Public Forum Position Paper

Standardized Testing
December 2015
If you enjoyed this fine book from PARADIGM RESEARCH, you'll be glad to know that we offer a complete menu of affirmatives, disadvantages, counterplans, kritiks, LD & Public Forum research, and much more.

Shop our online store anytime - www.oneparadigm.com

Catalogue is available on website

The Paradigm NSDA Public Forum Position Paper
December 2015

Copyright © 2015 by Paradigm Research, Inc. All rights reserved.

First Edition Printed In The United States Of America

For information on Paradigm Debate Products:

PARADIGM RESEARCH
P.O. Box 14924
Minneapolis, MN 55414
Phone: 651.307.6980
Toll Free / Fax: 844.277.2536
Web: www.oneparadigm.com
E-mail: service@oneparadigm.com

All rights are reserved. This book, or parts thereof, may not be reproduced by any means - graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems - without the written permission of the publisher. Making copies of this book, or any portion, is a violation of United States and international copyright laws.
Standardized Testing: Index

Pro: Testing Desirable
Testing Desirable: Topshelf ................................................................. 9
Testing Desirable: Accountability ......................................................... 13
Testing Desirable: Data—General ......................................................... 15
Testing Desirable: Data—Accountability ............................................. 16
Testing Desirable: Data—Annual Testing Key .................................... 17
Testing Desirable: Data—Comparisons ............................................. 18
Testing Desirable: Data—School Improvement ................................... 19
Testing Desirable: General ................................................................. 20
Testing Desirable: Knowledge Confidence ........................................ 21
Testing Desirable: Learning Methods ................................................. 23
Testing Desirable: Memory—General ................................................. 25
Testing Desirable: Memory—Forgetting ............................................ 28
Testing Desirable: Multiwarrant ......................................................... 29
Testing Desirable: SAT ................................................................. 30
Testing Desirable: School Reform ...................................................... 32
Testing Desirable: Student Success .................................................... 33
Testing Desirable: Teacher Quality / Evaluation .................................. 35
Testing Desirable: Answers to “Budget Tradeoffs’ ................................ 37
Testing Desirable: Answers to “Cheating Concerns” ............................ 38
Testing Desirable: Answers to “Creativity / Critical Thinking” ............. 39
Testing Desirable: Answers to “Curriculum Narrowing” ....................... 40
Testing Desirable: Answers to “Disparate Outcomes” ......................... 41
Testing Desirable: Answers to “NCLB Fails” ....................................... 42
Testing Desirable: Answers to “Results Unreliable” ............................ 43
Testing Desirable: Answers to “Student Stress” ................................... 44
Testing Desirable: Answers to “Teach to the Test” ............................. 45
Testing Desirable: Answers to “Teacher Opposition” ........................... 46
Testing Desirable: Answers to “Teachers Adequately Evaluated Now” .... 47

Con: Testing Undesirable
Testing Undesirable: Topshelf ............................................................. 48
Testing Undesirable: Bias—General ..................................................... 51
Testing Undesirable: Bias—Class ......................................................... 53
Testing Undesirable: Bias—Cultural ..................................................... 54
Testing Undesirable: Bias—Race ......................................................... 56
Testing Undesirable: Cheating Incentives .......................................... 57
Testing Undesirable: Commodification ............................................... 59
Testing Undesirable: Corporatization / Neoliberalism ........................... 60
Testing Undesirable: Counterproductive ........................................... 62
Testing Undesirable: Creativity / Critical Thinking ............................... 63
Testing Undesirable: Disparate Impact ............................................... 65
Testing Undesirable: Finland Proves .................................................... 66
Testing Undesirable: Gaming ............................................................ 67
Testing Undesirable: General ............................................................ 68
Testing Undesirable: Gifted Children .................................................... 71
Testing Undesirable: Intelligence Limits ............................................. 72
Testing Undesirable: Invalid Results ............................................... 73
Testing Undesirable: Multiwarrant ..................................................... 74
Testing Undesirable: Narrowed Curriculum ....................................... 76
Testing Undesirable: Other Mechanisms Superior ................................ 77
Testing Undesirable: Reading Tests ................................................... 79
Testing Undesirable: Resource Tradeoffs .......................................... 80
Testing Undesirable: SAT—General .................................................... 82
Testing Undesirable: SAT—Specific Mechanisms ................................ 84
Testing Undesirable: SAT—Wealth ..................................................... 87
Testing Undesirable: Standards Lowering ........................................... 89
Testing Undesirable: Student Motivation ............................................ 90
Testing Undesirable: Student Stress ................................................... 91
Testing Undesirable: Teacher Evaluations ......................................... 92
Testing Undesirable: Teacher Retention ............................................ 93
Testing Undesirable: Teaching to the Test ......................................... 94
Testing Undesirable: Time Tradeoff .................................................. 96
Testing Undesirable: Answers to “Academic Achievement” .................. 99
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing Undesirable: Answers to “Accountability”</td>
<td>100</td>
</tr>
<tr>
<td>Testing Undesirable: Answers to “Data Valuable”</td>
<td>101</td>
</tr>
<tr>
<td>Testing Undesirable: Answers to “Information Retention / Memory”</td>
<td>104</td>
</tr>
<tr>
<td>Testing Undesirable: Answers to “Modeling / Spillover”</td>
<td>105</td>
</tr>
<tr>
<td>Testing Undesirable: Answers to “Score Increases”</td>
<td>106</td>
</tr>
<tr>
<td>Testing Undesirable: Answers to ‘Spurs Reform’</td>
<td>108</td>
</tr>
</tbody>
</table>
Standardized Testing: Introduction

Resolved: On balance, standardized testing is beneficial to K-12 education in the United States.

Taken straight from the headlines, this month’s topic asks students to assess the merits of standardized testing at the element and secondary level. The topic is highly salient, given both the Obama administration’s recent directive asking schools to cap the amount of time devoted to assessment to two percent of instruction time and debates in congress about whether to reform federal education policy, which is heavily-focused on utilizing test data in school reform efforts.

The Obama administration’s announcement is only the most recent development in a growing political debate the desirability of standardized testing and other accountability measures. In fact, many observers believe that political and public momentum has shifted against the strategies deployed by the No Child Left Behind (NCLB) legislation:

That said, this is a perilous moment for reauthorization because of the backlash against standards, testing, and accountability. The effort to put “the standardized testing machine in reverse,” in the words of New York mayor Bill de Blasio, has diverse bastions of support. These include; conservatives who object to the seemingly ever expanding reach of the federal government into K-12 public education; concerned parents of children in well-regarded, often suburban schools, who believe that test-prep activities have narrowed the curriculum and put undesirable pressure on their children; progressives such as de Blasio, who see the challenges of public education as best addressed by more funding for schools and broad efforts to eliminate poverty rather than by holding schools or teachers accountable for results; and, teacher unions that are doing what unions are expected to do by trying to protect the less effective of their members from the consequences that follow from exposing their ineptitude in the classroom. [Grover J. Whitehurst, Martin R. West, Matthew M. Chingos and Mark Dynarski, “The Case for Annual Testing,” BROWN CENTER CHALKBOARD n. 93, 1—8—15, www.brookings.edu/research/papers/2015/01/08-chalkboard-annual-testing, accessed 11-8-15]

Students debating the topic will benefit from understanding the history and motives of standardized testing initiatives in the United States. The following quotation provides a useful synopsis of how standardized testing has come to dominate education policymaking in our country:

Were the mounting suspicions a result of failing schools? Or, were there other factors at work? Standardized achievement tests had their beginning not in public schools, but within a branch of the military during World War I (Popham, 2001). The war brought about a great need for officers. Consequently, army officials asked the American Psychological Association to develop a group-administered test that would help identify the recruits most likely to succeed in the Army’s officer training program. Around that time, the use of the Stanford-Binet tests, which produced what became known as IQ scores, became widespread. In 1917, a committee convened and developed 10 different subtests that were designed to “discriminate among test-takers with respect to their intellectual abilities” (Popham, 2001, p. 42). These were known as the Army Alpha Tests and were given to a norm group that would be used as a comparison mark for more than a million men getting ready for combat. Those ranked high would be selected for officer training, while those ranked lower would be relegated to the battlefield. Within a short time following World War I, new educational tests were copyrighted that mimicked the Army Alpha in its measurement strategy. Another factor contributing to the spread of standardized testing was the far-reaching Elementary and Secondary Education Act (ESEA). In 1965, President Lyndon Johnson, concerned with the growing numbers of children living in disadvantaged homes, issued ESEA as he launched his war on poverty. For the first time, large amounts of federal money were awarded to states in order to help them bolster children’s learning. These funds were specifically designed to offer assistance to schools that served large numbers of socially disadvantaged children. An addendum to the ESEA, offered by Senator Robert Kennedy, required states that received federal funds to evaluate and report on the effectiveness of their programs, namely via standardized achievement tests. These tests were based on the Army Alpha to discriminate among test takers. The tests available at the time, however, included the Metropolitan Achievement Tests and the Comprehensive Tests of Basic Skills, and bore no direct relationship to the skills and knowledge being promoted by any particular ESEA program. And yet, the government spent thousands of dollars in encouraging their use. The notion that a standardized achievement test could evaluate the success of various
We have thus picked a topic that is timely in a political sense and is highly relevant to student debaters.

Although opposition to NCLB-testing could not do with additional tests and accountability systems, A Nation At Risk would attempt to do. The report made recommendations in areas of content, standards, expectations, time, teaching, leadership, and fiscal support, and standardized tests became a pivotal part of evaluating the quality of education within each school in the country. By the late 1980s, most states required some type of mandatory testing; by 1991, students who completed high school took, on average, 18-21 standardized tests in their career, with the majority of them occurring in the K-5 years (Perrone, 1991, p. 133). In 1994, President Clinton issued his landmark education package, Goals 2000: Educate America Act (P.L. 103-227). This act provided resources to states and communities to ensure that “all students reach their full potential.” It established a framework by which to identify “world-class academic standards, to measure student progress, and provide the support that students may need to meet the standards.” Central to this act was a National Standards Board and a call for voluntary testing in grades 4, 8, and 12 to ensure that standards were being met. In 2002, the U.S. Congress signed into law President George W. Bush’s No Child Left Behind (NCLB) initiative, which has been the most far-reaching education act since the War on Poverty in 1965 (No Child Left Behind Act, 2001). Testing children in 4th, 8th, and 12th grades is now mandatory. Accountability systems that require assessments to prove children’s growth in academic subjects are mandatory. Tests are not simply what teachers give at the end of the year. They are now attached to high stakes, such as grade retention, admittance into special programs, graduation, admission into college, and whether or not schools remain open and teachers get to keep their jobs. Today, because of NCLB, all 50 states have some form of standardized testing whereby students are tested every year, beginning in the 3rd grade. In many states, 1st- and 2nd-graders are also tested. And, in some states, kindergartners are tested regularly as well. Large numbers of children are given standardized tests in two three-hour increments within a one- to two-week period each spring. The purpose of today’s standardized achievement tests remains much the same as it was with the Army Alpha (Popham, 2001). The test-takers’ scores are compared to a pre-determined norm group to discriminate among them and determine rank. Today, it continues to be the mission of a standardized test-maker to develop a set of items that allows for making accurate comparisons among test-takers and then rank-ordering those who take the test. Standardized testing, as it gets more all-encompassing, has become a nightmare of huge proportion in the United States. As Alfie Kohn (2000) states, “Standardized testing has swelled and mutated, like a creature in one of those old horror movies, to the point that it now threatens to swallow our schools whole” (p. 1).

Although opposition to NCLB-style testing is growing, the new set of Common Core standards will necessitate substantial testing.

The Common Core State Standards represent the most significant change to the standards-based assessment movement since the No Child Left Behind (NCLB) Act required all states to adopt standards and administer standardized tests over a decade ago. But unlike the federal mandates of NCLB, Common Core is a voluntary, state-led effort—albeit one that has received support from the federal government. The new standards in mathematics and English language arts (ELA) have been adopted by 45 states and the District of Columbia, representing 85 percent of American students. Eight states claim to have already fully implemented Common Core, 20 are in the midst of implementation this year (2013-14), and most of the remaining states expect to implement the standards next year (2014-15). [Matthew M. Chingos, STANDARDIZED TESTS AND THE COMMON CORE STANDARDS: YOU GET WHAT YOU PAY FOR? Brown Center on Education Policy, Brookings Institution, 10—13, p. 3]
Most of the terms in the resolution should be relatively non-controversial. “On balance” requires each side to establish an evaluative criteria to determine whether testing has improved education in the United States. Many teams will likely default to some kind of cost-benefit analysis (do the benefits outweigh the drawbacks) in order to determine whether such exams are net good or net bad for the country. Two unanswered question raised by the resolution are “for whom?” and “for what purpose?”, which could lead to some very interesting debates about the purpose and focus of our education policy “United States” should be uncontroversial. Standardized testing, defined in the first quote, refers to exams designed to assess student learning and achievement. Although the resolution does not restrict debates to discussion of tests taken by K-12 students per se, there is little reason to believe that exams outside those taken by elementary and secondary students are relevant to the topic. One interesting question is whether our assessment of the topic should include a discussion of the SAT, ACT, and Advanced Placement (AP) exams. Although these tests are utilized primarily for college admissions and/or to earn college credit, their importance does exert substantial influence over the curriculum in many high schools.

The case for testing is built on the argument that the data derived from tests are vital in measuring, and improving, academic outcomes in our nation’s schools. Supporters of the current testing regime maintain that testing data enables a number of accountability schemes, including those included in No Child Left Behind and being expanded through various value-added measurement programs. Such assessments have been vital in identifying best education practices, identifying shortcomings in our current systems, pressuring failing schools and districts to improve their performance, and providing parents and students with the information necessary to make informed decisions. Annual testing is particularly important because it enables comparisons of school performance, both across time and between school districts. Testing data allows school leaders, teachers, parents, the public, and policymakers assess the effectiveness of current instructional techniques. Schools that meet high standards and continue to improve their performance can continue to refine their current instructional philosophy, while schools that fail to meet certain benchmarks are forced to change their instructional focus and strategies. Supporters of the ‘testing and accountability’ approach offer two primary sets of arguments to justify testing. First, they point to increases in test scores nationally, both for all students and among children of color and poor children, arguing that demonstrable metrics prove that testing-based accountability can close the achievement gap while raising the average for all students. Second, testing supporters point to the soundness of the logic behind the ‘accountability = reform’ argument, highlighting critical reforms and achievement gains in former ‘failing’ schools. Testing proponents also highlight evidence that achievement is increasing among various student populations. Evidence explaining these arguments is available on the “Testing Desirable: Topshelf” section. We have also included evidence answering most of the likely objections to testing. One thing you should know when preparing your defense of the resolution is that it seems to be much easier finding strongly-worded evidence opposed to testing—most teachers, administrators, and education organizations have significant concerns about testing, so you should prepare yourself for a wide array of objections.

Students should have a much easier time crafting arguments for con positions, with a number of pretty compelling arguments in the literature of school testing. First, the test results are quite readily attacked on a number of levels, undermining any case for the use of standardized testing in measuring school performance or guiding education policy. Not only is there very strong evidence indicating that many schools have manipulated testing data (so-called ‘cheating’, including recent scandals in Atlanta and Washington DC), but there is some solid data showing that tests scores have declined for some groups of children. States have even responded to low test scores by lowering the standards for ‘acceptable’ levels of academic achievement, confounding much of the rationale offered for standardized testing in the first place. Second, there are many critics who persuasively argue that the testing instruments used in many states do not provide a meaningful measure of academic progress. The reliability of the test results is also increasingly threatened by the growing number of students who are “opting out” of the exams, meaning that the tests are not assessing the entirety of the student population. Third, a growing chorus of critics maintain that standardized testing has significant negative effects on poor students and students of color, the very groups that many testing proponents claim are supposed to be helped by the accountability measures allegedly enabled by standardized testing. Some testing detractors have even argued that most standardized tests are simply “affluence tests,” since the results very closely match the socioeconomic
status of the students being examined. Our testing focus also potentially short-changes our more gifted students, whose needs are ignored as schools focus their resources and class time on improving the test scores of their lower- (but not lowest-) performing students. In fact, testing-enabled accountability measures encourage schools to focus on students who are close enough to the “acceptable” line that they can be nudged enough to pass the exam, meaning schools have strong incentives to ignore students with serious learning deficits. Some schools have even been accused of coercing low-achievers to drop out of school in an effort to bolster flagging test scores. Fourth, testing standards and accountability measures have driven many schools and instructors to ‘teach to the test,’ focusing on rote memorization of a relatively narrow range of subjects at the expense of meaningful learning in the arts, social studies, and other key parts of a comprehensive education. The instructional strategies themselves may hinder the development of vital critical thinking skills. Fifth, testing measures so much pressure on teachers, and there is such a strong stigma from being associated with a ‘failing’ school, that many people are choosing to avoid the teaching profession, or are leaving the ranks of teachers. This effect works directly counter the states goal of many accountability measures, namely ensuring that teachers are highly qualified. Finally, there is strong evidence that the tests used by most states are biased against some groups of children, and that a testing focus is a poor measure of academic achievement. Any of these claims would make a pretty strong closing argument, and most of them outline in the evidence on the “Testing Undesirable: Topshelf” block. The extensions include supporting evidence for these claims, a number of other objections to the current use of standardized testing, and responses to the most likely pro arguments about the desirability of standardized testing. Although the following passage is meant to be humorous, it largely summarizes many of the objections to our current test-focused education policy:

The Onion, a satirical newspaper, recently took their shot at high-stakes standardized testing in an article listing what the editors saw as testing’s "pros and cons." Their number one "pro" testing argument is that "Every student" is "measured against same narrow, irrelevant set of standards." Other "pro testing arguments" were that the high-stakes tests hold "teachers personally accountable for success of large, monolithic testing organizations;" "Western tradition of critical thinking" is "best embodied in bubble-sheet format;" and "Repeated testing carefully develops teachers' cheating skills." There major "con" argument is that "There are easier ways to measure parents' income." They probably should have added that other forms of "cruel and unusual punishment" are outlawed by the United States Constitution. [Alan Singer, social studies instructor, Hofstra University, “The Fallacy Behind High-Stakes Testing,” HUFFINGTON POST, 5—26—15, http://www.huffingtonpost.com/alan-singer/the-fallacy-behind-highst_b_7441676.html, accessed 11-6-15]

Best of luck!
Testing Desirable: Topshelf

1. Standardized tests are justified, even with their limitations—there is no justification for restricting them


The idea of less testing with the same benefits is alluring. Yet in practice it would actually roll back progress for America’s students. Yes, test quality must be better than it is today. And, yes, teachers and parents have a right to be alarmed when unnecessary tests designed only for school benchmarking or teacher evaluations cut into instructional time. But annual testing has tremendous value. It lets schools follow students’ progress closely, and it allows for measurement of how much students learn and grow over time, not just where they are in a single moment. It also allows for a much more nuanced look at student performance. For example, rather than simply looking at average overall school performance, where high performers frequently mask what’s happening to low achievers, No Child Left Behind focuses attention on the progress that groups of students are making within schools — a level of analysis that is possible only with annual data. To be confident that the test results aren’t pulled up or down by a few students and to minimize year-to-year variability, states usually consider only groups of at least 30 or 40 students. States are also able to average results over multiple years or across grades. So, for instance, a school with 10 Hispanic students in each of grades three, four, five and six adds up to a group of 40, providing some accountability for what is happening to those students. The grade-span approach would eviscerate the ability to look at particular groups of students within schools. Instead of having multiple grades over which schools could compile results, each school would be held responsible only for the performance of students in a single grade. Not only would this lower the quality of the data, but it would also raise the stakes of the tests: If you think the stakes are too high now, imagine being a fifth grader in a school where your score determines the results of the entire school. Worst of all, under this approach, far fewer schools would be on the hook for paying attention to historically disadvantaged groups of students. A school with 10 Hispanic students in each grade would no longer be held accountable for whether those students were making sufficient progress, because the 10 fifth graders wouldn’t be enough to count as a meaningful population size. To get a sense of how many students could become newly “invisible,” consider public elementary schools in Washington, D.C. Applying the same minimum group size currently used for entire schools to the fifth grade only, about half of the city’s 119 elementary schools with fifth graders taking math tests would not be held accountable for the progress of low-income or African-American students, because there aren’t enough of them in that grade to constitute a reliable sample size. For that same reason, less than 10 percent of schools would be responsible for Hispanic students or English language learners, and not a single elementary school would be accountable for the progress of students with disabilities. Coincidentally, the push for limiting testing has sprung up just as we’re on the cusp of having new, better tests. The Obama administration has invested $360 million and more than four years in the development of new tests, which will debut this spring. Private testing companies have responded with new offerings as well. Today’s eagerness to jettison our commitment to leave “no child behind” is a shame, not just because better tests are on the horizon, but also because it worked. Fourth and eighth grade achievement scores of black, Hispanic and low-income students have never been higher. High school graduation rates are at an all-time high. And researchers repeatedly link No Child Left Behind’s emphasis on traditionally underperforming groups to real improvements in schools around the country. The conversations that No Child Left Behind sparked are not easy, but they are essential. That’s why it’s exactly the wrong time to accept political solutions leaving too many of our most vulnerable children hidden from view.
Testing Desirable: Topshelf [cont’d]

2. A move away from annual testing will make it harder to evaluate school performance


Second, and perhaps not part of the implicit motive for the “test-once” proposal, the reliability of information about school performance is weakened, particularly for the “vulnerable groups” about which Darling-Hammond and her co-authors want scores to continue to be reported . What would happen to estimates of student achievement at the school level if we switched from testing three grades to only testing one grade (e.g. from testing in grades 3-5 to only testing in grade 4)? We examine this question with data from the 2009, 2010, and 2011 school years in 62 elementary schools in a large urban school district with which we are conducting collaborative research. We find that the estimates of school performance become less reliable when using test scores from only one grade instead of three. The hit on the reliability of year-to-year estimates in school performance is only moderate when considering estimates for all students in the school but becomes unacceptably large when reporting for separate subgroups. These findings are illustrated in the following figure by plots of the variance accounted for in school-level estimates of student achievement from one year to the next (e.g., 2009 to 2010) for all students vs. Hispanic students using data for grades 3-5 vs. only grade 4.[v] The question is: how much of the spread of school-level achievement test scores in a particular year can be predicted from the scores of those same schools in the previous year? We don’t expect schools to remain at the same place in the distribution of test scores from one year to the next because some schools get better and some don’t. But there is a lot of evidence, including that presented in the figure below, that the year-to-year performance of schools relative to other schools within the same district is relatively stable. If we have a measure of school performance that bounces around a lot from one year to the next over a large number of schools, that measure is likely sending a very noisy, unreliable signal of actual school performance. [figure omitted] The figure above shows clearly that once we start using only a single grade to calculate the performance of a school for subgroups of students, the drop off in the reliability of the estimates of school performance is so large as to make such estimates practically meaningless. For instance, only about 10% of the differences among schools in a given year in the math performance of Hispanic students can be predicted from the school-level scores of 4th grade Hispanics from the previous year. In contrast, over 45% of the differences among schools in the performance of Hispanic students from one year to the next can be accounted for when the results from testing in three grades are used.

3. Testing is vital to accountability—this evidence also answers the most common objections

Grover J. Whitehurst, Martin R. West, Matthew M. Chingos and Mark Dynarski, “The Case for Annual Testing,” BROWN CENTER CHALKBOARD n. 93, 1—8—15, www.brookings.edu/research/papers/2015/01/08-chalkboard-annual-testing, accessed 11-8-15. Test scores matter for any form of accountability, including market-based accountability in which parents choose the schools their children attend and funding follows students to their school of choice. Proponents of charter schools and open-enrollment in public schools will find that the informational fuel of their favored version of school reform will evaporate without valid information on annual student gains. The removal of the requirement of annual testing will, necessarily, all but eliminate school-based accountability for the learning of subgroups of children because, as Whitehurst and Lindquist have shown, testing only samples of children or only one grade of children often leads to sample sizes for subgroups such as English learners and blacks that are too small to generate reliable estimates for the school as a whole. Thus, those concerned with equity should strongly support annual testing in multiple grades. Unless individual students are tested in adjacent grades, as they move through school, which requires annual testing, it is impossible to measure gains in student achievement from one year to the next. This has three consequences. First, schools that serve a disproportionate share of disadvantaged children won’t be credited for their success in improving the academic abilities of their students because improvement won’t be visible, only status. Thus, if children are only tested in 6th grade, the elementary school that moves its students from the 10th percentile in math to the 40th percentile from 3rd to 6th grade will look exactly the same as the school whose students performed at the 50th percentile in grade three and fell to the 40th percentile in grade six. For the same reasons, it will be impossible to differentiate teachers based on their ability to generate gains in student learning in their classrooms. Such value-added measures require a test-based estimate of the difference between how much math or language skills individual students have when entering vs. graduating from a teacher’s classroom. This requires annual testing. We have learned over the last decade just how important differences in teacher effectiveness are to student outcomes. The ability to collect and use this information to support improvements in teacher preparation, professional development, and personnel actions will be lost without annual testing. Finally, eliminating annual testing would prevent researchers and policymakers from judging the effectiveness of new education programs in which the research design depends on knowledge of students' recent achievement. By hampering our ability to learn about what’s working and what’s not, jettisoning annual testing would have a negative effect on the rate of improvement in achievement over time.
4. **Testing helps us assess whether students are really learning what they need to know**


In the field of education, knowledge is power. A recent New York Times article outlined potential new legislation in Texas, House Bill 866, which would allow elementary school students who excel at state reading and math exams in the third and fourth grade, to skip exams in those subjects in the fourth, sixth, and seventh grades. As an educator and researcher, with a doctorate in education and a specialty in curriculum and instruction, as well as a Ph.D. in psychology, I have spent the last two decades studying the way children learn. Therefore, I felt it necessary to respond to the New York Times article, and speak to the value of standardized testing. We must not fear that which can offer us the best possible opportunity to transfer information in the most effective way. One important measure for that transfer is the standardized test. Such testing gives the teacher important diagnostic information about what each child is learning in relation to what he has been taught. Only in this way can the teacher know if the student needs intervention and remediation; if the curriculum matches the course requirements; or if the teaching methods needed are in some way lacking and require adjustment.

5. **Testing provides many benefits—we need tests to ensure that we have quality schools**


Furthermore, the standardized test gives valuable insight into broader issues, such as the standard curriculum important to grade level requirements, and an education reference point for fair and equitable education for all children in all schools – district by district and state by state. This can also lead to better teaching skills, as teachers will be held accountable to help their students meet these standards. Moreover, student growth can be a very significant outcome of standardized tests, for though a child may return a low score he may show a growth pattern that is positive. These tests are but one tool that a teacher uses to diagnose her students’ teaching needs, so that an individual and child-centered curriculum can be developed. The standardized test is an objective and critical measure of achievement in skills, knowledge, and abilities, and must pass the criteria of measurement validity, reliability, and bias, as well as an awareness of the test’s potential limitations in scoring.

6. **Tests are good—they help motivate teachers and students**


The chief problem with U.S. schools apparently isn’t high dropout rates or underqualified teachers but standardized testing. This is the only conclusion that can be drawn from the push by parents and teachers in Buffalo, Philadelphia, Seattle and elsewhere to help students opt out of taking standardized tests. Members of this burgeoning anti-test movement fail to grasp testing’s valuable role in motivating and guiding students and teachers. Preparing young Americans for success in the global economy will require our schools to improve, not abolish, academic standards.

7. **Federal-based accountability measures improve academic achievement**


Federal school-based accountability is different in that the evidence demonstrates that it has had meaningful albeit relatively modest impacts on student learning, concentrated in mathematics in the worst schools and for the lowest performing students. Roughly half the states had consequential accountability systems in place prior to NCLB, and all do today. The function of NCLB’s accountability mandates can now be carried out by states through their own systems, perhaps supplemented by a limited federal focus on schools that fail at basic functions, e.g., schools in which significant percentages of students do not acquire basic competencies in reading and math in elementary and middle school, or do not graduate from high school. States could comply with such a federal requirement by identifying a basic competency cutoff score on their state test. Such a cutoff would be lower than current proficiency targets and realistically obtainable for nearly all students. This would allow schools in which nearly all children will acquire basic competencies as a matter of course to focus their instructional efforts where they wish without fear of running afoul of the federal accountability system.
Testing Desirable: Topshelf [cont’d]

8. Federal action is vital—information about school quality is a public good, and if the government does not provide the information, it will not be available


Information on school performance in education is a public good, meaning that individuals cannot be effectively excluded from using the information once it exists. Because it is impossible to prevent consumers who have not paid for the information from consuming it, far too little evidence will be produced if it is not required by the federal government. Further, only local authorities can collect information on school performance from test scores and other local data, but their narrow self-interests are not usually served by making that information easily accessible and useable by the public. Only federal requirements will achieve that end. Finally, evidence on school performance does not merely need to be produced; it needs to be of high quality. But gathering and auditing data are almost pure public services. That is why even when information on school or company performance is treated as a private good to support more informed consumer choice (e.g., college search sites that require a fee for access, or stock market services that sell advice on individual stocks), the information that customers pay to access is derived overwhelmingly from federal sources. In short, federal support for gathering and disseminating information on school performance is easy to justify. If the federal government doesn’t support it, it will not happen.
Testing Desirable: Accountability

1. Strong accountability systems are vital to improving school performance

Eileen Reed, Janie Scull, Gerilyn Slicker, and Amber M. Winkler, DEFINING STRONG STATE ACCOUNTABILITY SYSTEMS: HOW CAN BETTER STANDARDS GAIN GREATER TRACTION? A FIRST LOOK, Thomas B. Fordham Institute, 4—12, p. 8-9. 4) A system of rewards and consequences to drive improvement at the school and district levels Strong state accountability systems engage the multiple entities involved in educating children, including districts, schools, educators, and students, and incorporate both carrots and sticks to spur positive action and behavior. At the district and school level, incentives may include “honor roll” or similar “blue ribbon/distinguished” status, financial rewards, increased autonomy in operations/spending, and other regulatory relief (e.g., automatic renewal of district accreditation) in exchange for meeting heightened district accountability provisions and/or as a reward for performance or growth. State accountability systems should also require targeted interventions into low-performing schools. These may include replacement of staff, charter conversion, state takeover, contracts with outside entities to operate schools, and automatic school closure after consecutive years of failure. (Interventions may also be catalyzed via “parent trigger” provisions.) Such consequences need to be real (or have “teeth”), so the state’s system should avoid introducing loopholes by which low-performing schools or districts can evade tough sanctions; water down or delay sanctions; and/or restrict sanctions to a subset of the lowest-performing schools.

2. Accountability, through testing, is vital to push poor schools to improve

Marcus A. Winters, staff, Manhattan Institute, “Weakening Accountability in No Child Left Behind Rewrite Would Be a Mistake,” REAL CLEAR EDUCATION, 7—16—15, www.manhattan-institute.org/html/weakening-accountability-no-child-left-behind-rewrite-would-be-mistake-6308.html, accessed 11-5-15. Congress is still debating the details, but it seems likely that there will be some major changes to the federal education law known as No Child Left Behind. Some modifications are warranted. But it would be a mistake to substantially weaken the educational accountability established by NCLB. Doing so would hurt the neediest kids most. The fundamental principles of accountability -- that student achievement is worth measuring and that failing schools should be sanctioned – are as important today as ever. There are many great public schools. But some schools continue to systematically fail their students. And the bad schools aren't distributed evenly. Kids in neighborhoods with lower average income and heavy minority populations are more likely to end up in an ineffective school than are wealthier, whiter students whose parents can afford private school tuition or can move to more attractive areas. Without the threat of accountability, there is little reason to suspect that such schools will pursue meaningful changes.

