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

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>1</td>
</tr>
<tr>
<td>Definitions</td>
<td>9</td>
</tr>
<tr>
<td>Standardized testing defined. LZ</td>
<td>9</td>
</tr>
<tr>
<td>No Child Left Behind Act. ABH</td>
<td>9</td>
</tr>
<tr>
<td>Topic Analysis One</td>
<td>11</td>
</tr>
<tr>
<td>Pro Evidence</td>
<td>20</td>
</tr>
<tr>
<td>Testing Allows Teachers to Help Students</td>
<td>21</td>
</tr>
<tr>
<td>Standardized testing identifies areas where students need improvement. ABH</td>
<td>21</td>
</tr>
<tr>
<td>Standardized testing helps teachers assess knowledge. LZ</td>
<td>21</td>
</tr>
<tr>
<td>Non-test based solutions just aren’t accurate at measuring knowledge levels. LZ</td>
<td>23</td>
</tr>
<tr>
<td>Teachers Support Standardized Testing. ABH</td>
<td>24</td>
</tr>
<tr>
<td>Testing Good</td>
<td>25</td>
</tr>
<tr>
<td>Standardized testing motivates students to study. LZ</td>
<td>25</td>
</tr>
<tr>
<td>Testing doesn’t produce rote learning and feedback prevents this problem. LZ</td>
<td>25</td>
</tr>
<tr>
<td>Testing doesn’t necessarily trade off with other classroom activities. LZ</td>
<td>26</td>
</tr>
<tr>
<td>Standardized testing enhances retrieval processes. LZ</td>
<td>27</td>
</tr>
<tr>
<td>Studies prove – lots of testing increases memory skills. LZ</td>
<td>28</td>
</tr>
<tr>
<td>Testing helps identify gaps in knowledge. LZ</td>
<td>29</td>
</tr>
<tr>
<td>Testing facilitates further learning. LZ</td>
<td>29</td>
</tr>
<tr>
<td>Testing enhances learning during further study sessions. LZ</td>
<td>30</td>
</tr>
<tr>
<td>Test improves organization of knowledge. LZ</td>
<td>31</td>
</tr>
<tr>
<td>Testing doesn’t just shoehorn knowledge – it facilitates application to new situations. LZ</td>
<td>32</td>
</tr>
<tr>
<td>Evidence suggests testing is even better than restudying. LZ</td>
<td>32</td>
</tr>
<tr>
<td>Testing improves retention even on non-tested material. LZ</td>
<td>33</td>
</tr>
<tr>
<td>Testing improves metacognitive accuracy. LZ</td>
<td>34</td>
</tr>
<tr>
<td>Students do not self test themselves, so testing is a valuable classroom tool. LZ</td>
<td>35</td>
</tr>
</tbody>
</table>
### Testing prevents proactive interference. LZ ................................................................. 36
### Testing Closes Achievement Gaps ........................................................................... 37
  - Testing key to identify racial achievement gaps – first step to fixing them. ELT ......... 37
  - Testing ensures students with disabilities are not left behind. ELT ............................. 38
### Federal Standards Do Not Detract From Education .................................................... 39
  - Most student testing time is spent on district-level exams. ...................................... 39
  - Federally-overseen standardized testing is a crucial enabler of parental choice. DAT ...... 40
### State-level Standardization Improves Education .......................................................... 41
  - States with well-implemented holistic standards boost student achievement. DAT .......... 41
  - While federal control is unnecessary, states are adept at comprehensive testing. DAT .... 42
### Testing As Part of A Larger Educational Framework ................................................. 43
  - Teachers find standardized test information constructive .................................... 43
  - Testing provides sorely-needed data. DAT .................................................................... 44
  - Standardized test scores make it possible to predict student life outcomes. DAT .......... 45
  - Standardized tests are a piece of modern evaluation systems. DAT .......................... 46
### Testing Data is Key To School Choice ...................................................................... 47
  - Annual testing is key to prevent bad accountability systems from being implemented. ELT ...... 47
  - A majority of people support standardized testing. ELT ........................................... 47
### Con Evidence ............................................................................................................. 48
  - Fixation on standardized testing detracts from real teaching and learning. LZ .............. 49
  - Standardized testing is toxic to a culture of learning. LZ .......................................... 50
  - A school with high test scores loses the incentive to improve. LZ .............................. 50
  - There are a laundry list of reasons why standardized testing is bad. LZ ....................... 51
  - Standardized testing hurts teacher creativity. LZ ....................................................... 52
  - Artificially ranking people with tests hurts younger children. LZ ............................... 52
  - Standardized testing hurts the creative potential of education – it turns us into machines that only maximize wealth. LZ ................................................................. 53
  - Even worse, standardized testing actually makes kids less advanced in learning. LZ ........ 54
  - Extensive studies prove – standardized testing impairs children. LZ .......................... 54
Measuring the wrong thing isn’t benign – it results in problems like gaming the system and high stress among others. LZ .......................................................... 55

