.

Kohn asserts that traditional teaching techniques dominate United States classrooms, but are especially present in urban schools. According to Kohn, “urban schools” is a thinly masked phrase to signify children of color belonging to families with lower socioeconomic statuses. One writer commented, “The ghetto has been a hotbed for the basics” (Kohn, 1992, p. 9). According to Kohn, minority children are the most likely students to be taught low-level curriculum and to be taking multiple-choice standardized tests. For example, “Success for All” is a national drill-and-skill program that emphasizes ninety uninterrupted minutes of reading instruction and assessment every eight weeks to emphasize individual accountability. Students are grouped based on their current reading levels and assigned to read at home for a designated time each night. Parents are required to sign a form stating that they witnessed their child reading for the required amount of time. The most surprising aspect of this program is that schools are encouraged to adopt it for their kindergartners, who would attend school the entire day (Kohn 1992).

“Success for All” seems like an aggressive approach that has the potential to assure that children from all backgrounds will meet the expectations of a standard
performance level. However, the program mostly provides mind-numbing instruction that alienates the students at an early age and bores them to a point where they do not enjoy school. A journalist from the *Atlantic Monthly* and a supporter of "Success for All" simultaneously and most likely unknowingly provided support for the discontinuation of the program when he stated that it is used "almost exclusively in poor schools" (Kohn, 1992, p. 10), and that is may be "punitive to local school boards, principals, and teachers—but they had it coming" (p. 10).

The mayor of Jersey City, New Jersey, also advocated the use of "constant drill and repetition" (Kohn, 1992, p. 10) with minority children and stated, "It’s not that hard to give answers if someone just told you. They memorize back and know and get used to a lot of A’s on quizzes" (p. 10). According to this politician, "Those schools are best for certain children" (p. 10). Those certain children to whom the mayor is referring happen to come from families of color with a lower socioeconomic status. The only conclusion we can draw about the status of minorities in education is that they do not benefit from the more traditional type of classroom in the same way of children belonging to the majority.

The claims of innovative teaching styles running rampant in our education system, then, comprise a key element in the rise of standardized testing. Kohn explains that the standardized testing phenomenon becomes even clearer when we look at those who support raising standards in our schools and compare them to those who oppose the movement. The demand for tougher standards in schools is perhaps most prevalent among businessmen and politicians who have fixed ideas on the aims and role of a school
Research versus Rhetoric 11

(Kohn 1992). These individuals tend to believe that schools can single-handedly create a world-class market of intelligence, viewing children as their most valuable commodities. However, schools can only do so much for children, especially when a child’s community and nation are not quite as supportive as would be ideal.

Kohn (1992) explains that raising standards is a popular idea among Democrats and Republicans alike, and that it enjoys the support of virtually all corporations, the President, most of the governors, and almost all education officials. With support such as this, it is no surprise that standardized testing has become so widespread in the United States. On the other end of the spectrum are the extreme right-wing individuals who remain suspicious and rebellious toward any type of federal government interference (Kohn 1992). What is missing from these two extremes is the viewpoint of the students themselves who actually take the tests. Test-taking, then, goes beyond the educational and extends clearly into the political and commercial worlds.

As soon as we realize that some of the biggest supporters of standardized tests are business leaders, it becomes clear that they might not always share the same goals as educators and parents. The main goal of any type of corporation is to produce a profit and to provide financial return for its owners. It is easy to see how this might not interact with and support the goals of our schools. Kohn (1992) believes that any overlaps that occur between the aims of corporations and the objectives of education are purely coincidental, and that these overlaps tend to be minimal. In many cases schools are discussed and managed like businesses. Much of the jargon that used to be reserved for the corporate world is seeping into American schools. Moreover, presidential
commissions on education are usually headed by executives of large companies. When James B. Hunt, Jr. was the governor of North Carolina, he stated, "We must tell the business community that, if it wants better employees and higher profits, it must be involved in what the schools teach and how they teach it" (Kohn, 1992, p. 15). It is tempting for businesses to become more involved in the schooling of their future employees. However, what is good for business is rarely good for students.

The important question we need to ask ourselves is "Are our schools really failing?" (Kohn, 1992, p. 17). Kohn believes they are not. Many researchers would like us to believe we are not meeting the educational level of schools in other countries. Studies are released almost every week that supposedly prove the inadequacy of United States students' knowledge compared to previous years and other schools. However, many of these studies contain grossly exaggerated and misleading results, all of which are based upon test measurement and quantitative assessment. Specific details such as which schools were included in the studies, on which exams the students were being tested, and on what subjects are crucial elements that contribute to the understanding of the results (Kohn 1992). These same details are often omitted from studies in which the aim is to shock and to produce feelings in the reader that the United States is failing its students, or rather, that our students are failing the United States.

In fact, results released by the National Assessment of Educational Progress (NAEP) do not show the extreme drops in students' abilities that other studies do. The NAEP was created by the federal government to assess the abilities of children at the ages of nine, thirteen, and seventeen. Recent studies from the NAEP show that little has
Research versus Rhetoric 13

changed over the last few decades (Kohn 1992). However, the same politicians who claimed our students are not as educationally prepared as they used to be now state that it is not good enough that students have not changed all that much in the past years. Never mind that their previous claims had not been based on fact.