3. Strong accountability measures are critical—we should resist efforts to undermine them

Marcus A. Winters, staff, Manhattan Institute, “Weakening Accountability in No Child Left Behind Rewrite Would Be a Mistake,” REAL CLEAR EDUCATION, 7—16—15, www.manhattan-institute.org/html/weakening-accountability-no-child-left-behind-rewrite-would-be-mistake-6308.html, accessed 11-5-15. In a 2012 paper, my co-author Joshua Cowen and I found evidence that New York City schools that were issued a failing grade under the district's accountability system responded by making substantial improvements the following year. Our results were similar to another study of New York City's school grading policy by Columbia University economists Jonah Rockoff and Leslie Turner in 2010. But New York City's experience demonstrates how the anti-accountability sentiment is threatening even the most effective localized systems. Despite evidence that the policy has been effective, one of the first things that the de Blasio administration did upon taking office was to scrap the city's school grading system. The city now has a much weaker accountability policy that focuses on providing detailed information about several aspects of the school rather than identifying and sanctioning those that are low performing. There are those on both sides of the political aisle that are frustrated with federal accountability through NCLB. But it is worth remembering the reason this bipartisan law came about in the first place: Strong accountability policies offer important leverage over schools that aren't adequately serving their students. The research suggests that today's schools are better off for expansive accountability. We should work to make accountability systems even better, not find ways to derail them.
4. Testing can impose needed accountability—demonstrated by teacher focus group data

Theme #5: Accountability is useful. This final theme was the most surprising because many criticisms of testing are related to the accountability that comes with the exams. That is, while many educators welcome the data and information that standardized assessments provide, few will claim an affinity for the consequences that are sometimes attached to student performance on these tests. Nevertheless, teachers in our focus groups even made positive comments that the testing programs encourage teachers to feel an even greater responsibility that students learn the material. One teacher noted that tests "hold accountable" those teachers who "are just there to get summers off and an 8-to-3 job." From another school, a teacher said: "We all need accountability. . . . [It] helps us be better teachers and not just take a day off because we don't feel like it and let [the students] watch a movie." At a third school, a teacher said, "I absolutely believe in accountability and having [the students] know those benchmarks before they go on to another grade.” And from a fourth school, a teacher pointed out that "the only way someone might say [that testing makes teaching] less desirable is because they're [other teachers] going to work harder.” As a specific example of how testing encourages teachers to be accountable for ensuring that each student makes progress, one teacher noted that "I've totally changed my philosophy of teaching. I used to stand up there and teach the best of my knowledge, and if they didn't get it, 'Sorry, we've got to move on.' But now it's, 'OK, these kids didn't learn it; what's the plan?'" These teachers used to hold themselves accountable for delivering good lessons; now, thanks to the testing regime, the teachers can go one step further by assessing whether the lessons translated into improved student learning.

5. Testing information is vital to accountability, and thus education effectiveness

Stijepko Tokic, “Examining the Big Picture Regarding the Importance of the No Child Left Behind Act: Is It Worth Giving a Chance?” THURGOOD MARSHALL LAW REVIEW v. 32, Spring 2007, p. 322-323.
B. Heavy Emphasis on High-Stakes Standardized Testing is Essential because it Provides Control and Guidance for Authorities, While it Does Not Have a Negative Impact on Student Performance 1. Control Educators most commonly use achievement and aptitude tests to evaluate school programs, report on students' progress, diagnose students' strengths and weaknesses, select students for special programs, place students in special groups and certify student achievement. Without standardized tests, it would be practically impossible to measure progress, to show the need for improvement and to keep all schools accountable. The purpose of state assessments required under the Act is to provide an independent insight into each student's progress, as well as each school's. This information is essential for parents, schools, districts and states in their efforts to ensure that no child- regardless of race, ethnic group, gender or family income-is limited to the progress of a consistently low-performing school. After each standardized testing, the results are piercingly clear and it can be easily determined how students performed, how teachers performed and how the school performed. For example, standardized testing is beneficial for teachers because overall poor results could indicate that the curriculum needs to be reviewed and aligned with the content upon which state standards are based or that teachers need to modify their instructional methods. Furthermore, annual tests show principals exactly how much progress each teacher's students have made, the strengths and weaknesses of students-in terms of the whole school-so they can use this information to guide decisions about program selection, curriculum arrangement and professional development for teachers. Therefore, heavy emphasis on standardized exams is important because it provides control and guidance for authorities.
**Testing Desirable: Data—General**

1. **Testing data is useful—provides with an objective view of student skills**


   At their core, standardized exams are designed to be objective measures. They assess students based on a similar set of questions, are given under nearly identical testing conditions, and are graded by a machine or blind reviewer. They are intended to provide an accurate, unfiltered measure of what a student knows. Now, some have argued that teachers’ grades are sufficient. But the reality is that teacher grading practices can be wildly uneven across schools—and even within them. For instance, one math teacher might be an extraordinarily lenient grader, while another might be brutally hard: Getting an A means something very different. Teacher grading can be subjective in other ways, including favoritism towards certain students, and it can find its basis in non-achievement factors like classroom behavior, participation, or attendance. But when students take a standardized exam, a much clearer view of academic mastery emerges. So while standardized exams are not intended to (and should not) replace the teacher grade book, they do provide an objective, “summative” assessment of student achievement. Standardized assessments of achievement can be used for comparison and accountability purposes, both of which are discussed in turn.

2. **Testing provides valuable data and the federal government is the only entity that can make it happen**


   The performance of the nation’s education system is critical to our future and to the lives of the students who experience it. The fundamental responsibility of schools is student learning. Valid estimates of student learning that strongly predict later life outcomes can be derived from annual academic tests. Much depends on the continued collection and dissemination of such information. Only the federal government is in a position to see that it happens. Congress can reauthorize ESEA, retain the requirement for annual tests that yield measures of student growth, and satisfy a diverse set of political factions if it focuses on its responsibility to see that valid information on school performance is available for all to use while pulling back from previous efforts to insert the U.S. Department of Education into roles that were previously reserved to states and school districts.

3. **Tests provide a lot of useful data—teachers agree**


   Theme #1: Tests provide useful data. Almost all teachers agreed that tests provide useful data showing whether students have grasped certain concepts. As one teacher put it, "I do think it has been very helpful in our building for when they walk in on day one to know this kid really doesn't know this, this kid didn't understand ... and that data has certainly helped us to know our kids much better." Another teacher said, "as soon as we get our class lists, we'll have those results too... I actually sat down with my kids and said, 'OK, this is where we were messing up.'" Indeed, these comments exemplify a belief shared by many proponents and opponents of testing -- more data is one clearly positive outcome of the recent emphasis on student testing.
### Testing Desirable: Data—Accountability

1. **Data from standardized tests are the best too we have to foster accountability**


   Like it or not, standardized exam data remain the best way to hold schools accountable for their academic performance. To its great credit, Ohio is implementing a cutting-edge school accountability system. The accountability metrics include robust measures often referred to as “student growth” or “value-added” measures, along with conventional proficiency results and college-admissions results. All of these outcome measures are based on standardized test results. The information from these accountability measures enables policymakers to identify the schools that need intervention, up to closure. For example, the charter school automatic closure law uses state exam results—both school-level value added and proficiency—to determine which schools must close. In addition, districts can go into state oversight via the Academic Distress Commission if they are low-performing along test-based outcomes. Another use of standardized testing data is coming in the area of deregulation. One priority bill being considered in the Senate (SB 3) would give “high-performing” districts certain flexibilities and freedoms from state mandates. How are these high performers identified? Answer: Through state accountability measures, based on standardized test scores. Outside of standardized test results, no objective method exists for policymakers to identify either poor-performing schools needing intervention or high-performing schools deserving rewards. Consider the alternative: Who would want policymakers to intervene in a school based on their “gut feeling” or reward a school based on anecdotes? Statewide standardized exams are essential for upholding a fair and objective accountability system. In a utopian world, one could wish away standardized tests. All schools would be great, and every student would be meeting their potential. But we live in reality. There are good schools and rotten ones; there are high-flying students and pupils who struggle mightily. We need hard, objective information on school and student performance, and the best available evidence comes from standardized tests. Policymakers need to be careful not to undermine the integrity of the state’s standardized tests.

2. **Decreasing tests will make it very hard for us to track school performance**


   In many schools there simply aren’t enough members of individual “vulnerable subgroups” in the tested grades and subjects to provide an estimate of school performance with respect to those groups that is sufficiently reliable to serve as a basis for school accountability and intervention. This is the well known minimum “n” size issue that bedevilled the measurement of adequate yearly progress under the provisions of NCLB. We see in the figure the effects of really small “n” size when only one grade is used to evaluate the performance of a school with subgroups – in many cases there aren’t enough students being tested to generate stable estimates. Whatever the arguments for less testing, reducing the frequency of testing along the lines of the proposal by Darling-Hammond and her co-authors would eliminate the possibility of evaluating individual teachers or schools with metrics that include gains in student achievement. And it would create such high levels of unreliability in the estimates of school level performance for vulnerable subgroups as to undermine any attempt to ensure that these students receive the quality education that they have been historically denied.
Testing Desirable: Data—Annual Testing Key

1. Annual testing data is vital measuring student progress


What impact would a move to less frequent testing have on states’ abilities to provide parents with meaningful information on school quality and to hold accountable schools that contribute the least to student learning? We address this question using roughly a decade of detailed, student-level records from all public elementary schools in Florida and North Carolina. We examine how schools would look if they were judged based only on their average test scores in a single grade—as might be the case under a grade-span testing regime—compared to how they can be judged using measures based on the growth in student test scores from year to year, which is only possible with annual testing. The type of growth measure we use in our analysis has been validated using data where students are randomly assigned to schools, thus increasing our confidence that it captures the true impact of schools on student achievement. The measure is conceptually similar to the growth measures currently used in some state accountability systems, such as student growth percentiles.

2. Multi-year data is vital to do fair assessments of school quality—annual testing is needed


Growth-based measures and demographic-adjusted average test scores identify similar types of schools as low-performing, as shown by the fact that the red dots (low-income schools) are spread out in the figure, as opposed to clustered on the left-hand side. However, the two measures do not identify the same schools as low-performing; the correlation between the two measures is only 0.56. About seven percent of all schools would fall in the bottom 15 percent of schools based on their adjusted scores, but not based on their growth. The opposite is true as well, with schools with very poor growth avoiding sanction based on mediocre (but not terrible) average scores. Demographic-adjusted average test scores also do a worse job at identifying schools where students learn the least, with the average growth rates of bottom-15% schools based on this metric closer to that of the average score measure than the growth-based measure. The reason growth measures outperform demographic adjustments when judging school quality is straightforward: although student characteristics such as family income are strongly correlated with test scores, the correlation is not perfect. In sum, our results confirm that using average test scores from a single year to judge school quality is unacceptable from a fairness and equity perspective. Using demographic adjustments is an unsatisfying alternative for at least two reasons. In addition to providing less accurate information about the causal impact of schools on their students’ learning, the demographic adjustments implicitly set lower expectations for some groups of students than for others.
**Testing Desirable: Data—Comparisons**

1. **Testing data is important—allows for cross-pupil and cross-district comparisons**


   The very objectivity of standardized exams yields comparability of student achievement, a desirable feature for parents and practitioners alike. Most parents, for example, would like to know whether their child is meeting state benchmarks, or how she compares to statewide peers. Statewide standardized exams give parents this important information. Meanwhile, school-shopping parents have every right to inspect and compare the standardized test results from a range of schools, including charters, district schools, and STEM schools, before selecting a school for their child. School practitioners also use statewide test results to benchmark their students’ achievement across school and district lines. For instance, the principal of East Elementary could compare the achievement of her students against those attending West Elementary, the district average, the county average, and the statewide average. How do her students stack up? Only a statewide standardized test could tell. Interestingly, proposals have been floated to allow schools to select their own assessment—a pick-your-own-assessment policy. This is a flawed idea and should be rejected. It would undermine the comparability principle of statewide testing. First, to be clear, standardized exams are not the all the same. Consider an obvious example: Ohio’s old state tests and the PARCC exams are both standardized exams, yet they are as different as night and day. Meantime, a pick-your-own-assessment policy would open a Pandora’s box of confusion over how to interpret the results. Imagine that Columbus City Schools selects NWEA as its testing vendor and reports an 80 percent proficiency rate. Now let’s say Worthington City Schools (suburban Columbus) selects PARCC and reports a 50 percent proficiency rate. Should we infer that Columbus students are actually achieving at higher levels than Worthington? Or is the test just different? Based solely on these test data, we’d have no clue. State assessment policy should not amount to a Choose Your Own Adventure for districts and schools. Instead, Ohio legislators must continue to implement a single, coherent system of standardized exams that provides comparable results.

2. **Data is important—it helps us differentiate performance between districts**

   Eileen Reed, Janie Scull, Gerilyn Slicker, and Amber M. Winkler, DEFINING STRONG STATE ACCOUNTABILITY SYSTEMS: HOW CAN BETTER STANDARDS GAIN GREATER TRACTION? A FIRST LOOK, Thomas B. Fordham Institute, 4—12, p. 8.

   3) Annual determinations and designations for each school and district that meaningfully differentiate their performance Stakeholders need to know how schools and districts compare with one another. States need to characterize and differentiate among schools and districts based on student achievement in a valid and reliable manner. These designations should not obscure the truth (i.e., via inflated performance or categories that permit high percentages of schools to fall disproportionately into the top tiers). They should be user-friendly—we prefer the A-to-F grading system. And the state should integrate state and federal accountability systems so that designations do not conflict. School and district designations should be derived, at least in part, from measures of individual student growth (either normative growth or growth to standard). Those designations should also be informed by the performance of specific student groups (e.g., race/ethnicity, income, English language learner status, and special- needs status) or “super subgroups.”
Testing Desirable: Data—School Improvement

1. Data is vital to driving improvement in school performance

Eileen Reed, Janie Scull, Gerilyn Slicker, and Amber M. Winkler, DEFINING STRONG STATE ACCOUNTABILITY SYSTEMS: HOW CAN BETTER STANDARDS GAIN GREATER TRACTION? A FIRST LOOK, Thomas B. Fordham Institute, 4—12, p. 6-8.

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement. Robust and timely diagnostic and outcome data on school, student, and adult performance enable educators, policymakers, parents, students, and the broader public to take appropriate action on the decisions that concern them (e.g., teacher hiring, school choice, classroom interventions, selection of turnaround model, just to name a few). Though we appreciate the need for a wide variety of data to assess sundry outcomes at all levels, we target here the data of most concern: student performance. States should release annual aggregate student outcome data in an accessible manner (meaning clear and user-friendly). In short, states should publish results for every school that includes a snapshot of how many students are reaching various achievement levels (such as basic, proficient, advanced, etc.), how many students are making expected progress over time, and whether or not the school itself is improving from year to year. (Of course, states should also require that schools provide reports to parents on their own child’s performance and progress.) In addition, the state should collect and report data on multiple measures of student performance whenever possible. These might include attendance, graduation rates, on-track indicators, postsecondary matriculation/retention or remediation rates, course completion (transcript data), attainment of industry certification, and performance on SAT, ACT, and AP/IB exams. State databases should also permit school-level comparisons. States should also require that the data above be disaggregated by subgroups within schools of race/ethnicity, gender, free and reduced-price lunch status, special-needs status, English language learner status, and so on. States might also choose to substitute one or more of these groups with a low-performing subgroup or high-performing subgroup (i.e., a “super subgroup”).

2. We need tests, perhaps even more, to effectively track school achievement


Surely if the nation is going to go to the trouble and expense of testing students to see how they and their schools are performing, we want the signal sent by those test scores to be as reliable and actionable as possible. It is psychometrically straightforward that test reliability is strongly correlated with the amount of information that is sampled. Thus, we knew before we conducted our empirical exercise that using one grade to estimate the performance of elementary schools would produce less reliable information than using three grades – the question was only how much. It follows that if we want the most reliable signal from test score data, we should test more not less. Any psychometrician worth his or her salt will tell you this. Paradoxically, it may also be the case that many of the unintended negative consequences of high stakes student testing would be reduced if we tested more not less. Consider what would happen to the pervasive test-prep sessions that consume weeks of class time in many schools leading up to the end-of-the-year test if students, instead, spent an hour or so monthly being tested on content drawn from their lessons in the previous few weeks. Under this scenario the high stakes tests blend into the tests and quizzes that good teachers have always given their students regularly, and that research shows without a doubt increase student learning.[vii] Yes, you heard it here: The solution to too much testing is even more testing.
Testing Desirable: General

1. **Tests will work as long as teachers and students recognize their importance**

Stijepko Tokic, “Examining the Big Picture Regarding the Importance of the No Child Left Behind Act: Is It Worth Giving a Chance?” THURGOOD MARSHALL LAW REVIEW v. 32, Spring 2007, p. 323-324.

In addition to the conclusion that test anxiety did not impact performance on standardized exams significantly, the study provided other significant findings. First, student attitude towards testing will predict performance on achievement tests because those students who value their test performance score higher than those who do not value it. Second, students of teachers who placed heavy emphasis on standardized exams will receive higher performance scores. The testing required by the Act is “considered to be ‘high-stakes testing’ because its results dictate serious measures including student retention and evaluation of schools.” Thus, each exam is extremely valuable. Basically, the Act forces students and teachers to value each exam. The study shows that if students and teachers value standardized tests, a positive impact on student performance can be expected. Therefore, it is reasonable to expect that the Act’s emphasis on students and teachers placing high value on standardized exams will have a positive impact on student performance.

2. **On balance, testing has done more good than harm**


Without a doubt, and in the main, testing has done more good than harm in America’s schools. My Fordham colleague Andy Smarick is absolutely correct to argue that annual testing “makes clear that every student matters.” The sunshine created by testing every child, every year has been a splendid disinfectant. There can be no reasonable doubt that testing has created momentum for positive change—particularly in schools that serve our neediest and most neglected children.
Testing Desirable: Knowledge Confidence

1. Testing results help students figure out what they do and do not know—improves studying / outcomes


One of the reasons that tests are unappealing to some students and to their overweening parents is that tests fairly reveal what we do and do not know. This feedback can violate the positive feelings we hold about ourselves and our abilities, which are often inappropriately optimistic, especially in the classroom (Hacker, Bol, Horgan, & Rakow, 2000). This violation causes students to rate instructors more poorly (Isley & Singh, 2005) and to generate complicated but unsupported theories about supposed learning styles that their classrooms are failing to support (Pashler, McDaniel, Rohrer, & Bjork, 2009). What is of particular concern is the way that such inappropriately tuned self-assessments influence study behavior. There are two ways in which having a poor calibration between what we know and what we think we know can be harmful. The first concern is that inappropriate overconfidence will lead us to study less than is warranted (Bandura, 1993). Why would a student who feels that they have mastered the material continue to study? The second is that poor insight into what we have or have not yet mastered can lead to poor decisions about how to allocate our study when we do choose to study. Spending additional time on already mastered aspects of the curriculum or spending very little time on poorly mastered aspects is a non-optimal use of learners’ time.

2. Tests help students appropriately tune their confidence about what they do and do not know


Confidence. Overconfidence in one’s abilities and judgments is ubiquitous across domains (e.g., Fischhoff, Slovic, & Lichtenstein, 1977; Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999). Experts in a domain are not immune to this bias; on the contrary, they may be even more susceptible. Soccer experts, for example, have been shown to predict the outcome of World Cup matches with the same accuracy as non-experts, but with much higher confidence (Andersson, Edman, & Ekman, 2005). It is easy to see what the costs of such overconfidence might be: a failure to gather more evidence prior to committing to a decision, an unwillingness to consider the opinions of others, making inappropriate wagers, and the like. Analogous costs are apparent for the overconfident student: insufficient study time, poor prioritization of study techniques, and so on. And, indeed, students who exhibit more overconfidence in their assessment of mastery on a set of definitions do indeed reveal poorer performance in later exams on those materials (Dunlosky & Rawson, 2012). On the other side of the coin, students with higher grade-point averages (GPAs) exhibit lower overconfidence when predicting performance in an upcoming exam (Grimes, 2002). There are techniques that are known to reduce overconfidence. Most relevant for our discussions here is the role of feedback in ongoing learning and testing. Expert forecasters who are forced to directly compare their predictions with the outcomes of the events they are trying to predict often exhibit exquisitely tuned calibration. A terrific example is provided by weather forecasters, who get all kinds of feedback from bosses and angry citizens when their predictions are wrong. Consequently, weather forecasting is now an exceptionally accurate endeavor (Murphy & Winkler, 1984). Contrast this with sports or political forecasting, where the outcomes are available but often discounted when they fail to confirm one’s predictions or ignored entirely in the hubbub of the next major event (Silver, 2012). Tests provide the opportunity for students to tune their confidence in their understanding and mastery of course materials to appropriate levels. Students who receive immediate feedback on the accuracy of their responses by a computerized testing system reveal much more enhanced calibration of confidence than students who do not receive feedback (Zakay, 1992). Although standardized tests do not often force the students to directly compare their confidence in material prior to the exam with the outcome of that exam, later review of that exam can serve that function, as can exams that require students to make decisions about which answers they choose to submit for grading (e.g., Higham, 2007) or how precise an answer to submit (Higham, 2013). Students who experience multiple cycles of studying material, making judgments about their ongoing learning, and taking a test exhibit improved judgment accuracy over cycles (Kelemen, Winningham, & Weaver, 2007). In addition, computerized learning environments that include the opportunity to take tests and to assess one’s performance reveal an advantage on later tests over control conditions that do not include such forced assessments (Metcalf, Kornell, & Son, 2007).
Testing Desirable: Knowledge Confidence [cont’d]

3. Tests help students sharpen their focus onto learning what they do not already know


Tests inform us of what topics are important and where our learning is deficient. One of the most direct ways in which tests promote learning is by motivating students to study. The benefits of this effect can be controversial when it is believed that the test measures unimportant skills or when teachers focus on the test to the exclusion of other materials, two common criticisms of the current standardized tests for the Common Core. But the curriculum for the Common Core, as well as its attendant tests, is fluid and likely to experience considerable development. Students who take regular quizzes in the classroom are more likely to attend unrequired meetings (Fitch, Drucker, & Norton, 1951) and exhibit better class attendance (Wilder, Flood, & Stromsnes, 2001), both of which are known to increase student achievement. Moreover, tests with a clear agenda can focus teachers’ and students’ activities onto materials that are broadly considered to be valuable. Students learn very rapidly from tests how to direct their study to important materials. They learn to ignore materials that are not likely to be tested, to relate materials in such a way that conforms to the expected nature of the test (Finley & Benjamin, 2012), and to study more for tests that they expect to be difficult (Meyer, 1936). Tests also sharpen focus on important but unmastered materials (Thiede & Dunlosky, 1999). Having the opportunity to attempt to retrieve information from memory on a test—where the learner does not have easy access to the answer—is highly diagnostic of one’s level of mastery for that material (Benjamin, Bjork, & Schwartz, 1998) and consequently highly related to the success of future attempts to retrieve the material (Nelson & Dunlosky, 1991). Study following a test is more effective in part because it enhances the extent to which learners spend more time with the more difficult, unmastered materials (Soderstrom & Bjork, 2014), a strategy that is highly effective (Tullis & Benjamin, 2011).
Testing Desirable: Learning Methods

1. They overstate their case—tests can be used as tools of learning


Until recently, a great deal of the most influential psychological theory on testing began and ended with these very questions, providing guidance on how to develop, implement, and score tests possessing these desirable characteristics. More recently, primarily in the last few decades, evidence has emerged out of memory research within cognitive psychology, a different historical tradition that has placed little focus on individual differences and more on mental processes common to all. This work indicates that tests have value not previously considered by test designers. The purpose of this article is to briefly review this body of research. The key point we will argue for is that the benefits of testing are not limited to those arising from good assessment: There is an important potential role for tests as tools for learning and not just tools for assessment. Following our review, we revisit the question of standardized testing and provide some recommendations for ways in which the beneficial consequences of testing can be maximized without compromising their use for assessment purposes.

2. Testing improves our ability to learn new concepts


Testing also boosts our ability to learn concepts. In one example (Jacoby, Wahlheim, & Coane, 2010), subjects were required to learn about different families of birds. One group was given four study blocks with pictures of birds and their family names; the other was given only one study block and three blocks on which they were shown only the picture and asked to retrieve the family name. Feedback was provided after their response. The novel aspect of their procedure was a later test that included entirely new pictures of birds from the same families. The group that had had quizzes was actually better able to sort those new pictures of birds into the appropriate families, indicating that quizzing had done more than simply enhance their memory for the previously studied birds. Rather, it seems to have actually improved their knowledge of the categories that differentiated among the birds. Tests encourage the kind of thinking that is essential not just for retention but also for mentally organizing the acquisition of new material. One consequence of the recent rapid growth in testing research is that some conclusions are still in flux, and the boundary conditions of some effects of testing are still undiscovered or under debate. The benefits of testing on generalization is an area for which this is particularly true—there are reports of failures to generalize as well (Tran, Rohrer, & Pashler, 2015). What can be stated right now with clarity is that there are likely conditions under which the type of learning engendered by testing will generalize effectively, though the range of those conditions is still under active exploration, and the extent of the benefit is yet unknown. At the very least, remembering is a precondition for generalization—we can be quite sure that conditions under which students remember less of what they have learned are not apt to lead to effective generalization and inference to new situations.
Testing Desirable: Learning Methods [cont’d]

3. Test results help teach students which learning methods are effective and which are ineffective


Tests help us learn which learning techniques are effective and which are ineffective. One bottleneck to effective student learning is the widespread use of poor learning techniques. Many of the techniques that have been identified as highly effective in basic research on learning and memory (for reviews, see, for example, Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Pashler et al., 2007) are ones that are dispreferred by students. Most people, for example, eschew distributing practice (Baddeley & Longman, 1978), interleaving multiple to-be-learned skills (Simon & Bjork, 2001), and—most poignantly for this review—testing (Karpicke, 2009)!

When we take a test, we discover which of our study endeavors have been successful and which have not been. This appears to be particularly true when learners make judgments about those techniques (or about materials learned with those techniques) prior to the test (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin, 2003; Fiechter, Benjamin, & Unsworth, in press), and then have the opportunity to see the consequences of those varying techniques play out in memory performance. The benefits of quizzing oneself are, for example, apparent on a later test, and learners change their assessments of that technique when guided to observe the gap between their assessments of how testing will influence memory and their actual later performance (Tullis, Finley, & Benjamin, 2013).

Overall, it has been found in educational settings that opportunities for self-assessment enhance performance in the classroom (Dochy, Segers, & Sluijsmans, 1999). In combination with performance tests, self-assessments can help students gain a better appreciation for (a) their overall level of competence, (b) what they do and do not currently know, and (c) which study techniques are serving them most effectively. These techniques comprise a skill that has been recognized as critical for effective learning and transfer for more than half a century: “learning to learn” (Postman, 1969).
Testing Desirable: Memory—General

1. **Testing improves memory—the effects are cumulative, improving results on later tests**


   Are the benefits of testing simply a consequence of changes in motivation or desire to learn? Some authors have suggested that asking a question can enhance a learner’s curiosity to know the answer (Berlyne, 1966). The benefits of quizzing are roughly the same regardless of how much learners are paid for their correct responses (Kang & Pashler, 2014). These results are inconsistent with the idea that motivation plays a major role in producing the benefits of testing. However, motivation does play a major role in how people direct their study and other more indirect ways in which the experience of taking tests can influence memory. These are important points we will return to in greater detail later in this article. When taken together, these results help us understand why students who take more tests in the classroom tend to perform better on later exams (Bangert-Drowns, Kulik, & Kulik, 1991). Most of the benefits come from the first few tests, indicating that it does not require much compromise in the allocation of class time to administer periodic tests. In addition, students of all abilities appear to benefit from the opportunity to take tests (Pan, Pashler, Potter, & Rickard, 2015). As we will see below, these benefits are not limited to enhanced memory for the tested material. We will review additional research that indicates a positive role for testing for other, untested material, as well as for other aspects of cognition and motivation.

2. **Testing helps improve information retention—‘testing effect’**


   What are those benefits? Decades of research on the “testing effect” have documented that calling up a fact or a concept from memory actually helps us remember it better the next time. Foreign language learners who tested themselves on vocabulary words remembered 80 percent of the words they studied, according to one experiment, while learners who used other study methods remembered only about a third of them. Likewise, students learning from a science textbook who were queried about what they read retained about 50 percent more of the information than students who studied in other ways.

3. **Testing shapes what we know—it helps improve memory**


   Although we often conceive of memory as something like a storage tank, and a test as a kind of dipstick that measures how much information is inside, that’s not actually how the brain works. Every time we pull up a memory, we make it stronger and more lasting — so that testing doesn’t just measure what students know, it changes what they know. Reading over material to be learned, taking notes, even engaging in teacher-favored techniques like concept-mapping: none of these practices are as effective as testing at improving retention.

4. **Proper testing improves our memory**


   When we take a test on which we are asked to retrieve and produce previously learned information, successfully recalling that information increases our ability to retrieve it again later. A good example of the advantages of testing is provided by Roediger and Karpicke (2006; Experiment 2). In their experiment, subjects read text passages, and then were either given three opportunities to re-read the passage, two additional re-reading opportunities followed by a quiz in which they tried to recall as much as they could from the passage, or three quizzes with no re-study opportunities. On a final test 1 week later, the latter group remembered the material best—despite having had only one opportunity to read the passage! Quizzing during a review phase has been shown to improve memory for other types of materials as well, including foreign-language vocabulary (Carrier & Pashler, 1992) and simple facts (McDaniel & Fisher, 1991).
5. Testing enhances information retrieval—clustering effects


Testing also increases the effectiveness of the way in which we choose to access and organize the tested information. For example, when studying a list of categorized materials, quizzes increase both the number of categories that are reported on a final test and the number of items from each of those categories (Zaromb & Roediger, 2010). These beneficial effects are probably due to the fact that testing promotes clustering of similar items during the test, a retrieval strategy that is very effective (Mulligan, 2005).

6. Multiple types of tests have been shown to enhance memory


Other research has found beneficial effects of testing for a variety of different testing formats. Taking either a short-answer or a multiple-choice practice quiz enhances memory on a later test, even when the later test is in a different format than the quiz. The benefits are especially prominent if the quiz includes feedback (LaPorte & Voss, 1975). However, it does appear overall that short-answer quizzes increase later retention to a greater degree than multiple-choice quizzes (cf. Glover, 1989; Kang, McDermott, & Roediger, 2007).

7. Testing’s memory-enhancing effects are valid in typical classroom-settings


These results generalize to actual classroom settings. Students who take periodic tests on material remember that material better on later exams, and the enhancement is greater for short-answer than multiple-choice tests. This result has been shown in college students (McDaniel, Anderson, Derbish, & Morrisette, 2007), sixth-grade students (Roediger, Agarwal, McDaniel, & McDermott, 2011), and eighth-grade students in science (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011) and history (Carpenter, Pashler, & Cepeda, 2009). It is evident for a wide range of materials, including biology (McDaniel et al., 2011), social studies (Roediger et al., 2011), general science (Roediger & Karpicke, 2006), biographical materials (Gates, 1917), and spelling (Forlano, 1936).
8. Making errors during a test improves our memory for correct answers


Making errors during a test enhances memory for correct answers. One concern that people have with testing is that test takers will make errors and that the process that leads to those errors will become engrained and will prevent the learner from acquiring the correct solution. Interestingly, this does not appear to be the case; in fact, making errors may even have tangible benefits for learners. In one representative experiment, Kornell, Hays, and Bjork (2009) asked learners to answer unanswerable questions about made-up events. After the subject provided an answer, learners were given the “correct” answer by the experimenter. Those subjects remembered the “correct” answers better than a group that was given the answer but not given the opportunity to make a mistake prior to being given that answer. Even when tests require people to construct explanations for scientific observations, being required to produce what were generally erroneous speculations did not reduce subsequent learning from feedback (Kang et al., 2011). A similarly beneficial effect of making errors can also be seen following a pretesting phase, where learners answer questions about the material before they study it at all (Richland, Kornell, & Kao, 2009). All of this is not to say that making errors willy-nilly is good for learning overall. Benefits of making errors are only apparent when the errors are meaningfully related to the learning material. So, having someone estimate the age of a person in a picture, for example, does lead to enhanced memory for the correct answer when that correct answer is given as feedback (McGillivray & Castel, 2010), but guessing a random, unrelated associate to a provided word does not (Huelser & Metcalfe, 2012). Errors that are made with high confidence are even more likely to be successfully remedied by feedback with the correct answer (Butterfield & Metcalfe, 2001). This result is unexpected because it stands to reason that the errors we make with high confidence are the ones we believe most strongly, and such strong beliefs should be more resistant to change. But this reasoning appears to be wrong. In general, feedback that is surprising draws our attention (Butterfield & Metcalfe, 2006) and improves our memory (Fazio & Marsh, 2009). This effect has been shown in young children (Metcalfe & Finn, 2012), young adults, and older adults (Eich, Stern, & Metcalfe, 2013), indicating generalizability to a wide variety of populations.
**Testing Desirable: Memory—Forgetting**

1. **Testing decreases the rate at which we forget material**


The cognitive benefits of testing are not like a single shot in the arm. Taking a test improves memory for the material, and it also decreases the rate at which we forget that material. What this means is that the benefits of testing are even greater when looking at longer-term retention. In the study with texts reviewed earlier (Roediger & Karpicke, 2006), the benefits of multiple quizzes were largest at a 1-week delay after the original study event.

2. **Testing slows forgetting better than does close reading**


All of this is particularly noteworthy because, counterintuitively, there are not many cognitive interventions that appear to slow the rate of forgetting. Studying material more leads to a higher initial degree of learning but does not slow forgetting (Anderson & Schooler, 1991; Hellyer, 1962). Employing a “deep” level of processing—in which the learner is encouraged to think about the meaning of the to-be-learned information—does not slow forgetting (Nelson & Vining, 1978). Yet, testing slows forgetting (Carpenter, Pashler, Wixted, & Vul, 2008), sometimes considerably (Wheeler, Ewers, & Buonanno, 2003), which may make it an ideal technique for promoting long-term, durable learning.

