

MEASURING STUDENT
ACHIEVEMENT: A STUDY OF
STANDARDIZED TESTING & ITS
EFFECT ON STUDENT LEARNING

Jeremiah Gawthrop

29 April, 2014

**Measuring Student Achievement:
A Study of Standardized Testing and its Effect on
Student Learning**

Jeremiah Gawthrop

Professor Phillip Escamilla

29 April, 2014

Table of Contents

ABSTRACT..... 3

INTRODUCTION TO STANDARDIZED TESTING..... 3

THE ROLE OF EDUCATION IN AMERICA 4

PROS OF STANDARDIZED TESTING 5

CONS OF STANDARIZED TESTING 7

THE DEVELOPMENT OF GOVERNMENT OVERSIGHT IN EDUCATION 8

THE EFFECT OF IMMIGRATION & NATIVISM ON PUBLIC EDUCATION 9

THE FIRST ASSUMPTION: PSYCHOMETRICS AND TEST OBJECTIVITY 11

ANALYSIS OF BIAS IN ASSESSMENT TESTING..... 14

THE SECOND ASSUMPTION: ASSESSING STUDENT KNOWLEDGE 15

DESCREPECENCIES BETWEEN TESTS & TEACHERS..... 17

ELIMINATION OF IMPORTANT TEST ITEMS..... 19

CAUSATION FACOTRS..... 21

EFFECT OF SOCIAL EXPECTATIONS ON TESTING..... 24

THE TWO ASSUMPTIONS ARE FALSE 29

POLICY PROPOSAL: SHRINKING THE FOCUS OF STANDARDIZED TESTING 30

FINAL THOUGHTS 33

APPENDICES 35

BIBLIOGRAPHY 39

ENDNOTES..... 42

ABSTRACT

This paper will question the two assumptions of standardized assessment testing. By reviewing the historical development of standardized testing, it will be made known that standardized tests are not entirely objective or unbiased. This study will also identify how standardized test data is misapplied, specifically in how it is inappropriately used to determine education quality. Additionally, it will be revealed how the misuse of standardized testing and social expectations can affect minority groups. Finally, some solutions will be proposed on how to better apply standardized testing, in order to more accurately predict academic ability.

INTRODUCTION TO STANDARDIZED TESTING

The use of standardized testing is supported by two fundamental assumptions, those being: (1) standardized tests are designed objectively, without bias and (2) standardized tests accurately assess a student's academic knowledge. These assumptions have convinced school officials, to use test data, as the main criteria in determining a student's academic ability and for creating curriculum. Because legislators also believe test data is a reliable indicator of student ability, these tests have become an integral part of the education process and are often used in drafting education policy, such as the No Child Left Behind Act and Race to The Top. This paper will test the two assumptions, determining if standardized tests can objectively and without bias, reveal academic ability and if so, whether or not school officials and politicians are utilizing this data properly. If the goal of education is to help children succeed in life, it is important to be sure of the method used in gauging that goal.

THE ROLE OF EDUCATION IN AMERICA

Education has played an important role from the very inception of America as a country. President George Washington said in his farewell address, “promote... as an object of primary importance, institutions for the general diffusion of knowledge.”¹ Other Founding Fathers like Thomas Jefferson, John Jay, James Madison, John Adams and others, also forcefully advocated for universally available education.² Especially in the modern era, with the rapid development of new technologies, having a strong education has become an essential attribute for success.

The modern world is one that depends less on physical strength and brawn and more on conceptual ideas and theoretical frameworks. The President of the United States no longer leads the military from the frontlines. Instead, developed nations are typically led by academics, whose political battles are more ambiguous than traditional war. Additionally, the most successful careers are those that require years of dedication to academic studies and impressive degrees from prestigious universities. The modern job market demands an educated workforce, meaning that the length of your resume and the extent of your education, determine the measure of your success.³ Without at least a high school diploma, it is nearly impossible to find a job that will pay above the poverty line.

Education is a necessary part of any democratic society. While education was not mentioned in America’s founding documents, with the sacred beliefs of freedom and liberty, it has become abundantly clear that without education, such notions as liberty will disappear. The Declaration of Independence spells out the dream and vision of the Founding Fathers saying that the inherent rights of humanity are, “Life, Liberty, and the pursuit of Happiness.”⁴ Education is the only tool that can ensure the dream of the

Measuring Student Achievement

Founding Fathers remains attainable. Education is the bridge that leads people out of poverty, into prosperity. It is the shield between the U.S. citizenry and the machinations of compromised politicians and bureaucrats.

The value of education in American culture is what makes the topic of standardized testing so important. Historically, government influence in education has risen in conjunction with assessment testing. If education serves as a shield between the people and the government, than any aspect of education assessment, championed by the government, must be scrutinized. Standardized testing is supported by both school and government officials, because it is believed to provide helpful data for policy and curricular creation. It is necessary then, to question exactly how this data is helpful, and then determine if the evidence supports the supposition.

PROS OF STANDARDIZED TESTING

The legal definition for standardized testing is, “A test administered and scored in a consistent or standard manner... administered under standardized or controlled conditions that specify where, when, how and for how long children respond to the questions. In standardized tests, the questions, conditions for administering, scoring procedures, and interpretations are consistent. A well designed standardized test provides an *assessment* of an individual’s mastery of a domain of knowledge or skill.”⁵

The purpose of standardized testing is to assess a student’s knowledge base, in an academic domain, such as science or mathematics. When taking a standardized test, it is assumed that the substance of the test and the administering of the test will be the same for all takers. With this uniformity, a certain measure of fairness and objectivity is achieved and it is believed that elements of bias are removed. Identical tests, with

Measuring Student Achievement

identical degrees of difficulty and identical methods of grading, are propagated as the most fair and objective means, of assessing how a student is progressing in their learning.

Because standardized tests are created to be unbiased and objective, they supposedly ensure that the score a student receives, is an accurate measurement of ability and progress. *Validity* and *reliability* are critical components test makers need, in order to create assessment tools, which create usable inferences about the knowledge and skill of students in a particular area.⁶ The validity of a test is determined by how well the test measures, what it was designed to measure; how accurate the results are.⁷ Whereas reliability refers to whether or not the results of the tests are consistent; students achieve similar scores no matter how many times they take the test.⁸ Tests that are proven to provide both valid and reliable inferences are then norm-referenced, which means the student's knowledge and skills can be compared to a national sample of students in the same grade level. The efficiency and affordability of standardized tests, for evaluating teachers and students, led to these tests becoming the primary tool used by legislators and administrators, in evaluating the effectiveness of schooling on children, as well as to provide data to better manage school systems and develop education curriculum.⁹

Standardized testing is a central part, at all levels, of the current education process. Universities use standardized tests to assist in selecting applicants. For students in grades K-12, testing plays a critical role in evaluating and classifying students, as well as identifying educational strengths and weaknesses, throughout their compulsory education. Testing also shows teachers their own weaknesses and provides insight on how to better structure lesson plans and focus areas. Moreover, standardized state requirements demand accountability from the teachers to the school, the parents and

Measuring Student Achievement

government. Standardized testing is a cheap and efficient method of measuring whether or not schools are achieving the state standards, sometimes forcing schools to revise their curriculum and testing programs so that they can reach these standards.

CONS OF STANDARIZED TESTING

The primary function of standardized testing is to provide information, specifically calibrated to be helpful to legislators, school officials, university recruiters and other administrative positions, all of which operate from outside the classroom.¹⁰ Mass-producing assessment tests that are valid, reliable and norm-referenced make it relatively easy for policy makers to accumulate data on students. This is interesting, since the second key assumption about standardized testing, is that its primary function is to determine a student's academic standing. However, test data is certainly more useful to administrators than students, because a competent teacher can determine a student's proficiency level based off homework, quizzes or classroom participation. If standardized tests are not necessary to determine a student's academic level, it raises an important question, whether test results (versus other sources such as teacher input), are the best source for determining policy or curriculum changes. Since standardized tests can only assess, not determine, a student's academic status, the argument is made that it is dangerous for policy makers to rely predominantly on the data provided by these tests.