Another way researchers can mislead the readers of their studies is by publishing only those studies that yielded results supporting their claims of failing schools and students. Many other studies that do not show much change in the students’ abilities are not published, because they are less exciting to the public (Kohn 1992). Whether or not our students compete successfully with students from other countries is as debatable as standardized testing itself. However, what is very clear is that we are more focused on comparing ourselves to others than ensuring that each student in the United States receives quality education. Of course, it is fitting that a nation so focused on test scores and other numerical measures would want to see how we compare to other nations. Kohn (1992) and many educators stress the fact that we may be missing the point of education. In other words, we should not be so focused on comparisons with other schools and other children and should instead take care of our students’ needs.

Now that we know some background information about why we use standardized tests in the way that we do and to the extent that we do, we can examine the tests themselves. Kohn (1992) states that there is something to be said about the schools most of us attended, because so many of us forget to evaluate the basic premise of standardized test scores: standardized test scores should be valid indicators of educational quality. Until we judge a test’s appropriateness for its aims, objectives, and designs, we cannot
begin to debate on the success of student performance. Even the individuals who support
the current uses of standardized tests do not have much knowledge about the mechanics
of any of the tests. Currently, most standardized tests do not assess students’ deep-seated
knowledge about a subject. They simply illustrate whether or not a student has
memorized a certain number of words or facts (Kohn 2000). A test fails to become a test
when it does not sample a student’s knowledge, but instead represents the full extent of
the students’ knowledge.

Frequently, students are not able to provide a correct answer for a particular test
item because the test item was worded in a way that was different from the way the
teacher had worded it in a drill session before the test. One example of this phenomenon
occurred when John Dewey visited a classroom in order to observe classroom instruction
and its results. With the teacher’s permission, he took over a geography lecture. Dewey
reportedly asked the students what they would find if they dug a hole in the earth. As
faces stared up at him in confusion, the teacher suggested that he may not have asked the
question in the right way. She asked the class “What is the state of the center of the
earth?” (Kohn, 1992, p. 20). Much to Dewey’s amazement, the class stated in unison,
“igneous fusion” (p. 20). This story illustrates the inflexibility of students’ abilities when
they are given a bit of information in one particular way in a drill session during test
preparation. Most of the time, students are not able to generalize what they know to other
situations. As Emily Dickinson once stated, these students may possess “the facts, but
not the phosphorescence of thought” (Kohn, 1992, p. 20).
Besides the rather facile arguments for the efficient administration of standardized tests, these instruments persist because they are relatively inexpensive (Kohn 1992). Individuals who believe that they have persisted because the tests are effective only need to remember that we are a country that still believes spinach contains more iron than any other vegetable. Such rote learning proved erroneous and embarrassing when it was discovered that this finding occurred because of an erroneous shift of a decimal point. Like most of our “correctional” facilities and welfare programs, standardized testing continues to be widespread despite its effectiveness.

Corporations make enormous amounts of money for preparing and grading standardized tests and even make money from the teacher’s guides that they sell to help teachers raise their own students’ test scores. Kohn (1992) believes that a general rule of thumb about standardized testing could be that the more efficient the test is, the less effective it tends to be. Since schools are provided with a quick and easy way of assessing the abilities of their students, they would rather not adopt more time-consuming, costly methods of assessment even if they are more valid. Linda McNeil, an educational theorist and author, has even stated that “measurable outcomes may be the least significant results of learning” (Kohn, 1992, p. 75). Moreover, the more negative consequences we attach to low standardized test scores and the more rewards we attach to high standardized test scores, the less valid the scores become. When test anxiety has become a subfield of educational psychology, we know that we have gone too far.

Other than the problems with the test itself, we need to remember that there are actual breathing humans taking it. One friend of Kohn’s remembers filling in the test
bubbles on a standardized test to form the shape of a Christmas tree. This man’s boredom as a child landed him in a remedial class as a result even though he had no problems performing in class and was regarded by his teachers as a bright child (Kohn 1992). Many students who are forced to take standardized tests simply do not care about the tests for good reason. Others make themselves sick agonizing over the test’s consequences. Either way, the scores these students receive do not adequately reflect their abilities and achievements. However, we still hold celebrations when test scores improve a few points from the last year, and we still punish ourselves when we do not receive high test scores.

Other factors that contribute to a student’s test score is the socioeconomic status of the student’s family. One educator stated that we could save a lot of time and money by asking, “How much money does your mom make?” (Kohn, 1992, p. 77). While this statement may be an oversimplification, it is naïve to believe that factors such as mood and socioeconomic status do not play an important role in standardized test-taking. As we will see in further discussion of the cultural bias that exists in standardized test-taking, children belonging to minority groups receive much of the harmful consequences of unfair testing. For instance, many educators believe that individual test items can be culturally biased. Another area of test-taking that can provide cultural bias is the language in which the test is administered (Kohn 2000). These examples of cultural bias simply cannot be ignored and provide some of the most compelling evidence for striving for a change in how high-stakes tests are administered and relied upon today.
Kohn (1992) explains that the worst kind of test is the norm-referenced test, which compares one student to another and provides no information about the abilities of the student. Most states used norm-referenced tests as opposed to criterion-referenced tests that measure a student against a particular standard. With norm-referenced tests, 10% of students will score in the top 10% and 50% will score below the median. Even if the students’ scores are fairly similar, the results still show the same distribution no matter the pettiness of the differences. Students who have taken these tests will never know what percentage of questions they answered correctly or incorrectly, information that would probably have been a more useful measurement tool than the assessment of the percentage of students that scored higher or lower than them. Kohn (1992) states that using norm-referenced tests to assess how much the student learned is “like measuring temperature with a tablespoon” (p. 78).