3. **Reduced forgetting is a significant benefit of tests**


It may even be the case that the benefit to memory from testing is due entirely to the reduced forgetting it engenders. When a test is administered immediately after the review phase in an experiment, performance is often superior following a re-study than a quiz event. However, this advantage is short lived: after a relatively short time interval, the benefits of quizning are apparent. So, although quizning may not be the study regimen of choice for a student who is doing last-minute cramming, it is a better way to promote long-term retention.
**Testing Desirable: Multiwarrant**

1. **Testing affords many benefits—we may need reform, but we should keep assessments**


   But before we retreat to the pre-NCLB era of grade-span testing, assess samples of students instead of all students, or revert to some other testing-light position, let’s at least recall some of the benefits of annual testing of all kids. It makes clear that every student matters. It makes clear that the standards associated with every tested grade and subject matter. It forces us to continuously track all students, preventing our claiming surprise when scores are below expectations. It gives us the information needed to tailor interventions to the grades, subjects, and students in need. It gives families the information needed to make the case for necessary changes. It enables us to calculate student achievement growth, so schools and educators get credit for progress. It forces us to acknowledge that achievement gaps exist, persist, and grow over time. It prevents schools and districts from “hiding” less effective educators and programs in untested grades. Now, I’m not saying that people shouldn’t be aggravated by federal intrusiveness or unfitness of school classifications and interventions. But these concerns (and others) argue for additional performance metrics, different methods of assessing and classifying schools, and new approaches to school improvement. Jettisoning annual testing is a punishment that doesn’t fit the crime. Let’s continue to critically analyze the consequences of NCLB and ESEA waivers (as folks like Anne Hyslop are doing ably). And let’s explore new ways of executing state governments’ responsibility to ensure high-quality schools for all kids. But let’s also avoid tossing the baby with the bath water. Remember, there’s a reason why today’s era of standards, assessments, and accountability exists: The era that proceeded it wasn’t working for entirely too many kids.

2. **Testing and other federal accountability measures are important for multiple reasons**


   There are many things that need fixing in NCLB, certainly including its unrealistic accountability provisions (under which every child in the nation was expected to be proficient in reading and math by last year and schools were threatened with restructuring for failing to make adequate progress towards this goal). But it would be a serious mistake for Congress to treat standards, testing, and accountability as a single target to be taken out with a shotgun blast. Each member of the triumvirate can exist on its own and has a different impact on school performance. And each has a different rationale and political basis in our federal system of education. What follows is a case for retaining in ESEA annual testing requirements that produce information on the growth in student achievement from one year to the next, while eliminating most of the provisions related to standards and accountability, control over which would revert to the states and school districts. The argument has four parts: 1) federal control of standards and accountability is unnecessary whereas the provision of information on the performance of schools is a uniquely federal responsibility; 2) test scores are valid indicators of student learning that matters for important long-term outcomes and therefore provide essential information on school performance; 3) several important functions for managing and improving education depend on annual measures of student achievement growth; 4) most of the political opponents of standards, testing, and accountability should favor the retention of annual testing shorn of federally dictated standards and accountability.
Testing Desirable: SAT

1. The SAT is useful in predicting whether students will finish their college degrees—data from the SUNY system probe


Thus, by comparing graduation rates at SUNY campuses that raised the SAT admissions bar with those that didn’t, we have a controlled experiment of sorts that can fairly conclusively tell us whether SAT scores were accurate predictors of whether a student would get a degree. The short answer is: yes, they were. Consider the changes in admissions profiles and six-year graduation rates of the classes entering in 1997 and 2001 at SUNY’s 16 baccalaureate institutions. Among this group, nine campuses raised the emphasis they put on the SAT after 1997. This group included two prestigious research universities (Buffalo and Stony Brook) and seven smaller, regional colleges (Brockport, Cortland, New Paltz, Old Westbury, Oneonta, Potsdam and Purchase). Among the campuses that raised selectivity, the average incoming student’s SAT score increased 4.5 percent (at Cortland) to 13.3 percent (Old Westbury), while high school grade-point averages increased only 2.4 percent to 3.7 percent, a gain in grades almost identical to that at campuses that did not raise their SAT cutoff. Yet when we look at the graduation rates of those incoming classes, we find remarkable improvements at the increasingly selective campuses. These ranged from 10 percent (at Stony Brook, where the six-year graduation rate went to 59.2 percent from 53.8 percent) to 95 percent (at Old Westbury, which went to 35.9 percent from 18.4 percent). Most revealingly, graduation rates actually declined at the seven SUNY campuses that did not raise their cutoffs and whose entering students’ SAT scores from 1997 to 2001 were stable or rose only modestly. Even at Binghamton, always the most selective of SUNY’s research universities, the graduation rate declined by 2.8 percent. The change is even more striking if we compare experiences of three pairs of similar SUNY campuses that, from 1997 to 2001, took sharply divergent paths. First, Stony Brook and Albany, both research universities: over four years, at Stony Brook the average entering freshman SAT score went up 7.9 percent, to 1164, and the graduation rate rose by 10 percent; meanwhile, Albany’s average freshman SAT score increased by only 1.3 percent and its graduation rate fell by 2.7 percent, to 64 percent. Next, Brockport and Oswego, two urban colleges with about 8,000 students each: Brockport’s average freshman SAT score rose 5.7 percent to 1080, and its graduation rate increased by 18.7 percent, to 58.5 percent. At the same time, Oswego’s freshman SAT average rose by only 3 percent and its graduation rate fell by 1.9 percent, to 52.6 percent. Finally, Oneonta and Plattsburgh, two small liberal arts colleges with 5,000 students each: Oneonta’s freshman SAT score increased by 6.2 percent, to 1069, and its graduation rate rose 25.3 percent, to 58.9 percent. Plattsburgh’s average freshman SAT score increased by 1.3 percent and its graduation rate fell sharply, by 6.3 percent, to 55.1 percent. Clearly, we find that among a group of SUNY campuses with very different missions and admissions standards, and at which the high school grade-point averages of enrolling freshmen improved by the same modest amount (about 2 percent to 4 percent), only those campuses whose incoming students’ SAT scores improved substantially saw gains in graduation rates. Demeaning the SAT has become fashionable at campuses across the country. But college administrators who really seek to understand the value of the test based on good empirical evidence would do well to learn from the varied experiences of New York’s state university campuses.

2. This is not about material equality—we need to be focused on valuing everyone’s unique attributes


I haven’t used the word “meritocracy” to describe this because it doesn’t apply. Merit has nothing to do with possessing a high IQ. It is pure luck. And that leads to my reason for writing this. As long as we insist on blaming inequality of academic outcomes on economic inequality, we will pursue policies that end up punishing children whose strengths do not lie in academics. We will continue to tell them that they will be second-class citizens if they don’t get a college degree; to encourage them to accumulate student debt only to drop out or obtain a worthless degree. Worse, we will prevent them from capitalizing on their other gifts of character, grit and the many skills that the SAT doesn’t test. What we need is an educational system that brings children with all combinations of assets and deficits to adulthood having identified things they enjoy doing and having learned how to do them well. What we need is a society that has valued places for people with all combinations of assets and deficits. Both goals call for completely different agendas than the ones that dominate today’s rhetoric about educational and economic inequality.
3. The SAT is not an affluence test—the outcomes they describe are driven by parental IQs


Spring is here, which means it’s time for elite colleges to send out acceptance letters. Some will go to athletes, the children of influential alumni and those who round out the school’s diversity profile. But most will go to the offspring of the upper middle class. We all know why, right? Affluent parents get their kids into the best colleges by sending them to private schools or spending lots of money on test preparation courses. Either way, it perpetuates privilege from generation to generation. The College Board provides ammunition for this accusation every year when it shows average SAT scores by family income. The results are always the same: The richer the parents, the higher the children’s SAT scores. This has led some to view the SAT as merely another weapon in the inequality wars, and to suggest that SAT should actually stand for “Student Affluence Test.” It’s a bum rap. All high-quality academic tests look as if they’re affluence tests. It’s inevitable. Parental IQ is correlated with children’s IQ everywhere. In all advanced societies, income is correlated with IQ. Scores on academic achievement tests are always correlated with the test-takers’ IQ. Those three correlations guarantee that every standardized academic-achievement test shows higher average test scores as parental income increases. But those correlations also mean that a lot of the apparent income effect is actually owed to parental IQ. The SAT doesn’t have IQ information on the parents. But the widely used National Longitudinal Survey of Youth contains thousands of cases with data on family income, the mother’s IQ, and her children’s performance on the math and reading tests of the Peabody Individual Achievement Test battery, which test the same skills as the math and reading tests of the SAT.

4. SAT scores are more strongly linked to parental IQ than income


Let’s throw parental education into the analysis so that we can examine the classic indictment of the SAT: the advantage a child of a well-educated and wealthy family (Sebastian, I will call him) has over the child of a modestly educated working-class family (Jane). Sebastian’s parents are part of the fabled 1%, with $400,000 in income, and his mother has a college degree. But her IQ is only average. Jane’s family has an income of just $40,000 and mom has only a high-school diploma. But mom’s IQ is 135, putting her in the top 1% of the IQ distribution. Which child is likely to test higher? Sebastian is predicted to be at the 68th percentile on the PIAT. Jane is predicted to be at the 78th percentile. If you want high test scores, “choose” a smart but poor mother over a rich but dumb one—or over a rich and merely ungifted one. There is nothing new in this analysis. The relationship between IQ and income was first documented decades ago. But people refuse to confront it because it exposes an unwelcome reality: The Sebastians of contemporary America usually have smart parents as well as affluent, well-educated ones. The more strictly that elite colleges admit students purely on the basis of academic accomplishment, the more their student bodies will be populated with the offspring of the upper-middle class and wealthy—not because their parents are rich, but because they are smart. No improvement in the SAT can do away with this underlying reality.
Testing Desirable: School Reform

1. We should be lauding testing—it can force reforms that improve our low-performing schools


The authors even manage to portray one of high-stakes testing’s greatest strengths--its ability to expose and quantify the failure of inner-city schools to provide a decent education to poor and minority students--as an oppressive menace. Acknowledging the persistence of an “achievement gap between wealthy, mostly white school districts and poor, mostly minority school districts,” they accuse high-stakes exams of “testing poor students on material they have not had a sufficient opportunity to learn.” Back through the Looking Glass: one could hardly imagine a more conclusive argument in favor of high-stakes tests, which expose the shameful failure of inner-city schools to provide their students with decent educational opportunities. If disadvantaged youngsters are held back because they fail a high-stakes test, but pass the next year (or the following year) after having learned basic reading and math skills that they lacked before, one might expect a sense of accomplishment: high-stakes testing has forced the school to provide those students with a real education.

2. Testing is important—we need it to raise education standards


To address U.S. students’ international achievement gap, the National Governors Association, in partnership with the Council of Chief State School Officers, a nonpartisan organization of public school officials, helped create a set of nationwide achievement goals known as the Common Core State Standards. These voluntary benchmarks in English language arts and math reflect what young Americans will need to know if they are to compete with students from China, Singapore, Finland, South Korea and elsewhere. Instead of speaking out against the Common Core, critics of standardized testing should see this reform effort as an opportunity to make testing better. Educators have an indispensable role in creating tests that do justice to student achievement while rewarding skilled teachers. Raising standards should be the primary goal of education reform. Those who argue against standardized testing are not only misguided but are also leading U.S. schools and students in precisely the wrong direction.


Testing Desirable: Student Success

1. Scores are valuable—useful in measuring prospects for future student success


Scores that students receive on standardized tests administered in schools are strongly predictive of later life outcomes that are of great value to those students and the nation, after controlling for all the other observable characteristics of those students that are associated with later success. What’s more, gains in test scores that result from interventions such as being assigned to a particularly effective teacher or attending a school facing accountability pressure also predict improvements in adult outcomes. In other words, how much students learn in school makes a big difference in their lives, and standardized tests capture valid information on this. As such, information on school performance that does not include information on student learning as measured by standardized tests will be badly compromised, like information on the performance of a publicly traded stock that does not include its historical returns. Recent work by economists Raj Chetty, John Friedman, and Jonah Rockoff on teacher effectiveness utilizes data from test score data in reading and math in grades 3-8 in New York City linked to IRS records for the same students as they became adults. Our focus here is on these linked records and what they tell us about the predictive power of test scores, rather than on the story they tell about teacher effectiveness that was the focus of the Chetty et al. study. The school records of the study sample provide test scores as well as a rich set of control variables, including student variables (e.g., gender, ethnicity, special education status, record of suspensions, and limited English proficiency) and school variables (e.g., class size, teacher experience). The tax records include individual earnings, information on college attendance, and child dependents (from which mothers who were teenagers when they gave birth could be identified). Without controlling for other student characteristics and school variables the association between student test scores in grades 3-8 and later outcomes is huge, but it could reflect, for example, the impact of ethnicity or limited English proficiency independent of student learning. With all controls in place the most important of these omitted variables are accounted for. The association is still very large and most plausibly a function of the academic knowledge that is being assessed on the standardized tests.

2. Scores predict future success—we need federal data collection


The figure below (generated by the authors of this piece from information in Appendix Table 3 of Chetty et al.) represents the relationship between earnings at age 28 for students from the 5th to the 95th percentile in test scores in reading and math in grades 3-8, after adjusting earnings for the influence of all the control variables mentioned above. To compare two points on the graph: relative to individuals who as students were at the 30th percentile, individuals who were at 70th percentile in test scores in elementary and middle school were earning 13.6% more as young adults. To repeat, the association is net of the other variables that entered into the prediction as controls. Some of these controls, such as special education status, capture the impact of student knowledge as measured on standardized tests, and thus bias downward the association that is represented in Figure 1. Nevertheless, the estimated impact on earnings is very large. Similarly large impacts are found on college attendance, the quality of college attended, the quality of the neighborhood of residence, and giving birth as a teenager among females. Some of these relationships surely reflect the impact of students’ innate ability on their adult outcomes, but many studies have found that interventions that impact test scores also have impacts on later-life outcomes, including class-size reduction, school accountability, charter schools, and exposure to highly effective teachers. Given these strong predictive and causal relationships, who among us that wants to improve education would choose to ignore how much students are learning in school as measured by standardized tests? Unless the federal government mandates the collection and dissemination of this information, we will be bereft of one of the best indicators we currently have of school performance.
Testing Desirable: Student Success [cont’d]

3. They are wrong—a testing emphasis will not lower student performance

Stijepko Tokic, “Examining the Big Picture Regarding the Importance of the No Child Left Behind Act: Is It Worth Giving a Chance?” THURGOOD MARSHALL LAW REVIEW v. 32, Spring 2007, p. 323.

2. Heavy emphasis on high-stakes standardized testing might actually have a positive effect on student performance

The standards and assessment requirements under the Act are very rigid because it requires annual review of all students in grades three through eight and at least one evaluation in grades ten through twelve. These reviewing procedures apply to all students in all public schools. This increased emphasis on standardized testing in the Act has raised concern over the impact of standardized testing on students. Mulvevon, Stegman, and Ritter conducted a study to investigate what impacts student performance on standardized exams beyond specific knowledge. The key finding of their study suggest that test anxiety did not significantly impact performance on standardized exams. The study further suggests that although trauma associated with testing probably exists, overwhelmingly most students reported no additional anxiety due to standardized testing. Moreover, the study found that seventy-five percent of students reported high anticipation during standardized test week. Also, students reporting the greatest anxiety performed higher on standardized exams in comparison to those with lower anxiety. Therefore, the study suggests that the Act’s heavy emphasis on standardized testing should not have a negative impact on student performance, including traditionally low performing minority students.
Testing Desirable: Teacher Quality / Evaluation

1. Tests are useful because they can be used to help improve teacher quality


There is growing bipartisan support in favor of using student standardized test scores to improve teacher evaluations. Poor evaluations could then be used to remove the system’s worst teachers. Such "value-added" analyses of teacher quality measure each instructor’s independent contribution to student-learning during the school year. Policymakers who support using value-added measures to identify and remove ineffective teachers span the political spectrum. Republicans such as Gov. Chris Christie have voiced support, as have Democratic New York Gov. Andrew Cuomo and President Obama. Research suggests that, if used properly, value-added measures such as test scores could drastically improve the quality of education provided to students. A nationwide study shows such measures have demonstrated far greater accuracy at predicting future effectiveness than have conventional measures of teacher quality, such as years of experience and advanced degrees. A rich evaluation system would assess a teacher’s performance using both student test scores and observations of classroom performance. The subjective classroom analysis would help to flag cases when test scores wrongly identify good teachers as ineffective. School systems should use the improved evaluation to identify and remove their worst teachers. Teachers deemed ineffective for more than one year by a robust evaluation system should be removed from the classroom. And the policy should apply to all teachers, rather than just new applicants: No student should have to suffer simply because a teacher has been employed by the school system for more than three years and, therefore, can’t be fired without an overly burdensome process.

2. Teacher quality is the key variable in education outcomes—we should be working hard to improve it


Research over the last two decades has confirmed what most parents already knew: Teacher quality is any public school’s most important asset. Taking that simple and obvious premise seriously means working to identify and remove ineffective teachers. A bipartisan group of lawmakers in New Jersey and nationwide is pursuing this path. Students are harmed when they are taught by bad teachers. Research shows that being assigned to an ineffective teacher can reduce a student’s learning during the school year by as much as a grade level. Anyone who understands the importance of education won’t be surprised to learn a recent study by economists at Harvard and Columbia universities showed that assignment to one or another teacher is related to later life outcomes, such as the likelihood of early pregnancy, the chances of college attendance and lifetime earnings. Studies have consistently found that all public school systems have both high- and low-performing teachers. This is not just a problem for struggling urban districts. No matter where you live, there is a chance your child will be assigned to an ineffective teacher.

3. Annual testing is necessary to implement value-added metrics


Currently the ESEA requires that elementary and middle school students be tested in every grade beginning in grade 3. What would happen to the quality of information that would result if these assessments were cut back to one grade? First, as is probably intended by Darling-Hammond given her position on teacher evaluation,[ii] the ability to evaluate teachers based on test score value-added would vanish. This is because value-added calculations at the teacher level depend on the difference between the test scores of a teacher’s students at the end of the school year and the test scores of those same students at the end of the previous school year. The annual gain in test scores of the teacher’s students, with some additional statistical information, is the teacher’s value-added. Teachers whose students show greater gains have added more value to their students’ achievement than teachers whose students show lesser gains. Value-added can’t be calculated for many teachers as it is,[iii] but in the tested grades and subjects in which it can be estimated, it provides an important point of validation for other more widely deployed measures such as classroom observations.[iv] It is also the basis for calculating the school’s value-added, e.g., the test score gains between 3rd and 5th grade for all the students attending a particular elementary school relative to the gains in other elementary schools. But with only one test at each of three school levels, gains can’t be calculated for teachers or schools. The result is that value-added and the useful information it provides is out the door.
4. **Standardized tests are important tools in improving teacher quality despite their limitations**


Test scores are important because they’re objective measures of the schooling outcome. It’s appropriate to emphasize student achievement on math and reading tests because these are the building blocks for success, and far too few students attending public schools today adequately possess these basic skills. Developing new tests and the right methods for analyzing them can be costly. But their potential contribution to improving teacher quality $\bar{A} \bar{A} \bar{A}$ the single most important school-based factor for fostering student learning $\bar{A} \bar{A} \bar{A}$ far outweighs the upfront cost. Of course, test-score analysis can’t tell us everything we want to know about a teacher’s performance. Using it in isolation to evaluate teachers creates bad incentives and can miss a great deal of what makes a teacher effective. But research shows that evaluations of a teacher’s contribution to her student’s test scores this year is a far better predictor of how much her future students will learn than are the factors prioritized by the current system: years of experience and possession of advanced degrees. Failing to utilize such important and accessible information about a teacher’s effectiveness is scandalous. Standardized tests are imperfect measures of student achievement, and the statistical analyses that utilize such tests are imperfect tools for evaluating teachers. But despite their limitations, standardized tests provide important information about teacher quality that we should use to improve our terribly flawed system for evaluating teachers. New York City’s movement toward increased use of test scores to evaluate teachers is a step in the right direction.

5. **We need assessments to measure teacher effectiveness—this is the first step in improving quality**


But of all the current reform proposals related to teacher quality, the most fundamental are those that would change the way public-school teachers are evaluated. The reason is straightforward: Any reform policy linked to teacher quality requires first being able to accurately assess teacher performance. It would be impossible to compensate teachers based in part on their performance, or to improve the system’s ability to assign tenure, without an evaluation tool capable of distinguishing between the system’s best and worst teachers.

6. **Teacher quality is vital successful educational outcomes**


Great teachers matter enormously to the success of their students. While this seems like common sense, it has also been borne out by decades of empirical research: Several studies show that to which teacher a student is assigned makes a huge difference in determining how much that child will learn in a given school year. And many of the benefits of good teachers are more long-lasting. A new study by researchers at Harvard and Columbia, for instance, shows that a student assigned to a great teacher is less likely to have an early pregnancy, is more likely to attend college, and will earn a higher salary as an adult. In response to such findings, policymakers across the nation are aggressively pursuing policies to address teacher quality. At the federal level, teacher-quality reforms played a central role in President Obama’s Race to the Top program. And several states — including Colorado, Florida, Indiana, Connecticut, New Jersey, and New York — have recently adopted or seriously pursued reforms that would connect crucial employment decisions, such as those regarding salary and tenure, to teachers’ performance (rather than to seniority).

7. **Teacher quality plays an enormous role in education outcomes**


Policymakers on both sides of the aisle have either recently realized the truth or have finally become emboldened to state the obvious: There are some bad teachers in our schools. This conclusion has been supported by a robust body of research, which consistently finds that the differences among bad, average, and great teachers are substantial. For example, a study by Stanford University economist Eric Hanushek using data from Texas public schools found that, on average, students assigned to teachers whose quality is at the 25th percentile tend to make about half of a grade level’s worth of reading achievement gains during the course of a school year. In contrast, if those same students are assigned to a teacher at the 75th percentile, they tend to gain about one and a half grade levels in reading during the year. The difference between a child’s being assigned to one teacher or another can thus reasonably be as much as a grade level’s worth of learning during the school year. Studies have found similar results using data from school systems in several other states.
Testing Desirable: Answers to “Budget Tradeoffs”

1. The cost tradeoff argument is not persuasive—tests costs are a drop in the bucket


Testing critics usually point to estimates of total spending on assessments; a commonly cited figure—$1.7 billion spent by states each year—comes from a report I wrote in 2012. But what these claims always miss is that, however calculated, spending on testing is barely a drop in the bucket of a public education system that spends over $600 billion per year. If testing were eliminated entirely, what could schools do with the $1.7 billion saved? Very little, it turns out. Teacher salaries could be increased by one percent or pupil-teacher ratios could be reduced by 0.1 students. The $34 per student spent by states on federally and state-mandated tests simply isn’t very much in a system that spends about $10,000 per student. Put in the context of the NEA position, $34 per student would not buy very much early childhood education—only eight hours of preschool per student in Florida to be exact.

2. Cost arguments against testing are penny wise and pound foolish—they help level the playing field


A retreat from annual testing would be penny-wise and pound-foolish. Testing makes it possible to measure the performance of students, teachers, and schools, and enables research aimed at improving student learning. And annual tests help level the playing field between schools, enabling policymakers to judge schools based on how well they serve their students, rather than the type of students they serve. Critics of testing raise some legitimate concerns about the quality of tests. But those who deride “bubble tests” can’t have it both ways, as states generally can’t buy higher quality tests at a lower price. Upgrading assessments to make them more suitable for use in high-stakes decisions, such as by lessening the focus on multiple-choice questions, likely entails increased costs (especially for scoring). States are required to administer uniform tests, but whether they spend enough to get a high-quality test has become a politicized decision. In my 2012 study, I found many examples of states that spend well below the average. Those truly concerned about the prevalence of low-quality tests might ask Congress to restrict a small amount of federal education funding, such as $30 or $40 per student, to be spent only on assessments. This minor tweak to NCLB can be accomplished without an increase in total spending, and would insulate state policymakers from political pressure to compromise on assessment quality at the expense of students.

3. Tests are cheap, and the data is forcing the establishment to reform


A clear sign that the federal No Child Left Behind Act is fulfilling its main purpose in shaking up the ossified public-education establishment is how angry educators are about it. In August, Connecticut joined the growing chorus of states denouncing NCLB as an “unfunded mandate.” Connecticut’s attorney general filed a lawsuit arguing that that rich state couldn’t possibly afford to provide annual tests in basic reading and math skills to kids in the third through eighth grades, as the law requires. This is silly, since testing costs little. Educational firms sell off-the-shelf tests for as little as $2 per student; an elaborate one, written especially for a state’s curriculum, might cost the princely sum of $30 per kid. A pencil still costs less than a nickel. Connecticut already tests students in grades four, six, and eight and says that’s sufficient. But you only get one chance to educate a child. If a student is falling behind, parents and teachers shouldn’t have to wait two years to find out.
Testing Desirable: Answers to “Cheating Concerns”

Cheating concerns do not de-justify testing


Another oft-heard argument is that standardized tests drive educators to cheat. Teachers and administrators in the Atlanta public school system, for instance, were indicted this year in an alleged scheme of inflating their students’ test scores to avoid sanctions and secure performance-based bonuses. Not surprisingly, some education advocates were quick to blame the scandal on the tests themselves. It should be noted that most teachers are honest, dedicated professionals. But even if this sort of fraud were rampant, it would be absurd to fault standardized tests. As Thomas J. Kane, director of the Center for Education Policy Research at Harvard University, noted this spring, such a reaction would “be equivalent to saying ‘O.K., because there are some players that cheated in Major League Baseball, we should stop keeping score, because that only encourages people to take steroids.’ ”
Testing Desirable: Answers to “Creativity / Critical Thinking”

1. Tests do not sap teacher and student creativity—focus groups prove


Theme #3: Test-prep does not necessarily sap creativity for teachers or students. Several teachers disagreed with the notion that testing forces teachers to replace creative lessons with dreary test-prep. One math teacher said, "I used to think and really had a strong opinion that it caused me not to be a creative teacher, but I've changed now." Another teacher, one who was often critical of testing -- said that "true creative people" will "find a way to be creative regardless of what the framework is." When asked if tests discourage creativity, one teacher claimed that "I think it's done just the opposite... I think adding open response where you used to have tests that were just strictly multiple-choice has forced teachers to be more creative." From another school, a teacher said, "In science and math, we're expected to use manipulatives as much as possible and do hands-on as much as possible... I do so much more now, so much more." As such, many teachers praised the Arkansas tests for encouraging more creative and critical thinking by students. Due to the many open-response questions in math and English, students "have to understand it [the material] to get the points on that [test]; they can't just make a guess and maybe get it right." One teacher remarked that "as an older teacher," the introduction of benchmarks and tests "totally changed the way we taught because [formerly] we were just answering questions; we didn't have them think about, 'why is this the answer?'" In this way, the Arkansas tests encourage both teachers and students to be creative and to think more deeply about questions and answers. We were certainly surprised to find that teachers voicing the "testing prevents creativity" criticism were distinctly in the minority. One teacher in particular summed up the opinions of many other teachers: 'I think that [if] there's somebody who says, 'Oh all you do is teach to the test and you're not being very creative,' I think they haven't been in the classroom in a long time? either that or they've been in a classroom or a school where their principal is not doing their job." In other words, whether testing saps creativity depends on the teachers and principals. They could respond to the perceived pressure by resorting to dull lessons premised on rote memorization, but they could also seek creative strategies for ensuring student learning. For many teachers in our Arkansas sample, the testing regime fostered greater creativity.

2. Testing improves our inferential reasoning capabilities


Of course, remembering what is taught is only a small part of the process of becoming educated in a discipline. Being able to generalize and draw new inferences on the basis of the learned material is critically important if we want students to apply their learning to new situations. And there is evidence that quizzing facilitates the generalization and application of knowledge as well. In one study (Chan, McDermott, & Roediger, 2006), subjects read a set of passages and were either quizzed on their memory for selected facts from those passages or given extra study time. Prior quizzing enhanced memory for the non-quizzed (as well as the quizzed) aspects of the texts when compared with re-studying. The benefit persists after a rather long interval (7 days; Chan, 2010) and even when the final test material is quite distant from the original material being quizzed (Butler, 2010). In general, these benefits are most pronounced in cases in which the learners took what the authors called a broad approach to retrieving responses during the practice quiz: When they thought widely about lots of details relevant to the terms in the question—even if those details were not directly related to the sought-after answer—the benefits of quizzing on related but un-quizzed material were most pronounced. So it appears that testing benefits learning in part because of the way that it motivates learners to think about relations among learned facts, and in part because it encourages effective reorganization. Such a result is consistent with the finding reviewed earlier that short-answer tests benefit learners more than do multiple-choice tests, as short-answer tests presumably offer more opportunities for broad thinking. It is also consistent with the well-accepted finding in education that asking and answering “deep questions”—ones that focus on relations, logic, and causation, for example—dramatically benefit student learning (King, 1994).
Testing Desirable: Answers to “Curriculum Narrowing”

- Tests can be used to improve the curriculum


We will argue here that an important but overlooked part of the conversation about how those standards and the tests should be developed is the question of what happens to the student who is taking a test. Tests, we will show, provide an opportunity to enhance as well as measure American students’ education. We argue that a greater understanding of the benefits of taking tests can serve to defuse some of the concern over standardized testing and enable more concrete progress toward the shared goal of improving American education outcomes.
Testing Desirable: Answers to “Disparate Outcomes”

1. Testing is a civil rights issue—it is necessary to hold schools accountable


   Last week, after leaders of the new Republican-led Senate made it clear that rewriting No Child Left Behind is a top priority, Sen. Patty Murray, D-Wash., stood on the Senate floor and called standardized testing a civil-rights issue. “We know that if we don’t have ways to measure students’ progress, and if we don’t hold our states accountable, the victims will invariably be the kids from poor neighborhoods, children of color, and students with disabilities,” she said. Civil-rights advocates want to make sure that any rewriting of No Child Left Behind, the 2001 version of the Elementary and Secondary Education Act maintains regular testing and reporting of test score data. Yet even as they gear up for a debate over education standards, we don’t actually know how well standardized tests measure student learning.

   Nineteen groups, including the NAACP and the Children’s Defense Fund, recently released a statement backing the law’s core testing requirement. “ESEA must continue to require high-quality, annual statewide assessments for students in grades 3-8 and at least once in high school,” Wade Henderson, head of the Leadership Conference on Civil and Human Rights, said at a Senate hearing on Wednesday. Tests should track students’ progress toward state standards, said Henderson, and those standards have to align with what students need to know to succeed in college or in the workforce. Without data to show that students are on track, it could be all too easy for disadvantaged children to receive a substandard education.

2. Lack of data would make it harder to identify education gaps that harm minority students


   Some civil rights advocates have voiced similar concerns about accountability systems that rely exclusively on growth measures, which could allow schools serving disadvantaged students to avoid sanction even if their students’ academic progress is insufficient to close achievement gaps. This is a legitimate concern, and policymakers may want to strike a balance between average scores and growth when deciding where to focus improvement efforts. However, not administering the annual tests required to produce student growth measures would make it impossible to distinguish those schools where students learn very little from those that perform well despite difficult circumstances.
**Testing Desirable: Answers to “NCLB Fails”**

1. **NCLB has improved school quality despite its flaws**


   That's not to say that NCLB itself is flawless. It was an obviously imperfect scheme from the start. The law increased the role of the federal government in education to an uncomfortable level. Its requirement that every student reach proficiency by 2014 was always absurd. More important, but less well-appreciated, the law actually does a poor job of identifying ineffective schools. NCLB judges a school's quality by the percentage of its students with test scores above an arbitrary threshold – which is highly correlated with student backgrounds – rather than measuring the progress that its students make in a given year, which researchers consider to be a more accurate measure of the school’s influence on educational outcomes. And yet, despite its flaws, the best evidence to date suggests that NCLB has made schools better. A convincing recent study by Stanford University's Thomas Dee and University of Michigan's Brian Jacob found that NCLB produced substantial test score improvements. The gains were particularly meaningful among historically low-performing students.