We should measure things that aren’t as profitable but more beneficial like creativity. LZ .................. 55

Standardized testing takes time away from other subjects. ABH ...................................................... 56

Testing Prevents Students From Learning Life Skills. ABH ............................................................... 57

Standardized tests undermine students’ drive. DAT ........................................................................ 58

Teachers focus less on social studies and science. ABH ................................................................. 59

Testing decreases students’ motivation to learn. ABH ................................................................. 60

Effective assessment requires room for self-assessment. ABH .................................................... 61

Testing cuts out science from curriculums. ELT ......................................................................... 62

Standardized Testing decreases success in college. ELT ............................................................ 63

Standardized testing incentivizes bad education reform. ELT ...................................................... 63

Ensures we lose the best teachers – good models support teacher autonomy. ELT .......................... 64

Doesn’t Measure ............................................................................................................................. 65

Common core standards don’t actually measure learning. LZ ......................................................... 65

Because of the limitations of standardized testing, it makes a poor measurement of actual learning. LZ 66

Trying to measure quality of education with standardized testing is like measuring temperature with a tablespoon. LZ ....................................................................................... 68

Because large corporations own standardized tests, there is no real standardization among different educational objectives. LZ .................................................................................. 68

One size fits all just doesn’t work – it can’t align with what is being emphasized in a particular classroom. LZ ........................................................................................................ 69

Robust studies prove – tests do not cover what is actually taught. LZ ............................................ 69

Mismatches are very rarely recognized, but they definitely exist. LZ ............................................... 69

Standardized tests cannot allow for meaningful comparisons among students because of limited collections of comparisons. LZ ............................................................................ 70

Standardized tests suffer from confounding factors. LZ .................................................................. 72

What should educators do instead? They should learn more about standardized tests, carry out an educational campaign against standardized tests, and arrange better assessments. LZ .............................................. 73

Standardized testing just can’t measure learning – especially true when public education is a market driven activity. LZ ......................................................................................... 74
There are systemic errors within standardized testing. ABH ................................................................. 75

Standardized Testing errors have severe negative consequences. ABH................................................................. 76

Testing has negligible effect on student achievement – comparisons with other countries fail. ELT ..... 77

Teaching to the test makes results useless. ELT................................................................. 78

Standardized tests measure relative achievement but not actual learning. ELT................................................................. 79

Standardized testing decreases the quality of education over the year. ELT................................................................. 80

Unfair ............................................................................................................................................. 81

Standardized testing is racially unfair. LZ................................................................................................. 81

Standardized testing creates self-fulfilling prophecies that disadvantage racial minorities. LZ ............ 81

Research clearly demonstrates that tests unfairly advantage certain racial groups – only ending reliance on standardized testing solves. LZ ........................................................................................................... 82

The standardized testing gap is huge between the poor and the wealthy. LZ ........................................... 82

The history of standardized testing is racially biased. LZ ................................................................. 83

Colleges are already acknowledging the harms of testing, but K-12 still uses them and they are pervasive in the system. LZ................................................................................................................................. 83

Standardized testing is definitely racially biased. LZ................................................................................................. 83

The schools affected by standardized testing are mostly minority schools. LZ ............................................... 84

Standardized testing prevents education equity. LZ ......................................................................................... 85

Standardized tests are designed for white, middle class kids. LZ ................................................................. 86

Factors outside of the classroom impact test scores. ABH ................................................................................................. 87

Testing disproportionately targets urban schools. DAT......................................................................................... 88

Health greatly impacts test scores. ABH ................................................................................................................................. 89

Students’ socioeconomic status influences test scores. ABH ................................................................................................. 90

Standardized tests mean students with disabilities miss out on life skills. ELT ........................................... 90

Testing fails to capture achievements of students with learning disabilities. ELT ........................................... 91

The way tests are written guarantees students with disabilities will fail. ELT................................................................. 91

Standardized Testing Hurts Teachers ........................................................................................................... 92

Teachers lose control of the curriculum. ABH ................................................................................................................................. 92

Testing robs teachers of resources. DAT ................................................................................................................................. 93

Constantly shifting standards makes it impossible for teachers to provide resources. DAT................................................................. 94
Teachers lose control of both content and scope. DAT .......................................................... 95
Standardized test data is too inaccurate to be a reliable teacher performance gauge. DAT .......... 96
Standardized tests disproportionately harm teachers of low-income students. DAT .................... 97
The use of standardized tests to evaluate teachers has perverse impacts on needy districts. DAT .... 98
Test-based teacher evaluations pervert teachers’ incentives. DAT ........................................... 99
Teachers feel pressured by the media. ABH ............................................................................. 100