Finally, Kohn points to the impact standardized tests have on children and young adults. In a culture that is already so competitive, standardized tests, especially norm-referenced tests, encourage children to be ruthless in their fights to score higher than others. Many children who will grow up with the heavy use of standardized tests we have today will most likely confuse real intelligence with trivia. Children today will also grow up in a world in which finding the “right” answer is only a matter of time (Kohn 1992). Originality does not exist nor is it wanted. Standardized testing will most likely promote a black-or-white viewpoint in children in which there is always a right answer as if it were God-given. Likewise, anything that is not right is wrong. However, in the real world, there are multiple truths. This type of thinking will prove to be maladaptive and
will actually inhibit future generations all in the name of tougher standards. Therefore, raising standards through high-stakes testing will eventually do the opposite of its aims.

Kohn’s ideas on education have permeated the minds of many parents and educators. He has led many people to question the validity and consequences of high-stakes testing in spite of the movement’s relentless popularity and support. Previously, parents and educators viewed the tests as nearly perfect assessment tools that should not be questioned. They did not possess even a basic understanding of the mechanics of standardized tests and believed that important decisions regarding the usage of standardized test scores should be decided by local, state, and federal governments. Kohn’s theories on high-stakes testing have been gaining support, however, and have been providing parents and educators with the tools and vocabulary they need to become involved in the standardized testing debate. Peter Sacks is another educational theorist who builds on Kohn’s ideas and who finds other hazards of the testing movement.
Chapter Three

Sacks (1999) believes that the standardized testing movement persists in the United States because Americans are obsessed with measuring mental capacities, more so than in other countries. In many other industrial countries, standardized tests used for admission into colleges and universities require the prospective students to write essays and to demonstrate their classroom knowledge. Americans are much more concerned with IQ than are people living in the rest of the world. What is even more damaging than our obsession with IQ is the narrow definition of intelligence in terms of logical-analytical abilities. Still, one of the biggest mistakes Americans make is a belief in innate intelligence, and that this inborn trait can be represented as a single numerical score (Sacks). Proof of our unhealthy affinities for mental measurement can be seen when we browse the bookshelves at any bookstore. Standardized test-takers can choose from a multitude of testing guides such as *Cracking the SAT 3000* and *Up Your Score: The Underground Guide to the SAT*. The standardized testing movement has become an unstoppable force in the United States and the focus of the accountability bandwagon.

Sacks (1999) observes that Americans place a higher value on the potential to achieve than on actual achievement. It is apparent that we are more focused on what students can do and not what they have done in their educational careers. Other countries tend to emphasize educational achievement rather than potential; as a result, their assessments tend to be more accurate. It is presumptuous to believe that we can predict the future of a child on the basis of one test score. Many children are underestimated when they are tested at an early age. This may cause parents and everyone else to lower
their expectations of children based on low test scores (Sacks 1999). A self-fulfilling prophesy then ensues, because we expect less from these children and consequently provide them with less educational opportunities. Therefore, the same individuals who claimed these children would not be successful in almost all aspects of life can bask in the glory of being right all along.

Sacks (1999) addresses three widely-held myths in education. The first myth claims that schools in the United States are in serious trouble and are failing our children. Like Kohn, Sacks states that we receive a skewed representation of education when we remain focused on the fervor and excitement of novel findings about education instead of on overall historical progress. Sacks calls the accountability movement a “mind-boggling social experiment on the nation’s schools” (p. 82). Before the National Commission on Excellence in Education published A Nation at Risk in 1983, our nation’s schools were supposedly failing. Students were not learning enough and were not learning the right subjects; schools were operating with uneven and inadequate standards. Now that the secret of our failing schools has been revealed and demonstrated, we should be seeing significant progress (Sacks 1999). With new accountability measures for all schools, the level of education in the United States should have increased in a meaningful way. However, this is not the case.

In March of 1979, 85.6% of twenty-five to twenty-nine year olds had completed high school. By the mid-1990s, 86.9% of this group had completed high school (Sacks 1999). This increase is hardly statistically significant. If our nation’s schools are actually failing, we would expect that holding schools accountable and raising standards would
have real effects on the students' achievements. Another example of less than thrilling educational progress comes from international comparisons. In 1985, in both the United States and Japan, 23.2% of adults had graduated from college. *A Nation at Risk* had described Japan as a major threat to American achievement. The former West Germany was also identified as being an educational threat to the United States even though the college graduation rate in West Germany was only 13.5% in 1985 (Sacks 1999). Still, some may conclude that our education system has been in peril for decades and that we have not seen significant improvement after *A Nation at Risk* because we are just as bad now as we ever have been. Sacks replies that “a belief in crisis and inferiority has been bred into American schools for nearly 200 years” (p. 84).

The second myth we have about American education is that the United States economy is in trouble because of our inferior education system (Sacks 1999). It is misleading and naïve to believe that the quality of the United States, or any economy, is solely the direct result of educational reform. However, many people blame schools when the economy is poor as if it were the sole reason for economic downfalls. Sacks found that the same people who claim United States schools are failing also tend to exaggerate about the failing economy. A 1998 analysis of world competitiveness by the International Institute for Management Development showed that the United States was the most internationally competitive country, and stated that the United States had always been “historically exceptional” (Sacks, p. 86).

Sacks also found that schools were not praised when the economy improved. Stanford education professor Larry Cuban explains this by stating that “the myth of better
schools as the engine for a leaner, stronger economy was a scam from the very beginning" (Sacks, p. 87), and that "stimulating economic growth depended far more on fiscal and monetary policies than turning around schools" (p. 87). Some economists still argue that lags in productivity over the years result from fewer skilled workers who graduate from high school. However, the same lags in productivity occurred throughout Europe and Japan (Sacks 1999). It is difficult, if not impossible, to explain productivity lags by pointing to failing schools since it did not occur solely in the United States.