The price and efficiency of using standardized testing, to accumulate vast amounts of information, is quite appealing to administrators, who require such information to make policy decisions. Standardized tests have been increasingly used, "to make major decisions about students, such as grade promotion or high school graduation, and schools. More and more often, they also are intended to shape curriculum and

Measuring Student Achievement

instruction.”¹¹ It is assumed that newer tests have overcome the flaws of past tests and are accurately able to measure important data that is worth “testing to”. However, this argument completely ignores the real-world limitations to what a standardized test can actually do.¹² Tests are created to assess a student’s knowledge base; meaning test results are not representative of the student’s total academic ability.

THE DEVELOPMENT OF GOVERNMENT OVERSIGHT IN EDUCATION

The foundation for a public education system was laid in the colonial and post-Revolutionary periods. Up until the late 1800s, American schooling largely followed the private philanthropic tradition of minimal public services.¹³ Some federal policies, such as the Land and Northwest Ordinances of 1785 and 1787, required a public system of education, but the administration and oversight was entirely private.¹⁴ ¹⁵ “Up until the mid-19th century, education consisted of private schools run by paid tutors, state-chartered academies and colleges with more formal programs of instruction, benevolent societies, and church-run charity schools – in sum, a ‘hodge-podge’ reflecting the many.”¹⁶

The Constitution of the United States makes no mention of education, which is likely due to the fact that the early American education system was almost exclusively privatized. However, as the 19th century dragged on, the amount of incoming immigrants began to increase exponentially, causing the population of students to grow and the focus of schooling to change. Religious, educational and civil leaders began taking note, that lacking in education led to ignorance and moral delinquency.¹⁷ These leaders began advocating for public schooling so that poorer children, who did not have access to church-run charity schools or common pay schools, would receive an education.¹⁸

Measuring Student Achievement

Between 1820 and 1860, American cities grew at a faster rate than at any other time in the nation's history and that growth continued into the 20th century.¹⁹ Approximately 125,000 immigrants came to the U.S. in this period, mainly from Europe.²⁰ This population explosion created new challenges for the education system; according to census statistics, public school enrollment increased from 6.8 million, in 1870, to 15.5 million by 1900, and at this time 80% of children, between the ages of 5 and 17, were enrolled in some sort of school.²¹ This incredible growth in student population, throughout the 19th century, led to several federal policies, such as the issuing of land grants in 1841 and 1848, where Congress granted surplus revenue and over 77 million acres of public land as donations, to support schools.²² In 1862, the First Morrill Act was passed, otherwise known as the Great Land Act, which donated more public lands for states to create, at least one college, with the purpose of teaching agriculture, mechanic arts and industrial education.²³ In 1867, the Department of Education was created, to start collecting data on schools and teachers, in order to help the states establish more effective school systems.²⁴ The Second Morrill Act followed this, in 1890, giving the new Office of Education the responsibility of administering support for the system of land grant colleges, created in the First Morrill Act.²⁵

THE EFFECT OF IMMIGRATION & NATIVISM ON PUBLIC EDUCATION

The United States experienced a period of mass migration leading into the twentieth century. People came from all over fleeing crop failure, famine, looking for jobs and property, but mostly searching for the promise of economic opportunity and freedom.²⁶ Throughout the course of the 19th century, millions of people immigrated to the U.S., mostly from Germany, Ireland and England.²⁷ These newcomers disrupted the

Measuring Student Achievement

homogeneous culture of the young United States and nativism began to spread; especially in New York City, where an estimated 70% of all immigrants entered the U.S.²⁸ New Yorkers felt threatened by the surge of immigrants, bringing with them foreign culture and worse, different religion. The Catholic Irish especially caused uproar in the rigid Puritan communities of New England.²⁹ The ant-Catholic sentiment would remain a dominant theme of American culture well into the 20th century, but all of this domestic turmoil pressed forward the idea, that the government needed to Americanize these new immigrants.

Although it is difficult to establish a causal link between demographic and educational changes, the argument can be made that the population growth and increased heterogeneity of American culture, necessitated the creation of institutions, particularly universal schooling, to *Americanize* the masses.³⁰ The 20th century social philosopher, Hanah Arendt, said, “In America, as a matter of fact, education plays a different and, politically, incomparably more important role than in other countries. Technically, of course, the explanation lies in the fact that America has always been a land of immigrants; it is obvious that the enormously difficult melting together of the most diverse ethnic groups—never fully successful but continuously succeeding beyond expectation—can only be accomplished through the schooling, education, and Americanization of the immigrants’ children.”³¹

Often these Americanization laws, which arose during the 1800s, are looked on favorably, because they were specifically designed to target the poorest parts of society and ensure that the children would receive an education. However, closer examination will show that while policies, focused on educating the poor, were partly borne out of

Measuring Student Achievement

genuine interest in bettering the children, equally influential was the desire to keep these children from continuing to follow the lifestyle model, taught by immigrant parents and deemed harmful, or un-American. Prior to the influx of immigrants, parents had been given complete authority over determining how a child was to be educated. Under this method of total parent authority, an 1850 census showed that only 1 out of 10 people self-identified as illiterate.³² Nevertheless in 1851, the Massachusetts Teacher published an article which said that in order to fix the problems of the Irish immigrants, “the great remedy is EDUCATION... the parents are unfit guardians of their own children. If left to their direction the young will be brought up in idle, dissolute, vagrant habits, which will make them worse members of society than their parents are; instead of filling our public schools, they will find their way into our prisons, houses of correction and almshouses.”³³

This push for more government oversight and less parental control was not because parents had been doing a bad job (again, the illiteracy rate was incredibly low) but because it was believed by many nativists, that immigrants, like the Irish, were inferior and if their children were not forced into public education, separated from parental influence, these children would grow to model the beliefs and actions of their parents; to the degradation of American society. The influence of this early nativist mentality would directly affect the creation of standardized testing by introducing biases. These original biases challenge the first assumption of standardized testing.

THE FIRST ASSUMPTION: PSYCHOMETRICS AND TEST OBJECTIVITY

The introduction of psychology-based intelligence testing normalized dependency on numerical objectivity; so that administrators and the public would need objective measurements of student intelligence, in order to feel that the results were legitimate.³⁴

Measuring Student Achievement

Out of this emerged the study of psychometrics, a field of study concerned with the theory of psychological measurement, including the measurement of knowledge abilities, attitudes, personality traits and educational measurement.³⁵ Psychometricians created the assumption that assessment tests are objective and they continue to be responsible for developing and devising psychometric or, standardized tests.³⁶

Science became a crucial aspect of education bureaucracy because it could legitimize school administrators' actions and requests for financial assistance. A prominent figure in the common school movement, Henry Barnard, said that, "Crucial to educational bureaucracy was the *objective and efficient classification*, or grading of pupils."³⁷ Based off this sort of mentality, mid-19th century school reformers argued this information could only be obtained through standardized achievement testing, which would more rationally and efficiently classify student achievement. Scientific expertise was now needed to help streamline the business of education, by determining not only how to track student progression, but also to determine a student's intelligence through tests; subsequently placing those students on appropriate learning tracks, based off the results of those tests.³⁸

In *The Definition of a Profession: the Authority of Metaphor in the History of Intelligence Testing, 1890-1930*, JoAnne Brown analyzes psychologists from this era, where psychology first began to be accepted in education. Psychologists purported to have a superior understanding of student potential than teachers did. Some prominent psychologists from this era like Robert Mearns Yerkes, Lewis Madison Terman, and Henry Herbert Goddard argued that teachers' firsthand knowledge and personal relationship with their students, rendered teachers' evaluations subjective and biased,

Measuring Student Achievement

giving them an inaccurate interpretation of classifying students.³⁹ Psychologists could create objective achievement tests, giving school administrators access to numerical truth, to guide academic interpretations, showing legitimate reasons for why education funding was needed and how it was being effectively used.

It was during this period that government dependence on standardized testing was born. The profession of psychology was largely legitimized, by making teachers look biased and subjective, and establishing itself as the only profession with the expertise to properly manipulate and create tests that would provide objective and accurate results.⁴⁰ However, psychologists were taking individual test results and manipulating them into subjective norms and merely called this norm an *objective* standard.⁴¹ The subjective norm, in many of these tests, aligned with nativist biases.