2. **Most of their objections are to the other mandates, not the tests themselves**


   Most of the opponents of federally imposed standards, testing, and accountability should be in favor of federally imposed annual testing—shorn of standards and accountability Conservatives, generally, want to rein in federal control of education while driving bottom-up reforms by empowering parents with greater choice of where to send their children to school. Choice is empty without valid information on school performance (like going online to choose a restaurant for dinner and finding no reviews), and student learning is the most critical school function on which customers need performance data. Conservatives should favor a federal role in collecting and disseminating this information. And it doesn’t have to be the same test across the nation to provide this information, or even a single end-of-the-year test as opposed to a series of tests given across the year that can be rolled-up into an estimate of annual growth. All that is required is something that tests what a school intends to teach and is normed to a state or national population. Progressives have a strong commitment to educational equity and adequacy for historically disadvantaged populations. They think that funding is critical, but nearly all understand that how the money is spent and to what ends is equally important. One of the undeniable successes of NCLB was to expose to public scrutiny the failures of many of our public schools to adequately educate disadvantaged subgroups. If information on student learning from annual testing disappears, so too will the attention to the needs of subgroups that are illuminated through annual testing. Progressives should support annual testing for reasons of equity. Concerned parents are reacting to test prep regimes for annual tests, not the tests themselves (which take no more than a day of school time to administer). If the federal targets for test scores and associated sanctions are jettisoned, so too should much of the test prep regime. Test scores become, then, one among several forms of information on school performance that parents should value and consume. Parents who are concerned about their children’s schooling should want to know how their school of choice is performing on state tests, as well as the satisfaction of parents and students who are served by the school, the experience and effectiveness of its teachers, the extent to which the school prepares its students for the next step in their education journey, the school’s extracurricular activities and degree of student engagement, and other factors that people care about and can be made available for public scrutiny. Surely, such parents no more want to be in the dark about a K-12 school’s academic performance than they would want to ignore the quality of the college to which their child will eventually seek enrollment.

3. **Getting rid of testing is the wrong way to fix the issues with No Child Left Behind**


   An exclusive reliance on student performance levels, on the other hand, is perhaps the principal shortcoming of the much-maligned accountability system mandated by No Child Left Behind. Under that system, whether a school makes Adequate Yearly Progress is determined primarily based on the share of students scoring at proficient levels in math and reading in a given year. But a key reason Congress mandated such a system in 2002 was that many states were not yet administering annual tests, and many of those that did lacked the capacity to track the performance of individual students over time. Eliminating the annual testing requirement would therefore recreate the conditions that led to the adoption of a mistaken accountability system in the first place. Policymakers thus face a stark choice: require annual testing or settle for low-quality and potentially misleading information on school quality.
- The information provided by high-stakes tests is reliable—correlations with other tests


This report tackles that important policy issue by comparing schools’ results on high stakes tests with their results on other standardized tests that are not used for accountability purposes, and thus are “low stakes” tests. Schools have no incentive to manipulate scores on these nationally respected tests, which are administered around the same time as the high stakes tests. If high stakes tests and low stakes tests produce similar results, we can have confidence that the stakes attached to high stakes tests are not distorting test outcomes, and that high stakes test results accurately reflect student achievement. The report finds that score levels on high stakes tests closely track score levels on other tests, suggesting that high stakes tests provide reliable information on student performance. When a state’s high stakes test scores go up, we should have confidence that this represents real improvements in student learning. If schools are “teaching to the test,” they are doing so in a way that conveys useful general knowledge as measured by nationally respected low stakes tests. Test score levels are heavily influenced by factors that are outside schools’ control, such as student demographics, so some states use year-to-year score gains rather than score levels for accountability purposes. The report’s analysis of year-to-year score gains finds that some high stakes tests are less effective than others in measuring schools’ effects on student performance. The report also finds that Florida, which has the nation’s most aggressive high stakes testing program, has a very strong correlation between high and low stakes test results on both score levels and year-to-year score gains. This justifies a high level of confidence that Florida’s high stakes test is an accurate measure of both student performance and schools’ effects on that performance. The case of Florida shows that a properly designed high stakes accountability program can provide schools with an incentive to improve real learning rather than artificially improving test scores.
Testing Desirable: Answers to “Student Stress”

We can adapt the exams to deal with stress-based concerns


The third argument is that high-stakes testing places too much pressure on students. This objection is not without some merit. Having visited schools in other countries where a single five-day examination can determine a student’s future, I understand how tests can sometimes constitute cruel and unusual punishment. But surely there is a sensible middle ground between such brutal practices and full-scale abandonment of standardized testing. Finding that middle ground has never been more important, as U.S. students continue to fall far behind their international peers. In its most recent report, the World Economic Forum ranked U.S. math and science education 52nd in the world. A 2009 evaluation of students in 34 developed nations found that U.S. 15-year-olds were outperformed in science by students from 12 countries. The results were worse in math: Students in 17 countries outperformed U.S. students.
Testing Desirable: Answers to “Teach to the Test”

1. **The negative effects of ‘teaching to the test’ are exaggerated**


   Opponents of standardized tests typically rely on three basic arguments. First, they contend that these exams detract from the larger goals of education by encouraging teachers to “teach the test.” In a certain sense, however, teaching the test is the whole point. Exams are instruments for measuring student proficiency. And, as I’ve learned during my career in the business world, measuring something is often the best way to maximize or improve it. Economist Dan Ariely of Duke University has said: “CEOs care about stock value because that’s how we measure them. If we want to change what they care about, we should change what we measure.” If an exam effectively gauges a student’s mastery of U.S. history or English grammar, then teaching the test is simply a matter of helping students develop that knowledge. Teachers who feel that a test ignores something essential should commit to fixing the test, not condemning the entire practice of testing.

2. **Students still learn even if some instructors are ‘teaching to the test’**


   The key to good standardized testing is to test twice a year, both spring and fall, to prepare children and to communicate to them that this is just one vehicle to help assess their achievement. Therefore, it is valuable to get the students’ feedback, so that they can have a voice in the process and be invested — not only in the remediation phase, but also in future test outcomes. Then there are alternative assessment models such as classroom performance, teacher observation, subject testing and communication, which offer a balanced and equivalent evaluation in high stake decisions such as tracking, high school graduation, or retention. Finally, it is true that some teachers will find the need to teach to a standard test. But even in this instance, it has been shown that children can get the benefit of the overall curriculum requirements for their grade level. So let's not throw the baby out with the bathwater, but rather use the test format as a catalyst to both examine teaching pedagogy and evaluate student performance.
Testing Desirable: Answers to “Teacher Opposition”

1. Testing can improve teacher collaboration—focus groups prove

Theme #4: Testing can lead to collaboration. Many teachers claimed that the Arkansas exams encouraged teacher collaboration to ensure that students are well-prepared. As one teacher said, testing “does give us a common goal, and we can work together.” From a different school, a teacher asserted that “I feel like we have a lot more open policy teacher-to-teacher now because everybody's reaching for the best way... It's just so worth it to collaborate with your other teachers on better ideas and seeing what they do that you can do better. Teachers at one school had adopted a professional learning community (PLC) model (DuFour 2004). Several teachers in that school said that they used to hate testing, but that they now thought testing had introduced a greater sense of collaboration in ensuring that each child learned everything that he or she was supposed to know. Teachers at the PLC school reported that they regularly gathered to decide on the “eight to 10 essential standards and try to focus on just those.” Most teachers rose to the challenge that the testing provided by teaming together to enhance student achievement.

2. Many teachers find tests to be useful—help structure the curriculum

Theme #2: Testing and standards help create a road map for the year's instruction. Many teachers noted that before testing, it was easy to teach idiosyncratically -- perhaps spending "six weeks on the dinosaur unit and just totally ignoring]" other topics. With increased focus on testing, however, teachers have focused on matching their instruction to a coherent set of standards. Thus, one math teacher said that while she had initially "hated" the Arkansas benchmark tests, she has since changed her mind: "I'm OK with it now, to be honest; I see where knowing the standards and knowing what's going to be tested can help me plan the whole year and make sure I've covered everything." One additional benefit of standardization ? a "lifesaver," as one teacher put it, is that districts can collaborate in setting a consistent schedule for teaching the state standards, which means that "if a child moves from [neighboring town] to our school... they should be right where we are." This point assumes that a state's tests are aligned with the state's curricular standards; without such alignment, schools would face conflicting signals and would therefore have a difficult time creating a consistent road map. Most teachers agreed that the Arkansas tests do a fairly good job of matching the Arkansas standards. This finding is also in line with an argument made by many advocates for testing and standards ? that standards foster

3. Focus groups show that many teachers think that tests are valuable

What happened next was the interesting part. After the surprisingly short introductory period in which teachers voiced complaints, teachers began talking about good aspects of the Arkansas exams. One teacher followed the other, with no prompting from the researchers leading the focus groups. In the end, teachers said many good things about various aspects of the testing process and, overall, gave a very positive impression of the effects of the annual assessments on classroom teaching. After we sifted through all of the comments from all of the teachers at all of the school sites, five positive themes emerged. The consensus of teachers with whom we spoke was that the tests provide useful data, that the testing regime helps create a road map for the year's instruction, that the standards and tests don't sap creativity or hinder collaboration, and, perhaps most surprising, that the accountability
Testing Desirable: Answers to “Teachers Adequately Evaluated Now”

1. Most teachers are not meaningfully evaluated


The lack of a rigorous teacher-evaluation method is hardly a new problem in American education: It long pre-dates collective bargaining and powerful teachers' unions. Consider that, as recently as 2006, only 0.89% of New York City's teachers received "unsatisfactory" ratings on their official evaluations. This is roughly in keeping with the results from nearly a century ago: A 1914 study of New York City's teacher-evaluation system found that only 0.5% of elementary-school teachers evaluated in the city's schools were considered to be deficient in instruction, and only 0.9% were considered non-meritorious in discipline. The author of that long-ago study argued that New York's evaluation system would surely have to change in order to more accurately assess teachers' effectiveness. But it wasn't until earlier this year that the city finally began taking some incremental steps toward improvement — when Mayor Michael Bloomberg proposed a new evaluation tool that would incorporate teachers' observed performance in the classroom as well as their contributions to students' standardized-test scores. Meanwhile, most of the country is still languishing under the old rubber-stamp evaluation system that does nothing to meaningfully distinguish between excellent teachers and terrible ones. According to most evaluation systems in use today, there are in fact hardly any low-performing public-school teachers in the United States; in any given school system, nearly all of the teachers are rated as effective. Indeed, a 2009 report by the New Teacher Project looked at outcomes from teacher evaluations in 12 school districts across the United States and found that these systems rated fewer than 1% of their teachers as "unsatisfactory."

2. Teachers do not receive meaningful evaluations— their effectiveness is often rubber-stamped


That discussion must begin with why it is that, even though everyone knows some teachers don't belong in the classroom, the current system continues to rubber-stamp them as qualified. After all, in any profession, trade, or job, there will be some duds and some superstars. That such distinctions should exist among the nation's 3.3 million public-school teachers seems self-evident. And yet the current system nevertheless officially rates all teachers as being equally effective. One significant part of the problem is that, precisely because poor teacher evaluations are given so rarely, the designation implies that the recipient is not merely "unsatisfactory" but egregiously incompetent. This is often a much stronger message than a principal intends to send, and he is then left with no way to distinguish on an official evaluation between an ineffective teacher who could use some remediation and a hapless one who should find a new career. The default is just to list the teacher's performance as "satisfactory" or higher. This practice hurts not only schools and students, but also struggling teachers in need of constructive feedback. Another major problem is the basis for teacher evaluations: The current system relies entirely on limited observations of a teacher's performance in the classroom. Typically, a principal sits in on a teacher's classroom session at a time both have agreed on and assesses the teacher's performance according to a protocol, evaluating the teacher's classroom-management techniques, the set-up of the room, and the quality of his lesson plan.
Testing Undesirable: Topshelf

1. High stakes tests hurt students—13 independent reasons


As we enter the March Madness of testing season, many parents and teachers have become increasingly concerned that the high-stakes attached to so many tests are actually harming our students and schools. There is particular concern about the disproportionate impact high-stakes-testing may be having on our poorest students, most struggling students, English Language Learners, and students of color.

So what are the “high-stakes” for students in high-stakes testing? Examples we’ve been hearing from parents and educators across Pennsylvania include: Lost learning time: There’s less time for learning with testing and test prep (for example, Pittsburgh students now take 20-25, or more, high-stakes tests a year, with new tests this year in art and music). Reduced content knowledge: Students are learning how to take high-stakes-tests, but cannot demonstrate subject mastery when tested in a different format. [Koretz, 2008]

Narrowed curriculum: With a focus on reading and math scores, students lose history, world languages, the arts, and other programs. Shut out of programs: Stakes exclude students when test results count as extra weight in magnet lotteries or for entrance to gifted programs or advanced courses. Diverted resources: Schools that perform poorly on high-stakes-tests are labeled “failures” and sometimes have resources taken away from them. The hundreds of millions of dollars spent on testing in Pennsylvania are not available for classroom education. School closures: Schools labeled as “failing” on the basis of test scores can be threatened with closure. These schools are usually in communities of color. Loss of curiosity and love of learning: Bubble tests are developmentally inappropriate for the youngest learners. Emphasis on “skill drill and kill” fails to stimulate children’s imagination and limits their natural curiosity.

Blocked access to facilities: Many schools find their computer labs taken over by testing for weeks on end and not available for learning. Harmful stress: Children are pressured to not only demonstrate their knowledge but to represent the effectiveness of their teachers and their schools. Teachers are reporting children throwing up, losing control of their bowels, and increased commitments for psychiatric and anxiety issues. Internalized failure: Struggling students who are forced to repeatedly take normed tests (which are designed to fail a certain portion of test-takers) begin to believe they are “bad” or “worthless” students who cannot succeed in school. Grades: Some high-stakes tests are included in students’ grades. Graduation requirements: The NAACP has protested Keystone graduation exams, saying they force too many children out of school on the basis of a single score. Altered school culture: Schools are usually in communities of color. Loss of curiosity and love of learning: Bubble tests are developmentally inappropriate for the youngest learners. Emphasis on “skill drill and kill” fails to stimulate children’s imagination and limits their natural curiosity.

2. Our obsession with tests is undermining our entire education system


While we believe strongly in the promise and potential of the Common Core, and we've seen good implementation firsthand, these standards will be meaningless if policymakers keep reducing them to a test score. Even worse, we are losing the promise and purpose of public education -- and the joy -- by trying to reduce everything about teaching and learning, whether for students or teachers, to a number or algorithm. Test scores and black-box algorithms can't help children critically think and problem solve. These strategies don't help children build trusting relationships and instill confidence and persistence. And they ignore the countless other ways educators nurture and develop our children. The fixation on testing and data over everything else is a fundamental flaw in how our nation approaches public education. No other nation in the world tests every student nearly every year. And no other nation relies so heavily on a test score to rank and sort teachers. And we continue to move in the wrong direction. Just look at what's happened with the over-reliance on tests and value-added methodology (VAM). VAM is an incomprehensible formula, at least to those who don't have a Ph.D. in advanced statistics, which attempts to predict how a teacher's students will score in the future by using past test scores and other various assumptions -- and then compares that prediction to actual results. Like predicting the weather, VAM is subject to many factors that influence the final result. That VAM score is then used to sort, rank and evaluate teachers. The AFT has always been leery about VAM -- and we've said since day one that VAM should never be the singular measure of student learning used to evaluate teachers. In fact, I questioned the fairness, accuracy and reliability of value-added metrics in a 2007 New York Times column. We have enough evidence today to make it clear that not only has VAM not worked, it's been really destructive and it's emboldened those seeking to turn public education into a numbers game.
3. High-stakes testing incentivizes stripping everything else out of the curriculum—it leaves our students unprepared to face the challenges that life throws at them


However, the problem arises when these assessments are used as virtually the sole indicators of schools’ successes and failures. Noddings (2007) describes the resulting dilemma: “Today, even elementary teachers complain that the pressure to produce high test scores inhibits the work they regard as central to their mission: the development of caring and competent people” (69). Because the stakes have reached disproportionate levels, educators are often forced to abandon all things unrelated to the test and consequently lose sight of what is important: the whole child, who is not simply composed of intellect but is emotional and spiritual as well. Greene (2007) speaks of the broader scope of education and cautions that if not put into perspective, the current preoccupation with academic rigor and standards may be for naught: “The existential contexts of education reach far beyond what is conceived of in Goals 2000. They have to do with the human condition in these often desolate days, and in some ways they make the notions of world-class achievement, benchmarks, and the rest seem superficial and limited, if not absurd” (32). What good is it if our students are academically successful, yet lack what is necessary to cope with the more difficult life issues they will face and the interpersonal skills needed to coexist in the global economy for which they are ultimately being prepared? By limiting their educational experience to only one facet of their being, we are unable to adequately nurture their personal growth, which is equally important to their success in life.

4. High-stakes test only serve to narrow the curriculum, undermining innovation and forcing teachers to ‘teach to the test’


According to Common Core claims that it is based on the idea that "students should be able to think critically rather than just memorize material for tests. But according to a report on Business Insider, "Common Core and the tests tied to those standards might prevent students from achieving that goal. Those rigorous tests could discourage teachers from being creative and force them to teach to the test" because teachers are being evaluated based on mandated improvement in student test scores, what is also known as a Value-Added Model. The report quotes Michael Benezra, a legislative director for the Massachusetts Senate, who told Business Insider "The reliance on testing pigeonholes the teachers to teach only to the test . . . "I think it's kind of counterintuitive to students getting the big picture because they're required to test so much. In order to perform well on the test, you have to memorize things. ... You can say we're trying to get them to think more critically and read closely ... but at the end, the students take a test, they don't write a long essay where they're forced to think deeply about the issue."

5. Standardized tests hurt the entire education system


Disagree and take this opportunity to point out other disadvantages that attend our obsession with standardized tests: they are unfair to whites and bad for society at large. Standardized tests damage mainstream social institutions in three ways. They encourage test-takers—that is to say, most people—to cultivate a narrow form of intelligence. They relegate many people whose intelligence is not narrow to low-level jobs. And they contribute to the slow decline of the societies that rely on testing to select undergraduates, lawyers, firefighters, and police. This decline has been especially perceptible in sectors such as education and law, where narrow intelligence is a serious handicap.
6. **High-stakes testing undermines the curriculum—multiple reasons**


High-stakes testing not only negatively affects motivation and learning, it also undermines the curriculum. Because of the increased pressure on teachers for their children to do well on standardized tests, the curriculum has been narrowed. The curriculum, and thus instructional time, has shifted to only those areas that are to be tested. In many instances, the time given to art, music, creative writing, physical education, and recess has either been reduced or dropped altogether in favor of more intensive drilling on the test subjects (Amrein & Berliner, 2003). With the advent of Reading First grants, specific curriculum and materials used to teach are now being mandated, which narrows the curriculum even further. Low-performing schools can apply for these federal monies; in order to receive the grant, however, the schools must use government-approved materials and teachers must be trained by government-approved providers (Garan, 2004). No longer is teachers’ professional judgment about curriculum and instruction valued. It has been replaced with curriculum deemed valuable by the federal government as a means to achieving high scores on standardized tests. Teachers report that the pressure to do well on the tests hinders their instructional practice (Pedulla, Abrams, Madaus, Russell, Ramos, & Miao, 2003). They are forced to teach in ways that are not developmentally appropriate and do not promote critical thinking and decision-making. Rather, instruction has become mundane and boring as children complete worksheets on basic facts and memorize items for the test. Instruction has been reduced to teaching to the test. The very instructional strategies that should be used to create and promote democratic values in the classroom are now replaced with mundane skill-drill-kill exercises whereby children do not think for themselves, critically examine possibilities, or take risks. The very heart of democracy has been stripped from our public schools in the name of high-stakes test scores.

7. **Standardized tests produce results that only reflect access disparities in the student population**


A second line of criticism is that standardized testing (and scoring) is insensitive to the diversity of our student population. The diversity is not only in diverse ethnic and racial backgrounds. Different socio ethnic backgrounds as well make a crucial difference (Hedges & Nowell, 1998). Students from educated families have an advantage. Their parents provide books and other educational paraphernalia at home. Such families’ expectation of their children, and support for them (emotional, aspirational and in various other forms) set them apart from students of low socio economic and un- or ill-educated families (Amrein & Berliner, 2002). Differently advantaged or disadvantaged students call for different approaches to teaching them. Standardized testing makes them impossible.

8. **High-stakes tests are a mistake—risk ruining people’s lives over a single exam**


In Aldous Huxley's classic novel Brave New World, he paints the bone-chilling picture of a dark future of oppression. Citizens are divided into castes at birth, their jobs and economic positions predetermined. The future of the individual is taken out of his or her control and thrust squarely into the hands of the establishment. Because of the volatile political environment at the time it was written, Huxley's book struck a chord with his contemporaries and now, in 2013, parts of it still seem all too relevant. As we continue to see the impact high-stakes testing has had on our youth, especially youth of color, it's hard to say that lessons from this cautionary tale from the 1930s don't still apply today. In fact, the model currently used for standardized exams was first created by Frederick Kelly back in 1914, 18 years before Huxley ever put pen to paper. Across the country today, our children's futures are being decided by this antiquated method of evaluation. From No Child Left Behind, to Race to the Top and now to the Common Core, standardized tests are becoming more and more high-stakes. Grade promotion and graduation are often dependent on test scores; teachers and administrators receive financial bonuses for high scores but can be reassigned or even fired for reporting low marks; schools are either shamed or commended through public rating systems; and more and more, schools are being closed or converted to charter schools because of unsatisfactory test results. Indeed, in the name of "accountability," a single negative performance can unfairly jeopardize a young person's education and career prospects, derail a teacher's career and put a Scarlet Letter on a school. In cities and towns nationwide, students, teachers, parents, communities and others are banding together to say enough is enough. They are tired of seeing young people -- and especially students of color -- being deliberately pushed out to raise test scores. They are tired of racially biased assessments dictating their diplomas. They are tired of having the creativity sucked out of their curriculum and of having to follow unrealistic pacing guides. They are tired of always being on pins and needles wondering whether their school is next on the chopping block because of its low test scores. Frankly, they are tired of seeing an outdated model dictate every aspect of the modern education system.
Testing Undesirable: Bias—General

1. Tests cannot be viewed as reliable—they are biased against women and students of color


Standardized tests can not accurately measure intellectual merit because racial and gender stereotypes interfere with the intellectual functioning of those taking the tests, according to Stanford psychology Professor Claude Steele. Steele reported his findings at the annual convention of the American Psychological Association on Saturday, Aug. 12, in New York City. The meeting also featured a task force report on what is known about genetic links to intelligence, and three other sessions devoted to the controversy over The Bell Curve. In that 1994 book, Richard Herrnstein and Charles Murray argued that at least some part of group differences on IQ tests is inherited, reflecting some innate intellectual inferiority of African Americans. In a symposium about several research projects at Stanford, the University of Michigan and the State University of New York, Steele detailed experiments on factors that can depress the academic performance of women and African Americans in college environments. His seven-year research project bears on three currently controversial issues: How to interpret racial differences in test performance. The way selective universities and others interpret standardized test scores in implementing affirmative action policies in admissions; and The degree to which racial differences in college performance can be eliminated by appropriately designed schooling. In laboratory testing at Stanford and in a field program at the University of Michigan, Steele found that a dynamic that he calls "stereotype vulnerability" may be responsible for depressed performance. He also found that the performance gaps between men and women in mathematics, and between whites and African Americans as expressed in test scores, grades, and dropout rates, can be eliminated with appropriately designed affirmative action programs. "These findings demonstrate another process that may be contributing to racial and gender differences in standardized test performance, a process that is an alternative to the genetic interpretation suggested in The Bell Curve," Steele said before leaving for the conference. "And they show that group differences in school achievement can be reduced substantially by programs that emphasize challenge instead of a 'dumbing down' remediation." Steele also said the findings "underscore the danger of relying too heavily on standardized test results in college admissions or otherwise. The research shows that societal stereotypes can systematically depress the test performance of some groups more than others, even when those groups enter the test situation with equal knowledge." Steele's research - conducted since 1991 in the psychology department at Stanford with Joshua Aronson, now an assistant professor at the University of Texas at Austin, graduate students Joseph Brown and Kirsten Stoutemyer, and with Steven Spencer at the State University of New York at Buffalo - is supported by grants from the Russell Sage Foundation and the National Institute of Mental Health. It is the latest entry in a century-long controversy over alleged intelligence differences among groups such as European, African and Asian Americans, or women and men. Psychologists periodically argue over whether group differences on standardized tests stem from genetic differences and are thus more difficult to eradicate, or from environmental differences between groups, which are easier to change. Still others argue they merely reflect bias in the tests. "To this set of explanations, our findings add a new possibility - that stereotype vulnerability and its differential impact on groups in the immediate testing situation" are responsible for a difference in performance, Steele wrote in a paper prepared for the annual meeting of the psychology association, an organization of 132,000 researchers, educators, clinicians, consultants and students organized into four dozen sub-fields of psychology. (Stanford Psychology Professor Philip Zimbardo participated at the convention in a symposium on shyness; psychology graduate student Lisa Stallworth presented research she did with Stanford Professor Felicia Pratto in another session on militaristic and nationalist attitudes.)

2. Tests do not measure merit—are inappropriate for female and minority test takers


Studies by Professor Claude Steele at the University of Michigan and Stanford found that standardized tests do not accurately measure intellectual merit because racial and gender stereotypes contribute to "stereotype vulnerability" that interferes with the intellectual functioning of female and minority test takers. Steele concluded that "societal stereotypes can systematically depress the test performance of some groups more than others, even when those groups enter the test situation with equal knowledge."
3. Disparate results demonstrate that we just need to quit relying on the tests


What these recent developments make clear is that instead of setting different educational benchmarks for groups based on race or income, it may simply be time for us to stop relying so heavily on standardized tests to begin with. Though opinions differ as to why, on K-12 achievement tests and college entrance exams, lower income students, as well as black and Latino students, consistently score below privileged white and Asian students. These gaps persist despite decades of research and numerous studies attempting to explain and then close them. One theory suggests that students with grandparents who have graduated from college always score higher, suggesting that the tests unfairly penalize students who are the first in their family to attend college. Whatever the explanation, it is difficult to reconcile why we rely on such tests when we know that they so heavily advantage some and disadvantage others.
Testing Undesirable: Bias—Class

1. The wealth gap in test results is enormous


And if the standardized testing gap between racial minorities is bad, it’s nothing compared to the gap between the poor and the wealthy. For example, one recent study by the Annie E. Casey Foundation found that the gap for achievement test scores between rich and poor have grown by almost 60% since the 1960s and are now almost twice as large as the gap between white students and children of other races. The playing field is far from level when we continue to use tests where we know at the outset that wealthy students will do better than less wealthy students and white and Asian students will outperform blacks and Latinos.

2. Many standardized tests are really measuring the socioeconomic status of students


One thing it seems standardized tests are exceptionally good at measuring is socioeconomic status. In Class and Schools, Rothstein argues that this is because wealthier students have parents who can spend more time with them and more money on enrichment programs for them. He also writes that wealthy students also generally have better health and more housing stability than their lower-income peers, both of which also lead to higher achievement. Many of these shortcomings are inherent to these types of standardized tests. But some problems with the tests administered to children in many states are easily avoidable. A 2013 investigation by Heather Vogell of The Atlanta Journal-Constitution found that problems like poorly-worded questions, missing pages in exam booklets, and malfunctions in answer-sheet scanners were commonplace in high-stakes standardized tests administered in states across the country, and that “the vast majority of states have experienced testing problems—some repeatedly.”

3. Our tests are biased—do not account for diversities across areas and student populations


To begin with, “Are standardized tests reliable?” America is a vast country with very diverse cultures. Since standardized testing is used on a wide variety of children throughout the nation, it does not take into consideration that “one test does not fit all”. Standards and commonalities within a state are used to develop standardized tests. The culturally diverse areas and common biases in a region are not considered. Even when using statistical tools to reduce bias, guarantees of a bias-free test form or content is not assured. Yet, despite the built-in bias the tests continue to be used by testing organizations. Even with differences in the item content and claims by test-makers that the same things are being measured, the result usually produces very mixed results (Kumeh, 2011; Poham, 2001).
Testing Undesirable: Bias—Cultural

1. Cultural bias means that these approaches cannot increase achievement—we are teaching the wrong students to the wrong test


School districts employing this approach feel that it will lead to improved standardized test scores. Unfortunately, it may also prove detrimental to inner-city students in the long run because institutions of higher learning and many employers value critical thinking and independent thought. If inner-city students are not taught how to think in this manner, while peers at suburban schools are, then they are placed at a severe disadvantage. If the testing procedures are culturally biased, then the testing requirement represents bad public policy because the tests are not accurately measuring student achievement. An investigation into whether the tests are culturally biased, however, is beyond the scope of this Note. Even if, arguendo, the tests are not culturally biased, the tests still represent bad public policy because many schools have adopted a policy of teaching to the test, which comes at the expense of critical thinking and independent thought.

2. Testing requirements will only hurt inner-city schools—test itself is culturally biased, too expensive, and hurts the schools that need the most help


The testing requirement of the NCLBA is fundamentally flawed and will ultimately have a negative impact on inner-city schools. The requirement is bad public policy because it takes a “tough love” approach that, like the “War on Drugs,” results in increased sanctions but does little to remedy the underlying problem. Additionally, it is bad public policy to evaluate schools and students on a possibly culturally biased test and to force school districts to adopt instructional strategies aimed at “teaching to the test.” The punitive sanctions imposed on schools failing to meet standards create problems such as impediments to hiring and retaining quality faculty, overcrowding, diverting Title I funding from needy schools, encouraging the development of manipulative strategies, and contributing to decreased morale at underperforming schools. Finally, the testing requirement is flawed due to the exorbitant cost of implementation. This money would be better spent on decreasing class sizes, modernizing facilities, and increasing faculty salaries.

3. Test-based punitive sanctions will only widen the achievement gap, and tests themselves are culturally biased


The NCLBA testing requirement constitutes bad public policy for several reasons. For one, the punitive sanctions imposed on poorly performing schools will do little to close the achievement gap, akin to the “War on Drugs,” resulting in an increased amount of drug-related arrests but not a reduction in illicit drug use. Additionally, a considerable body of literature argues that the tests mandated by the Act are culturally biased in favor of mainstream white culture. If they are indeed biased in this regard, then it is bad public policy to attach so much weight to the results of tests that do not accurately measure all students’ achievement equally. This is an especially salient issue for inner-city schools because a majority of students from these schools come from minority backgrounds. Finally, even if the tests are not culturally biased, the requirement still represents bad public policy because it has prompted many school districts to adopt instruction methods aimed at “teaching to the test.”
4. High-stakes testing undermines educational performance—encourages poor learning methods, tests themselves may be racist


Critics of the NCLBA, on the other hand, argue that the high-stakes nature of the standardized tests that the Act mandates result in overly choreographed classes emphasizing mechanical, rote memorization, and repetition in an effort to "teach to the test." School districts adopting this model often require teachers to take a commanding, authoritarian approach to instruction and allow little room for creativity, collaboration, and community. Moreover, critics argue that this model, which many districts have adopted in response to the NCLBA, may also be racist. They argue that teachers at predominantly white schools that are not in jeopardy of failing under the Act's guidelines have the latitude to employ methods emphasizing creativity, independent thought, and student input. Minority students who do not benefit from classrooms employing critical thought are not only put at a disadvantage, but are stigmatized by the conveyance of the message that they should not strive to be critical thinkers but instead should merely regurgitate information. This stigmatizing effect, critics argue, is exacerbated when their schools are labeled and characterized as "failing" under the NCLBA.

5. Culturally biased tests are very damaging to students


In inner-city schools with a high percentage of non-native English speakers and non-standard English speakers, culturally biased tests can have particularly damaging effects. For instance, if students do not comprehend the scope of the questions being posed due to a lack of familiarity with standard English, they are less likely to answer correctly. For instance, this type of scoring inaccuracy (or anomaly) may occur when a native Spanish speaker responds to a testing prompt asking him to describe his parents by writing about his cousins. The student may respond in this manner because he read the word "parents" and assumed it had the same meaning as "parientes," the Spanish word for "relatives." The test, therefore, may end up as more of an elaborate spelling and vocabulary exam, rather than an accurate assessment of students' abilities. Considering that the purpose of the tests is to determine students' abilities, it would be appropriate to provide the option of having tests written and administered in the student's native language.