The third myth Sacks (1999) addresses concerns the relationship between greater school accountability and higher achievement. Currently, standardized tests scores are heavily considered in several important decisions such as promotion to the next grade level, granting high school diplomas, rewarding student performance, assessing school and teacher effectiveness, issuing "skills guarantees" for high school students, and accrediting schools. Sacks gathered data from the National Assessment of Education Progress (NAEP) and found that a state with a high-stakes testing program is more likely to display below-average achievement levels than a state with low or no stakes. States that have low-stakes testing programs are more likely to reach national achievement levels than states with high-stakes testing programs. The most striking finding, though, is that a state without any testing programs has the best chance of any other state of reaching above-average achievement levels. The National Center for Fair and Open Testing duplicated Sacks' results and noted that "proponents have based their rationale for high-stakes testing on ideology, not evidence" (p. 93). Of course, the results of these two studies have to be measured carefully for fear of making the same mistakes as others when poor education is blamed for economic downfalls. Many other factors contribute to
students' performance on standardized tests in any given state. However, there is more evidence to support a negative correlation between high-stakes testing and educational achievement than there is to support a causal relationship between failing schools and economic decline.

Sacks (1999) asserts that perhaps the most devastating outcome of the standardized testing movement is the increase in high-stakes testing for very young children. Sacks found that Manhattan was the epicenter of the standardized testing movement for young children in 1999. Many preschool and elementary schools in Manhattan require that prospective students take rigorous standardized tests in order to be welcomed into the schools. Not all of these schools are private (Sacks 1999). The standardized testing movement in the United States has caused many people to revert back to the ideology of children as being "little adults." One Manhattan school is called Hunter College Elementary, reflecting the perspective of children as little adults. The two intelligence tests that were required in 1999 in Manhattan private and some public preschools were The Wechsler Preschool and Primary Scale of Intelligence and a version of the Stanford-Binet Intelligence Test used at Hunter College Elementary (Sacks 1999). Both of these tests are used today, and they had been used for about 20 years when Sacks studied them. Sacks refers to them as the "Baby Boards" (p. 35). He states that while these tests are less wide-spread than standardized tests for older children, they are equally damaging (Sacks 1999). They may even be more damaging because their negative effects are far more subtle than other tests, and parents may be less inclined to question and protest. These tests are used as screening devices to differentiate between the gifted
and the mediocre. Such a view is alarming when we remember that we are talking about children and not some kind of well-oiled test-taking machines.

Most of these Baby Boards are extremely developmentally inappropriate for very young children. For instance, the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R), also known as the “Wippsie-R,” is a standardized intelligence test consisting of ten subtests which require over an hour to administer. Many toddlers are not able to sit long enough to complete the test the first time and have to be tested again. The Wippsie-R includes five verbal tests and five performance tests. Toddlers are expected to display amazing verbal skills and knowledge of information to receive an adequate score for many private preschools. Some of the concepts included in the Wippsie-R measure toddlers’ abilities to find similarities in words and concepts, assessments of hypothetical situations, and learning a few dozen words (Sacks 1999). It is notable that most developmental psychologists believe that people do not possess the abilities to think hypothetically until adolescence. However, the developmental inappropriateness of the Wippsie-R has not stopped private and public preschools from relying on its scores for admission.

Many of the test items on the Wippsie-R measure the speed in which the child completed the task rather than the correctness of response. For instance, one item requires toddlers to reproduce patterns with flat blocks from models and then from pictures. High performance on this item is mostly based on the time of completion instead of accuracy (Sacks 1999). However, developmental psychologists mostly agree that children, especially young children, learn at their own paces. They state that a child
who can complete a task faster than another child is not necessarily smarter. Time-based test items are especially damaging for toddlers who tend toward perfectionism and who do not leave a task unless it is completed to their high standards. Most Manhattan preschool faculty who use standardized intelligence tests for admissions make the official statement that these test scores are just one way to assess prospective students. They claim that they do not emphasize test scores over other factors in considering an applicant (Sacks 1999). Nonetheless, one school official at Grace Church School stated:

The test score is not for the children we take, it’s for the children we turn down. I don’t want to feel that so much is based on my reaction to that child. I wanted something else to substantiate what I saw. Maybe the chemistry wasn’t right. For the children we have to turn down, I look to have this impartial, objective analysis (Sacks, p. 38).

This school official put a new perspective on early intelligence testing for admission. Even though she attempted to justify testing by saying that it provides an objective measure of assessment, she also confirmed the belief that school officials may claim a child’s test scores prevented admission although the actual reason was a lack of a correct “chemistry” with that child. It enables officials a way to reject a child that may have come from a family of lower socioeconomic status or of a non-Caucasian race without being called prejudiced. Officials can simply present the child’s test scores and claim that they were not sufficient for admission into their school.