These supposedly objective tests were actually geared towards native-born Americans, from a specific social class and an anticipated base knowledge. Immigrant students were not expected to know this information, but were to be taught it in public schools. Consequently, immigrant children were placed in a lower academic category than native-born students. Evidence of this has been found in recent research, showing that the SAT is highly predictive of white students but does not do so well in predicting other ethnic groups.⁴² Appendix F is a table that shows the correlation of SAT scores to subgroups in the First Year GPR (FYGPA) of college. The results show that the SAT predicted white students with correlations from 0.46 to 0.51 and were less predictive for underrepresented groups, with correlations ranging from 0.40 to 0.46 for FYGPA.⁴³

This disparity between different ethnic groups, in regard to educational success, has its roots in these *assumedly* objective tests that actually (perhaps unintentionally),

Measuring Student Achievement

placed the poorest students into lower academic categories; not because they had less intelligence, but because the tests were biased against them. Nativist influence and the idea of manipulating subjective test results into a seemingly objective norm, undermine the first assumption of standardized testing. If assessment tests were originally created with a bias, it is to be expected it still happens today.

ANALYSIS OF BIAS IN ASSESSMENT TESTING

Modern psychologists and test makers have worked to eliminate biases, but the existence of tester prejudice remains. The issue is that a system fully open to flaws was introduced and normalized as an inherently fair and unbiased system. However, standardized achievement testing is not an infallible system; it is vulnerable to misuse; the gravest of which, ties into the issue that affects modern education.

There exists a disparity, between the original purpose of standardized assessment testing and the modern expectation. Achievement testing was originally used and proposed by psychologists, as a means of placing students, according to their academic capabilities. Standardized testing *evolved* to incorporate not only what group children should be classified in, but also to gauge what children were learning.⁴⁴ Always a leader in education development, Massachusetts had the first reported use of a written examine, to determine the shortcomings of state schools.⁴⁵ Under the guidance of, Secretary of the State Board of Education in Massachusetts, Horace Mann, Mass. switched over from oral to written testing. This was because written testing was much cheaper, efficient and objective in grading and an overall better method for classifying, evaluating and comparing large groups of students over a greater area.⁴⁶

Measuring Student Achievement

As the methods of testing became increasingly quantitative and scientific in appearance, public faith in them grew. However, a fundamental concept was introduced that is often overlooked. Although the use of tests for monitoring students was logical, the idea of monitoring education, to evaluate student learning, was born in the minds of individuals who were already convinced that the level of education was substandard.⁴⁷ As has been previously noted, the education system, established during the colonial and post-Revolutionary era, had been very effective in creating a literate nation. Still, these future-minded thinkers were convinced that psychological science had proven that intelligence was something that could be measured and tested. Because these early tests were biased, test scores supported predictions that certain student populations would need greater help. Therefore, these standardized tests were created with the preconceived notion that education was in need of reform and consequently, the information obtained by these tests, confirmed said hypothesis.

THE SECOND ASSUMPTION: ASSESSING STUDENT KNOWLEDGE

The chief problem with the second assumption of standardized tests is that the data collected through assessment tests, can only provide a glimpse of a student's total academic ability. Administrators seem to use tests, not just as one of many tools, but the final determining factor, in making decisions concerning education. One of the anticipated benefits of testing is that students and teachers get better feedback on how they are doing. However, this assumes that the test measures every factor of the student's life. There is an incredible amount of information that a child, at any age, is likely to know, creating a significant difficulty for test makers, who have to limit the content domain of testing, otherwise the tests would be too long.⁴⁸

Measuring Student Achievement

A standardized test has to be created in such a way that it yields valid norm-referenced information on students over a large content base. Test results do not have valid and reliable norm-referenced comparisons, unless there is a high score variance; meaning a 50/50 disparity between the students who take the test and get correct answers. Test makers avoid questions that students get the correct answer for, too many or too few times because that would be considered an inappropriate question for that academic level. By carefully focusing on items in tests that discriminate optimally, test creators form assessment tools that do an excellent job at providing relative comparisons of students' content mastery nationwide.⁴⁹

Using this methodology, it is possible to determine students' relative strengths and weaknesses in different subject areas. It also can be used to determine which areas, in different subjects, a student is strongest. For example, a standardized math test can have 20 questions devoted to Algebra, Geometry and Trigonometry respectively. By splitting up questions in this way, a rough estimate can be made of the students' ability in these areas. Still, the results are only rough estimates of student ability, since the number of questions necessary for an accurate measurement would be much greater; but such a test would take too much time to complete.⁵⁰ These norm-referenced tests do a remarkable job at what they are supposed to do - that being to provide a longitudinal study of student achievement. Yet, these results are not nearly as specific, or accurate, as psychometricians make the scores out to be. These tests should be solely regarded as rough approximations of a student's knowledge base and should not be used to evaluate any other criteria. Unfortunately, the results of assessment tests are often used incorrectly.

Measuring Student Achievement

The assumption that assessment tests provide an accurate measure of student knowledge is still accurate, but it must be clarified with an understanding that it is only a general measurement. Nevertheless, because assessment testing is so effective at accumulating data on student knowledge, school officials and policy makers, who use the data, act as if assessment testing is equally effective at identifying statistics for different criteria. One significant way that the data, acquired through assessment testing, is used erroneously, is seen through the idea that a student's test scores directly reflect the quality of the education the student received. Student test scores do not automatically carry over into explaining the quality of the education. It seems logical that low tests scores are equivalent to bad schooling; but that is not a correct assumption. UCLA Emeritus Professor W. James Popham, proposes that there are three main reasons for why test scores do not necessarily predict education quality: (1) discrepancies between the information in the test and what the teachers teach, (2) a psychometric tendency to eliminate important test items and, (3) confounded causation.⁵¹

DESCREPENCIAS BETWEEN TESTS & TEACHERS

There are four private companies who dominate the testing market: Harcourt Educational Measurement, CTB McGraw-Hill, Riverside Publishing and NCS Pearson.⁵² According to a report in 2011, from the Educational Marketer, Harcourt, CTB McGraw-Hill and Riverside Publishing write 96 percent of the exams administered at the state level, while Pearson controls the scoring of standardized tests.^{53 54} Since these are for-profit businesses, they are under considerable pressure to sell standard achievement tests; but run into problems because of how diverse the curriculum is in the United States.

Measuring Student Achievement

Because different states have different standards, these standardized tests have to be created in a way that encompasses the wide variety of curriculum, from every single school, in the United States. Further complicating things is the fact that some schools (like charter schools) have special permission to further localize curricular decisions.^{55 56} On a very general level, state standards are similar, for example, in how they all focus on the major subjects like math and science. But on the level that counts for assessment testing, the instructional level, the educational objectives and the timeframe for learning those objectives are very different. One school might teach a different section of Algebra, at a different time than another school, but the standardized test will assume that every school has reached that same level in Algebra. If the results of standardized tests are seen as an assessment, which is a sample of how much the student has learned, it will accurately assess that the child does not know a certain area of Algebra. However, if you take that assessment and extrapolate it out, under the assumption that the quality of education was low, that conclusion could be utterly wrong.

By definition, a standardized test is a one-size fits all sort of thing, but that does not work in a system with widely varying curriculums. A test cannot offer questions that are perfectly aligned with all the different curriculums, in every school, in the United States. Even if a common curriculum were to be implemented (as Common Core is attempting to do), where every state and school had the same curriculum, that still does not mean that it would be the best curriculum for every student, or that those students would learn that curriculum at the same speed. There would still be wide variations between schools and standardized test results would remain unable to provide a complete picture of student performance.

Measuring Student Achievement

An intensive study done by Michigan State University, in 1983, illustrated a disconnect that occurs between what is taught at the state and local level and what is tested in standardized achievement tests.⁵⁷ The study looked at five nationally standardized achievement tests, in mathematics, and analyzed their content for grades 4-6. Operating on the assumption that textbooks influence classroom instruction, the study also looked at the content of four widely used textbooks for the same grade levels.⁵⁸ The research revealed that between 50 - 80 percent of the mathematical knowledge that was expected in the standardized tests, were not properly addressed in the textbooks.⁵⁹ The Michigan State researchers said, “The proportions of topics presented on a standardized test that received more than cursory treatment in each textbook was never higher than 50 percent.”⁶⁰ The manuals that accompany standardized tests have descriptive material explaining what information is in the test.⁶¹ These descriptors show that usually, the information in the test is fairly general. The material has to be general in order to work for a nation that has different curriculums. Yet these general descriptions leave room for assumptions that the teaching and testing are in alignment; although in reality they are not. Standardized tests are great at providing a rough estimate for where a student is at academically, but they should not be used to evaluate teachers or school quality.