6. The tests are poorly considered—display a high degree of cultural bias


The tests may also be biased due to students' cultural backgrounds and environments. An example is a third grade urban student who has never left the city being asked on a test to describe a scene easily recognizable to a student from a mainstream white cultural background, such as a camping ground. If this student scores poorly on this problem it is unclear if it is due to a lack of exposure to different environments or to a lack of ability. If the tests do not take these factors into account, they are ineffective in accurately assessing students' abilities and the efficacy of instruction at their schools. The tests, arguably, have difficulty measuring forms of ability that are not valued by white, mainstream America. Additionally, many inner-city students come from backgrounds that do not consider success on standardized tests to be important. Students raised in such an environment, therefore, are less likely to perform well on the test yet schools are held responsible for the pervasive attitudes of students' cultural backgrounds. The testing requirement may also severely stigmatize inner-city students. The tests administered pursuant to the Act are used not only to evaluate schools but also to evaluate student progress. Students labeled as "failing" under a testing framework that may not accurately assess their abilities, may lower their expectations for academic success. As a result, they may be more likely to be inattentive in the classroom, avoid homework assignments, avoid school during test time, or even drop out of school. Such consequences further increase the difficulty inner-city schools have in meeting standards.
Testing Undesirable: Bias—Race

1. Standardized tests results on reflect inequities in our education system


Let’s assume this argument is made in good faith, rather than as a cover for pursuing a standards-and-testing agenda for other reasons. Moreover, let’s immediately concede the major premise here, that low-income minority students have been badly served for years. The problem is that the cure is in many ways worse than the disease—and not only because of the preceding eight facts, which remain both stubbornly true and painfully relevant to testing in the inner city. As Sen. Paul Wellstone, D-Minn., put it in a speech delivered last spring: “Making students accountable for test scores works well on a bumper sticker, and it allows many politicians to look good by saying that they will not tolerate failure. But it represents a hollow promise. Far from improving education, high-stakes testing marks a major retreat from fairness, from accuracy, from quality, and from equity.” Here’s why. *The tests may be biased. For decades, critics have complained that many standardized tests are unfair because the questions require a set of knowledge and skills more likely to be possessed by children from a privileged background. The discriminatory effect is particularly pronounced with norm-referenced tests, where the imperative to spread out the scores often produces questions that tap knowledge gained outside of school. This, as W. James Popham argues, provides a powerful advantage to students whose parents are affluent and well-educated. It’s more than a little ironic to rely on biased tests to “close the gap” between rich and poor. *Guess who can afford better test preparation. When the stakes rise, people seek help anywhere they can find it, and companies eager to profit from this desperation by selling test-prep materials and services have begun to appear on the scene, most recently tailoring their products to state exams. Naturally, affluent families, schools, and districts are better able to afford such products, and the most effective versions of such products, thereby exacerbating the inequity of such testing. Moreover, when poorer schools do manage to scrape together the money to buy these materials, it’s often at the expense of books and other educational resources that they really need. *The quality of instruction declines most for those who have least. Standardized tests tend to measure the temporary acquisition of facts and skills, including the skill of test-taking itself, more than genuine understanding. To that extent, the fact that such tests are more likely to be used and emphasized in schools with higher percentages of minority students (a fact that has been empirically verified) predictably results in poorer-quality teaching in such schools. The use of a high-stakes strategy only underscores the preoccupation with these tests and, as a result, accelerates a reliance on direct-instruction techniques and endless practice tests. “Skills-based instruction, the type to which most children of color are subjected, tends to foster low-level uniformity and subvert academic potential,” as Dorothy Strickland, an African-American professor at Rutgers University, has remarked.

2. Our testing focus benefits white and Asian students at the expense of others


Do standardized achievement tests unfairly advantage white and Asian students and disadvantage the rest? According to a group of educational organizations and civil rights groups the answer is yes. The recently filed a complaint with the U.S. Department of Education pointing out that black and Latino students in New York score below whites and Asians on standardized tests so consistently that although they are almost 70% of the overall student body, they are only 11% of students enrolled at elite public schools. As a result, the complaint argues that New York City is in violation of the 1964 Civil Rights Act because schools rely on a test that advantages one racial group over another.
Testing Undesirable: Cheating Incentives

1. High-stakes testing makes cheating inevitable—too much is on the line


The Ohio state auditor is investigating the practice of “scrubbing,” or dropping students from attendance rolls so they don't count against test scores. The former El Paso superintendent is in prison for using truant officers to encourage at-risk students to drop out. Other testing scandals have popped up in Mobile, Dallas, Houston, Detroit, Baltimore, St. Louis, and East St. Louis, Illinois. And everywhere USA Today looked in a 2011 investigation, they discovered statistically improbable aberrations in test scores, identifying 1,610 examples of anomalies that an Arizona State University professor compared to "a weight-loss clinic where you lose 100 pounds a day.” Organized, systematic cheating is the inevitable result of attaching high stakes to standardized tests, and it will continue as long as we're invested in the illusion that the system is working. The latest example came late on Good Friday when an Atlanta grand jury indicted 35 teachers, administrators and principals under laws meant to target the mafia. Dr. Beverly Hall, the since-retired superintendent of Atlanta schools, is facing charges of "racketeering, theft, influencing witnesses, conspiracy and making false statements.” In 2009, she was named National Superintendent of the Year and praised by Education Secretary Arne Duncan. Atlanta business leaders tried to get former Gov. Sonny Perdue to back off his investigation. Everyone wanted it to be true. The real problem with uncovering test-cheating scandals isn't that they're hard to find, it's that it's hard to get education officials to look. This is a glaring hole in No Child Left Behind.

2. The tests backfire—they only encourage cheating


The pretense that we can measure everything using sophisticated algorithms that no one can explain, that the profit motive or punishment are the only or best way to motivate human behavior, or that when people fail or are left behind it is because of their own weaknesses, are perverting American culture and transforming this country into a nation of high-anxiety cheaters always searching for an edge. It is easy to put the blame for cheating on teachers and administrators, as they did in Atlanta. But cheating on high-stakes tests in the United States is endemic and systemic. The school reform “Texas Miracle” that helped propel George Bush to the presidency was based on falsified data. According to a Government Accountability Office report evidence of organized institutional cheating was confirmed for at least one standardized test in 33 states in the school years 2010-11 and 2011-2012 alone. Thirty-two of the states decided to cancel, invalidate, or nullify test scores because of the suspected cheating. It is as if the tests are designed to turn us all into cheats. The Fiscal Times claims “corruption is as American as apple pie.” I agree, but add when money is on the line - a good reason to oppose the outsourcing of education to private for-profit companies. Tom "The Deflator” Brady, Lance Armstrong, Alex "A-Roid" Rodriguez, Marion Jones, Roger Clemens, and Barry Bonds, Brian Williams, Wall Street hedge fund managers, Bernie Madoff, and Bill Clinton - cheaters all. American society likes to dismiss individuals as cheaters but ignore the incentives created by a highly competitive society where the rewards for crossing the line without getting caught are astronomical and after all, everybody cheats. Just last week, four of the world's largest and most powerful banks, Citigroup, JP Morgan Chase, Barclays, and Royal Bank of Scotland pleaded guilty to conspiring to manipulate the value of world currencies.
3. **High-stakes testing only encourages cheating by teachers and administrators**


Are we placing such a high emphasis on these tests that teachers and school administrators are forced to cheat in order to save their jobs? Nationwide, stories indicating tampering with tests have been popping up in the news. After Michelle Rhee started “tying student scores to principals’ and teachers’ employment,” one school in Washington DC was suspected of changing students’ test scores in order to show improvement. Although this cheating dates back to 2008, some of the details in the case were just recently released. The principal of the DC school was walking by an office when she saw three staffers with students’ test booklets. The principal “noticed that the erasers were down and the pencil points were up.” She told her superiors about the situation and was quickly urged to stay quiet and to “respect the legacy that had been built [at their school]” (Toppo, 2013). There have been prominent cheating scandals in other cities as well. Schools in Atlanta, Georgia have been under investigation for several years. Recently, a once prominent National Superintendent of the Year, Beverly Hall, and other Atlanta school professionals were indicted by a grand jury. At first glance, one might wonder how such a prominent woman, as well as her colleagues, could be involved in a test-tampering scandal. One theory is that the educators cheated to reap financial rewards associated with higher student test scores (Brum-back, 2013). Teachers at a school in Brooklyn, NY, were also accused of cheating in a similar fashion in May of 2012 (Morales, 2012). Furthermore, in an article titled “California Education Rankings: 23 Schools Stripped of API Ratings for Cheating” (2012) the Huffington Post reported that the California Department of Education took action against schools suspected of cheating and exhibiting questionable test results. With so much on the line for standardized test scores, what can we really expect from our teachers and administrators? Are some of them merely cheating so that their low-income schools will get the proper funding they need to succeed?
Testing Undesirable: Commodification

1. Standardized testing turns children into commodities


Standardized testing is a reality with which all educators must contend. Although the laws enforcing such assessments do so under the premise that students will thereby be assured an equal opportunity for academic success, they overlook a critical point—students are human beings with needs that reach beyond what is measured on a test. In this article, the author examines the impact of standardized testing on the educator’s ability to adequately address these needs and questions whether the focus of education has turned to treating children as commodities rather than helping them to develop not only their intellect but also their emotional and social selves.

2. Our current testing focus objectifies students—crowds out efforts at student development beyond “human resources”


Many in education speak of the importance of fostering a sense of community among students and incorporating character education into schools, yet we do not always allow youngsters the opportunity to be children and to share who they are as human beings because of the pressure imposed by standardized testing. Yet, are we not running a greater risk by denying them the opportunity to learn about and from one another as people? Greene (2007) speaks of how the overwhelming focus on standards and testing has actually served to restrict student growth in terms of how they perceive themselves within the educational context and the types of students they become as a result: “[Students] find themselves described as ‘human resources,’ rather than as persons who are centers of choice and evaluation . . . [so] perhaps it is no wonder that the dominant mood is one of passive reception” (33). Instead of moving forward with our teaching by helping students to become active agents in their learning, we instead stunt their development by limiting their instruction to content and skills. The objectives of achieving academic excellence and nurturing the development of knowledgeable and compassionate human beings do not have to be, and should not be, mutually exclusive: “To have as our educational goal the production of caring, competent, loving, and lovable people is not anti-intellectual. Rather, it demonstrates respect for the full range of human talents” (Noddings 2007, 65). Although giving them the best preparation possible is unarguably the responsibility of every educator and administrator, students must feel valued for more than the economic potential that they hold in the global market.

3. We need to encourage teachers to move beyond tests and help students learn how to succeed in life


Yet, social time alone will not address all of the needs described earlier. Teachers must be encouraged to use the curriculum to not only teach the skills and content necessary to achieve on the test but also to achieve in life. “Teachers can be very special people in the lives of children, and it should be legitimate for them to spend time developing relations of trust, talking with students about problems that are central to their lives, and guiding them toward greater sensitivity and competence across all the domains of care” (Noddings 2007, 69–70). When students are able to see their teachers as people and share in experiences with them that reach beyond skills and standards, they are more receptive to the instruction given in the classroom. Educators can also use the curriculum to teach tolerance and recognition of those shared experiences that run through all humanity. Greene (2007) suggests that teachers use the arts and literature in a way that transcends their current use as simply tools to teach children those skills assessed on the test. She points out that interaction with the arts not only nurtures a student’s journey of self-discovery, but it also promotes a deeper learning. All of this is directly related to developing what is today described as the active learner, here conceived as one awakened to pursue meaning. . . . Encounters with the arts nurture and sometimes provoke the growth of individuals who reach out to one another as they seek clearings in their experience and try to live more ardently in the world. (Greene 2007, 37) Furthermore, by studying literature, students can come to recognize the common threads that run through all human experience, transcending age, race, and religious differences, which will promote tolerance and help them to function in a global society.
Testing Undesirable: Commodification [cont’d]

4. We need to allow space in schools for socialization—this is also a key part of the education process


Several years ago, the four academic teachers with whom I work in a cluster and I took our seventh graders to a morning activity known as “breakfast club,” run by special-needs students. This club not only gave these mildly and severely learning-impaired children the opportunity to interact with their peers, practice social skills, and handle simple money transactions, it also gave the cluster students the opportunity to interact with friends whom they did not regularly see because of scheduling. The children thoroughly enjoyed the opportunity to eat breakfast with each other and, dare I say it, their teachers. They were well behaved, knowing that this was a privilege that occurred only twice a month. Time passed quickly, as it often does when people are enjoying themselves, and we remained in the cafeteria for a little longer than usual. The principal at that time became quite concerned when she walked in and saw that we were still there, which she made clear by her question: “How are you going to make up this instructional time?” I quickly did the math, dividing the extra fifteen minutes that we had squandered by the remaining seven academic periods left in the day, yet I, like my colleagues, remained silent. The question was valid, but the answer, once again, depends on how one defines instructional time and educational goals. School must be about achieving a balance between developing the intellectual, emotional, and social selves of each individual. Allowing children a reasonable amount of time to socialize with one another not only addresses these needs but also aligns with the natural disposition of the human being. “Intellect, emotion, and spirit depend on each other for wholeness. They are interwoven in the human self and in education at its best, and we need to interweave them in our pedagogical discourse as well” (Palmer 2007, 72). The time the children spent that morning interacting without a keyboard in front of them or a phone between them may be more valuable than those irretrievable fifteen minutes. Perhaps the time may even have increased the camaraderie between them, thereby enriching ensuing discussions and facilitating the ease and efficiency with which they worked on learning activities within the classroom. The fact is that, on that day at breakfast club, we, as teachers, made a choice. We chose to invest those fifteen minutes of academic time in nurturing the social and emotional development of our students—so that one day they are able to successfully function in society and willing to learn from and listen to people different from themselves because perhaps they have learned standards of humanity in addition to the standards on an assessment.
Testing Undesirable: Corporatization / Neoliberalism

- Testing fuels the corporatization of education—this is the biggest threat to our education system


The biggest threat to education today is the corporate education reform movement—what many of us call “Ed Deform.” It is also the biggest threat to teachers' working conditions. Changes in education legislation are creating new government-funded markets for education entrepreneurs. Spending is being shifted away from teacher salaries, benefits, and pensions and into standardized tests, curriculum, and technology. To maximize this investment opportunity teachers must be reskilled away from deciding on content, assessing students, and tailoring education to meet diverse students' needs and interests. This reduces the room for teachers to implement, for example, the demands of anti-racist advocates and concerned parents for "culturally relevant curriculum" or, indeed, anything that deviates from relevant test-prep skills. Standardized test scores provide a simple metric for measuring "productivity" against teacher labor costs. One example of this Taylorist dynamic is New York City's new "Advance" Teacher Evaluation system.


**Testing Undesirable: Counterproductive**

1. **Our focus on high-stakes testing has backfired—it only worsens the problems that it is intended to address**


   Although standardized testing was issued with good intentions, very few can argue that it has yielded the intended effects. The push for standardized testing has gotten stronger, but has there been any positive results? As of 2012, the US educational programs ranked 17th out of 50 other countries. The US was out-ranked by Finland, South Korea, Hong Kong, Japan, Singapore, the UK, Canada, and Germany, to name a few. From these results it is interesting to note that the country that placed first, Finland, has been the most outspoken against standardized testing. Finland, along with many other countries, has instead put an emphasis on having good teachers and trusting those teachers with their students’ academic growth (Gayathri, 2012). The cultures which respect teachers and place education high on their priority list are performing better in the classroom and producing thoughtful students and citizens for the progress their country. Gayathri (2012) also noted that “Having a better [teacher] is statistically linked not only to higher income later in life but to a range of social results, including lower chances of teenage pregnancy and a greater tendency to save for their own retirement.” If better teachers are what we need, why are we instead focusing our attention on standardized test scores? Why don’t teachers have a strong support system to help them pursue challenging goals, persevere in the classroom, and cope with adverse conditions (Clark, 2012; Darvin, 2012)? The typical standardized test does not only deter students from learning, but deters teachers from teaching and actually causes many novice teachers to quit the profession (Darvin, 2012).

2. **Current high-stakes testing is counter-productive—multiple reasons**


   In this paper we have not tried to evaluate the grounds for the novice teachers’ antipathy to the NCLB Curriculum. It would seem obvious however if test scores are the only measure to determine teacher accountability, test scores are bound to become the teacher’s sole concern especially for the novice teachers concern about their reputation and performance. Teaching to the test is the natural result which leads to the neglect non testing areas thus gravely limiting the scope of learning. Whatever their validity the novice teachers concerns should be heeded by policy makers. Testing has not given conclusive evidence of bettering student performance. The evidence is quite mixed: students from low socio economic families continue to lag behind students of higher income families (Amrein, & Berliner, 2002; Hedges & Nowell, 1998). Indeed in some areas the testing has made performance worst. This raises a very important issue. The burden and responsibility for improving our children learning have mostly fallen on the teachers, with little policy attention to multitude of other factors that impinge on education (Clark, 2012).
Testing Undesirable: Creativity / Critical Thinking

1. Standardized tests inhibit the type of education—critical thinking skills—that are necessary to compete in a global economy


The most common criticism of standardized testing is that teachers find themselves “teaching to the test” instead of teaching the various content and skill areas of the curriculum. In recent years, standardized tests have become the predominant tool used to determine a student’s progress, to promote or retain a student at the current grade level, and to identify if a learning disability exists. The main problem with standardized tests is that they inhibit the kind of education that matters the most, preparing young people with “higher order thinking skills” to compete in a global economy. Does “teaching to the test”, an integral part of standardized tests, really increase student capabilities and knowledge, or does it simply put more pressure on teachers and students? Teachers want their students to excel on their standardized tests for both their benefit, as well as the benefit of their students. High scores become even more important because the school district uses individual school test scores to evaluate each school. In many cases, school ratings are now linked to funding and teacher evaluation. Novice teachers are the next generation of educators who will be teaching school children. These enthusiastic, optimistic young professionals have a unique perspective that has not been tainted by the educational bureaucracy. In this paper some novice teachers who were presently teaching voiced their concerns and opinions against standardized tests.

2. Being forced to ‘teach to the test’ cuts out learning critical thinking skills—such work receives ‘short shrift’


The standardized tests and to score well in them have become an all-consuming force in the schools. On the scores depend the school ranking, in the district, even in and funding region, and these are widely covered by the media. Persistent low scores may attract severe penalties for the school. Pressure builds up in the school board and percolates down to the teachers. The teachers under pressure concentrate on teaching to the test to better course as well as their own reputation (Wallace & Irons, 2010). Inevitably this leads to the neglect of other dimensions of learning beyond the testing areas. Since the tests are on reading, math and science other subjects come to be neglected. Citizenship inculcating “good” habits and development of a sustained capacity for learning are given short shrift. For our interlocutors this is the most dispiriting aspect of teaching experience. The values they learned in college seem completely naïve and idealistic in the “real world”.

3. Standardized testing focus stifles creativity and innovation


Then, last month, I was disappointed to see Mr. Duncan reaffirm his support for annual standardized testing of all American schoolchildren in grades 3 to 8 and in high school. This announcement runs counter to our pledge to be future-ready. The age of standardized testing has de-emphasized creativity and innovation by overly relying on test performance as a criterion of school and student success. This emphasis has resulted in limiting school curricula, robbing students of experience with the arts and other non-tested subjects. Mr. Duncan has said that "testing issues today are sucking the oxygen out of the room in a lot of schools." His reaffirmation for the need for continual annual testing contradicts this previous statement. Let me be clear, my colleagues and I embrace assessment. It is essential to inform instruction and allow educators to respond to the needs of their students. However, it should be done daily to appropriately challenge and support each student. Likewise, broader periodic assessments provide children with multiple ways to demonstrate what they know and can do. Standardized tests do not acknowledge the developmental differences in children. When we endorse them we subscribe to the belief that all children learn the same way and at the same rate. Likewise, standardized tests fail to measure the skills that employers have identified as essential for success now and in the future: communication, collaboration, critical thinking and creativity.
4. Our testing focus crowds out the teaching of critical thinking skills


Standardized tests are one of many resources that can be used to evaluate student progress. When used alone, standardized tests do not present a clear picture of student knowledge and skills. One major problem is when teachers begin “teaching to the test”. Most teachers would say that their main job is to foster a love for learning (Heubert & Hauser, 1999). Teachers accomplish this task by encouraging their students to think critically and take their knowledge and skills outside the classroom and into adulthood. Standardized test scores promote rote memorization at the expense of critical thinking skills, pressuring teachers to spend most of their instructional time teaching testing material (Heubert & Hauser, 1999). Because of the risk of lower test scores, teachers rarely deviate from testing curriculum even if they have to eliminate other important subject matter content. As a result those items in the curriculum are considered “unworthy” and remain uncovered (Ezer, Gilat, & Sagee, 2010; Hom, 2003).

5. The types of questions used on the tests do not encourage or reward critical thinking


Amrein & Berliner (2002) reported that despite the ongoing changes in standardized testing for decades, the structure of the test has remained the same. High-stakes testing uses multi-choice formatting with only one correct answer. Therefore, it is reasonable to believe a student may know the answer, but he or she may be unable to articulate it in the way that it is written on the paper. These kinds of test questions do not promote critical thinking skills and the ability to solve real-world problems. Instead they promote test specific curriculum using outdated instructional methods (Kohn, 2004).

6. Test pressures can turn class time into mindless cramming sessions


But many educators say that the current tests used to measure student progress—and hold schools and teachers accountable—don’t measure learning well. “Standardized tests measure the wrong things,” Stephen Laz-er, a teacher at Harvest Collegiate High School, told the packed hearing room. They measure mindless repetition of facts, he said. Pressure to raise student test scores can turn classes into cramming sessions. Laz-er said he spends the entire month of May training his students to pass the Regents, New York state exams. “The learning and opportunity gap widens” when students with low scores spend so much time on test prep, while students with high scores can take on more complex assignments, he said.
Testing Undesirable: Disparate Impact

1. The negative effects of high-stakes testing are most felt by low-income and minority students


Our next question is who is hurt the most by these practices? The numbers indicate that students from low-income and minority-groups suffer the most from high-stakes testing through failure to pass to the next grade level and remediation programs. They are subjected to curriculum using rote memorization and practice drills, the same outdated teaching methods that got them to this point. Then we wonder why so many low-income and minority groups suffer from low self-esteem and in many cases fail to graduate from high school (Hom, 2003). Conversely, upper and middle income white students are often guided to higher level coursework and honors programs where more current teaching methods are implemented to enhance critical thinking skills. Hom (2003) believes that minority groups, culturally diverse populations with limited English language skills, and those students with learning disabilities suffer the most as a result of high-stakes testing.

2. Schools adapt to tests by adopting curriculum that only exacerbates disparities between different student populations


Furthermore, this kind of standardized test encourages school districts to implement rote and drill bundled programs. These low-level learning experiences lend themselves to more low income children spending time completing worksheets and falling further behind the more affluent students who are getting practical experiences that help develop a real understanding of the material (Shepard, 2000).
Testing Undesirable: Finland Proves

1. We should model Finland and move away from high-stakes testing


Lavonen suggested if the U.S. wants to mimic Finnish success, it should consider adopting the nation's philosophy on testing. "We need more decision making and assessment at the local level. We need less standardization and national testing," said Lavonen, a professor of physics and chemistry at the University of Helsinki who was visiting Teachers College, Columbia University with several colleagues. "We need less test-based accountability." The Finnish government does occasionally test a random sample of a specific grade and subject in order to insure that the country is meeting its education goals. Lavonen helped design a high school science exam taken by a sample of Finnish students in 2011. The last time high school science had been tested was 2001. Overall, students answered an average of 58 percent of the questions correctly. While there were some troubling findings, such as a gender gap favoring boys in physics and girls in biology, Lavonen said everyone was generally pleased with the results. There was an extremely high correlation between a student's score on the exam and the end-of-semester grade he or she received, which Lavonen said indicates that teachers are grading well. The test also included many questions to measure students' attitudes about science - how well they'd learned it and how interesting and relevant it was to them. And while there are no annual standardized tests there are still ways that the school system checks for quality. Progress is monitored at both the local and municipal level in a variety of ways, including assessments throughout the school year. But the design and timing of any exams are left up to the teacher. Lavonen, for instance, helped create an online tool for science teachers to develop tests and quizzes as they saw fit. Some might never use it, instead relying on informal checks as they interact with students. It all comes back to what the Finnish visitors described as a "culture of trust," where teachers are given flexibility and autonomy. "Everything they decide themselves; how they teach and what they teach," Lavonen said. (Finland does have a national curriculum, however, that teachers must work within.)

2. Standardized tests are not necessary for us to have high-achieving schools—Finland roves


As the United States focuses more on using tests as a means of holding educators and school districts accountable, Finland—which is one of the top performers on international tests--has gone in the opposite direction. In the U.S., states give annual high-stakes exams that determine whether schools must undergo reforms, in some cases whether students can pass to the next grade level or graduate from high school, and increasingly whether teachers can receive tenure and keep their jobs. Yet the U.S. tends to rank in the middle on international tests. In Finland, by contrast, the few tests students take are low stakes, said Finnish educator, Jari Lavonen in a presentation on Thursday in New York. Assessments are used as a tool for professional development and to help teachers gauge student growth, never for accountability. Yet, despite a lack of practice, when Finnish students do take standardized exams, they tend to excel. The country ranks consistently near the top in math, reading and science in the Programme for International Student Assessment (PISA), which is a standardized test taken by students in dozens of countries. The Finnish school system has become the envy of less successful nations around the world, including the United States.
Testing Undesirable: Gaming

1. The tests can be gamed—you simply need to study the textbooks written by the major testing companies


This is because standardized tests are not based on general knowledge. As I learned in the course of my investigation, they are based on specific knowledge contained in specific sets of books: the textbooks created by the test makers. All of this has to do with the economics of testing. Across the nation, standardized tests come from one of three companies: CTB McGraw Hill, Houghton Mifflin Harcourt, or Pearson. These corporations write the tests, grade the tests, and publish the books that students use to prepare for the tests. Houghton Mifflin has a 38 percent market share, according to its press materials. In 2013, the company brought in $1.38 billion in revenue.

Pennsylvania currently has a multi-million-dollar contract with a company called Data Recognition Corporation (DRC) to grade the PSSAs. DRC works with McGraw-Hill as part of a consortium that has a $186 million federal contract to write and grade standardized tests for the rest of the country. McGraw-Hill, meanwhile, also writes the books and curricula schools buy to prepare students for the tests. Everyday Math, the branded curriculum used by most Philadelphia public schools in grades K—5, is published by McGraw Hill. Put simply, any teacher who wants his or her students to pass the tests has to give out books from the Big Three publishers. If you look at a textbook from one of these companies and look at the standardized tests written by the same company, even a third grader can see that many of the questions on the test are similar to the questions in the book. In fact, Pearson came under fire last year for using a passage on a standardized test that was taken verbatim from a Pearson textbook.

2. People readily game the standardized testing system


All true (except for the one test being reliable). Given high stakes and the accompanying pressure, people will game a system. And it is all too true that grades vary widely because of four factors: a teacher's conception of achievement, a teacher's sense of equity and rigor, a teacher's ability, and the composition of students. But people are already gaming standardized testing, sometimes criminally. And, at a basic level of competency, a grade or an evaluative report would give us as much information as we now get from standardized tests.
Testing Undesirable: General

1. Testing and standards now dominate everything that happens in schools


At the same time, America’s students and teachers still face a daily barrage of testing jargon and impossible-to-fulfill standards from administrators heeding the recommendations of education reformers, who favor making instructional decisions based on data analysis from standardized tests rather than developmentally appropriate practice. It permeates everything we do in school. And it hinders the education students receive. As the authors of the CGCS report remark: “We think it is worth noting that most tests that schools administer don’t actually assess students on any particular content knowledge.”

2. We need to stand up and speak out against the current uses of tests— they are hurting our schools and students


While discussion has certainly increased in both content and intensity and alternatives have been explored, we have not seen a significant change in the use (and, in many cases, the misuse) of testing. With the advent of the No Child Left Behind law, enacted in 2002, quality developmentally appropriate teaching and learning practices have taken a backseat to the more focused attention on low-level skills that can be assessed easily on a standardized multiple-choice test. Standardized tests are now used to hold up children and schools for comparison; the scores are used to discriminate rather than diagnose, punish rather than reward. Equally disturbing is the misuse of these tests and these tests alone-to unjustly hold teachers and schools accountable and then punish those who have not met adequate yearly progress, as deemed by people other than those working with children on a daily basis (e.g., politicians). When Vito Perrone updated ACEI’s position paper on testing in 1991, he claimed that a testing moratorium was even more important than it had been in 1976; I believe the need has continued to grow. Excellent teachers are leaving the profession out of frustration. High-quality schools that serve children from disenfranchised homes are being closed and children displaced. Important skills that schools once taught, such as critical thinking, discussions, and problem solving, are being replaced by low-level, fill-in-the-blank worksheets and drills. The gap between the poor and the rich is ever-widening, and there appears to be no end in sight. It is past time for teachers, schools, administrators, parents, and the public to stand up and let their voices be heard. It is past time for all involved to put a stop to the misuse of tests in all grades, particularly in the primary years.

3. Standardized tests are fundamentally flawed—they don’t assess individual progress, and instead are only used to measure districts


It’s good to hear that some leaders in Washington hope to alter the No Child Left Behind Act so that it gets rid of the standardized testing requirement. The landmark 2002 federal law did a fine job of teaching educators and administrators to speak the political language of accountability, but since its passage in 2002 and subsequent modifications, it has done little to actually improve education in the United States of America. Standardized testing is good for assessing the progress of a child’s education and determining where to go next. That is, if the test is a diagnosis-type test. However, standardized testing -- the types done to comply with NCLB -- is seriously flawed. These tests aim to determine how a classroom, a school or a district stands in educating children. They are flawed partially because they encourage schools to teach to the test and partially because they fail to consider the contribution parents make or, more pointedly, fail to make. Teaching to the test only builds memorization skills, rather than understanding context, opinions and social aspects.
Testing Undesirable: General [cont’d]

4. Our current approach to reading tests turns them into a ‘minefield’ for students and teachers—multiple reasons


Even our best schools serving low-income children—public, parochial, and charter alike—have a much harder time raising ELA (English language arts) scores than math. This is unsurprising. Math is school-based and hierarchical (there’s a logical progression of content to be taught). But reading comprehension is cumulative. The sum of your experiences, interests, and knowledge, both in and out of school, contribute to your ability to read with understanding. This is why affluent children who enjoy the benefit of educated parents, language-rich homes, and ample opportunities for growth and enrichment come to school primed to do well on reading tests—and why reading scores are hard to move. Teacher quality plays a role, but note how fourth-grade NAEP math scores have risen over the years while reading has remained flat, even though the same teacher usually handles both subjects. This suggests that our teachers, when they know what to teach, are stronger than we think. In math, standards, curriculum and assessments are closely aligned (there’s no surprising content on math tests). By treating reading as a collection of content-neutral skills, we make reading tests a minefield for both kids and teachers.

5. Our testing focus is making our schools hostile to the students who need it the most


The United States is a nation built on the ideals of freedom and equality, a nation of principles and responsibilities. It is a nation of wealth and power, a land of opportunity where democratic values are honored and people strive to help children succeed. At the same time, it is a nation of the poor and indigent, a land where families are impoverished and disempowered. For those children who come from homes that struggle under the burden of financial stress and poverty, the fruits of democracy can appear out of reach. For many children, U.S. public schools have become places where separateness is evident, where those who do not have are expected to achieve the same standards as those who have. School, the very place where democratic values should be taught and practiced, is being eroded in the name of standardized testing and accountability. As a result, many voices have been silenced.

6. Students react poorly to compulsory learning


4. Students are less interested in whatever they’re forced to do and more enthusiastic when they have some say Once again, studies confirm what we already know from experience. The nearly universal negative reaction to compulsion, like the positive response to choice, is a function of our psychological makeup. Now combine this point with the preceding one: If choice is related to interest, and interest is related to achievement, then it’s not much of a stretch to suggest that the learning environments in which kids get to make decisions about what they’re doing are likely to be the most effective, all else being equal. Yet such learning environments continue to be vastly outnumbered by those where kids spend most of their time just following directions.

7. Standardized testing fails—does not account for different learning styles


Thirdly, standardized testing is insensitive to individual students’ learning style. It is well established in learning theories (Sternberg, 1998) that there are great variation in the way students learn. The pressure to teaching to the test and improve scores disable the teachers to be attentive to such differences. Teaching to the tests results in standardized teaching. The resulting poor scores of such students may stigmatize capable students for the rest of their academic lives and the teachers as well for their “poor” performance (Sternberg, 1998). Fourthly, not all students do well in tests. Quite capable students may do poorly in tests, thus bringing blame on themselves as well as on their teachers (Wolf, LeMahiue, & Eresh, 1992).
Testing Undesirable: General [cont’d]

8. The tests themselves are flawed—risk producing a ‘trainwreck’

Charlotte Danielson, a noted academic and author who is a strong supporter of Common Core, was one of the early educators to express concern about the validity of the high-stakes testing regime. According to Danielson, "I'm concerned that we may be headed for a train wreck there. The test items I've seen that have been released so far are extremely challenging. If I had to take a test that was entirely comprised of items like that, I'm not sure that I would pass it—and I've got a bunch of degrees. So I do worry that in some schools we'll have 80 percent or some large number of students failing. That's what I mean by train wreck." That was in March 2013, two years before the opt-out movement really took off.