Kohn and Sacks believe that the current uses of standardized tests in the United States have become inappropriate to say the least. They have found that much of the
rhetoric surrounding standardized testing has not been supported by research. Both theorists have conducted their own research studies and have discovered that the costs of standardized testing far outweigh the benefits. However, the question remains about what we can do to change the way high-stakes testing is used in American schools.
Chapter Four

James Popham is yet another educational theorist who, like Kohn and Sacks, has become disenchanted with the design and uses of standardized tests. He approaches the matter in an extremely straightforward, pragmatic manner. The major strength of his argument against the current standardized tests lies in his systematic critique of the actual tests as opposed to falling into the debate of whether standardized testing is right or wrong. Whether standardized testing is right or wrong can be debated endlessly until we arrive at the same place we started. However, it is very difficult to argue with the mechanics of the tests. Popham systematically submits standardized tests to an item-by-item analysis and proves to be a refreshing voice of reason in the midst of so many moral arguments. In this way, Popham is an invaluable asset to the world of education and more specifically to the debate of standardized testing. His theories about educational assessment coincide nicely with those of Kohn and Sacks while providing educators with an alternative to standardized testing as it is used today. According to Popham, standardized tests should not be discounted. However, he does maintain that test scores should not be relied upon as much as they are today for so many important decisions since many of the tests are not valid measures of students’ abilities. Used in the right way, Popham (2001) believes standardized tests can provide an important assessment tool for parents and educators. As we will see, Popham suggests some of his own ideas for creating better assessment tools.

Popham (2001) explains that the major problem with standardized tests lies with the public’s opinions of them. Many educators mistakenly lump standardized tests with
the Internal Revenue Service and lie detector tests because they believe such things are beyond their comprehension. Standardized tests are viewed as some mystical force that an average person could not possibly understand. Popham believes this explains why standardized testing has prevailed as long as it has. With terms like psychometrician, it is no surprise that most people are afraid to object to standardized testing for fear of looking stupid in front of such learned individuals, whom Popham lovingly calls "measurement people" (p. 45) or "test folks" (p. 45). However, the mechanics of standardized tests are not all that complicated. Once we learn about the origins of the standardized testing movement, we can easily understand the goals and methods used in current standardized tests.

Standardized tests were first introduced into the United States during World War I, when an army of immense proportions was necessary to serve as a valuable force. In order to assess which individuals were intelligent enough to serve as officers, the U.S. Army needed a way to test a number of recruits at one time in an efficient manner. Alfred Binet's individually administered intelligence tests would not be practical because there were so many men to be tested. Therefore, Robert M. Yerkes, the president of the American Psychological Association, was called upon to design a test that could provide efficient assessment. Yerkes and eight testing specialists created ten different subtests which became known as the "Army Alpha" (Popham, 2001, p. 41). More than 1,700,000 took the tests during World War I, and the Alpha was able to compare each man to a norm group of men who had previously taken the tests.
Even though the tests represented the first use of multiple-choice test items on such a large scale, each item was analyzed to ensure that the test would be valid. This first standardized test was an aptitude test as opposed to an achievement test. It tested the men’s innate abilities and mental capacities as opposed to what they were taught by an instructor. The consensus on the Alpha was that it succeeded in separating men into categories according to their mental abilities (Popham 2001).

Soon after the war, people began to submit requests to the United States Copyright Office to copyright educational tests. All of these new tests shared quite a resemblance with the Army Alpha tests used in World War I even though some were designed to be aptitude tests and others were meant to be achievement tests. Most standardized tests today aim to reproduce the strength of the Army Alpha: the ability to allow fine-grained comparisons among the test-takers and to do so accurately. It is this element ingrained in standardized tests today that renders them useless when it comes to making instructional decisions based on test scores (Popham 2001).

The history of standardized tests illuminates the debate about whether we should use them as an indicator of a school’s effectiveness. Knowing the original aims and uses of these tests, we can now take a look at the elements of current standardized tests that mislead parents and educators. Popham (2001) cites three shortcomings of current standardized tests that should call for a reassessment of their value.

The first shortcoming Popham (2001) recognizes is the mismatch between what is taught in the schools and what is tested. Standardized testing incorporates a “one-size-fits-all” approach in which students from schools in Kentucky are tested over the same
material as students from schools in New York. This proves to be a major downfall of
testing. The curriculum of schools in one part of the country is rarely identical to the
curriculum taught in another. Different regions of the United States emphasize different
parts of the curricula. Popham found that at least half and sometimes even more of what
is tested in a particular state or district is not in the curriculum for that area. This is more
than a modest difference. One study conducted by D.H. Freeman and his colleagues in
the early 1980s found that with every standardized test used in elementary schools around
the country, at least half of the content was not stressed in textbooks. In some cases, 80%
of the material was not addressed in any of the textbooks provided to the students
(Popham 2001).

Other times, the test material is addressed in the classroom but in a completely
different way that it is presented on the standardized test (Popham 2001). For instance,
some standardized tests use partially completed graphic organizers such as charts or
dgraphs to evaluate the students' reading comprehension. The students are expected to fill
in the blanks on the diagram. For many students, this type of design to assess reading
comprehension is completely foreign. Other tests use cloze methods in which the student
has to fill in missing words in paragraphs from which every fifth word was left out.
These mismatches between instruction and standardized tests are extremely important
when interpreting test scores. However, test items are hardly ever subjected to an item-
by-item analysis (Popham 2001). According to Popham, this procedure never happens,
and it results from educators' awe-struck attitudes toward standardized tests. To them, it
is useless to scrutinize each test item individually because they will never understand
how the test works anyway.
The second shortcoming of standardized tests can be attributed to their origins. Most standardized tests today work to achieve the same success that was possible with the Army Alpha, which was to allow fine-grained comparisons among test-takers (Popham 2001). This is achieved through a mechanism know as score-spread. A large score-spread, or test scores that show wide variation, is created most effectively when about half of the test-takers have answered a particular item correctly. A p-value is assigned to each item that indicates how many answered the item correctly. If half of the test-takers answered a particular item correctly, the p-value would be .50. Items with p-values ranging from .40 to .60 contribute the most to a test’s score-spread. Items with p-values of more than .60 are deemed too easy and items with p-values of less than .40 are deemed to be too difficult. An item with a p-value of 1.00 is an item that all students answered correctly and that does not contribute to score-spread at all.