ELIMINATION OF IMPORTANT TEST ITEMS

The second reason why standardized achievement tests should not be used to evaluate education quality is because these tests provide meaningful comparisons based off a small collection of items.⁶² In order to get an accurate average result, test designers want to have test items that spread out the total test scores. This means that they do not want to ask question that everyone gets right; they want an even spread of right and

Measuring Student Achievement

wrong answers. Items or questions that are answered correctly, 40 to 60 percent of the time, provides a proper spread.⁶³ A test item that is answered correctly, by 90 percent or more of the test takers, does not work with test efficiency. Since there is a limited amount of questions and length, while a test is being developed, it is unlikely to keep questions that the majority of the sample testers get correct. Because they are looking for questions that provide an even spread of right and wrong results, achievement tests can be considered a moderate difficulty test.

The results of testing to the middle, are that in trying to create score variance, questions that students would most likely get correct, are removed from the test. The questions children are most likely to know the answer too, are the subjects a teacher deemed the most important and therefore spent the greatest amount of time making sure the children understood it. Accordingly, the better job that a teacher does at teaching an important knowledge or skill, the likelihood of that question appearing on the test diminishes.⁶⁴

For example, teachers across the country/state may determine the most important part of a science class is to learn the elements on the periodic table, whereas the less important part is to learn about nuclear fission. Assuming these teachers do a good job, during the course of creating a test, a test developer could discover that 90 percent of the sample testers have a good knowledge of the elements and so to increase score variance, questions about the elements are taken out of the test and replaced with questions on nuclear fission. If the achievement test remains focused on evaluating student knowledge, the test would accurately show that more students lack knowledge in nuclear fission. If anyone pursuing curriculum reform, incorrectly tried to assume the quality of education

Measuring Student Achievement

with these test results, it would not take into account how well the teachers taught students the periodic table and would instead say that overall, teachers are doing poorly.

This process of eliminating test questions plays into both the assumption of test objectivity and that assessment tests accurately reflect student knowledge. A test cannot be perfectly objective when test developers need half the test takers to get incorrect answers. If the majority of students get a specific question correct and consequently, that question is removed, the test results will continue to accurately reflect a *portion* of student knowledge but will not provide a *complete* measurement. If school officials or legislators make decisions based solely off test results, it would be an improper use of the test data.

CAUSATION FACOTRS

The third reason for why achievement tests should not evaluate the quality of education is standardized achievement tests are heavily swayed by three causative factors: what a child learns in school, a student's inherent intellectual capacity, and what a child learns outside of school.⁶⁵ Only one of these factors, what a child learns in school, has anything to do with the quality of a child's education. It is inherently illogical, to attempt to evaluate the quality of a child's education, based off the child's natural ability or out of school learning, things that have nothing to do with the classroom environment. Again, the focus of an achievement test is to determine how much a student knows, that can include native ability and out of school learning, but test results do not explicitly categorize these groupings.

Appendices C, D & E are examples of different questions that evaluate what is learned in the classroom, native ability and out of school learning. Appendix C asks a

Measuring Student Achievement

mathematical question, which provides a valid inference about a 3rd graders ability to pick a number sequence that matches a verbal representation of a subtraction problem.⁶⁶ This is a problem that would certainly be addressed in a classroom. The student's answer, on this type of question, would accurately evaluate the student's knowledge and possibly the quality of the classroom environment. However, these questions still could not definitively reflect the student's learning, due to an endless number of possibilities as to why the child did not get the question right i.e. mental inability, lack of sleep on the day of that lesson, failure to do homework etc.

Appendix D presents a question that is meant to evaluate a student's innate intellectual ability. Sadly, everyone is not born with the same intellectual ability; an unpleasant but infallible fact. The question in Appendix D requires an inherent level of deductive ability. A student with greater intellectual aptitude would be able to look at answers A, B, and D and come to the realization that they are not very good methods of conserving resources; subsequently choosing the correct answer, of C. However, not only does this require greater intellectual ability, this sort of question is also based on the assumption that academic intelligence only has one form. Recent studies by educators and psychologists have begun to show that there are actually, multiple forms of intelligence.⁶⁷ A child that has less aptitude in quantitative or verbal tasks may possibly have greater interpersonal intelligence (some examples of this can be seen through the replacing of the IQ test with the EQ test).⁶⁸ Achievement tests do not test for these different types of intelligence, meaning that students, with a different type of brain function, will do poorly on a question specifically designed to measure the standard intelligence. The reasoning behind having questions that provide an advantage, to only a

Measuring Student Achievement

certain part of the testing population, refers back to the need for variance in test results; assuring that less than 90 percent of the testers will get the answers correct.

Appendix E gives an example of a question on a standardized test that evaluates what a student may or may not have learned from outside of the classroom. While these questions are not subjective in their answers, they are completely subjective on the student's home life and create a distinct discrepancy between economically advantaged and disadvantaged students. Students with a higher socioeconomic background have a more stimulating environment and greater access to resources that increase the likelihood of knowing the correct answer. Appendix D asks which of the following items is not a fruit, but that question is unlikely to have been covered in a classroom. The seemingly obvious answer would be celery, but a disadvantaged student, with a poor background, may have never been exposed to certain foods, and would have no way of knowing the correct answer.

The disturbing fact is that achievement tests have a considerable amount of questions, with answers based on out of classroom experiences. To illustrate why this is wrong, imagine an economically disadvantaged school. If a substantial number of the questions, on the standardized tests, are questions like Figure D, the likelihood of students scoring high is lessened. Meanwhile, a school in a high-income community is likely to have students who score much higher, simply based on a different socioeconomic status.

One of the biggest reasons why socioeconomic factors are tied to test scores is because of questions that are designed to assess knowledge, learned outside of the classroom.⁶⁹ Again, the reason why these questions are used is because there is a high

Measuring Student Achievement

variance between rich and poor children and asking questions like this ensure that 90 percent of students will not get the answer correct, spreading out the student test scores so that accurate, norm-referenced interpretations can be made. Even more than race, the biggest difference between good and bad scores is wealth. A recent study by the Annie E. Casey Foundation found that the gap in achievement test scores, between the rich and the poor, had risen by nearly 60 percent since the 1960s, a gap that is almost twice as large as the gap between white and other ethnic children.⁷⁰

EFFECT OF SOCIAL EXPECTATIONS ON TESTING

There is an interesting dichotomy facing the modern public education system. Amidst growing concerns and protests that standardized testing is damaging, the frequency and use of testing has increased.⁷¹ American universities are particularly known for a heavy reliance on standardized testing, requiring applicants to pass the Scholastic Aptitude Test (SAT), Graduation Record Examination (GRE), the Law School Admission Test (LSAT) or the Medical College Admission Test (MCAT). Like all assessment tests, these are meant to gauge the likelihood of a student to succeed in college. This means that these collegiate achievement tests are attempting to determine a student's merit and whether or not that student has earned admittance into a school. However, these tests do not measure past academic achievements. Therefore, standardized tests are better at measuring how well someone is likely to do, or merit, but never takes into account how much a person may or may not deserve to get into a college.

To some, this may seem like a useless question. If college achievement test scores are attempting to gauge merit and determine a student's proficiency level, it does not matter whether or not someone deserves admittance, because they are not as capable as

Measuring Student Achievement

the person who scored better on the test. However, as was previously examined, standardized tests are filled with questions that are not based off academic knowledge and they are not precise indicators of intellectual ability. Defenders of merit based testing will argue that college openings are scarce and it is in society's best interest, to utilize those spots, with students who will be the most productive.⁷² Still, this sort of plan where only the highest test scores are accepted, do not guarantee a good balance between academic ability and effort. Someone actively involved in community projects may score lower on the GRE, than a student who spent his whole life in the library studying, but that does not mean he has less merit and therefore should not be accepted into the school.