9. Effective public schools are vital to the health of our democracy

“Public education is the latchkey that can open the door to a land of opportunity; it is the cornerstone of our nation’s democratic system of government” (Popham, 2001, p. 4). From the earliest days of the United States and into the early part of the 20th century, public education was revered and applauded for its success in educating children of all backgrounds. It was esteemed as the necessary tool for the country to continue thriving, as informed citizens are critical to a true democracy. The essential value of the public school in a democracy, from the beginning, was to ensure an educated citizenry capable of participating in discussion, debates, and decisions to further the wellness of the larger community and protect the individual right to ‘life, liberty, and the pursuit of happiness.’ (Glickman, 1993, pp. 8-9) Public school, by its very nature, was intended to secure for children their place in a democratic society with the knowledge, understanding, and tools necessary to make decisions for the good of all its citizens. During the late 1960s and early 1970s, an increasing rumble of discontent concerning the nation’s schools began. What are our schools doing? What are our schools teaching? What are our children learning? Where is our money going? Are our children being prepared to take their place in the competitive world here and abroad? With these questions came suspicion and then distrust in the nation’s teachers and schools. In order to combat the mounting disregard for public education, the federal government took steps that would change the landscape of education in the United States and threaten the very fabric of democracy.
Testing Undesirable: Gifted Children

1. Our testing focus distracts us from providing for the needs of more gifted students

Marcia Gentry, Associate Professor and Associate Director, Gifted Education Resource Institute, “No Child Left Behind: Neglecting Excellence,” ROEPER REVIEW v. 29 n. 1, 2006, pp. 24-27.

Then it occurred to me that I would also be preaching to the choir, as most educators are aware of the negative impact of NCLB, especially gifted education professionals whose programs and students have already become marginalized in its wake. This realization led me to question why we as educators have allowed this legislation and the thinking that underpins it, to, in effect, rise to power. Perhaps, over the past two decades, we have spent too much time and energy trying to conform and comply with the reform of the month and too little time and commitment to the art and science of educating children and youth for life-long learning so that they can become productive citizens in our democracy. In doing so, we have bought the idea that education in America needs the federal government to fix it, and now we have NCLB. By conforming and playing the score boosting game, we are denying an entire generation of children quality education while we scramble to raise test scores. NCLB has created an environment in which school administrators have no incentive to concentrate on educating gifted children or developing talents of any children. It will take years before we know the true reaches of this legislation, but early studies and observations are not encouraging (e.g., Amrein & Berliner, 2002; Moon, Callahan, & Tomlinson, 2003; Nichols & Berliner, 2005).

2. Testing does nothing to address the needs of gifted children

Marcia Gentry, Associate Professor and Associate Director, Gifted Education Resource Institute, “No Child Left Behind: Neglecting Excellence,” ROEPER REVIEW v. 29 n. 1, 2006, pp. 24-27.

Since NCLB fails to address the clear educational needs of gifted children and educators' responsibility to develop and meet their needs, school districts across the country have cut already small gifted programs and reallocated the funds to remedial programs due to fear of sanctions if students' scores aren't increased (Golden, 2004). As described in the National Excellence Report (US Department of Education, 1993), gifted children require special services partly because the regular curriculum does not challenge them, and many have already mastered half of the required curriculum. Sadly, gifted child education has lost ground since 1993 in the public schools, due in part to the emphasis on remediation without similar language to support educationally necessary enrichment and acceleration. Yet, it is precisely the type of education that occurs in gifted programs that has the potential to increase meaningful student learning and move students well beyond proficiency (Kaplan, 2004).
Testing Undesirable: Intelligence Limits

1. Standardized tests fail us—they only test one kind of intelligence


Scientists know that intelligence includes more than the narrow range of abilities that paper-and-pencil tests measure. Professor Subotnik acknowledges this, but insists that testing other traits, such as empathy and practical judgment, would be costly and difficult. By testing for measurable skills, employers and admissions officers can at least be sure that a candidate possesses some of the desired skills. “If a job requires two skill sets bumping and skipping[,]” he asks, “and if only jumping can be successfully tested, does equity really require that the measurable skill be left untested?”. But with complex tasks, the skills making up a good practitioner may not be so easily divided. They may not even be additive but mutually dependent, requiring the right balance to operate at an optimum level. For example, a political leader with great intelligence but no moral sense might be dangerous. A lawyer with highly developed analytical ability but little practical judgment might have a career marked by blunder after blunder. An inventor with a superb command of physics or chemistry but little head for marketability may generate one useless patent after another. And so on. Rewarding one skill exclusively may not be like having half a loaf of bread, better than none at all.

2. The SAT is biased towards particular types of intelligence


B. Primary Reliance on the SAT is Misplaced Because the SAT Creates Inequalities Predicated upon Type of Intelligence Possessed. The SAT creates inequality because it elevates certain forms of intelligence over others. Although the SAT purports to test innate academic ability and college preparedness, instead the SAT actually selects two of the many forms of intelligence and uses them as a proxy for general intelligence and college preparation. These two selected forms of intelligence—mathematical and linguistic—are transformed by the SAT into the gatekeepers of elite college admissions, merit-based scholarships, and access to a college education generally. Because a college education, especially at an "elite" university, is tied to relatively high future earnings and societal status, the SAT has the power to create social stratification predicated upon the type of intelligence with which an individual is endowed. This subsection argues that there exist at least eight forms of intelligence, and by elevating two forms of intelligence above the others, the SAT may create a form of intelligence social stratification and inequality. Howard Gardner proposed the theory of multiple intelligences. Multiple intelligence theory posits that there exist at least eight forms of intelligence. The list of potential intelligences is open, and Gardner has proposed a ninth potential form of intelligence. The eight forms of intelligence are discussed below, and, for simplicity, Table 2 summarizes these forms of intelligence. [table omitted]

3. The SAT only tests for very limited forms of intelligence


Unfortunately, the SAT does not attempt to test all (or even many) of the intelligences possessed by humans. Instead, the SAT tests, at least facially, only linguistic intelligence (in the English sections of the SAT) and logical-mathematical intelligence (in the math sections of the SAT). Because the SAT tests only logical and linguistic intelligence for college admissions purposes, students who possess these forms of intelligence are systematically given preference over students possessing the other forms of intelligence described above (and the other forms yet to be isolated). There is no reason to believe that James Clerk Maxwell was more important than Martin Luther King, or that a high school student with intelligence in English is more important than a high school student with intelligence in music. However, the SAT is flawed because it would elevate Maxwell over Dr. King and the English student over the music student. Because future status and financial benefits flow from a college education and a college education is contingent upon certain performance on the SAT (or the similar ACT), which elevates two intelligence forms over myriad other forms, the SAT may create inequality and social stratification predicated upon which form of intelligence with which a person is endowed.
Testing Undesirable: Invalid Results

1. **Test scores are not valid—they force curriculum changes that invalidate the tests’ capacity to determine whether students are meeting learning objectives**

   Sumita Bhattacharyya, Mary Junot and Hillary Clark, Nicholls State University, “Can You Hear Us? Voices Raised Against Standardized Testing by Novice Teachers,” CREATIVE EDUCATION v. 4 n. 10, 2013, pp. 633-639, p. 635. Do standardized tests accurately measure our students’ level of ability or does coaching to improve these test scores taint the validity of the results? As we understand it, the general purpose of standardized testing is not to measure on what level a student can read, but the student’s comprehension of the given material. The test is merely an instrument of instruction, and should not be the focus of the evaluation. But the problem comes in when standardized tests are used to determine curriculum. The process of preparing the students for the test reduces the time available for instruction and narrows the curricular topics and methods of instruction. This in turn limits the instructional materials that a teacher can use especially if they are not similar to standardized testing formats. Studies suggest that even though test scores tend to improve when students are “taught the test”, the overall level of student learning does not improve (Shepard, 2000; Smith & Fey, 2000). Tenth grade teacher in her 2nd year expressed the same frustration, “Unfortunately we feel helpless making changes. We will lose our job or government funding if the NCLB Acts are not satisfied.” Seventh grade teacher in his 3rd year of teaching said the same thing but in a different way. He said, “School administrators engage the staff for a few times in a year discussing ways to improve test scores, publicly let teachers know how their students performed compared to other. Thus we feel intimidated.” As a result, it is becoming increasingly questionable as to the validity of standardized testing.

2. **Tests are not objective—they do not account for many out-of-school factors that affect student performance**

   Sumita Bhattacharyya, Mary Junot and Hillary Clark, Nicholls State University, “Can You Hear Us? Voices Raised Against Standardized Testing by Novice Teachers,” CREATIVE EDUCATION v. 4 n. 10, 2013, pp. 633-639, p. 635. Standardized tests are only objective in terms of scoring, because they are machine tallied. The type of test, the content and structure, the number of test items, the choice of the right response, the instructions given, and the test results and their use are all decisions made by subjective administrators (Boser, 2000). Therefore, how can standardized tests be truly labeled as objective? Additionally, non-instructional factors play an important role in student achievement on standardized tests. Standardized tests tend to overlook family home situations, parents’ educational backgrounds, communities, and poverty rates, in an attempt to put all students on equal education levels (Sambar, 2001). Po-ham (2001) argues that standardized assessments measure general knowledge that is gained in most middle- and upper-class homes where resources are readily available. Similarly, a novice English teacher noted that “... standardized tests are just so unfair. The students I teach from high- and middle-class families have more knowledge and skills than my students from low-income families. It became obvious that standardized test scores are comparable to technology in the home and the avail-ability of books”. A standardized test is said to be completely reliable only if the results are the same the next time the test is taken. But an individual’s score on a standardized test can change each time the test is taken due to the student’s physical, mental, or emotional state. This leads to inaccurate student test scores (Boser, 2000).

3. **Standardized tests are not valid—they do not reflect the diversity of factors that affect student achievement**

   Sumita Bhattacharyya, Mary Junot and Hillary Clark, Nicholls State University, “Can You Hear Us? Voices Raised Against Standardized Testing by Novice Teachers,” CREATIVE EDUCATION v. 4 n. 10, 2013, pp. 633-639, p. 635. The next question is how valid are these tests? Standardized test results have major implications in a student’s present life and possibly future career—all from one test score. Schrag (2000) reports that there are many variables determine a student’s test score. The particular test, how many times a student takes that test, whether test instructions are complete, and the comfort of the test setting are a few of the variables that can influence a test score. When asked the teacher who is in her 2nd year of teaching became emotional School achievement is a function of different factors. Only certain parts are under the influence of schools but not all. “... Schools can influence the quality or quantity of instruction, motivation and positive learning environment. But how about their family and home environment? The standardized scores don’t consider the complexity of achievement. Thus standardized scores lack in proper interpretation. It practically oversimplifies the nature of the scores.”
Testing Undesirable: Multiwarrant

1. Tests promote cheating, and deny children a well-rounded education


I talked to a first-grade teacher some time back who told me she was retiring, and when I asked her what made her decide that it was time to leave the classroom, she said, "When they handed me the script." In light of recent news developments, I began to wonder: what if they had also handed her an eraser? I ask that question not to be provocative, but as a sad commentary on the corruptive influence high-stakes tests have had on our students, teachers and schools. An Atlanta Journal-Constitution and Cox newspapers investigation found "suspect” scores on standardized tests in school districts across the country. The paper's analysis does not prove that any widespread cheating occurred; it merely points to unusual test score swings. It would be unfair to pass premature judgment on any of these school systems. The story will undoubtedly trigger a thorough review -- as it should. As educators, we should never condone cheating of any kind for any reason. Likewise, we should never miss an opportunity to underscore the possible consequences when real learning and effective teaching are sacrificed at the altar of high test scores. Under No Child Left Behind, students as young as 6 or 7 years old are subjected to weeks of preparation for high stakes tests. Because math and reading are the dominant testing subjects, history, civics, music and art are squeezed out of the school day. For the last 10 years we have shortchanged countless children because of NCLB's overemphasis on standardized multiple choice tests. Or as Gary Miron, professor of education at Western Michigan University, wrote: "The bigger problem is a more serious type of cheating -- one that's perfectly legal and apparently acceptable. Students are being cheated of a broader education..." Tests shouldn't be used to punish schools, as is the case under NCLB, or to pigeonhole students or their teachers. Educators aren't alone in being fed up with narrow, punitive accountability measures. Parents also want well-designed, timely assessments that monitor individual student progress across a range of subjects and skills, one of the key findings in a new study by the Northwest Evaluation Association. We should use assessments to help teachers improve their practice, help students evaluate their own strengths and needs, and focus help on the students and subjects that need attention. Let's get back to the core purpose of public education -- ensuring students have access to a great education that prepares them for lifelong learning and success -- and leave the pressure cooker for pot roasts.

2. High-stakes testing is fundamentally counterproductive—three reasons


Personally I do not have a problem with "federal overreach." I would like to see Republican governors and legislatures forced to provide adequate educational funds and programs for minority students in their states. I would also like to drive Pearson and the other testing companies out of the education business, but that is not a good enough reason to end the high-stakes testing regime. However, educators know there are at least three valid educational reasons to stop the high-stakes testing regime that drives American miseducation. 1. We know that high-stakes testing transforms curriculum into test prep that undermines any test validity. 2. We are witnessing how high-stakes testing and the penalties for poor performance create a nation of cheaters not learners or producers. 3. We know that high-stakes testing actually generates anxiety that undermines student performance and learning.

3. Schools are narrowing the curriculum and lowering standards because of testing pressures


Schools are cutting back on teaching science, social studies, and art to become proficient in math and reading tests by NCLB's 2014 deadline. The New York Times' Sam Dillon reports. This is the reason the US lags so far behind other countries when it comes to science proficiency, researchers told The Hechinger Report. Also to make the goal, more than half of states have lowered their standards to redefine "proficient."
4. The current emphasis on standardized tests is highly counterproductive—8 different factors prove


Standardized testing has swelled and mutated, like a creature in one of those old horror movies, to the point that it now threatens to swallow our schools whole. (Of course, on “The Late, Late Show,” no one ever insists that the monster is really doing us a favor by making its victims more “accountable.”) But let’s put aside metaphors and even opinions for a moment so that we can review some indisputable facts on the subject. Fact 1. Our children are tested to an extent that is unprecedented in our history and unparalleled anywhere else in the world. While previous generations of American students have had to sit through tests, never have the tests been given so frequently, and never have they played such a prominent role in schooling. The current situation is also unusual from an international perspective: Few countries use standardized tests for children below high school age—or multiple-choice tests for students of any age. Fact 2. Noninstructional factors explain most of the variance among test scores when schools or districts are compared. A study of math results on the 1992 National Assessment of Educational Progress found that the combination of four such variables (number of parents living at home, parents’ educational background, type of community, and poverty rate) accounted for a whopping 89 percent of the differences in state scores. To the best of my knowledge, all such analyses of state tests have found comparable results, with the numbers varying only slightly as a function of which socioeconomic variables were considered. Fact 3. Norm-referenced tests were never intended to measure the quality of learning or teaching. The Stanford, Metropolitan, and California Achievement Tests (SAT, MAT, and CAT), as well as the Iowa and Comprehensive Tests of Basic Skills (ITBS and CTBS), are designed so that only about half the test-takers will respond correctly to most items. The main objective of these tests is to rank, not to rate; to spread out the scores, not to gauge the quality of a given student or school. Fact 4. Standardized-test scores often measure superficial thinking. In a study published in the Journal of Educational Psychology, elementary school students were classified as “actively” engaged in learning if they asked questions of themselves while they read and tried to connect what they were doing to past learning; and as “superficially” engaged if they just copied down answers, guessed a lot, and skipped the hard parts. It turned out that high scores on both the CTBS and the MAT were more likely to be found among students who exhibited the superficial approach to learning. Similar findings have emerged from studies of middle school students (also using the CTBS) and high school students (using the other SAT, the college-admission exam). To be sure, there are plenty of students who think deeply and score well on tests—and plenty of students who do neither. But, as a rule, it appears that standardized-test results are positively correlated with a shallow approach to learning. Fact 5. Virtually all specialists condemn the practice of giving standardized tests to children younger than 8 or 9 years old. I say “virtually” to cover myself here, but, in fact, I have yet to find a single reputable scholar in the field of early-childhood education who endorses such testing for young children. Fact 6. Virtually all relevant experts and organizations condemn the practice of basing important decisions, such as graduation or promotion, on the results of a single test. The National Research Council takes this position, as do most other professional groups (such as the American Educational Research Association and the American Psychological Association), the generally pro-testing American Federation of Teachers, and even the companies that manufacture and sell the exams. Yet just such high-stakes testing is currently taking place, or scheduled to be introduced soon, in more than half the states. Fact 7. The time, energy, and money that are being devoted to preparing students for standardized tests have to come from somewhere. Schools across the country are cutting back or even eliminating programs in the arts, recess for young children, electives for high schoolers, class meetings (and other activities intended to promote social and moral learning), discussions about current events (since that material will not appear on the test), the use of literature in the early grades (if the tests are focused narrowly on decoding skills), and entire subject areas such as science (if the tests cover only language arts and math). Anyone who doubts the scope and significance of what is being sacrificed in the desperate quest to raise scores has not been inside a school lately. Fact 8. Many educators are leaving the field because of what is being done to schools in the name of “accountability” and “tougher standards.” I have no hard numbers here, but there is more than enough anecdotal evidence—corroborated by administrators, teacher-educators, and other observers across the country, and supported by several state surveys that quantify the extent of disenchantment with testing—to warrant classifying this as a fact. Prospective teachers are rethinking whether they want to begin a career in which high test scores matter most, and in which they will be pressured to produce these scores. Similarly, as the New York Times reported in its lead story of Sept. 3, 2000, “a growing number of schools are rudderless, struggling to replace a graying corps of principals at a time when the pressure to raise test scores and other new demands have made an already difficult job an increasingly thankless one.” It also seems clear that most of the people who are quitting, or seriously thinking about doing so, are not mediocre performers who are afraid of being held accountable. Rather, they are among the very best educators, frustrated by the difficulty of doing high-quality teaching in the current climate.
Testing Undesirable: Narrowed Curriculum

1. Teachers cut out other material to focus on things that are on the test


At the same time that researchers addressed the validity of test score gains, studies have also been done to examine the effect of high-stakes accountability pressure on curriculum and instructional practices. These studies, which involved large-scale teacher surveys and in-depth field studies, show that efforts to improve test scores have changed what is taught and how it is taught. In elementary schools, for example, teachers eliminate or greatly reduce time spent on social studies and science to spend more time on tested subjects.

2. Tests strip down the curriculum—teachers focus far too much on how to take the test


More significantly, however, because it affects how well students will eventually understand the material, teaching in tested subjects (reading, math, and language arts) is also redesigned to closely resemble test formats. For example, early in the basic-skills accountability movement, Linda Darling-Hammond and Arthur Wise found that teachers stopped giving essay tests as part of regular instruction so that classroom quizzes would more closely parallel the format of standardized tests given at the end of the year. In a yearlong ethnographic study, Mary Lee Smith found that teachers gave up reading real books, writing, and long-term projects, and focused instead on word recognition, recognizing spelling errors, language usage, punctuation, and arithmetic operations. Linda McNeil found that the best teachers practiced “double-entry bookkeeping,” teaching students both what they needed for the test and the real knowledge aimed at conceptual understanding. In other cases, test preparation dominated instruction from September until March. Only after the high-stakes test was administered did teachers engage the real curriculum such as Shakespeare in eighth-grade English. These forms of curriculum distortion engendered by efforts to improve test scores are strongly associated with socioeconomic level. The poorer the school and school district, the more time devoted to instruction that resembles the test.

3. The curriculum gets narrowed—non-tested subjects end up getting cut out


Because only math and reading test scores count towards a school’s “adequate yearly progress” under NCLB, schools have deemphasized, and in some cases completely stopped, teaching things like “social studies, literature, art, music, physical education, and other important topics where test scores do not result in judgments of school quality,” writes Richard Rothstein in his 2004 book Class and Schools. A 2006 study by the Center on Education Policy supported this claim, finding that since NCLB was passed, 71 percent of school districts cut back on subjects like history and music so they could spend more time on the tested subjects.
Testing Undesirable: Other Mechanisms Superior

1. Software-based assessment is superior—more data, less classroom disruption


This shift away from high-stakes testing towards software-based assessment will actually increase the number of data points available to teachers, as every in-class assignment will be evaluated through the use of technological tools without actually interrupting the instructional process. Right now, frequent testing can actually reduce instructional time, as the time a teacher spends proctoring an exam during which students are obligated to be silent is time a teacher is not actually working with students. Grading tests represents a further investment of time that could be spent more productively. Software-based assessment has the potential to free up a great deal of time for teachers and students. And it could greatly reduce “cramming,” in which students scramble to memorize facts and formulas without actually gaining a deeper, more durable understanding of course material.

2. Software-based approaches will actually improve transparency


High-stakes testing sounds tough, and I absolutely get why many education reformers gravitate towards tools and institutions that appear to toughen up the often rather soft world of K-12 education. But the explosion of computing power means that we can collect more data more often, and this enables us to do a better, finer-grained job of understanding what students actually know. It is important to note that while opposition to high-stakes testing and value-added analysis often seems self-serving — it is easy to see why ineffective teachers might resist accountability — moving towards embedded software-based assessment actually raises the level of transparency, by allowing us to monitor not just what happens on the day of a high-stakes test, but rather to see how students learn over time. While high-stakes testing lends itself to fraud, software-based assessment makes it, in theory at least, extremely difficult.

3. High-stakes testing has outlived its usefulness


I am increasingly convinced that while high-stakes testing may have once had an important role to play, it has outlived its usefulness. Rick Hess, Michael Petrilli, and Martin West, among others, have warned that mandating that teachers be evaluated on the basis of value-added analysis might stifle innovation, particularly as it relates to how schools deploy personnel. Value-added analysis built around the model of a single teacher leading a class of 25 students might make it impossible for schools to pursue more personalized instructional models, or models in which a team of professionals is responsible for raising achievement. It is easy to see why value-added analysis might represent an improvement over the status quo if we assume a fixed instructional model. But if we instead believe that we need to find entirely new ways to do things, and to let a thousand flowers bloom through charter schools and course-level instructional choice, the danger involved in locking ourselves into a fixed model becomes clear.
Testing Undesirable: Other Mechanisms Superior [cont’d]

4. **We no longer need high-stakes tests—other tools can provide the needed information**


But more importantly, we are seeing the emergence of new and better tools for assessing student performance, as Anya Kamenetz, author of a forthcoming book on the history and future of assessment, and a sharp critic of high-stakes testing, recently observed: The long term vision is software-based assessments that are “Embedded, low-anxiety, formative – not the assessment event that creates stress for teachers, principals, students, and families,” said Cameron Evans of Microsoft, and that serve as “durably predictive short term metrics,” said John Katzman of “real-world outcomes” like happiness, employability, and civic responsibility. “We have to think about assessment as something that should bolster education, not just measure it.” That is, instead of high-stakes tests administered at long intervals, Anya argues that we ought to move towards low-stakes software-based assessments that don’t need to be administered at all. Rather, they will be part and parcel of the work students do in the classroom. Anya’s vision of software-based assessment is closely aligned with the advent of the “flipped classroom,” a model in students view short video lectures at home while devoting their in-class time to completing exercises and projects, and interacting with teachers and fellow students.
Testing Undesirable: Reading Tests

1. Current reading tests, on balance, do more harm than good


But it’s long past time to acknowledge that reading tests—especially tests with stakes for individual teachers attached to them—do more harm than good. A good test or accountability scheme encourages good instructional practice. Reading tests do the opposite. They encourage poor practice, waste instructional time, and materially damage reading achievement, especially for our most vulnerable children. Here’s why: A test can tell you whether a student has learned to add unlike fractions, can determine the hypotenuse of a triangle, or understands the causes of the Civil War—and, by reasonable extension, whether I did a good or poor job as a teacher imparting those skills and content. But reading comprehension is not a skill or a body of content that can be taught. The annual reading tests we administer to children through eighth grade are de facto tests of background knowledge and vocabulary. Moreover, they are not “instructionally sensitive.” Success or failure can have little to do with what is taught. A substantial body of research has consistently shown that reading comprehension relies on the reader knowing at least something about the topic he or she is reading about (and sometimes quite a lot). The effects of prior knowledge can be profound: Students who are ostensibly “poor” readers can suddenly comprehend quite well when reading about a subject they know a lot about—even outperforming “good” readers who lack background knowledge the “poor” readers possess. Reading tests, however, treat reading comprehension as a broad, generalized skill. To be clear: Decoding, the knowledge of letter-sound relationships that enables you to pronounce correctly written words, is a skill. This is why early instruction in phonics is important. But reading comprehension, the ability to make meaning from decoded words, is far more complex. It’s not a skill at all, yet we test it like one, and in doing so we compel teachers to teach it like one. Doing so means students lose.

2. Even if testing is generally good, our current approaches to reading assessment fail


Students who score well on reading tests are those who have a lot of prior knowledge about a wide range of subjects. This is precisely why Common Core calls for (but cannot impose) a curriculum that builds knowledge coherently and sequentially within and across grades. That’s the wellspring of mature reading comprehension—not “skills” like making inferences and finding the main idea that do not transfer from one knowledge domain to another. As a practical matter, standards don’t drive classroom practice. Tests do. The first—and perhaps only—litmus test for any accountability scheme is, “Does this encourage the classroom practices we seek?” In the case of annual reading tests, with high stakes for kids and teachers, the answer is clearly “no.” Nothing in reading tests—both as currently conceived or anticipated under Common Core—encourages schools or teachers to make urgently needed, long-term investments in coherent knowledge building from grade to grade that will drive language proficiency. What could replace them? Options might include testing reading annually, but eliminating stakes; testing decoding up to grade four; or substituting subject-matter tests to encourage teaching across content areas. The best and most obvious solution would be curriculum-based tests with reading passages based on topics taught in school. But that would require a common curriculum—surely a nonstarter when mere standards in language arts are politically upsetting. Annual testing “makes clear that the standards associated with every tested grade and subject matter,” Andy writes. Again, I agree wholeheartedly. But reading is not a subject. It’s a verb. It’s long past time to recognize that reading tests don’t measure what we think they do. Accountability is essential and non-negotiable, and testing works. Just not in reading.
Testing Undesirable: Resource Tradeoffs

1. Testing is enormously expensive, and much of the money is wasted


President George W. Bush’s signing of the No Child Left Behind Act in 2002 ushered in the current era of high-stakes testing. The law required states to administer math and reading tests every year to students from third to eighth grade and imposed increasingly harsh punishments on schools that failed to make “adequate yearly progress” on these tests. By 2011, according to the Center on Education Policy, almost half of schools nationwide were labeled as “failing” because they could not make adequate yearly progress. The law has come with a hefty price tag for taxpayers. A 2012 study by the Brookings Institution determined that states spend $1.7 billion per year on testing, an enormous increase over the $423 million states spent in 2001 before NCLB, according to the Pew Center on the States. All of this money has fueled a booming testing industry, with companies like Pearson racking up billions in sales. A POLITICO investigation published on February 10, 2015 revealed that Pearson receives tens of millions in taxpayer dollars even though there is “little proof its products and services are effective. Now, with politicians on the left and the right dismissing NCLB as a failure, Congress is set to rewrite the law, and there is a vigorous debate over whether to keep the annual testing mandate. From the conviction of 11 teachers in Atlanta on racketeering charges following a high-profile cheating scandal to the rapid growth of the opt-out movement, evidence is mounting that the accountability system is broken.

2. Testing costs undermine efforts to decrease class sizes, which would boost achievement


Inner city public schools generally have the greatest problems with overcrowding. Indeed, inner-city schools are the schools most likely to be classified by the National Center for Educational Statistics (NCES) as "severely overcrowded." The NCES indicated in 1999 that overcrowding problems are so severe that in a number of classrooms there are twice as many students as desks and that this is most prevalent in inner-city public schools. If the federal government's primary purpose in passing NCLBA legislation was to close the achievement gap, as it contends, then federal resources would be better spent on alleviating the problem of school overcrowding. A considerable body of research highlights the superiority of smaller classes. Small schools have higher graduation rates, higher ACT and SAT scores, lower dropout rates, and more students going on to post-secondary education. According to the U.S. Department of Education, small schools also have fewer incidents of drugs, violence, crime, and alcohol and tobacco use. Of particular note, considering the federal government's emphasis on the use of test scores as indicators of students' and schools' performance, a Tennessee study revealed that students attending smaller classes in the primary grades outperformed their peers on reading and math tests and retained this lead through high school. Despite substantial evidence that smaller classes result in increased student performance, significantly fewer federal funds are directed towards decreasing class sizes when compared to those being spent on implementing the NCLBA testing requirement. Decreasing class sizes can be a costly enterprise, as it often involves building new schools or expanding and modernizing existing schools. Because public schools are primarily funded by local property taxes, it is difficult for many inner-city public school districts to pay for these improvements because they have weak tax bases. The allocation of federal, state, and local funding toward implementing NCLBA testing requirements only exacerbates this difficulty.
Testing Undesirable: Resource Tradeoffs [cont’d]

3. Testing costs trade off with technology use, which is better at increasing student achievement


3. The Cost of Implementing NCLBA Testing Creates Impediments to Increasing the Availability of Technology in Inner-City Schools

In general, students attending inner-city public schools use less technology for instructional purposes when compared to other students. The NCES reported that students attending schools with high minority and poverty enrollment rates were the least likely to use computers for a wide range of activities, including graphical presentations, multimedia presentations, word processing and spreadsheets, research using CD-ROM and the Internet, and corresponding with experts. Additionally, teachers at schools with high minority enrollment rates were the least likely to use computers or the Internet for gathering information at school, creating instructional materials at school, and communicating with colleagues at school. There is evidence that technology has a positive impact on student achievement, with some schools crediting their use of technology with gains in math and reading of up to twenty percent. The use of technology in the classroom also has a significant impact in a variety of different areas that are not necessarily measured by standardized testing. This includes the quantity and quality of student writing, as students benefit from the ease of typing and spelling and grammar checks; integrated learning, as access to the internet enhances teachers' ability to integrate different subjects in fewer assignments; co-operative learning, as teachers can use computers as learning stations and encourage co-operative projects; and application of different learning styles, particularly for students who are visual learners. Possibly most importantly, exposing students to technology in the classroom will aid in preparing them for their lives upon completion of a secondary education. At institutions of higher learning, a familiarity with technology has become important, as many college courses increasingly utilize technology. A familiarity with technology is also important in the working world because jobs are also becoming more technology-focused. As technology becomes increasingly ubiquitous in society, not exposing students to technology risks leaving them “in the global dust,” as Michigan state superintendent of public instruction Mike Flanagan fears. Similar to efforts to decrease class sizes, equipping schools with technology and then training faculty and students in the new technology is not an inexpensive undertaking. Many states contend that the NCLBA is underfunded, and states do not have unlimited education budgets. Therefore, because the Act mandates compliance from states seeking Title I funding, the testing requirement may compromise effective integration of technology in the schools due to re-allocation of limited funds that otherwise may have been spent on technology initiatives.

4. Testing costs trade off with resources better used in other areas—are very expensive


The cost of implementing the testing requirement is exorbitant. The federal and state funds required to implement the requirement would be better spent elsewhere. Alternatives include programs aimed at decreasing class sizes, implementing technology in classrooms, and attracting and retaining competent faculty. 1. The Cost of Implementing NCLBA Testing There is debate as to whether the NCLBA is underfunded or overfunded. Studies arguing that the Act is underfunded focus only on the “hard costs” of implementation, which include developing new tests, administering tests, professional training and development, and accountability information systems. The average of seven independent state studies estimating the hard cost of implementing the testing requirement is $ 11.3 billion. Former Secretary of Education Rod Paige contends that these costs are fully covered by the more than fifty percent increase in Federal Title I appropriations since the passage of the Act. Studies contending that the Act is underfunded are referred to as standards-based adequacy studies and take into account a more comprehensive list of resources needed to maximize achievement and do not address administrative costs. Based on the median amount of forty standards-based adequacy studies, the cost of complying with the testing requirement is estimated at $ 137 billion. Adding in the administrative costs of $ 11.3 billion, the total estimated cost is $ 144.5 billion. Regardless of whether the total cost is $ 11.3 billion or $ 144.5 billion, the costs associated with implementing the testing requirement are substantial. Federal funds used to pay for the hard costs of implementation would be better spent on alternative, more effective programs aimed at increasing overall achievement and closing the achievement gap. The non-hard costs of implementation, which are not paid for by the federal government but by the states, impede to states' abilities to implement alternative strategies.
Testing Undesirable: SAT—General

1. The SAT reinforces and exacerbates existing social inequalities


This Note argues that the current SAT produces and reinforces educational inequality, and demonstrates how such inequality could be reduced if all universities adopted the SAT-optional paradigm. This section will provide a brief overview of the background literature, methodology, and outline for this Note. The inequality and stratification literature often draws a distinction between ascription and merit (or ascription and achievement). Ascription is a process by which placement in society is beyond the control of the individual and is determined instead by status of parents, race, or gender.Merit, in contrast, is achievement tied to an individual's objective ability. The less congruence between merit and placement, and the greater the influence of ascription, the more unequal is a society. In order for inequality to persist, some mechanism must legitimize the ascription in a form acceptable to those marginalized. To this end, societies have attempted to use custom, legal, and ideological rationales for stable inequality. Some scholars have suggested that the SAT is such a mechanism for creating and maintaining stable inequality. While the SAT purports to test innate intellectual aptitude, the SAT may in fact reflect pre-test inequalities, including those based on social group status and on an individual's form of intelligence. If the SAT legitimizes inequality rather than objectively testing merit, then the importance of the SAT in the U.S. college admissions system should be diminished. This Note argues that (1) the SAT reinforces social group inequality by transforming unacceptable historic and persistent inequality into acceptable merit-based distinctions, (2) the SAT creates inequality based on form of intelligence possessed by elevating logical-mathematical and linguistic intelligence above all the other forms, and (3) the SAT-optional paradigm would help address these inequality problems by providing tests with more school support and reflecting diversity of disciplines and forms of testing. These arguments rest upon the premise that, in the United States, higher educational achievement carries with it a certain status level and that economic and societal returns flow from that perceived status. Importantly, it must be emphasized at the outset that even broad adoption of the SAT-optional paradigm would not suddenly transform an unequal education system into a system of complete equality. A multivariate problem such as educational inequality requires a multi-faceted solution. However, as this Note will demonstrate, de-emphasizing the SAT is an important, substantive first step toward equalizing educational opportunity.