Therefore, if the p-values of a particular item increase to more than .60, that item will be removed from the test the next year it is revised (Popham 2001). When this methodology is discussed in technical terms such as p-values, there seems to be nothing unjust about the process of test-making. However, we must not forget what this actually means for real students and real teachers. As one examines the consequences in thinking and learning, the teacher who explains a particular concept effectively is penalized, for the test item testing that concept will be removed from the test in the following years. Therefore, teachers are fighting a losing battle. Since most teachers believe in stressing the most important content in their classrooms, test items covering the most important content will be dropped from the test and only trivial test items will remain (Popham 2001). It is impossible for teachers to master these standardized tests because the content
is constantly changing, and anything that is taught well will not be reflected in the test in the following years.

As we have seen, mismatches between what is tested and what is taught and the never-ending crusade for score-spread have left teachers baffled. However, standardized tests carry with them another short-coming that may be the most detrimental. The third problem Popham (2001) found with standardized tests pertains to the issue of confounded causality. Communities have been known to celebrate increases in test scores from one year to the next and to punish themselves when test scores have decreased. Yet comparing one class of students to another is like comparing apples and oranges. Most teachers would agree that in some years, the students simply seem as if they do not grasp the material as well as students have from previous years.

Cohort effects can also cause great differences in the educational abilities of particular generations. It is a gross oversimplification to assume that improvements and setbacks in test scores must result from instructional changes. It is highly likely that many teachers hardly ever alter the ways that they teach. Any changes in test scores from year to year would then have to be due to other factors. Some policymakers have aimed to solve this dilemma by creating tests that assess the same group of students at the end of each school year. However, it is impossible to write tests that are equally difficult for each grade level (Popham 2001). If tests were sufficiently precise and valid for this feat, making comparisons from one year to the next or from one student to the next may be plausible. Nevertheless, our current standardized tests are not precise enough to adopt this challenge.
Another factor that is commonly confused with educational achievement is the socioeconomic status of the student. Since test items are not usually analyzed to judge their fairness or validity, it is not surprising that they are not analyzed for cultural bias. Many test questions are extremely easy to master for children who have been raised by families of higher socioeconomic status and whose homes contain certain luxuries such as televisions and books (Popham 2001). Families who are fairly wealthy are able to take trips to Disneyland and other vacation spots that enrich children’s experiences. On the other hand, children who come from families without a lot of money may have never been exposed to certain luxuries such as cable television, theme parks, and the internet. Socioeconomic differences are well-represented in test items on standardized tests because these items usually contribute to score-spread (Popham 2001). Some students answer the item correctly; some do not. It is not difficult to see how the goals of most standardized tests do not coincide with the education of students of a lower socioeconomic status.

Some items such as this one are difficult for children who come from homes in which only the basic goods needed to survive are present:

4.4 A 6th Grade Science Item

If you wanted to find out if another planet had mountains or rivers, which of these tools should you use?

A. a pair of binoculars
B. a telescope
C. a microscope
D. a camera
It is obvious that not every child will have had exposure to a telescope or even to any of these items. Moreover, this type of item is addressing certain inventions that are not usually discussed in sixth grade textbooks (Popham 2001).

Other test items that assess the student’s knowledge of certain vocabulary words are riddled with socioeconomic bias. For instance, one popular test item asks the student to choose from multiple definitions of the word “field.” The sentence in which the definition is supposed to be derived is “My father’s field is computer graphics.” The problem with this item is that not every father’s career could be described as a field. Some children’s fathers’ and mother’s places of employment are simply referred to as “jobs” by the public. The use of the word field to describe a career may be foreign to many children whose parents have low-paying, disreputable jobs. Popham has dedicated much of his time to scrutinizing single test items in popular standardized tests, as the individual items often explain high or low results for the respective student. He found that up to 65% of the test items were linked to socioeconomic status. The highest percentage of socioeconomic-related test items were found in the language arts portions of standardized tests (Popham 2001). The overall result is that schools with higher proportions of children of lower socioeconomic status will show lower test scores compared to their more homogeneous counterparts.

The final issue under the heading of confounded causality in standardized tests pertains to the students’ innate cognitive capacities. In other words, all children are not created equal. Genes play an important role in the ease with which children are able to acquire certain skills. Each child probably has one area of strength such as athletics or
language. Test items tend to tap into children's innate abilities, because differential responses lead to greater score-spread. Some items ask that students hold many pieces of information in the students' short-term memories at once (Popham 2001). However, children and young adults differ in their cognitive abilities to perform this task. One example of this type of item asks the students to find a "secret number" in overlapping shapes. The test item provides three pieces of information that the child must remember at the same time: "The secret number is inside the circle. It is also inside the square. It is NOT inside the triangle" (Popham, p. 68). This skill was most likely not addressed in the classroom because it is not useful in real life. Most teachers would agree that finding secret numbers in overlapping geometric shapes was not at the top of their curriculum. Nonetheless, this particular test item remains in many standardized tests just for that reason. Popham found that up to 55% of some standardized tests' items assessed innate cognitive abilities as the dominant factor. Obviously, such biased test items are not rarities in the midst of more appropriate test items.