Universities understand that test scores do not reveal the whole picture about applicants and look at other factors besides test scores. It seems like common sense, that universities would look at more factors, in a potential student, than simply the test scores; but in compulsory education this is not the case. Test scores are typically the determiner of everything in grades K-12 and as a result, this can create adaptation. Test questions that require out of school knowledge, significantly affect students who come from low socioeconomic backgrounds. The majority of these poorer students are minorities, either African American or Latino. The most provocative evidence on the negative effect of standardized testing is the tendency of African American students, to adapt to the expectations of standardized tests; these expectations being that black students will not do as well as white students.⁷³

The phenomenon of adaptation arises when people adjust to the expectations society places on them. Since there is a societal expectation that minority students will not do as well on standardized tests, these students become aware of that expectation and

Measuring Student Achievement

act accordingly. A conservative economist, Gary Becker, found that disadvantaged minorities made poor decisions about investing in their own future based on their perception, of their own capabilities, which were shaped by societal expectations.⁷⁴ For Becker, “the beliefs of employers, teachers, and other influential groups that minority members are less productive can be self-fulfilling” where members of disadvantaged factions will, “underinvest in education, training, and work skills” which subsequently make these groups less productive.⁷⁵ This illustrates how adaptation by minority groups, to societal and cultural expectations, creates a cost that effects the starting position of these groups; meaning they do not enjoy equal status in an “objective” standardized test.⁷⁶

The hypothesis that minority students are negatively affected by societal expectations was tested when the University of California, Berkeley, between 1981 and 1990, lowered their admission standards, to comply with affirmative action regulation.⁷⁷ During this same period, the number of first year African American students increased 32 percent. For many critics of affirmative action, the greater admission of African Americans showed how lowered standards resulted, in less qualified individuals, taking positions away from more qualified students. However, if the societal expectation of African American students was true and affirmative action was a mistake, by admitting inferior students, naturally these students would fail to graduate because they were not at the necessary academic level. But in fact, the graduation rate for African American students doubled and 62 percent of these students had graduated by 1996.⁷⁸ As African Americans refuted the adaptive image they held of themselves, they perceived themselves as being more capable.

Measuring Student Achievement

Social expectations are not something that randomly appear when a student applies to college. It is something that is developed over the course of the student's childhood. Students who grow up in disadvantaged homes, lack encouragement in school and do not test as well. Consequently, these students fit the social bias of failure and are likely to adapt to this expectation. As was described earlier, standardized tests need to be norm-referenced, requiring a wide variance in scoring, meaning that the questions on the test have to be questions that have an equal number of right and wrong answers. To reach this number, the questions must be geared towards a specific population. Here is where bias is introduced. These tests are designed towards a specific population of students. They must be in order to have a high score variance. This resembles the early development of education and written assessment, when Americanization was an important factor. The result was that these tests were biased, potentially racist, because a major goal of education was to teach immigrant populations how to be "American".

Today, that same sort of cultural bias still exists, but in a different form. By definition, cultural bias refers to where a test contains questions that are tied to a specific culture, where a student taking that test, from a different culture, would be at a disadvantage. Supreme Court Justice Sonia Sotomayor, the first Latina justice in the court's history, said that despite her academic achievement and intellectual ability, "[her] tests scores were not comparable to that of [her] classmates... There are cultural biases built into testing." ⁷⁹

In attempting to create a test that must have questions where students answer incorrectly, the test will be geared toward whatever population sets the cultural norms. For example, a study conducted by Roy Freedle, of the Educational Testing Service,

Measuring Student Achievement

identified something called “differential item functioning” or DIF.⁸⁰ A DIF is a question that has notable differences between different ethnic groups, even though these groups share the same educational background and skill sets. The study, conducted in 2010 and 2003, showed that there were no DIF problems in the mathematics section.⁸¹ What these studies confirmed, is that in some types of verbal questions, there was a DIF for black and white students. On some of the easier verbal questions the studies found that a DIF favored white students, whereas in the more difficult verbal questions, the DIF was in favor of black students.⁸² A likely reason for this is because the easier questions reflected cultural expressions that were dominant in a society, which just so happens to have a majority white population. What this means is that white students had an edge, not due to better education, study skill or aptitude, but because they were surrounded by these cultural expressions. This is an unfair advantage and a problem in a high stakes test like the SAT.

Kathleen Steinberg, a spokeswoman for the College Board, said that the organization disagrees with the 2003 and 2010 study saying, “We believe that our test is fair. It is rigorously researched, probably the most rigorously researched standardized test in the world.”⁸³ Steinberg said that the studies were discredited and that the different scores in subgroups were certainly present, but that they were a reflection of education quality and not a reflection of bias in the tests themselves. However, it has already been addressed how standardized tests are not capable of accurately measuring academic quality and apparently there were some issues because the SAT has recently undergone major changes.

Measuring Student Achievement

Although standardized tests may have advantages for one group over another, it does not mean racial discrimination is an intentional goal of test makers. Achievement tests are meant to evaluate knowledge base; how much a student knows. College achievement tests, like the SAT, are slightly different because they evaluate the aptitude and proficiency of a student, but still follow the same idea that what is being evaluated is the student's academic knowledge. There is nothing inherently racist with testing a student's out-of-class knowledge, nor is it inherently discriminatory to ask questions that are based off of cultural norms. The SAT and other achievement tests are fair. They do a good job at pointing out which groups of students have less out-of-class knowledge. It is difficult to argue that the College Board and Pearson and others, are intentionally discriminatory in asking these questions. Nonetheless, the presence of these questions is based off of a bias that does affect student test takers on a socioeconomic level.

THE TWO ASSUMPTIONS ARE FALSE

The initial premise of this study was to challenge the two assumptions of standardized testing. The first assumption, tests are completely objective and free of bias, was challenged by looking at the historical development of assessment testing in American education. The influence of nativism and Americanization was a significant reason for why legislators became involved in initiating education policy and standardized testing. When looking at the role of psychometricians, it was affirmed that standardized tests are objective and unbiased, in the sense that there is only one correct answer and all tests are graded the same; but in the process of test development, there is bias towards specific populations and the need for high variance in test answers does not allow for truly objective tests.

Measuring Student Achievement

The second assumption, tests accurately reflect student knowledge, was shown to be true. Assessment tests do a good job at assessing a student's mastery of the material covered in the test. However, the weakness is that assessment tests are not able to account for the wide range of variables, besides student knowledge, that could affect test results. In the pursuit of expediency and efficiency, combined with a trust in the objectivity of tests, the data acquired from test results are often inappropriately applied to different education policies and curricular reforms. A significant example was shown in the inability for test results, to accurately determine education quality. Drafters of education policy and developers of school curriculum primarily reference assessment test data to solve problems the data does not relate to.

POLICY PROPOSAL: SHRINKING THE FOCUS OF STANDARDIZED TESTING

In order for the education system to become more successful in equipping students, the government needs to relinquish some control. This study has revealed that as the role of administrative figures in education has increased, the use of testing has also increased, because it is efficient and relatively easy to accumulate large amounts of data. But if the goal of education is to equip students academically and prepare them for successful living, then new methods must be developed, even if they are less efficient and more difficult to implement. Standardized assessment testing should not be the primary factor in evaluating teacher or student performance, nor should it be the main method of creating education policy and curriculum.

The diversity of education, across the United States, is so incredibly vast, it is impossible for the federal or state governments, to provide standardized assessment tests that can accurately evaluate student ability and knowledge, on a national level. The PISA

Measuring Student Achievement

test, a widely respected international assessment test, consistently reveals that the U.S. equips students with a mediocre education, in comparison to other countries.⁸⁴ Something needs to change and obviously, that change is not going to come from more tests.

Assessment testing is not designed to help teachers; it is a tool for keeping the education system accountable to government. But the inability of standardized testing to accommodate for the wide variety of curriculum has fostered the dangerous idea that education itself should be standardized. That is a terrifying concept, because who should determine this standard and what will the standard be? A standardized education system would be the end of America's belief in freedom of expression.

The education system needs to become accountable to the communities in which they are located, not to the government. By making school funding and teaching jobs dependent on standardized test scores, the government could force the adoption of a common curriculum and adapt learning to fit a politically correct agenda. Certainly government has a role in preventing unequal treatment in education, but one of the benefits and curses of a democracy, is that people are allowed to learn and think differently. Government has historically employed standardized tests, which would work better with a uniform curriculum, but the implementation of such a curriculum would go against the Constitutional right of free expression. If government prefers to rely on a tool that cannot fix education problems, maybe it is time for a new approach.

Research has shown that parental involvement is arguably, the most important factor in determining a student's educational success.⁸⁵ No amount of test data will provide policy makers, or school officials, with a solution to replace parental participation. In fact, government sometimes exacerbates this issue, by supporting the

Measuring Student Achievement

false assumptions of standardized tests; fooling some parents into thinking their children are being cared for by the school system. Nothing will influence a child's academic success as much as the example his/her parents set⁸⁶; but for students without that support, a teacher may be able to fill the void. Therefore, assessment testing should be redesigned to be useful to the people directly involved in student development, the teachers and parents.