The Opportunity Costs of Standardized Testing ...................................................................... 101
Better data can be collected “on the fly”. DAT ........................................................................ 101
Standardized testing creates time-wasting redundancies. DAT ............................................... 102
Non-standardized testing provides a better assessment of future performance. DAT ................. 103
The time costs of standardized testing. DAT .......................................................................... 104
Comprehensive information already exists. DAT ..................................................................... 105
National standardization strikes down more effective state initiatives. DAT ........................... 106
Visualized: time spent on standardized assessments. DAT ..................................................... 107

Financial Costs of Testing ........................................................................................................ 109
Spending per student varies wildly across the nation, and is cumulatively massive. DAT ............ 109
Funds are essentially flowing from teacher salaries to private testing corporations. DAT .......... 110

Big Data and the Obsolescence of Standardized Testing ........................................................... 112
Case study: adaptive programs simultaneously teach and track students. DAT ....................... 112
Adaptive software can focus on improving low-scoring students’ performance. DAT ............... 113

Testing Creates Cycles of Failure ............................................................................................ 114
Students don’t get the feedback they need to succeed. DAT .................................................. 114
Standardized testing does not increase student performance. DAT ........................................ 115

Testing Hurts Early Learning ................................................................................................... 116
Standardized testing in early elementary grades has increased. ABH ..................................... 116
NCLB increased pressure on kids in kindergarten. ABH ......................................................... 116
Students get less exposure to art and science. ABH ............................................................... 117
Introducing tests too early damages children’s perception of school. ABH .............................. 117
Children need to build up self-confidence at a young age. ABH ............................................. 118
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
<td>Rigid standardized testing does not accurately measure the ability of young students. ABH</td>
</tr>
<tr>
<td>120</td>
<td>Tests Cause a Harmful Level of Stress</td>
</tr>
<tr>
<td>120</td>
<td>Students have harmful physical reactions to stress. ABH</td>
</tr>
<tr>
<td>121</td>
<td>Failure on standardized tests has severe emotional consequences. ABH</td>
</tr>
<tr>
<td>122</td>
<td>Stress from testing has long term consequences. ABH</td>
</tr>
<tr>
<td>123</td>
<td>Pro Counters</td>
</tr>
<tr>
<td>124</td>
<td>Popular Complaints Are Historically Unfounded</td>
</tr>
<tr>
<td>124</td>
<td>Studies suggest that the positive impacts of testing outweigh the commonly discussed harms. ABH</td>
</tr>
<tr>
<td>125</td>
<td>Taking tests does not suck up disproportionate amounts of time. DAT</td>
</tr>
<tr>
<td>126</td>
<td>Testing takes up only enough time to produce meaningful results. DAT</td>
</tr>
<tr>
<td>127</td>
<td>The Financial Cost of Testing is Negligible</td>
</tr>
<tr>
<td>127</td>
<td>The minimal monetary cost of standardized testing is worth the results. ABH</td>
</tr>
<tr>
<td>127</td>
<td>Standardized tests don’t need to be pervasive or costly. DAT</td>
</tr>
<tr>
<td>128</td>
<td>Cost savings of eliminating standardized tests are negligible. DAT</td>
</tr>
<tr>
<td>129</td>
<td>Districts can minimize costs even further. DAT</td>
</tr>
<tr>
<td>130</td>
<td>Gains in teacher quality outweigh costs. ELT</td>
</tr>
<tr>
<td>131</td>
<td>Teachers’ Performance is Measured Accurately</td>
</tr>
<tr>
<td>131</td>
<td>Standardized testing is a surprisingly good indicator of teacher performance. DAT</td>
</tr>
<tr>
<td>132</td>
<td>Value-added testing is the best way to evaluate teachers when combined with other measures. ELT</td>
</tr>
<tr>
<td>133</td>
<td>Testing Produces Crucial Data</td>
</tr>
<tr>
<td>133</td>
<td>Accounting for all other variables, test scores predict students’ future incomes. DAT</td>
</tr>
<tr>
<td>134</td>
<td>Standardized testing must be thorough and annual to produce useful results. DAT</td>
</tr>
<tr>
<td>135</td>
<td>Opting Out Bad</td>
</tr>
<tr>
<td>135</td>
<td>Opting out affects everyone. ELT</td>
</tr>
<tr>
<td>136</td>
<td>AT: Test too frequently</td>
</tr>
<tr>
<td>136</td>
<td>Frequent testing is key to good data. ELT</td>
</tr>
<tr>
<td>136</td>
<td>Frequent testing solves teaching to the test. ELT</td>
</tr>
<tr>
<td>139</td>
<td>AT: Teaching to the Test</td>
</tr>
<tr>
<td>139</td>
<td>Test scores are higher when teachers do not focus on test prep. ELT</td>
</tr>
</tbody>
</table>