After identifying the deficiencies of current standardized tests, one should attempt to replace these measurements with an alternative approach to testing. Popham (2001), a "recovering test developer" (p. 75), asserts that it is entirely possible to create large-scale tests that also illuminate instructional effectiveness. The first rule of a successful large-scale test is that it should highlight only a few priority outcomes that can be taught and assessed within the allotted time-frame. Popham observes that, when deliberations take place about what should be included in standardized tests, teachers and other curriculum experts often come with a "wish list" of content that they deem to be important. The problem is that the test-makers are expected to create a test that assesses such a wide
range of concepts that the test is awkwardly designed with too many objectives. When one creates a large-scale test, only those concepts crucial to the child's future should be included. After all, every test is designed to sample the child's knowledge, not to tap into every single concept. However, this is not what one content standard looks like:

The student uses concrete and graphic models to derive formulas for finding perimeter, area, surface area, circumference, and volume of two- and three-dimensional shapes, including rectangular solids, cylinders, cones, and pyramids (Popham, p. 84).

This objective actually consists of about ten standards thinly camouflaged as one.

Educators need to remain careful that one content standard is not a combination of several others that are hidden within the technical jargon. In addition, test items should be limited to only those concepts that are easily teachable by most teachers. It is a myth that every teacher can teach everything. Some of the more abstract concepts are not easily taught by most teachers but are still represented by numerous current test items in standardized tests today. Content standards should also be stated in clear, behavioral terms. Any vague langue such as "promote mastery of reading comprehension" should be identified and dropped in favor of more precise objectives such as being able to "identify the speaker's purpose." Limiting content standards in number and nature forces test-makers to think instructionally, a characteristic that is almost completely absent from test-making companies today (Popham 2001).
Popham (2001) suggests that, instead of focusing our attention solely on standardized test scores to determine teaching effectiveness, we create new methods in which teachers can still be held accountable. He states that the most successful way a teacher can demonstrate his or her effectiveness is to display the cognitive gains of a few students. Popham explains that gains in knowledge should be extensive and represent important outcomes. For instance, learning how to spell a few words by the end of the semester does not display a gain in knowledge that is important and extensive. Effective gains in students' learning are often overlooked as tools for assessing instructional effectiveness. Teachers who are able to create interest and excitement in their students for a particular subject such as math or science are most likely using effective teaching methods (Popham 2001). Gaining confidence when presenting a speech, for instance, is a favorable outcome when judging teaching methods.

Pretests and post-tests are a common assessment tool when gauging instructional effectiveness (Popham 2001). However, a factor known to educators as the reactive effect almost always confounds the results of post-tests. The reactive effect occurs when the student demonstrates gains on the post-test compared to the pretest, but the gains result from familiarization with the test (Popham 2001). Many students will consciously or unconsciously "know" the correct answer on the post-test because they remember the test question from the pretest. This can severely confound the results of the post-test, because the teacher cannot be certain if the cognitive gains can be attributed to instructional decisions.
Teachers should be aware that students know there will be a post-test while they are taking the pretest. Some students may try to remember which questions were asked on the pretest and may even locate some of the answers in a textbook before instruction has even begun. Many teachers believe that an equally difficult but different post-test will solve the problems of the reaction effect. However, this task is easier said than done. Most teachers find it extremely difficult to construct another test different from the first but equal in the level of challenge it provides in the time allotted (Popham 2001). The results gathered from the pretest/post-test method of assessment can be extremely misleading, and it is not likely that teachers will be able to avoid confounding variables.

Popham (2001) advocates for the use of the *split-and-switch design* to assess educational effectiveness. This design incorporates the pretest/post-test method in a way but avoids the problem of the reaction effect. Using the split-and-switch design, teachers construct two test forms that demonstrate fairly similar difficulty levels. The teacher then instructs one half of the class to take the first form as a pretest and the other half to use the second form as a pretest. After the designated period of instruction, the teacher once again splits the class into halves. The portion of the class that took the first form at the beginning of the period will now take the second form as a post-test and vice versa. After the testing has occurred, the teacher is now able to score the forms by mixing the pretests and post-tests but keeping the two forms separate. All of the tests are scored in a blind scoring method (Popham 2001). Finally, the teacher can compare the pretest and post-test scores on the first form to those of the second form. It is assumed that about half of the class will roughly represent the scores of the entire class. The reactivity effect does not occur because the students took two different forms of the same test. People who are
uninvolved in the testing process can serve as volunteer scorers to provide the most objective scoring possible. When there are more than 20 students taking the tests, the split-and-switch method is an effective way to judge how a teacher’s instructional methods affect the students. For classes consisting of fewer than 20 students, the traditional pretest/post-test may be more appropriate (Popham 2001). However, blind scoring would increase this method’s reliability.

As with most methods of assessment, there is the danger of subversion, or item-specific teaching (Popham 2001). This occurs when teachers willingly teach directly from each test item in order to inflate the scores on the post-test. Most teachers who adopt the split-and-switch design are genuinely interested in judging their effectiveness as teachers. However, some teachers are still tempted to create positive false effects to impress faculty and parents. School faculty can conduct random interviews with a few students in the classroom to provide a safeguard for subversion. Students should be guaranteed that their identities will not be revealed to the teacher while reporting any item-based teaching that had occurred in the classroom (Popham 2001).