By localizing assessment testing, rather than nationalizing it, the tests would more accurately reflect student ability, since they will be designed with specific curriculum in mind. If each school district were responsible for developing their own assessment tests, it would still not produce a full picture of student ability, but the issue of high score variance could be overcome. Instead of every assessment test being created for the national majority, localized assessment tests could be designed towards the local majority, meaning test scores would more accurately reflect the cultural norms of the community.

Alone, standardized testing is incapable of providing an accurate measurement of student ability. By narrowing the range of tests, it would be possible to reduce the amount of errors and target test results to the people directly responsible for the students; teachers and parents. However, this is one step in an arduous process to improve the process of student assessment. Different methods must be developed to partner with written assessment tests. Since written tests are limited to assessing student knowledge, perhaps an oral exam could be developed, to partner with written tests and gauge student's speaking ability and critical processing skills. Any plan separate from paper will be messy and difficult, simply because people are much more difficult to work with.

Measuring Student Achievement

Nevertheless, that is not a reason to give up on developing a new plan, when the current one is inadequate.

This policy proposal challenges government officials to relinquish control and trust the common people to take care of themselves. Policy that localizes the focus of assessment testing would provide the government with less useful data, but it would more accurately reflect student knowledge and be beneficial to teachers. The future of America's students is not a political bargaining chip. Policy should empower localized schooling, even if there is no political benefit from it. Instead of trying to fix problems from the top down, policy should put greater responsibility on the schools and teachers to implement change and increase parent involvement. Policy can only do so much; real change is only going to come from those directly involved in student learning. Therefore, any policy enacted, should work to place high caliber teachers in every school, rich or poor, and empower those teachers with the necessary tools to encourage parental involvement and truly assure student's academic success.

FINAL THOUGHTS

Standardized testing is a unique and effective tool for assessing student knowledge but it must be used appropriately. Standardized tests are imperfect and the greater the population the test is norm-referenced to, the more inaccurate the results will be. Government has a role to play in addressing problems in the education system, but using assessment test data as an all-purpose solution does more harm than good.

Education reform policy should rely predominately on input from teachers, even though that is a supposedly, less efficient process. Standardized testing may have a meaningful role to play, but by recognizing its limitations, better-suited methods can be implemented

Measuring Student Achievement

and may be the first step in ensuring that the education system truly equips all students, to be successful students.

APPENDICES

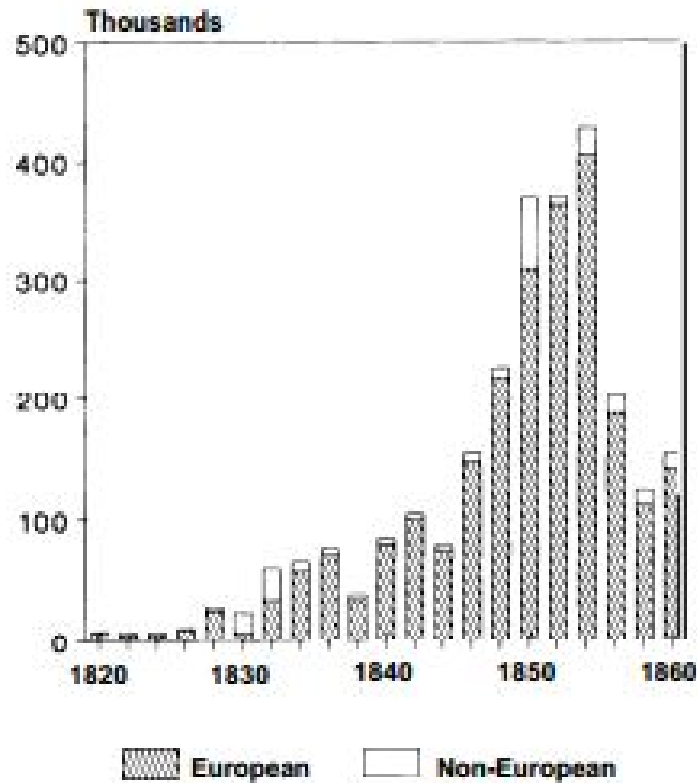
APPENDIX A

1855	<p>"COMPULSORY" ERA</p> <p>During this stage, government compels the establishment of school districts, taxation for government schools, curriculum and structure, and children's school attendance (O 160).</p> <p>Marked by decline of parental authority; children of certain ages compelled to attend school. (For a brief period in some states, it was illegal for children to attend non-government schools, even if parents could afford to pay tuition.)</p>	More than 6,000 private academies (high schools for occupational and college prep) exist in the US, with an enrollment of 263,000 students (O 164).
1857		National Education Association (NEA) founded with 43 members (O 53).
1865-1900		Influence over children and education shifts from parents to government authorities (C 82-83): <ul style="list-style-type: none"> • "Throughout the second half of the nineteenth century, education reformers, bureaucrats and teachers' organizations pushed to increase their powers" over children and public schools. • California State Superintendent of Public Instruction writes, "The child should be taught to consider his instructor, in many respects, superior to the parent in point of authority ... [T]he vulgar impression that parents have a legal right to dictate to teachers is entirely erroneous." • Wisconsin Teachers Association asserts that children are the property of the state.
1852-1913		Compulsory school attendance laws are enacted in all states (C 84).
1900		A majority of children, ages 6 to 13, are now enrolled in government elementary schools. By 1980, 99% of U.S. children attend government schools (O 164).
1909		First public junior high school established in Berkeley, California (O 181).
1916		American Federal of Teachers (AFT) founded (C 53).
1917		NEA's Commission on Reorganization of Secondary Education unveils proposal to restructure high schools (adopted 1918 as Cardinal Principals of Secondary Education.) Students sorted into 4 curricular patterns: <ol style="list-style-type: none"> 1. college preparatory or academic program 2. commercial or business program 3. industrial, vocational, home economics, and agricultural program 4. "a modified academic program for terminal students..." (O 169-170).
1922		Oregon revises compulsory school law to make it illegal for any child between the ages of 8 and 16 to attend a non-government school (C 122).
1925		US Supreme Court limits government's authority: "The fundamental theory of liberty upon which all governments in the Union repose excludes any general power of the State to standardize its children by forcing them to accept instruction from public teachers only. The child is not the mere creature of the State; those who nurture him and direct his destiny have the right, coupled with the high duty to recognize and prepare him for additional obligation" (US Supreme Court, <i>Pierce v. Little Sisters for the Poor</i>)
1944		Congress enacts the G.I. Bill to provide federal funds for college education of veterans at public, private, religious schools.
1950s		Minnesota enacts tax deductions for K-12 expenses. Parental deductions include tuition at private schools, transportation, textbooks, and other supplies.
1954		U.S. Supreme Court, in <i>Brown v. Board of Education of Topeka</i> , outlaws racial segregation in government schools (O 178).
1954		Maine revises its 100-year-old 'tuitioning' system to prohibit the use of public funds at private religious schools (H 72).
1958		Congress passes the National Defense Education Act, which provides federal funds to local public schools for science, math and foreign language instruction, as well as guidance counseling services (O 181).
1962	Vermont Supreme Court rules that the state's 100-year practice of allowing public funds to pay tuition at religious schools now violates the state's Constitution (H 168).	
1965	Congress enacts the Elementary and Secondary Education Act, providing federal funds for local public schools (O 181).	
1980	The U.S. Department of Education is elevated to Cabinet level status.	
Year(s)	Stages of American Education	Events in Education
1642	<p>"PERMISSIVE" ERA</p> <p>During the stage, government permits the organization of public schools subject to the approval of local voters (O 159).</p> <p>Marked by complete parental authority</p>	First education law enacted by Massachusetts General Court requiring parents and guardians of children to "make certain that their charges could read and understand the principles of religion and the laws of the Commonwealth" (O 147)
1749		Benjamin Franklin founds a private academy (a private secondary school) that offers a practical curriculum of a variety of subjects and useful skills. By the mid 1800s, many such private academies exist, offering a wide array of curricula and courses ranging from traditional Latin and Greek to very practical, utilitarian studies (O 154).
1821		First government-owned/operated public high school opens in Boston, MA (O 154).
1826	<p>"ENCOURAGING" ERA</p> <p>During this stage, government explicitly encourages the establishment of school districts and raising of tax revenues to support them. However, government did not require the establishment of schools (O 159).</p> <p>Marked by parental authority; children were not compelled to attend a public school</p>	Massachusetts passes a law requiring every town to choose a school committee, beginning the policy of organizing public schools into a system under a single authority. Connecticut follows suit shortly thereafter (O 159).
1827		Massachusetts enacts law requiring public high schools (O 180).
1830s		Seeking to win public support for government schools, chief advocate Horace Mann, who is appointed Secretary of Massachusetts Board of Education in 1837, assures Protestants that the public schools will regularly use the Protestant Bible (Coulson 81).
1836-1920		More than 120 million copies of McGuffey's readers, which emphasize the ideals of "literacy, hard work, diligence, and virtuous living," are sold (O 163).
Mid-1800s		Despite a trend toward building government-run elementary schools, Maine and Vermont continue their practice of 'tuitioning' students (i.e., allowing districts that decided not to own/operate public schools to use public funds to pay costs of a district student's schooling at a private, parochial, or religious school of parents' choice).
1850		US Census data reveal that only 1 in 10 people identified themselves as illiterate in the 1850 US census – half a century before public school attendance became common (C 84).
1850		The Maine Supreme Court declares it "legal for all students in the government schools to be compelled to read the Protestant Bible" (C 82).
1851	Mistrust of parents spreads in education leadership, as exemplified in this Massachusetts Teacher article: "In too many instances the parents are unfit guardians of their own children ... the children must be gathered up and forced into school" (C 79-80)	