2. Many groups perform more poorly on the SAT—multiple reasons


Many explanations have been suggested for why students from underprivileged social groups perform worse on average than their counterparts in privileged social groups. These explanations include: (1) family income disparity; (2) poor educational preparation; (3) tracking bias; (4) lack of role models; (5) peer ridicule; (6) stereotype threat; (7) subject matter bias; (8) segregation bias; and (9) lack of sufficient English language skills. This section will discuss each of these possible explanations in turn, with Table 1 summarizing these explanations.
Testing Undesirable: SAT—General [cont’d]

3. SAT reliance reinforces inequalities between groups


A. Primary Reliance on the SAT Is Misplaced Because the SAT Reinforces Social Group Inequality This subsection argues that American universities should not place substantial reliance on the SAT because the SAT reinforces social group inequality by purporting to test all students fairly but not actually doing so. 1. Certain Social Groups Perform Worse on the SAT than Other Social Groups The Educational Testing Service (ETS) claims the SAT helps minority applicants. This assertion by the ETS seems logical and the SAT appears facially as the height of meritocracy. In theory, if one is intelligent and prepared, and if the SAT actually tests intelligence and college preparedness, then the SAT should provide a means for even the poor or marginalized to reach the upper echelon of academic society. To use the Michael (disadvantaged) and David (advantaged) cases, Michael was presented with challenges David did not face. One could make the argument that if Michael was talented enough, his true talent would show through his strong SAT score. This would allow him to enter a prestigious university and he would be welcomed into the upper echelon of academic society. Although attractive in theory, in practice the SAT does a poor job of identifying talented, but underprivileged, students. Research studies have clearly demonstrated that certain social groups systematically perform worse than other social groups on the SAT. First, there exists a racial scoring disparity. The College Board, which administers the SAT, found in its own study that, "[t]here are substantial differences among groups in mean scores on standardized tests., with the largest gaps between white and African-American test-takers, followed by Hispanic test-takers." Using 2006 data, it was shown that whites score, on average, 17% higher than blacks, and Asian-Americans score 19% higher than blacks. More recent validity studies have confirmed the persistence of this racial scoring disparity. Thus, many minority students are admitted to colleges "in spite of their SAT scores, not because of them." Second, there exists a socioeconomic scoring disparity. Research studies have confirmed that individuals from lower socioeconomic backgrounds tend to perform worse on the SAT than those from higher socioeconomic backgrounds. Therefore, instead of alleviating social group inequality, research has shown that the SAT at best reflects the pre-existing social group inequality in society.
Testing Undesirable: SAT—Specific Mechanisms

1. The SAT only reflects existing inequalities—tracking bias


A third reason why certain social groups may perform relatively poorly on the SAT is tracking bias. Certain disadvantaged or minority students may be "tracked" early on, meaning they are pushed toward vocational pursuits instead of a college education. Suppose that the students David and Michael were of equal intelligence and had received equal preschool preparation. Once the two arrived at elementary school, however, David was placed in the academic track, while Michael was pushed toward more vocational courses. Over time, Michael may be encouraged to take metal shop or home economics instead of higher-level math. Michael would learn many skills that would be helpful in life, but ones that are less helpful on the SAT. David, in contrast, would take the highest-level math and English courses, and may have more practice sitting behind a desk with a pencil and paper for long periods of time. Moreover, Michael may feel relatively little incentive to prepare for the SAT if he expects to move directly into his vocation, possibly starting a lawn care business after high school. David, however, may feel a great deal of incentive to study if he aspires to work eventually in a field like international politics. In this case, Michael's lack of incentive to study would further intensify the impact of tracking on his SAT performance. Thus, if students from minority social groups are being disproportionately tracked toward vocational training and away from traditional college preparatory classes, this may help explain the relatively poor performance of certain minority social groups.

2. The SAT reflects current inequalities—exacerbated by the lack of role models for some groups


A fourth reason for the relatively poor SAT performance of certain social groups is the lack of sufficient role models. For better or worse, students are influenced by those they admire. David, who is rich, white, and was raised in an upper class community, probably has many role models who will encourage him to study hard, apply himself on the SAT, and maximize his potential in life. Even if David's parents do not value education, he can look to his teachers, people in his community, and especially his peers. Because David is a member of multiple social groups that traditionally succeed in the professional world and in academics, David may also rely on those who look like him on television or in newspapers as proof that it would be possible to succeed in a profession requiring strong academic skills. On the other hand, a student like Michael who was raised in a very poor, urban school may not have the support of a role model to help him succeed. Michael may not have had adequate parental support because his mother spent time in jail and did not have much success academically, likely due to disempowering forces that affected her. As a result, she could not be a sufficient educational role model for Michael. Additionally, because Michael grew up in a poor area with a good deal of crime, he may not be able to rely on others in his community (who themselves may not have had the opportunity to attend college) to be his academic role models. Likewise, Michael's peers might not provide the best examples of educational success and may instead encourage Michael to work rather than study. Michael may even have difficulty identifying with successful potential role models who he sees on television or in the newspaper because he does not believe that he could actually achieve what relatively few others in his social group have achieved (due to a history of discrimination and disempowerment). For these reasons, Michael and students like him may, on average, perform worse on the SAT over time than David and students like him. Therefore, a lack of sufficient role models might help explain the relatively poor SAT performance of certain social groups.

3. Peer ridicule problems dictate that SAT scores reinforce underlying social inequalities


Fifth, peer ridicule may help explain the disparity in test performance between privileged social groups and non-privileged social groups. Peer ridicule is a process whereby one's peers berate him or her for trying to act like part of a different social group than the group to which one belongs. In the African-American context, sometimes such berating consists of an African American being accused of "acting white." As such, a student like Michael may, in certain contexts, have difficulty finding support among peers for activities relating to studying and education. Because success in academic fields is often not associated with social groups to which Michael is a member (due to historic discrimination), Michael's peers may think that he is being a traitor to his group or trying to become part of David's social group. They may ridicule Michael for trying to act like a member of David's social group, and such ridicule may cause Michael to choose not to study as hard as he might, or to allocate more time to activities valued by his social peers. Thus, peer ridicule may help explain the relatively poor SAT performance of certain social groups.
Testing Undesirable: SAT—Specific Mechanisms [cont’d]

4. The SAT reinforces social inequality—stereotype threat

Sixth, stereotype threat may help explain the disparity in test performance between privileged social groups and non-privileged social groups. Stereotype threat is a psychological phenomenon suggesting that when a group is aware that a certain stereotype about the group exists in society, and members of the group are reminded that they are part of this social group and thus subject to that stereotype, they are more likely to become victim to acting in compliance with the stereotype. Thus, if a teacher informs students that they are "low achievers" or reminds them that negative educational stereotypes apply to them, then the students may actually perform poorly. For instance, research has shown that when African-American students are asked to identify themselves as African Americans before they take a standardized test, such students perform more poorly on the test than if the students are not asked to identify as African American. This happens because negative stereotypes exist about African Americans and test performance, and students are aware of such stereotypes. Knowledge of such stereotypes causes students to become nervous or to be primed such that they perform worse than they otherwise would have. The lead academics on the issue of stereotype threat, Claude Steele and Joshua Aronson, state, “we surmise that over the entire range of Black student [SAT] test takers, stereotype threat causes a significant depression of scores.” In the David and Michael cases, Michael might be constantly reminded of societal stereotypes aimed at the various minority social groups of which he is a member. If Michael is reminded of these negative societal stereotypes, he may perform worse on the standardized SAT than he otherwise would have. Therefore, stereotype threat may help explain the relatively poor SAT performance of certain social groups.

5. The SAT reinforces inequalities—subject matter bias

Seventh, subject-matter bias may help explain the disparity in test performance between privileged social groups and non-privileged social groups. Subject-matter bias means that the type of questions, context of questions, and issues and situations described in questions may favor certain social groups over other social groups. For example, consider the much reviled “regatta” question. Students were presented with an analogy question that required them to pick the answer choice reflecting the same relationship as “runner” to “marathon.” The correct answer was the relationship of “oarsmen” to “regatta.” This question has drawn attention because it is quite likely that a child who grew up in the inner city or was raised in a poor family has never seen an oar, let alone knows what a regatta is. Even though questions as biased as the “regatta” example have been mostly eliminated from the test, there are still questions that will favor certain social groups over others. A question about cricket, for instance, may be more easily answered by the sub-group of Americans who can afford to travel abroad or watch foreign sports on paid cable television. Moreover, certain schools with high African-American populations may rightly focus their curriculum on African-American history and culture. Because African-American culture is probably less reflected on the SAT than majority culture, African Americans who attend such schools will be disadvantaged by the test. In the Michael and David hypothetical, Michael may be much less likely to have been exposed to the issues and culture reflected in the SAT and, therefore, may perform worse than David. Therefore, subject-matter bias may help explain the relatively poor SAT performance of certain social groups.

6. The SAT only reflects the fact that many of our schools are still segregated

Eighth, segregation bias may help explain the disparity in test performance between privileged social groups and non-privileged social groups. Segregation bias means that certain minority groups may be harmed because they live in de facto segregated areas of the country, and the de facto segregation may prevent such students from regularly coming into contact with the majority culture reflected on the SAT. If a student is surrounded by friends and family who discuss certain sports, certain music, or certain books, that student is likely to be familiar with those sports, that music, and those books. Thus, if these are the sports, music forms, and books reflected on the test, then the student is more likely to perform well than someone less familiar with these sports, music forms, and books. In the Michael and David cases, David was clearly surrounded by friends, family, teachers, and others in the community who could introduce him to the type of culture reflected in the SAT. Michael, in contrast, is unlikely to have been introduced to the culture reflected on the SAT. Thus, Michael may perform worse on the SAT than David if all else is equal. Therefore, segregation bias may help explain the relatively poor SAT performance of certain social groups.
Testing Undesirable: SAT—Specific Mechanisms [cont’d]

7. Privileged groups do better on the SAT because of differential language skills


Ninth, poor English skills may help explain the disparity in test performance between privileged social groups and non-privileged social groups. It is intuitive that not speaking the language means that one cannot perform as well on the literature portion of the test as someone better at English. Even in math, the directions and question headings are in English. Moreover, the "Queen's English," which is reflected on the test, is not necessarily the type of English spoken by people on the street (and in certain areas such English is not spoken at all). Thus, even if a person speaks English, that person may have learned different vocabulary or grammar because that person lives in a community that uses different vocabulary or grammar. Because the SAT reflects the sort of "Queen's English" spoken in elite society, a student who is not exposed to that type of English will probably perform worse than a student who has been introduced to that form of high English. In the Michael and David hypothetical, David will certainly have learned the type of language skills useful for the test. David may have been exposed to the word "augment," while Michael was only exposed to the word "increase." David's parents may be meticulous about pronunciation and grammar, while Michael may not have been encouraged to speak "properly." Thus, if Michael does not speak the kind of English tested on the SAT, Michael may perform worse than someone who does speak the kind of English tested (such as David). Therefore, poor English skills may help explain the relatively poor SAT performance of certain social groups.
Testing Undesirable: SAT—Wealth

1. The SAT only reflects and reinforces existing social privilege structures


For these reasons, despite the fact that many believe the SAT to be an objective test of merit, the SAT does not in fact test all social groups equally. Because the SAT does not reflect a perfect congruence between talent and score--instead reflecting at least the pre-test inequalities discussed above--the test can actually provide a politically correct and fully acceptable means of discriminating against certain members of disempowered social groups. "Tests like the SAT con-vert the tainted advantages of birth and wealth into the neutral currency of 'merit,' enabling the fortunate to believe they have earned what they have merely been given." The privileged do not "see SAT results as proof that the poor are mistreated. They see them as proof that mistreatment of the poor is fair." The disadvantaged, in turn, accept that they deserve lower scores, believing they simply lack the intelligence needed to score better. The SAT's power to transform the illegitimate into the legitimate in a way accepted by disempowered groups reinforces and stabilizes societal inequality while encouraging historic social group disempowerment to continue.

2. SAT results reinforce inequality—are heavily influenced by income / wealth disparities


First, the income disparity between certain social groups influences SAT performance. Because racial minorities are more likely to be of lower socioeconomic status, it is possible that the correlation between racial minority status and lower socioeconomic status helps explain the SAT scoring disparity. Wealthy test takers have multiple advantages when taking the SAT. First, the rich have the necessary resources to take expensive preparatory classes. Test-preparation is a big business and companies such as Kaplan offer a range of services for test preparation. Students may attend live classes or they may attend over the Internet. In fact, Kaplan is so convinced a test preparation course will improve the score of students that the company offers a guarantee whereby students may receive a refund if their score does not improve. Thus, if relatively few students like Michael are able to afford these expensive test prep courses, and if these prep courses do indeed provide helpful preparation for the SAT, then this might help explain the relatively poor performance of certain minority groups. Second, wealthier students are probably less likely to be required to work outside of school during the time immediately preceding the SAT. A student like David would probably have very little incentive to work at a part-time job rather than prepare for the test in the months preceding the SAT. Even if a student like David chose to work, such an advantaged student could more confidently ask for reduced hours or a temporary leave in order to have the time necessary for test preparation. A disadvantaged student like Michael, for whom being fired may mean insufficient sustenance for himself or his family, might find a reduced paycheck an unacceptable hardship. If SAT preparation does indeed increase the likelihood of a better score, David's greater employment flexibility may create a disproportionate advantage in his favor.

3. Wealthy students are able to better prepare for the test


Third, wealthier students like David are more likely to have the necessary resources to take the exam multiple times than are poorer students like Michael. The ability to take the test multiple times provides several advantages. To begin, if a student like David is able to take the SAT several times, he might consider the first time he takes the test to be preparation for subsequent times, allowing him to become familiar with the question types, directions, and length and flow of the SAT exam. Further, if a student like David is able to take the exam multiple times, the student will probably be less nervous and feel more comfortable with all the procedures preceding the actual test (such as filling out the 'bubbles' with name, address, and other information). Moreover, the act of taking the test several times gives a student like David more opportunities to show what he knows. Perhaps during the first SAT administration, Da-vid was not feeling well, had just broken up with his girlfriend, or was overly nervous. By retaking the test, David may overcome such issues and may perform to his full potential. A poorer student like Michael, unable to afford it, would probably not receive multiple opportunities to take the SAT. Thus, there is reason to believe that to the extent minority status is correlated with generally lower socioeconomic status, lower family income may help explain why certain minority groups perform relatively poorly on the SAT. However, it is also important to note that socioeconomic status is not the exclusive explanatory factor. Empirical data has shown "[w]hites from families with incomes below $ 10,000 had a mean SAT test score that was 17 points higher than blacks whose families had incomes of more than $ 100,000." Therefore, while family income is a compelling factor for unequal social group SAT performance, other compelling factors also help explain the relatively weak SAT performance of certain social groups.
Testing Undesirable: SAT—Wealth [cont’d]

4. SAT scores only reflect the underlying inequalities in the quality of our schools


In addition to financial disadvantages, a second factor that may explain why certain social groups perform relatively poorly on the SAT is poor educational preparation. Certain areas of the country have public schools that are poorly staffed and resourced, when compared to average public schools. Students who come from these schools may not receive the same level of instruction in math and English, perhaps explaining why students coming from these schools perform worse on the SAT. The problem that under-staffing and under-funding of schools creates is compounded by the fact that many of these under-resourced schools have high racial minority populations, thereby adding to the race divide in SAT performance. In the Michael and David cases, it is quite likely that David received better educational preparation at the Lincoln School than Michael received in the likely understaffed and under-funded public school in the poor region where Michael lived. If lack of staff and resources hinder SAT performance, and certain students like Michael are raised in schools that lack sufficient staff and resources, this may cause students from certain social groups to perform worse in the aggregate than students from more privileged social groups.
**Testing Undesirable: Standards Lowering**

- **Our current testing regime only encourages states to lower standards**


  The perverse incentives of NCLB clearly have something to do with the downward movement of state standards. NCLB punishes a school when too few of its students make progress toward math and reading proficiency — beginning with extending to students the choice of other public schools to attend and culminating with completely restructuring the school. But the law has a gaping loophole: States get to define proficiency however they want. A state can thus meet NCLB targets by defining proficiency down. Toughening its standards, by contrast, handicaps its ability to meet the federal requirements. Of course, low standards have their own appeal. The lower the standard, the more students pass it, and lots of supposedly proficient students make an education system look as though it’s working well, even when it’s not. State governments love to tell constituents that students are doing great on standardized exams; the public usually just assumes that the criteria used on those exams are meaningful. States can get away with setting low standards in part because proficiency remains an inexact notion. In determining the difficulty of exam questions, states typically rely on the judgment of expert panels, whose members ask themselves: “Would a ‘proficient’ student answer this question correctly, or is it so difficult that a student who gets it right should be considered ‘advanced’?” The percentage of students who answered the question correctly also gets taken into account. Add to all this the fact that what constitutes an expert varies from state to state — and that the experts only make suggestions, which the state may ignore — and it’s clear that the process has an unavoidably subjective dimension.
Testing Undesirable: Student Motivation

1. Our testing focus hurts student motivation—they learn for reward, as opposed to doing so for a love of learning


The assumption surrounding current testing methods is that children will be motivated to learn when the associated rewards and consequences are made clear (Raymond & Hanushek, 2003). Yet, researchers have consistently found that an approach based on extrinsic rewards and consequences actually reduces children’s intrinsic motivation to learn (Amrein & Berliner, 2003; Good & Brophy, 1995; Kohn, 1993). Sheldon and Biddle (1998) boldly claimed that attaching high stakes to tests “obstructs students’ path to becoming lifelong, self-directed learners and alienates students from their own learning experiences in school” (p. 170). Because of high-stakes testing and the pressure that surrounds it, children are no longer engaged in enriching experiences for the pure joy of learning—experiences whereby they make decisions, explore options, make hypotheses, or problem solve. Extrinsic motivation, in the form of rewards and consequences, has replaced learning for the sheer pleasure of learning and the internal satisfaction that comes from a job well done. Children are now under increased pressure to perform on demand, memorize mundane facts and figures, and sit for long periods of time while listening to the teacher and/or filling in circles on a worksheet.

2. Testing demotivates—contributes to higher dropout rates

Bobbi A. Solley, Professor, Department of Elementary and Secondary Education, Middle Tennessee State University, “On Standardized Testing: An ACEI Position Paper,” CHILDHOOD EDUCATION, Fall 2007, p. 34.

Furthermore, an overreliance on extrinsic rewards and the subsequent lack of learning that follows has led to an increase in retention rates and an associated higher drop-out rate. In Louisiana alone, between 10 and 15 percent of 4th- and 8th-graders were retained in 2000 because of failure to pass the state’s high-stakes test (Robelen, 2000). And, in Florida, in the spring of 2003, more than 43,000 third-graders (25 percent of the total for that grade level) were not allowed to advance to 4th grade, due to their insufficient scores on standardized tests (Garan, 2004). Because of the correlation between retention and drop-out, motivation to learn and the desire to finish school has lessened (Goldschmidt & Wang, 1999). A study conducted by Nichols, Glass, and Berliner (2005) found that highstakes testing pressure is negatively associated with the likelihood that 8th- and 9th-graders will eventually enter and complete 12th grade. The use of standardized testing, along with the resultant system of extrinsic rewards and consequences, has had a negative effect on students’ motivation. As a result, students’ fear of failure has lessened their motivation to learn.
Testing Undesirable: Student Stress

1. The tests induce damaging fear and anxiety in many students

United States Secretary of Mis-Education Arne Duncan derisively dismissed parental protests against the high-stakes testing regime labeling opponents as "white suburban moms who -- all of a sudden -- their child isn't as brilliant as they thought they were, and their school isn't quite as good as they thought they were." Despite what Duncan says there are many scientific studies that have documented "performance anxiety," the impact of stress on performance. The nervousness students feel before a test can be so intense that it interferes with concentration and performance. Persistent fear and anxiety can also affect a young child's learning and development.

2. The stresses of testing bleed students of their love of learning

Bobbi A. Solley, Professor, Department of Elementary and Secondary Education, Middle Tennessee State University, “On Standardized Testing: An ACEI Position Paper,” CHILDHOOD EDUCATION, Fall 2007, p. 34.
Research by Glasser and Glasser (2003) indicates that stress increases the rate of aging and reduces the functioning of the immune system. The researchers also state that the worst kind of stress is caused when we have little or no control over our lives. As children are inundated with standardized tests, the resulting mundane methodologies of teaching in order to prepare for the test has both teachers and children feeling helpless. Sacks (1999) also talks of the dangers of test-driven classrooms: "Test-driven classrooms exacerbate boredom, fear, lethargy, promoting all manner of mechanical behaviors on the part of teachers, students, and schools, and bleed school children of their natural love of learning" (pp. 256-257).
Testing Undesirable: Teacher Evaluations

1. Adding new teacher evaluation mechanisms will only overstrain already stretched school budget


With these changes, as MORE has put it, teaching may "be reduced to a series of mechanical steps. Even the most skilled and veteran teacher, one whose experience informs their teaching style, will be forced to ignore their professional judgment when it conflicts with a supposedly 'objective' observation rubric." The new system will "pressure teachers to enforce a more narrow, lock-step curriculum."

Diane Ravitch observes that the evaluation system "will certainly produce an intense focus on teaching to the tests. It would also profoundly demoralize teachers, as they realize that they have lost their professional autonomy and will be measured according to precise behaviors and actions that have nothing to do with their own definition of good teaching." She goes on to say, "Evaluators will come armed with elaborate rubrics identifying precisely what teachers must do and how they must act, if they want to be successful." Furthermore, school districts will have to "hire thousands of independent evaluators, as well as create much additional paperwork for principals. Already stressed school budgets will be squeezed further to meet the pact's demands for monitoring and reporting."

2. So-called “value added measures” used to assess teaching are just junk science


Harry Braverman explains in Lahor and Monopoly Capitalism that for business, "every non-producing hour" someone is employed is a loss. Therefore, management pursues "complete, self conscious, painstaking, and calculating" control over the production process.' Facing stiff competition in the market, capitalists are driven to streamline production, splitting up skilled work into discrete tasks that can be executed by less skilled workers. This dynamic is "the underlying force governing all forms of work in capitalist society. "> Of course, for the most part the public sector does not directly face market competition, but is subject to political processes. Ed Deform seeks to bring market-type pressures to bear on teachers' labor. This requires a metric for measuring teacher productivity and quality, which is what "Advance" is designed to provide. Education experts like Diane Ravitch have branded "Value Added Measures"—formulas used to quantify teacher impact on student test scores—as "junk science." "Scientific management," created by Frederick Taylor in the 1880s, was the original junk science. As Braverman puts it, scientific management—Taylorism—does not seek to improve production in general, but adapts "labor to the needs of capital. It enters the workplace not as the representative of science, but as the representative of management masquerading in the trappings of science."
Testing Undesirable: Teacher Retention

- Tests increase teacher anxiety—many end up leaving the profession


Evidence from an extensive study conducted by Smith (1991) suggested that standardized tests raise the anxiety level of educators because of public notification, and sometimes parent ridicule, of school test scores. Since teachers are under so much stress to keep their jobs many of them are quitting. Lou Anne Johnson (2011) describes the reasons why teachers are quitting when she states, “They quit because they expect to teach, not shuffle endless paperwork, and spend weeks of their ‘free’ time learning how to teach to the test” (p. 29). Teachers think the standardized tests are used against them especially when their students receive a low rank on the test. Wallace and Irons (2010) confirm that belief by stating “school accountability and teacher effectiveness are often measured through student performance on high stakes tests” (p. 166). In addition, novice teachers are leaving the profession because they lack classroom resources, parental support, and professional development that all affect testing outcomes (Keogh, Garvis, Pendergast, & Diamond, 2012). Another novice teacher who is in her 2nd year high school chemistry teacher notes, “... school test rankings and scores are publicized to parents, local media outlets, on the internet, and throughout the school district. If the scores don’t indicate improvement, the school’s reputation is at risk. It can’t help but raise a teacher’s anxiety level. Recently we read in a newspaper that a teacher was caught while helping students with the answers during the test. The teacher was fired. Who is to blame?”
Testing Undesirable: Teaching to the Test

1. Testing pressures force low-resource schools to teach to the test, undermining overall learning

Again, there’s no denying that many schools serving low-income children of color were second-rate to begin with. Now, however, some of these schools, in Chicago, Houston, Baltimore, and elsewhere, are arguably becoming third-rate as testing pressures lead to a more systematic use of low-level, drill-and-skill teaching, often in the context of packaged programs purchased by school districts. Thus, when someone emphasizes the importance of “higher expectations” for minority children, we might reply, “Higher expectations to do what? Bubble-in more ovals correctly on a bad test—or pursue engaging projects that promote sophisticated thinking?” The movement driven by “tougher testing,” “accountability,” and similar slogans arguably lowers meaningful expectations insofar as it relies on standardized testing as the primary measure of achievement. The more that poor children fill in worksheets on command (in an effort to raise their test scores), the further they fall behind affluent kids who are more likely to get lessons that help them understand ideas. If the drilling does result in higher scores, the proper response is not celebration, but outrage: The test results may well have improved at the expense of real learning.

2. Teaching to the test confounds the entire purpose of getting an education

High-stakes standardized tests, and the new curriculum they have spawned, urge teachers to avoid thinking deeply about the messages we’re sending to students, and to ignore whether kids should even be taking the tests we administer. The aim of testing is to provide demonstrable, measurable evidence that work is being produced — that teachers and students are not playing when they should be lecturing, memorizing, studying. The idea is that a teacher’s job is to get information into the heads of students, and a student’s job is to write it out, unchaged, on a test. When we teach content and skills that will be on the state test instead of history and politics, language and literature, we neutralize the power of knowledge and undercut the possibility of building relationships with our students. By insisting that the structure of class remains unchanged day to day (and enforcing it through teacher observations and evaluation, as many urban school districts do), we lose the opportunity to be surprised by kids and to take a lesson in a completely unexpected direction. This inevitably leads to teachers designing lessons in which they hold the answers and the kids guess — instead of thinking critically and creatively. Every day we rob kids of the excitement of engaging with the human conversation captured in books.

3. The current regime makes it such that teachers and administrators focus on nothing but the test

Shortly before winter recess, the middle school faculty in a small, suburban Connecticut town received a foreboding e-mail from the principal imploring them to aid in controlling what appeared to be the inevitable onset of the madness known as “the approach of the holidays,” so designated by the subject line. At this point of the school year I write to beseech/beg/ plead/entreat/request that you not give in to the HOLIDAY HYSTERIA that oftentimes ensues when young people begin to fixate on the happenings related to the End [sic] of December. We must hold the line. We must dig in . . . I ask that you minimize in all ways anything related to the forthcoming holiday that is unrelated to the work directly tied to your curriculum. We must hold high academic and social expectations until the last minute of the last hour of the last day before vacation. Now is NOT a good time to relax and have some holiday fun! Although one may wonder what kind of heartless individual—aside from the Grinch—would communicate such disdain for the holiday season, the statement is less a reflection of the administrator as it is of the current state of education. Academic achievement is now limited to a single measure: the standardized test. As pressure mounts for all students to meet a specific standard on these assessments, more emphasis is placed on the preparation needed to meet this goal at the expense of all else. In the article “Teaching Themes of Care,” Noddings (2007) suggests that individuals such as the middle school principal are not alone in their apprehension: “Many otherwise reasonable people seem to believe that our educational problems consist largely of low scores on achievement tests” (64). Although narrowing the scope of education in such a way may seemingly simplify the work of teachers and administrators by allowing them to focus on a defined set of skills and standards, it poses a much more vital question: Are we, as educators, producing commodities or educating children?
4. **It only encourages teaching to the test—undermines learning objectives**


Teaching to tests has become a hindrance to actual learning and should be limited. President Barack Obama called for a 2 percent cap on classroom time spent on standardized testing. The rise of standardized testing and the amount of time spent teaching to the tests arose from a fear that Johnny wasn't learning to read, write or do math. Standardized testing became all the rage in an effort to determine the success of schools at their main mission. As it became the standard for how schools and teachers were judged, it became too big a focus of teachers in the classrooms. Teachers began spending more and more time preparing students for tests that actual learning fell by the wayside. Standardized testing became too important and our children are suffering for the misplaced priority. Educators, administrators, and lawmakers should reset priorities that enable actual teaching to be performed in class-rooms. The ultimate test of success is whether students can be successful in the real world. The emphasis on standardized testing does not help.
Testing Undesirable: Time Tradeoff

1. Tests currently occupy too much class time

Michelle Toh, journalist, “Why Obama Is Placing New Limits on Standardized Testing,” CHRISTIAN SCIENCE MONITOR, 10—24—15, www.csmonitor.com/USA/Education/2015/1024/Why-Obama-is-placing-new-limits-on-standardized-testing, accessed 11-5-15. President Obama on Saturday introduced new federal guidelines to curb the emphasis on standardized testing in public schools, signaling a turnaround on a subject that has in recent years grown to new heights of contention. Mr. Obama is now calling for standardized testing to take up to 2 percent of class time, compared to the 2.3 percent now used by the average eighth-grader, reports The Associated Press. Standardized tests became mandatory for 2001 for students in third grade and higher under No Child Left Behind, a law legislators voted to overhaul in July. Students currently spend 20 to 25 hours each school year taking standardized tests, according to a study of the nation’s 66 largest school districts also released Saturday by the Council of Great City Schools. While Obama cannot enforce the new policy in individual states or districts, the Department of Education has rolled out a plan that would help these local governments reevaluate their testing programs, The Wall Street Journal reports. "Learning is about so much more than just filling in the right bubble," the president said in a video announcement posted on Facebook. "So we're going to work with states, school districts, teachers, and parents to make sure that we're not obsessing about testing." The administration's recommendations are focused on three basic principles, he said. "First, our kids should only take tests that are worth taking," Obama said. "Tests shouldn't occupy too much classroom time or crowd out teaching and learning. Tests should enhance teaching and learning. And third, tests should be just one source of information." Obama's announcement Saturday "acknowledged its own role in the proliferation of tests," writes The New York Times. "I still have no question that we need to check at least once a year to make sure our kids are on track or identify areas where they need support," said Secretary of Education Arne Duncan. "But I can't tell you how many conversations I'm in with educators who are understandably stressed and concerned about an overemphasis on testing in some places and how much time testing and test prep are taking from instruction." The shift to a more moderate approach to testing comes as teachers' unions and parents have protested what they call "over-testing," and an increasing number of parents and students this year decided to "opt out" from required exams, The Christian Science Monitor reported. The Department of Education has for months been piloting alternative assessments to statewide standardized testing, the Monitor reported. A number of public school districts in New Hampshire, for example, were the first to debut a pilot program this year focusing on what's known as "competency-based learning," a tasks-based review of how to apply new subject material. "It fits into a much bigger conversation about ... how we can create a humane assessment system that's useful to teachers but also useful to states and the federal government for holding schools accountable," Julia Freeland, a research fellow at the Clayton Christensen Institute for Disruptive Innovation, told the Monitor.