Teaching effectiveness, though, is only part of the solution to assure intellectual integrity and progress; it is necessary to address the issue of school and district accountability. Principals and school boards usually have more time than teachers to carry out some of the more elaborate assessment methods. Also, their assessments are usually aimed toward providing accountability instead of instructional effectiveness (Popham 2001). In most school districts, a certain number of specialists are employed to carry out assessment methods in order to prove to the community and politicians that
their students are learning. Popham states that the only types of methods that actually benefit students are those that employ honesty and that build a strong case for accountability (Popham 2001). For instance, outside forces such as parents and policymakers will most likely have questions about the school's effectiveness.

School administrators are expected to present a wide variety of evidence of the school’s effectiveness. A single assessment tool that displayed favorable outcomes in the students’ knowledge is usually not enough to convince the public that the students are learning (Popham 2001). If administrators based their opinions about a school on a single assessment, it would be similar to basing a school’s effectiveness on tests scores on a single standardized test. The only factor that changes is that the school itself is providing the evidence. Also, administrators should resist the urge to present only those assessments showing favorable outcomes and should also be careful not to embellish assessment results. Finally, the school administrators should produce a report to display the results of assessment in a straightforward manner that parents and educators alike easily understand (Popham 2001). The report should be free of any technical jargon that could potentially discourage non-educators from becoming involved in education.

The assessment procedures employed by administrators to judge the effectiveness of a school could also make use of the split-and-switch design (Popham 2001). Basically, this design used on a larger scale produces reliable results. However, not all assessments have to be based on tests of any kind. Attendance rates and tardiness frequencies can also serve as a powerful tool for measuring a school’s success. Other reliable methods include the number of students who enroll in the school, the number of students who
attend college after graduation, and the frequency of participation in extra-curricular activities. Whatever the method of assessment that is used, a school district needs to approach the issue in a multi-faceted way (Popham 2001). Using one assessment procedure as the sole basis for decisions about the school's effectiveness lies at the root of the standardized testing problem and should be discouraged.

Kohn, Sacks, and Popham are not the only individuals who have become disillusioned with current uses of standardized tests. Many parents are equally dissatisfied, but do not feel qualified to speak out against standardized tests. Simply becoming educated on standardized test mechanics proves to be a useful tool in the standards debate (Popham 2001). However, action is required if we are to change the way standardized tests are being used to label and restrict children and adolescents. At the most basic level, individuals can express their ideas about testing to others, whether it is at a parent-teacher conference or the grocery store. Thus, people have so many myths about standardized testing that it would be beneficial for all to learn the basics.

The next step one could take would be to write a letter to the state superintendent and send copies to each member of the state's board of education. This letter could include a well-planned appeal for more instructionally illuminating tests and could identify elements that could be included in the test. Each state is expected to submit an official request for proposal (RFP) to test-making companies in order for the companies to provide that state with its own standardized test. The RFP addresses which standards that state considers sufficiently important to be tested. One way people can become active in the movement toward better assessment measures is to provide their states'
boards of education with specific suggestions as to the number and type of standards that will be requested in the RFP (Popham 2001). These people must keep in mind that they do not have to build the test; they only have to require that someone build a better test.

One of the most important factors in becoming active in search of better assessment measures involves implementing assessment literacy programs for teachers, administrators, community members, and parents (Popham 2001). Many of these people do not possess knowledge of the fundamentals of testing. This is not the overwhelming task that many test-makers assert. There are many resources such as books and videos for people who are interested in acquiring basic knowledge of standardized tests (Popham 2001). Parents can be an especially powerful force. They have first-hand knowledge of the stresses that high-stakes testing can bring to children. They are also witnesses to the future accomplishments of their own children in spite of less-than-exhilarating test scores. Parents have been ignored in the standardized testing movement, and they have been made to feel that their assessments of their children’s abilities are unreliable in the face of elaborate standardized testing methods. Unfortunately, parents’ voices may even be more credible than the outcries of educators, for many supporters of standardized tests claim that educators are simply complaining because of their schools’ lack of achievement.
Kohn, Sacks, and Popham have therefore posed important questions on the validity of standardized tests, and initiated the attempt to improve assessment measures for American schools. Their observations are invaluable to American education and explain how some well-intentioned programs can turn into unyielding barriers for many school-age children and young adults. All three theorists have found that buzzwords such as “accountability” and “standards” have colored our perceptions of the standardized testing movement. In the face of deficits in standardized testing knowledge, such words sound like legitimate ways to measure the success of our nation’s schools. However, when we strip away the accountability dogma, we find that the current uses of standardized tests do not meet their expectations. Standardized tests have not been shown to improve educational achievement: an increasing amount of research studies are duplicating this finding, and supporters of high-stakes testing are forced to rely on ideology to sustain their views. Like many other aspects of American society, theory is accepted as truth in place of empirical research.

However, the time has come to refocus our energies on what is beneficial for schools, and most of all, on what is beneficial for students. We need to take a long look at the goals of the standardized testing movement and compare them with the consequences we are finding with actual students today. Simply “meaning well” is not good enough, and we cannot afford to compromise children’s and young adult’s educational success just because the arguments for high-stakes testing “sound good.” Standardized testing rhetoric has the tendency to mask the actual negative effects of the
tests on students, and we should question any policymaker who states that standardized testing is improving the quality of education in the United States. School assessment procedures have been taken out of the hands of those who know about education and placed into the hands of those who have political motives. The most horrifying aspect of this change of hands is that children have been objectified and likened to commodities. It is essential that, when we debate higher standards, we remember that we are discussing living, breathing human beings whose lives are truly affected by ill-fitting educational practices such as the standardized testing movement.
References

National Public Radio.