(Gardner)

APPENDIX B

Figure 4-4—Annual Immigration to the United States:
1820-60



SOURCE: Office of Technology Assessment, based on data from U.S. Department of Commerce, Bureau of the Census, *Historical Statistics of the United States* (Washington, DC: 1975), pp. 105-111.

APPENDIX C

Figure 1. A 3rd Grade Standardized Achievement Test Item in Mathematics

Sally had 14 pears. Then she gave away 6. Which of the number sentences below can you use to find out how many pears Sally has left?

- A. $14 + 6 = \underline{\quad}$
- B. $6 + 14 = \underline{\quad}$
- C. $\underline{\quad} - 6 = 14$
- D. $14 - 6 = \underline{\quad}$

(Popham)

APPENDIX D

Figure 2. A 6th Grade Standardized Achievement Test Item in Social Studies

If someone really wants to conserve resources, one good way to do so is to:

- A. leave lights on even if they are not needed.
- B. wash small loads instead of large loads in a clothes-washing machine.
- C. write on both sides of a piece of paper.
- D. place used newspapers in the garbage.

(Popham)

APPENDIX E

Figure 3. A 6th Grade Standardized Achievement Test Item in Science

A plant's fruit always contains seeds. Which of the items below is not a fruit?

- A. orange
- B. pumpkin
- C. apple
- D. celery

(Popham)

APPENDIX F

Table 2

Correlation of SAT Scores and HSGPA with FYGPA by Subgroups (Minimum Sample Size ≥ 15)

Variable	k	n	SAT-CR	SAT-M	SAT-W	SAT	HSGPA	SAT, HSGPA	
Gender	Male	107	69,765	0.44	0.45	0.47	0.50	0.52	0.59
	Female	110	81,551	0.52	0.53	0.54	0.58	0.54	0.65
Race/Ethnicity	American Indian or Alaska Native	16	384	0.41	0.41	0.42	0.54	0.49	0.63
	Asian, Asian American, or Pacific Islander	82	14,109	0.41	0.43	0.44	0.48	0.47	0.56
	Black or African American	83	10,096	0.40	0.40	0.43	0.47	0.44	0.54
	Hispanic, Latino, or Latin American	86	10,486	0.43	0.41	0.46	0.50	0.46	0.57
	No Response	90	6,544	0.46	0.43	0.49	0.53	0.52	0.63
	Other	73	4,175	0.46	0.45	0.48	0.55	0.45	0.61
	White	109	104,017	0.48	0.46	0.51	0.53	0.56	0.63
Best Language	English Only	110	140,559	0.49	0.47	0.52	0.54	0.55	0.63
	English and Another Language	79	7,237	0.41	0.43	0.45	0.50	0.42	0.55
	Another Language	28	1,292	0.28	0.34	0.32	0.42	0.35	0.48
	Not Stated	44	1,171	0.39	0.37	0.45	0.59	0.47	0.69

Note: Pooled within-institution correlations are presented. Correlations are corrected for restriction of range using the 2006 national cohort. Computations were made within institutions for subgroups with at least 15 members. *k* = number of qualifying institutions and *n* = subgroup sample size. SAT is the multiple correlation for all three sections.

(Mattern)

BIBLIOGRAPHY

- Arendt, H. (1954). *The Crisis in Education*.
- Beck, G. (n.d.). *Response to the National Review Online article*. Retrieved from www.glennbeck.com:
<http://www.glennbeck.com/publish/uploads/2013/04/Common-Core-Facts-and-Response-to-NRO.pdf>
- Becker, G. S. (1993, June). The Economic Way of Looking at Behavior. *The Journal of Political Economy*, 101(3), 5.
- Colonial Education - Education for Boys and Girls*. (n.d.). Retrieved from Chesapeake Library:
http://www.chesapeake.edu/Library/EDU_101/eduhist_colonial.asp
- Congress, T. L. (2014). *Rise of Industrial America, 1876-1900*. Retrieved from Immigration to the United States, 1851-1900:
<http://www.loc.gov/teachers/classroommaterials/presentationsandactivities/presentations/timeline/riseind/immgnts/>
- Constitution, U. (n.d.). *Tenth Amendment*.
- Cultural Bias in Standardized Testing*. (n.d.). Retrieved from <http://culturalbiasinstandardizedtesting.weebly.com/what-is-cultural-bias-in-standardized-test.html>
- Curtis, C. (2013, January 30). *How Psychometric Tests are Developing and Why You Can't Do it Yourself*. Retrieved from onetest:
<http://www.onetest.com.au/home/blog/how-psychometric-tests-are-developed>
- Dubow, E. F., Boxer, P., & Huesmann, L. R. (n.d.). Long-term Effects of Parents' Education on Children's Educational and Occupational Success: Mediation by Family Interactions, Child Aggression, and Teenage Aspirations. *55 No. 3*. Retrieved April 2014, from <http://www.rcgd.isr.umich.edu/aggr/articles/Huesmann/2009.DubowBoxerHuesmann.EffectsParentsEd.MerrillPalmQuart.pdf>
- Education, U. D. (2014). *The Federal Role in Education*. Retrieved from ED.gov:
<http://www2.ed.gov/about/overview/fed/role.html?src=ln>
- Education, U. S. (2010, March). *A Blueprint for Reform*. Washington D.C. Retrieved April 2014, from <http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf>
- FCIT. (n.d.). *Classroom Assessment*. Retrieved from Florida Center for Instructional Technology: <http://fcit.usf.edu/assessment/basic/basic.html>
- Freeman, D. J., Kuhs, T. M., Porter, A. C., Schmidt, R. E., & Schwille, W. H. (1983). Do Textbooks and Tests Define a Natural Curriculum in Elementary School Mathematics? *Elementary School Journal*, 83(5) .
- Gardner, H. (1994). Multiple Intelligences: The Theory in Practice. *Teacher's College Record*, 95(4), 576-583.
- Institute, C. B. (n.d.). *A Brief History of Education in America*. Retrieved from CBLPI:
<http://www.cblpi.org/ftp/School%20Choice/EdHistory.pdf>
- Jacobs, L. A. (2004). *Pursuing Equal Opportunities: The Theory and Practice of Egalitarian Justice*. New York: Cambridge University Press.