2. Test preparation eats up valuable classroom time—particularly true for disadvantage students

Quinn Mulholland, staff, “The Case Against Standardized Testing,” HARPARY POLITICAL REIVEW, 5—14—15, http://harvardpolitics.com/united-states/case-standardized-testing/, accessed 11-9-15. The focus on test prep eats up time that could be spent doing hands-on projects and collaborative, interactive activities. Neely-Randall told the HPR, "In the beginning, we were doing all of these great projects and they were fluent readers and writers … and then all of a sudden, I had to stop everything to get them ready for a test." In fact, according to a report published in 2013 by the American Federation of Teachers (AFT), students in heavily tested grades can spend over 110 hours per year doing test prep, and as many as 50 hours per year taking the tests themselves, a total of roughly 15 percent of their instructional time. The schools that have been forced to devote the most time to test prep are those in the most disadvantaged communities, because they have to achieve the biggest increases in test scores under NCLB’s mandates. Robert Schaeffer, the public education director at the National Center for Fair and Open Testing, explained in an interview with the HPR, “In those kinds of schools, the curriculum becomes test prep: doing worksheets and practice tests and getting ready for the big test.” A report from the Center for American Progress substantiates Schaeffer’s claim, demonstrating that urban high school students spend as much as 266 percent more time taking standardized tests than their suburban counterparts do.

3. Test prep time undermines the quality of the education received by students

Quinn Mulholland, staff, “The Case Against Standardized Testing,” HARPARY POLITICAL REIVEW, 5—14—15, http://harvardpolitics.com/united-states/case-standardized-testing/, accessed 11-9-15. This increased focus on test prep has had a profoundly negative impact on the quality of education many students receive. Randi Weingarten, the president of the American Federation of Teachers, the second-largest teachers union in America, said in an interview with the HPR that the narrow focus on tested subjects causes students to become disengaged at school. “Most kids I know are so anxious about the high-stakes consequences of these tests right now that they hate school, but yet they can be really engaged if we engage them through music or through art or through projects.”
4. Testing simply overburdens students—they spend a huge amount of time taking tests


The number of standardized tests U.S. public school students take has exploded in the past decade, with most schools requiring too many tests of dubious value, according to the first comprehensive survey of the nation's largest districts. A typical student takes 112 mandated standardized tests between pre-kindergarten classes and 12th grade, a new Council of the Great City Schools study found. By contrast, most countries that outperform the United States on international exams test students three times during their school careers. In a video posted to Facebook by the White House on Saturday, President Obama pledged to take steps to reduce testing overload. In “moderation, smart, strategic tests can help us measure our kids' progress in school, and it can help them learn,” Obama said. "But I also hear from parents who, rightly, worry about too much testing, and from teachers who feel so much pressure to teach to a test that it takes the joy out of teaching and learning, both for them and for the students. I want to fix that.” The heaviest testing load falls on the nation's eighth-graders, who spend an average of 25.3 hours during the school year taking standardized tests, uniform exams required of all students in a particular grade or course of study. Testing affects even the youngest students, with the average pre-K class giving 4.1 standardized tests, the report found. The study analyzed tests given in 66 urban districts in the 2014-2015 school year. It did not count quizzes or tests created by classroom teachers, and it did not address the amount of time schools devote to test preparation.

5. We test too much—it should be capped at two-percent of class time


It portrays a chock-a-block jumble, where tests have been layered upon tests under mandates from Congress, the U.S. Department of Education and state and local governments, many of which the study argues have questionable value to teachers and students. Testing companies that aggressively market new exams also share the blame, the study said. "Everyone is culpable here,” said Michael Casserly, executive director of the Council of the Great City Schools. "You've got multiple actors requiring, urging and encouraging a variety of tests for very different reasons that don't necessarily add up to a clear picture of how our kids are doing. The result is an assessment system that's not very intelligent and not coherent.” Ahead of the study’s release, the U.S. Department of Education offered a mea culpa of sorts, issuing a 10-page “action plan” to states and local districts that spells out ways to reduce redundant and low-quality testing. The department pledged to make money and staff available to help and promised to amend some of its policies. "At the federal, state and local level, we have all supported policies that have contributed to the problem in implementation,” Education Secretary Arne Duncan said in a statement. “We can and will work with states, districts and educators to help solve it.” The agency is recommending that states cap the amount of time devoted to test-taking to no more than 2 percent of class time. A similar proposal is part of the bill pending in the Senate to replace No Child Left Behind. Casserly cautioned against an arbitrary limit, saying he is concerned that states would indiscriminately lop off tests to meet a federal testing cap. A better approach, he said, would be a coordinated effort among all players - federal, state and local - to come up with a more thoughtful system.

6. Even many test supporters agree that we give too many standardized tests now


Standardized testing has caused intense debate on Capitol Hill as lawmakers work to craft a replacement for No Child Left Behind. Testing critics tried unsuccessfully to erase the federal requirement that schools test in math and reading. Civil rights advocates pushed back, arguing that tests are an important safeguard for struggling students because publicly reported test scores illuminate the achievement gap between historically underserved students and their more affluent peers. But even testing supporters agree about an overload. "For those of us who support annual assessments, it doesn't mean we support this craziness,” said Kati Haycock, president of the Education Trust, an advocacy group focused on reducing the achievement gap. "There's a clear problem here," Testing tends to be concentrated between February and May. The council’s study found numerous examples of redundancy, with students often taking an end-of-course test, an Advanced Placement test and a final exam for the same course. In 40 percent of districts surveyed, test results aren't available until the following school year, making them useless for teachers who want to use results to help guide their work in the classroom, Casserly said.
 Testing Undesirable: Time Tradeoff [cont’d]

7. Tests suck up a ton of time and feed monopolistic testing companies

Allison White, co-founder of Port Washington Advocates for Public Education, is part of the parent uprising against high-stakes testing on Long island, New York. According to White, "They're not teaching kids. It's not just the time for the testing. It's weeks and months they spend prepping for the tests. I don't see any educational purpose for the individual kid.” White said. "If these tests are so important and the only way to measure whatever people pushing them claim they measure," White wants to know "why don't we require them in private schools?” She accuses the federal government of using the promise of federal dollars to "bribe states to adopt the Common Core.” White is also critical of the testing companies, especially Pearson. "Essentially, they're a monopoly. They make the tests, the test prep materials, the remedial materials you need if you fail the test. If more kids fail the test, you can convince the school to buy more remedial materials.”

8. NCLB has only encouraged schools to offer more intermediate tests

The study found no correlation between the amount of testing in a district and the way its students perform on the National Assessment of Educational Progress (NAEP), a federal test given every two years that is the only consistent measure of student achievement across state lines. "We can't assess our way to academic excellence,” Carvalho, of the Miami-Dade school system, said. While public schools have been administering standardized tests for generations, the current buildup began after Congress passed No Child Left Behind in 2001 and required states to test all students in math and reading annually from third grade through eighth grade, and once in high school. States that failed to make academic progress faced a series of consequences. States and districts responded by adding new tests during the school year to ensure students were on track. ”You prepare for the test to prepare for the test to prepare for the test,” said Robert Schaeffer of the National Center for Fair and Open Testing, a nonprofit organization critical of standardized testing. And, the study found, Obama administration policies have escalated the issue. To win a grant under the competitive Race to the Top program, or to receive a waiver from No Child Left Behind, states had to evaluate teachers based in part on student test scores. Since federal law required standardized tests only in math and reading in certain grades, states added tests in social studies, science, languages - even physical education - to have scores they could use to evaluate teachers. "Many of the appalling things reported on here are the direct result of the way the federal government has approached this,” said Marc Tucker, president of the National Center on Education and the Economy. ”The accountability system is what's driving this and it's fundamentally flawed.”
Testing Undesirable: Answers to “Academic Achievement”

1. Testing data availability does not improve student achievement—longitudinal data from multiple states proves


Further research by Amrein and Berliner (2002) indicated standardized testing is not the answer to improved testing scores. After analyzing data from 18 states where high-stakes testing was implemented, in all but one case, there was no significant increase in scores. Additionally, in one case the scores actually fell. This supports what some researchers believe, standardized tests yield few benefits to student learning while neglecting higher-order thinking skills. One novice teacher ex-pressed the same concern, “I ask them to look at the answer options to the question and then ask them to look for the answer from the text book. This backward strategy is really working for me and my students also feel confident.” Neil (2003) also reported similar cases where he found that students did not remember what they had read, even though they may have responded correctly to the test item, indicating once again the unreliability of the test scores. This is an indication of a lack of basic skills needed for success after high school.

2. The overemphasis on testing undermines learning even in the subjects on which students are tested


Yet even within the tested subjects of reading and math, an overemphasis on standardized testing hurts the quality of instruction students receive. The current system of accountability, which uses the same tests to measure trends in achievement and to evaluate teachers, necessarily promotes teaching to the test, according to Derek Neal, an economics professor at the University of Chicago. As he explained in an interview with the HPR, tracking progress “requires that these tests be designed in a way that they yield comparable scores across time. The demand for comparability … will inevitably lead to predictability, and once you have predictability, then the teachers are going to be tempted to coach, and not teach.”

3. There is little evidence that current testing improves learning outcomes

Bobbi A. Solley, Professor, Department of Elementary and Secondary Education, Middle Tennessee State University, “On Standardized Testing: An ACEI Position Paper,” CHILDHOOD EDUCATION, Fall 2007, p. 34.

Given the fact that high stakes are now being attached to all standardized tests, the amount of pressure placed upon children, teachers, and administrators to perform is overwhelming. When increased pressure is placed on individuals to perform, they naturally resort to doing the things that will earn the swiftest reward-in this case, higher test scores. But what does this mean for children’s learning? Are children learning more today because of mandated tests? Although each president since Lyndon Johnson has implemented some type of education package that included standardized tests and claimed its future success in creating better schools for our nation’s children, little evidence exists that children’s learning has actually improved because of these tests. Amrein and Berliner (2003) posited that if students were showing an increase in learning based on state tests, they should show an increase in learning on other independent measures as well. Those researchers examined four student achievement measures—the SAT, the ACT, advanced placement (AP) tests, and the National Assessment of Educational Progress (NAEP)-in 18 states. What they found, in terms of a connection to learning, was virtually nothing. “Nothing seemed to be happening on these measures of student learning. In fact, we can make a strong case that high stakes testing policies hurt student learning instead of helping it” (p. 35). A study by Nichols, Glass, and Berliner (2005) also indicates a weak correlation between high-stakes testing and learning. While they found some validity to the claim that math achievement increased as pressure from high-stakes tests became more prevalent, their findings also indicated that increased testing pressure produced no gains in reading scores at the 4th- or 8th-grade level when students took the National Assessment of Educational Progress (NAEP). Although those in power would have us believe that increased testing motivates students to learn more, research indicates that the correlation is weak at best and non-existent at worst. Testing does virtually nothing to support or increase student learning.
**Testing Undesirable: Answers to “Accountability”**

- Accountability and standards claims are simply wrong—we do not need testing to promote them


*Standards aren’t the main ingredient that’s in low supply. Anyone who is serious about addressing the inequities of American education would naturally want to investigate differences in available resources. A good argument could be made that the fairest allocation strategy, which is only common sense in some countries, is to provide not merely equal amounts across schools and districts, but more for the most challenging student populations. This does happen in some states—by no means all—but, even when it does, the money is commonly offered as a short-term grant (hardly sufficient to compensate for years of inadequate funding) and is often earmarked for test preparation rather than for higher-quality teaching. Worse, high-stakes testing systems may provide more money to those already successful (for example, in the form of bonuses for good scores) and less to those whose need is greatest. Many public officials, along with like-minded journalists and other observers, are apt to minimize the matter of resources and assume that everything deficient about education for poor and minority children can be remedied by more forceful demands that we “raise the bar.” The implication here would seem to be that teachers and students could be doing a better job but have, for some reason, chosen not to do so and need only be bribed or threatened into improvement. (In fact, this is the tacit assumption behind all incentive systems.) The focus among policymakers has been on standards of outcome rather than standards of opportunity. To make matters worse, some supporters of high-stakes testing have not just ignored, but contemptuously dismissed, the relevance of barriers to achievement in certain neighborhoods. Explanations about very real obstacles such as racism, poverty, fear of crime, low teacher salaries, inadequate facilities, and language barriers are sometimes written off as mere “excuses.” This is at once naive and callous, and, like any other example of minimizing the relevance of structural constraints, ultimately serves the interests of those fortunate enough not to face them.*
Testing Undesirable: Answers to “Data Valuable”

1. We can gather data about schools without standardized tests—other countries prove


In addition, as long as standardized-test performance is used as the primary method of judging the success of schools, it will be the primary educational focus of most schools, especially ones that struggle in our most challenged communities. Alternatives to annual standardized testing of all students to measure school performance do exist. Finland, for ex ample, which has one of the highest-ranked public education systems in the world, randomly selects schools for assessment. The results are confidential, so as not to be punitive in nature. Singapore, another nation universally lauded for its educational performance, invests heavily in professional development and mentoring of novice teachers as a proactive strategy, rather than using standardized tests to shame schools and teachers. Even China, which has traditionally valued standardized tests to determine the future success of its students, is realizing that this approach is folly and is working diligently to reduce the burden of testing and instead focus on learning experiences that accentuate creativity.

2. There is plenty of data that we use to evaluate school effectiveness


Every year in the United States, important data on student and school performance is gathered. Students pursuing post-secondary education participate in college entrance exams such as the ACT and SAT. Those who choose vocational fields are assessed with the NOCTI exam once they have concluded their programs. Across the United States, graduation and attendance data are gathered and schools are randomly selected to participate in the National Assessment for Educational Progress. This allows schools to measure and compare academic progress. This assessment, given in grades 4, 8 and 12, provides valuable data on the same subjects that are tested under the federal annual-testing guidelines. With the overdue reauthorization of the Elementary and Secondary Education Act on the horizon, education in America is at a critical crossroads. Rather than continue with an iteration of the act that brought us No Child Left Behind in 2000, I hope it is reauthorized in a way that captures the essence of the Future Ready Pledge. It is time for our government officials to display courage and do what is best for children. The rest of us must make sure our voices are heard as we demand that all children receive creative and engaging learning experiences that will best prepare them for the opportunities of the future. The question remains: Are we working to be future-ready?

3. Tests are not reliable measures of educational quality—they are not good proxies for actual learning


Furthermore, arguments like Murray’s assume that standardized tests are good proxies for student learning, which oftentimes is not the case. According to Stanford University professor of education Linda Darling-Hammond, “The tests we use, particularly the state standardized tests, are extremely narrow. Evidence shows that they measure almost exclusively low-level skills.” A 2012 study from the RAND Corporation backs this claim, finding that only three to 10 percent of elementary and middle school students in the United States were administered tests that assessed deeper learning skills. And even the low-level skills that the tests do measure can be impacted by how much sleep the student got the night before the test and whether the room where the student took the test was too hot or cold.
4. Even advocates concede that high-stakes tests are not a sound basis for evaluating students or teachers


Even people and groups generally supportive of Common Core are questioning the validity of the high-stakes testing, especially its use to evaluate teacher performance. The American Educational Research Association (AERA) and the National Academy of Education released a report co-written by prominent educational researchers including former Obama educational advisor Linda Darling-Hammond of Stanford University. According to the report's executive summary, a Value-Added Model "assumes that student learning is measured well by a given test, is influenced by the teacher alone, and is independent of other aspects of the classroom context. Because these assumptions are problematic, researchers have documented problems with value-added models as measures of teachers' effectiveness.” But the report concludes that Value-Added Models for measuring Teacher Effectiveness using student performance on high-stakes standardized assessments are "highly unstable, teacher evaluations "are significantly affected by differences in the students who are assigned to them, Value-Added ratings based on student performance on high-stakes standardized tests "cannot disentangle the many influences on student progress.” Translating from education jargon into plain English, this means their research findings show that high-stakes standardized tests are not valid for evaluating teachers or students.

5. Current tests simply do not measure the skills that our students need


Several studies have indicated that student assessment should include various forms of testing, class projects, self-reflections, research assignments, demonstrations, and displays (Ezer, Gilat, & Sagee, 2010). A 5th grade teacher insisted that Content related activities would give a better picture of student understanding (Hom, 2003; Neil, 2003; Sambar, 2001). An article by Anthony Rebora (2012) discusses how many teachers agree that standardized testing does not portray how much students know. He states: “Most teachers do not believe standardized tests have significant value as measures of student performance” (p. 14). Rebora further explains what teachers believe will accurately assess student knowledge: “teachers see ongoing formative assessments, class participation, and performance on class assignments as much more important measures of student learning” (p. 14).

6. Test scores are increasingly unreliable because of opt-outs


In April, hundreds of thousands of public school students from grades three through eight in New York take high-stakes tests in English Language Arts and Mathematics. Jeanette Deutermann’s son is not one of them. Deutermann, a Long Island parent and critic of New York’s test-based accountability system, told the HPR in an interview that she witnessed her school’s relentless focus on standardized testing transform her son, who is now in sixth grade, from someone who never complained about going to school to someone who intensely dreaded it. “He would be doing homework, and he would be sobbing, he’d be trying to wipe tears away while he’s trying to finish his homework so he could see the paper. He would constantly talk about how stupid he was, how ‘I can’t wait until I can drop out of school.’” Deutermann knew her son was not the best test-taker, but she also knew that his teachers were excellent, and that their evaluations depended on his test scores. So, for the past two years, Deutermann has opted her son out of the state tests in ELA and math. She explained, “This was not in opposition to the school, but in support.” She created a Facebook group to provide information to Long Island parents about opting their children out of tests, which now has over 17,000 members. She also became part of the steering committee for the advocacy group New York State Allies for Public Education, which opposes excessive standardized testing. This year, tens of thousands of parents across the country joined Deutermann in refusing to let their children take the tests, with over 200,000 in New York alone. The movement is gaining momentum in other states, too, with hundreds of students opting out of tests in New Mexico, 3,000 students in Florida, and over 40,000 in New Jersey. But the biased test outcomes resulting from high numbers of student opt-outs are just one way that standardized tests may provide an inaccurate measure for evaluating teachers, and they are by far the least insidious.
Testing Undesirable: Answers to “Data Valuable [cont’d]”

7. Test scores are not useful in measuring education outcomes—multiple reasons


5. Just because doing x raises standardized test scores doesn’t mean x should be done At the very least, we would need evidence that the test in question is a source of useful information about whether our teaching and learning goals are being met. Many educators have argued that the tests being used in our schools are unsatisfactory for several reasons. First, there are numerous limitations with specific tests. Second, most tests share certain problematic features, such as being timed (which places more of a premium on speed than on thoughtfulness), norm-referenced (which means the tests are designed to tell us who’s beating whom, not how well students have learned or teachers have taught), and consisting largely of multiple-choice questions (which don’t permit students to generate or even explain their answers). The third reason is the problems inherent to all tests that are standardized and created by people far away from the classroom — as opposed to assessing the actual learning taking place there on an on-going basis. This is not the place to explain in detail why standardized tests measure what matters least. Here, I want only to make the simplier — and, once again, I think, indisputable — point that anyone who regards high or rising test scores as good news has an obligation to show that the tests themselves are good. If a test result can’t be convincingly shown to be both valid and meaningful, then whatever we did to achieve that result — say, a new curriculum or instructional strategy — may well have no merit whatsoever. It may even prove to be destructive when assessed by better criteria. Indeed, a school or district might be getting worse even as its test scores rise. So how is it that articles in newspapers and education journals, as well as pronouncements by public officials and think tanks, seem to accept on faith that better scores on any test necessarily constitute good news, and that whatever produced those scores can be described as “effective”? Parents should be encouraged to ask, “How much time was sacrificed from real learning just so our kids could get better at taking the [name of test]?”

8. Tests are not necessary—teachers can readily identify which students are succeeding and which need additional help


There are two main arguments against using standardized tests to guarantee that students reach at least a basic level of academic competency. The first is radical: These tests are not necessary. The second—less radical and more familiar—is that, even if standardized testing were an efficient benchmark of basic skills, the costs associated with it are too high. Standardized tests are unnecessary because they rarely show what we don’t already know. Ask any teacher and she can tell you which students can read and write. That telling usually comes in the form of letter grades or evaluations that break down progress on skills. So trust the teacher. Publish grade distributions. Locally publish a compilation of evaluation reports. Release a state or national report reviewed and verified by expert evaluators with legislative oversight.
Testing Undesirable: Answers to “Information Retention / Memory”

1. Memorization-focused education does not help—most information is quickly forgotten


1. Much of the material students are required to memorize is soon forgotten The truth of this statement will be conceded (either willingly or reluctantly) by just about everyone who has spent time in school — in other words, all of us. A few months, or sometimes even just a few days, after having committed a list of facts, dates, or definitions to memory, we couldn’t recall most of them if our lives depended on it. Everyone knows this, yet a substantial part of schooling -- particularly in the most traditional schools -- continues to consist of stuffing facts into students’ short-term memories. The more closely we inspect this model of teaching and testing, the more problematic it reveals itself to be. First, there’s the question of what students are made to learn, which often is more oriented to factual material than to a deep understanding of ideas. (See item 2, below.) Second, there’s the question of how students are taught, with a focus on passive absorption: listening to lectures, reading summaries in textbooks, and rehearsing material immediately before being required to cough it back up. Third, there’s the question of why a student has learned something: Knowledge is less likely to be retained if it has been acquired so that one will perform well on a test, as opposed to learning in the context of pursuing projects and solving problems that are personally meaningful. Even without these layers of deficiencies with the status quo, and even if we grant that remembering some things can be useful, the fundamental question echoes like a shout down an endless school corridor: Why are kids still being forced to memorize so much stuff that we know they won’t remember? Corollary 1A: Since this appears to be true for adults, too, why do most professional development events for teachers resemble the least impressive classrooms, with experts disgorging facts about how to educate?

2. Memorizing information does not improve the students’ intelligence


2. Just knowing a lot of facts doesn’t mean you’re smart Even students who do manage to remember some of the material they were taught are not necessarily able to make sense of those bits of knowledge, to understand connections among them, or to apply them in inventive and persuasive ways to real-life problems. In fact, the cognitive scientist Lauren Resnick goes even further: It’s not just that knowing (or having been taught) facts doesn’t in itself make you smart. A mostly fact-oriented education may actually interfere with your becoming smart. “Thinking skills tend to be driven out of the curriculum by ever-growing demands for teaching larger and larger bodies of knowledge,” she writes. Yet schools continue to treat students as empty glasses into which information can be poured — and public officials continue to judge schools on the basis of how efficiently and determinedly they pour.
Testing Undesirable: Answers to “Modeling / Spillover”

- Modeling claims are wrong—high performing schools do so for reasons outside of instructional quality


That leads to the second argument. Even if standardized testing were not only desirable to give the public a picture of basic competencies, but also an efficient way to do so, the costs have been too great. Many have previously made cogent arguments (unrealistic definitions of achievement, skewed instructional schemes, inequitable curricular offerings, inevitable corruption, perverted charter school missions, alienation, disempowerment, and embarrassment of educators, etc.) in this vein, but let's think about a supposed example of success on this front—a school with the high test scores. In general, such a school has a compliant or affluent population. Test scores are a point of pride. The school has a good reputation. But, when you go in and observe, the teaching and learning do not impress. Never once have I looked at the test scores of this kind of school and thought, "How could I be more like them?" That's because success represented just a score on a narrow test of a limited band of achievement (a test, by the way, with content that I was not even legally allowed to talk about), and I couldn't see how looking at that score could help me in my day-to-day teaching. Even worse, I don't think the teachers at such schools have learned much from their good scores. If anything, the scores have prevented them from becoming better.
1. **Score increases only occur because students learn how to take the test**


   In 1991, I worked with several colleagues on a validity study to investigate more specifically whether increases in test scores reflected real improvements in student achievement. In a large urban school system in a state with high-stakes accountability, random subsamples of students were given independent tests to see whether they could perform as well as they had on the familiar standardized test. The alternative, independent tests included a parallel form of the commercial standardized test used for high-stakes purposes, a different standardized test that had been used by the district in the past, and a new test that had been constructed objective-by-objective to match the content of the high-stakes test but using different formats for the questions. In addition to content matching, the new test was statistically equated to the high-stakes standardized test, using students in Colorado where both tests were equally unfamiliar. When student scores on independent tests were compared to results on the high-stakes accountability test, there was an 8-month drop in mathematics on the alternative standardized test and a 7-month drop on the specially constructed test. In reading, there was a 3-month drop on both the alternative standardized test and the specially constructed test. Our conclusion was that “performance on a conventional high-stakes test does not generalize well to other tests for which students have not been specifically prepared.”

2. **The emphasis on tests has nothing to do with actual learning**

   Marcia Gentry, Associate Professor and Associate Director, Gifted Education Resource Institute, “No Child Left Behind: Neglecting Excellence,” ROEPER REVIEW v. 29 n. 1, 2006, pp. 24-27.

   Recently, I was asked to participate in a panel discussion concerning the No Child Left Behind act and its effects on gifted child education, an invitation that came with a charge to write a short, scholarly article on the topic. Without giving it too much thought, my response to the invitation was, “Sure, sounds like fun!” However, upon further reflection I realized that while short was no problem as there is little relevance for gifted children in NCLB, scholarly and No Child Left Behind created an oxymoronic conundrum—as clearly, NCLB is antithetical to scholarship. NCLB is a politically charged, top-down, hostile take-over of America's schools that has, in effect, ignored progress of individual children in favor of closing gaps and emphasizing perceived proficiency scores for schools and groups of children using questionable standards and measures of achievement. Little exists in the act to encourage schools, as they are held accountable to a throng of unfunded requirements, to develop individual differences, creative thinking, innovation, or individual potentials, some of the very things in our public education system that, in the past, have helped to make ours a great nation. Instead, states are being forced to create high-stakes tests to which educators must teach and on which groups of students must show “Adequate Yearly Progress” or face the demoralization of being labeled a failing school, and all the sanctions that accompany such a designation. And none of this has much to do with how much actual learning, quality instruction, and individual student progress occurs in the schools (Mayer, Mullins, & Moore, 2000).

3. **Tests force an emphasis on remedial drilling that does nothing to help students actually learn**

   Marcia Gentry, Associate Professor and Associate Director, Gifted Education Resource Institute, “No Child Left Behind: Neglecting Excellence,” ROEPER REVIEW v. 29 n. 1, 2006, pp. 24-27.

   Troubling for many reasons, this focus has made education punitive for students and teachers alike. Students are being forced into drill and kill type learning as well as being expected to attain “grade level” skills at the same rate as their higher-achieving peers. This approach is counterintuitive, as children learn best when they have elements of interest, challenge, choice, and enjoyment in their learning experiences—elements lacking in remedial based approaches. Tragically, if the remedial drills do actually raise test scores, it indicates the test measures low-level outcomes (Popham, 2001). As Tomlinson (2002) clearly articulated, proficiency is not enough, yet with NCLB proficiency is the goal and the focus, leaving the many students who can exceed proficiency in educational deprival. Already, many individuals have pointed out that the language of proficiency without language of excellence in NCLB puts at further risk the very students it purports to help—poor and minority students (e.g., Amrein & Berliner, 2002; Gallagher, 2004; Golden, 2004; Elmore, 2003; Neill, 2003; Tomlinson, 2002).
Testing Undesirable: Answers to “Score Increases” [cont’d]

4. Test scores are easily manipulated by schools—multiple mechanisms


There are myriad ways that test scores can be manipulated to make a student or a school appear to be doing better than they actually are. For example, states have lowered the scores students need to pass, according to a 2009 report published in the International Journal of Education Policy and Leadership. Thus, while it is true that the number of students scoring “proficient” on state standardized tests has risen, real student achievement has not necessarily improved. Low-stakes, diagnostic tests, which are not subject to this type of manipulation because they are not attached to rewards or punishments for teachers or schools, confirm this finding. Math and reading scores on the National Assessment of Educational Progress have improved incrementally at best, and have actually declined on the Programme for International Student Assessment. Schools and administrators have also pressured low-performing students to drop out or enter special education programs in order to raise overall test scores. According to a 2010 report from the civil rights organization Advancement Project, “the practice of pushing struggling students out of school to boost test scores has become quite common.” And a 2002 study from researchers at Arizona State University found a correlation between high-stakes testing and “higher numbers of low performing students being suspended before testing days, expelled from school before tests, or being reclassified as exempt from testing because they are determined to be either Special Education or Limited English Proficient (LEP).”
1. Meaningfully transforming our schools means getting rid of the tests


It all spirals out from here. Absurd standards build on absurd standards. Ridiculous data entry tasks build on ridiculous data entry tasks. The demands placed on teachers and students betray a skepticism that they can simply do their work. Even worse, these impositions are supposed to be fair and objective since everyone takes the same test, leading to the mistaken idea that if a student is failing, a lack of study skills or “grit” are to blame. Positive transformation of schooling will require doing away with all of that. It will take a radical departure from the standards and accountability regime — a break that Duncan, despite his recent comments, is not likely to entertain. But we don’t need his permission. What we need is a united movement to articulate an alternative to the global project of corporate education reform.

2. Overselling accountability measures undermines long-term support for needed educational reforms


Advocates drive good ideas to extremes when they oversell their promise and undermine their integrity. Unfortunately, this pattern is all too common. Problem One: Measures that are overly ambitious or poorly designed risk undermining popular support for sound and necessary reforms. No Child Left Behind (NCLB) took near-universal backing for tenets of accountability and deployed them in an overwritten federal statute that poisoned the NCLB brand. Indeed, EdNext polling in 2007 showed that describing the key precepts of NCLB without using its name drew 71 percent support, but the addition of the phrase “No Child Left Behind” reduced that figure by 14 points.

3. Overpromising reform effects stifles innovation and creativity


Problem Two: Overpromising. When they insist that ideas like school choice, performance pay, and teacher evaluations based on value-added measures will themselves boost student achievement, would-be reformers stifle creativity, encourage their allies to lock elbows and march forward rather than engage in useful debate and reflection, turn every reform proposal into an us-against-them steel-cage match, and push researchers into the awkward position of studying whether reforms “work” rather than when, why, and how they make it easier to improve schooling. Consider performance pay. Just recently a three-year randomized evaluation of a Tennessee merit-pay experiment funded by the federal government’s Teacher Incentive Fund found that bonuses tied to test scores didn’t lead to higher performance in middle-school math. “Study Casts Cold Water on Bonus Pay,” read Education Week’s headline, and the news was widely interpreted as a setback for attempts to link teacher compensation to classroom performance. Yet the most compelling rationale for merit pay is not any short-term bump in test scores, but rather its potential for making the profession more attractive to talented candidates, more amenable to specialization, more rewarding for accomplished professionals, and a better fit for the 21st-century labor market. Whether or not bonuses linked to test scores had any effect on measured achievement in the short run says absolutely nothing on this score. Yet, the lust for simple answers and for research that “proves” those answers right has led many would-be reformers to adopt and defend half-baked versions of pay reform. The primary goal of reform efforts should be to make it easier for problem solvers to gain access to and traction in the system, coupled with thoughtful public oversight of results. The impatient rush to “fix” teacher quality in one furious burst of legislating may instead lead to a situation in which promising efforts to uproot outdated and stifling arrangements become enveloped in crudely drawn and potentially destructive mandates. Rushing forward with statewide mandates to incorporate value-added assessments into teacher evaluation systems, for example, may wind up stifling innovation. Systems built around individual value-added calculations can stymie the smart use of personnel that reformers should encourage. Principals who rotate their faculty by strength during the year, or augment classroom teachers with online lessons, will find their staffing models a poor fit for evaluation systems predicated on linking each student’s annual test scores to a single teacher.
4. We can improve school quality without relying on standardized tests


Given all of these problems with standardized testing, it seems that the civil rights issue is too much testing, not too little. Instead of forcing low-income schools to spend millions of dollars and countless hours of class time preparing for and administering standardized tests that only serve to prove, oftentimes inaccurately, what we already know about the achievement gap, we should use those resources to expand programs in the arts and humanities, to provide incentive pay to attract teachers to areas where they are needed most, and to decrease class sizes, all things that could actually make a difference for disadvantaged students. This is not to say that America’s accountability system should be completely dismantled. Politicians and schools can de-emphasize testing while still ensuring high achievement. Student and teacher evaluations can take multiple measures of performance into account. The amount of standardized tests students have to take can be drastically reduced. The fewer standardized tests that students do take can incorporate more open-ended questions that force students to think critically and outside the box. Thirteen years after NCLB’s mandates were first set into place, the rhetoric used by politicians and pundits is sounding more and more like that which the same politicians and pundits used to endorse NCLB. Congress would be ill advised to try to use high-stakes test-based accountability to narrow the achievement gap and expect a different result than the aftermath of the 2002 law. It is time to acknowledge that putting an enormous amount of weight on standardized test scores does not work, and to move on to other solutions. Regardless of the outcome of the current debate, grassroots activists like Deutermann will continue to fight against harmful test-based accountability systems like New York’s. “This is an epidemic,” she said. “It’s happening everywhere, with all sorts of kids, from the smartest kids to the kids that struggle the most, from Republicans to Democrats, from kids in low-income districts to kids in high-performing districts. It doesn’t matter where you are, the stories are exactly the same.” “We may be passive when it comes to all the other things [corporate reformers] have interjected themselves into,” Deutermann warned, “but when you mess with our kids, that’s when the claws come out.”