Measuring Student Achievement

- Jaschik, S. (2010, June 21). *New Evidence of Racial Bias on SAT*. Retrieved from Inside Higher Ed: <http://www.insidehighered.com/news/2010/06/21/sat>
- Jernegan, M. W. (1919, May). The Educational Development of the Southern Colonies. *The School Review*, 27(5), 360-376.
- Jones, L. V., & Thissen, D. (2007). A History and Overview of Psychometrics. *Handbook of Statistics*, 26, 1-27.
- Leicester, U. o. (2010, November 5). *Parents' effort key to child's educational performance*. Retrieved April 2014, from Science Daily: <http://www.sciencedaily.com/releases/2010/10/101029121554.htm>
- LWV. (2011). *The History of Federal Government in Public Education: Where Have We Been and How Did We Get Here?* Retrieved from League of Women Voters: <http://www.lwv.org/content/history-federal-government-public-education-where-have-we-been-and-how-did-we-get-here>
- Marketer, T. E. (2011). *Archives*. Retrieved from Educational Marketer: <http://www.educationalmarketer.net/issuearchive/2011>
- Mattern, D. K., Patterson, F. B., Shaw, J. E., Kobrin, L. J., & Barbuti, M. S. (2008). Differential Validity and Prediction of the SAT. *College Board Research Report No. 2008-4*. New York, New York, USA. Retrieved April 2014, from https://professionals.collegeboard.com/profdownload/Differential_Validity_and_Prediction_of_the_SAT.pdf
- Monahan, T. (1998, December). *The Rise of Standardized Education Testing in the U.S.: A Bibliographic Overview*. Retrieved March 2014, from [torinmonahan.com: http://w.torinmonahan.com/papers/testing.pdf](http://w.torinmonahan.com/papers/testing.pdf)
- Obama, P. B. (2010, February 22). Remarks to the National Governors Association. Washington D.C. Retrieved April 2014, from <http://www.gpo.gov/fdsys/pkg/DCPD-201000114/pdf/DCPD-201000114.pdf>
- PBS. (2008, September 5). *Finance: How do We Fund Our Schools?* Retrieved from PBS: <http://www.pbs.org/wnet/wherewestand/reports/finance/how-do-we-fund-our-schools/197/>
- PBS. (n.d.). *General Article: The Progressive Movement (1900-1918)*. Retrieved from PBS: American Experience: <http://www.pbs.org/wgbh/americanexperience/features/general-article/eleanor-progressive>
- Peterson, R. A. (1983, September 1). *The Freeman*. Retrieved from FEE: Inspire, Educate, Connect: http://www.fee.org/the_freeman/detail/education-in-colonial-america
- PISA 2012 Results*. (2012). Retrieved April 2014, from OECD: <http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm>
- Popham, W. J. (1999, March). *Educational Leadership - Why Standardized Tests Don't Measure Educational Quality*. Retrieved from ASCD: <http://www.ascd.org/publications/educational-leadership/mar99/vol56/num06/Why-Standardized-Tests-Don't-Measure-Educational-Quality.aspx>
- Ravitch, D. (1988). *The Great School Wars*. Baltimore, Maryland: John Hopkins University Press.

Measuring Student Achievement

- Rooks, N. M. (2012, October 11). *Why It's Time to Get Rid of Standardized Tests*. Retrieved from TIME: <http://ideas.time.com/2012/10/11/why-its-time-to-get-rid-of-standardized-tests>
- Strauss, V. (2013, June 18). *The Common Core's Fundamental Trouble*. Retrieved from The Washington Post: <http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/06/18/the-common-cores-fundamental-trouble/>
- Test, F. (2007, August 17). *The Limits of Standardized Tests for Diagnosing and Assisting Student Learning*. Retrieved from Fair Test: The National Center for Fair and Open Testing: <http://fairtest.org/limits-standardized-tests-diagnosing-and-assisting>
- The Declaration of Independence*. (n.d.).
- The National Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform*. United States Department of Education, A Report to the Nation and the Secretary of Education.
- U.S. Congress, O. o. (1992, February). *Testing in American Schools: Asking the Right Questions*. Retrieved from Government Library: <https://www.princeton.edu/~ota/disk1/1992/9236/923606.PDF>
- University, G. (2013, June). *CEW Georgetown*. Retrieved 03 15, 2014, from RECOVERY: Job Growth and Education Requirements Through 2020: <http://cew.georgetown.edu/recovery2020>
- US Legal, I. (2014). *Standardized Test [Education] Law & Legal Defintion*. Retrieved from US Legal Defintions: <http://definitions.uslegal.com/s/standardized-test-education/>
- Vaseleck, J. (1993-1994). *Stop Working and Put Down Your Pencils: The Use and Misuse of Standardized Admission Tests*. Retrieved from Heinonline: (<http://heinonline.org/HOL/LandingPage?handle=hein.journals/jcolunly20&div=28&id=&page>)
- Ward & Trent, e. a. (1907-1921). *The Cambridge History of English and American Literature* (Vol. XIV). New York: G.P. Putnam's Sons; Cambridge, England: University Press.
- West, T. G., & Schambra, W. A. (2007, July 18). *The Progressive Movement and the Transformation of American Politics*. Retrieved from The Heritage Foundation: <http://www.heritage.org/research/reports/2007/07/the-progressive-movement-and-the-transformation-of-american-politics>
- WM.D.Swan. (1851, October). Immigration. *The Massachusetts Teacher*, IV(10), 1-5.

ENDNOTES

- 1 (Ravitch, 1988)
- 2 (Ravitch, 1988)
- 3 (University, 2013)
- 4 (The Declaration of Independence)
- 5 (US Legal, 2014)
- 6 (Popham, 1999)
- 7 (FCIT)
- 8 (FCIT)
- 9 (U.S. Congress, 1992)
- 10 (U.S. Congress, 1992)
- 11 (Test, 2007)
- 12 (Test, 2007)
- 13 (Ravitch, 1988)
- 14 (LWV, 2011)
- 15 (Institute)
- 16 (U.S. Congress, 1992)
- 17 (U.S. Congress, 1992)
- 18 (U.S. Congress, 1992)
- 19 (U.S. Congress, 1992)
- 20 (U.S. Congress, 1992)
- 21 (U.S. Congress, 1992)
- 22 (LWV, 2011)
- 23 (LWV, 2011)
- 24 (Education U. D., 2014)
- 25 (LWV, 2011)
- 26 (Congress, 2014)
- 27 (Congress, 2014)
- 28 (Congress, 2014)
- 29 (Ravitch, 1988)
- 30 (U.S. Congress, 1992)
- 31 (Arendt, 1954)
- 32 (Institute)
- 33 (WM.D.Swan, 1851)
- 34 (Monahan, 1998)
- 35 (Jones & Thissen, 2007)
- 36 (Curtis, 2013)
- 37 (U.S. Congress, 1992)
- 38 (Monahan, 1998)
- 39 (Monahan, 1998)
- 40 (Monahan, 1998)
- 41 (Monahan, 1998)
- 42 (Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008)
- 43 (Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008)
- 44 (U.S. Congress, 1992)
- 45 (U.S. Congress, 1992)
- 46 (U.S. Congress, 1992)
- 47 (U.S. Congress, 1992)
- 48 (Popham, 1999)
- 49 (Popham, 1999)
- 50 (Popham, 1999)
- 51 (Popham, 1999)

Measuring Student Achievement

- 52 (PBS)
- 53 (Marketer, 2011)
- 54 (PBS)
- 55 (Freeman, Kuhs, Porter, Schmidt, & Schwille, 1983)
- 56 (Popham, 1999)
- 57 (Freeman, Kuhs, Porter, Schmidt, & Schwille, 1983)
- 58 (Freeman, Kuhs, Porter, Schmidt, & Schwille, 1983)
- 59 (Popham, 1999)
- 60 (Freeman, Kuhs, Porter, Schmidt, & Schwille, 1983)
- 61 (Popham, 1999)
- 62 (Popham, 1999)
- 63 (Popham, 1999)
- 64 (Popham, 1999)
- 65 (Popham, 1999)
- 66 (Popham, 1999)
- 67 (Gardner, 1994)
- 68 (Popham, 1999)
- 69 (Popham, 1999)
- 70 (Rooks, 2012)
- 71 (Vaseleck, 1993-1994)
- 72 (Jacobs, 2004)
- 73 (Jacobs, 2004)
- 74 (Jacobs, 2004)
- 75 (Becker, 1993)
- 76 (Jacobs, 2004)
- 77 (Jacobs, 2004)
- 78 (Jacobs, 2004)
- 79 (Cultural Bias in Standardized Testing)
- 80 (Jaschik, 2010)
- 81 (Jaschik, 2010)
- 82 (Jaschik, 2010)
- 83 (Jaschik, 2010)
- 84 (PISA 2012 Results, 2012)
- 85 (Dubow, Boxer, & Huesmann)
- 86 (Leicester, 2010)