Note: The above table represents the table of contents from the document.
December 2015

Con Counters ......................................................................................................................... 140

AT: Unaccountable .................................................................................................................... 141

The system is already gamed and a simple report would give the same information as a standardized test. LZ ................................................................................................................................. 141

Grades differ anyways and publishing grades gives the same level of accountability. LZ .............. 141

AT: Closes Achievement Gaps ................................................................................................... 142

Curriculum narrowing from tests hurts disadvantaged students most. ELT .................................. 143

Accountability Doesn’t Mandate Standardization ......................................................................... 144

Scotland maintains high levels of accountability and consistency without testing. DAT .......... 144

Decentralization will actually increase accountability. DAT ....................................................... 145

Standardization Gives Publishers Monopolies ............................................................................. 146

Publishers make billions selling standardized test material that poor districts can’t afford. DAT .... 146

Poor students and districts essentially can’t afford the solutions to standardized tests. DAT ........ 146

Test Data Is Not Informative ....................................................................................................... 147

Value-added modeling of test data is wildly inaccurate. DAT ....................................................... 147

There are too many distortions for test scores to accurately gauge teacher performance. DAT .... 148

Test data merely mirrors socioeconomic status. DAT ................................................................. 149

Too many factors affect test performance for the data to be reliable. DAT ................................. 150

Some issues, like systemic inequalities, are obvious without test data. DAT ............................... 151

Tests measure only the least important learning goals. ELT ....................................................... 152

Successful Alternatives Exist ...................................................................................................... 153

Finland doesn’t use standardized tests. ABH ............................................................................... 153

Moving away from standardized testing was effective in Israel. ABH ........................................ 154

Teacher’s evaluations are more accurate than test scores. ABH ................................................ 155

Students could be assessed based on portfolios. ABH ............................................................... 156

Performance based assessments empirically more successful. ELT ........................................... 157

Standardized testing decreases graduation rates – incentives better. ELT ................................. 158

Cases ........................................................................................................................................... 159

Pro Case ....................................................................................................................................... 160

Introduction: .................................................................................................................................. 160
Contention One: Testing Improves Teaching ................................................................. 160
Contention Two: Standardized Testing Gives Students Control ........................................ 161
Contention Three: Testing Improves Retention............................................................... 161
Con Case .......................................................................................................................... 163
Introduction: ...................................................................................................................... 163
Contention One: Standardized testing doesn’t measure anything ........................................ 163
Contention Two: Standardized testing harms student outcomes ........................................ 163
Contention Three: Standardized testing harms teaching .................................................... 164
Definitions

*Standardized testing defined. LZ*


A standardized test is any examination that's administered and scored in a predetermined, standard manner. There are two major kinds of standardized tests: aptitude tests and achievement tests. Standardized aptitude tests predict how well students are likely to perform in some subsequent educational setting. The most common examples are the SAT-I and the ACT both of which attempt to forecast how well high school students will perform in college. But standardized achievement-test scores are what citizens and school board members rely on when they evaluate a school's effectiveness. Nationally, five such tests are in use: California Achievement Tests, Comprehensive Tests of Basic Skills, Iowa Tests of Basic Skills, Metropolitan Achievement Tests, and Stanford Achievement Tests.

*No Child Left Behind Act. ABH*


The No Child Left Behind Act of 2001 (NCLB) was a piece of legislation with a clear vision. It imagined a world where every student in the United States—particularly those belonging to historically disadvantaged groups—met basic standards of literacy and numeracy. It imagined a path to that world that did not require spending significantly more money, but rather focusing on directing resources more efficiently toward this basic goal. *In a foreshadowing of the era of “big data,”* it introduced a national mandate for standardized testing, imagining test scores as the key to understanding and addressing the shortcomings of the American education system.

Many teams will probably discuss NCLB as a major example of standardized testing, so it would be very useful to be familiar with the act.
Growing up, no one likes taking tests. Having spent dozens of hours of my own childhood bubbling in exam sheets, I had something of a visceral reaction to seeing this resolution. That said, it’s a surprisingly versatile one.

Oftentimes, personal perspectives are implicit when creating cases. For many topics, especially foreign policy ones, debaters’ (Western) perspectives often work their way into whatever advocacy they are pursuing. It does not help that, typically, resolutions specify U.S. interests. This one is, in a way, no different.

As a student, however, you are party to this resolution. And I have yet to find someone who, as a student, genuinely enjoyed standardized testing. Inherent to Public Form debate as a requirement for impartiality. You can’t win a tournament without having given your all to preparing both sides of this case. For this resolution especially, any apathy needs to be replaced with empathy. This means stepping out of your own shoes and into those of both all kinds of students—high or low income, minorities, low achievers, high achievers, etc—and everyone else with a stake in the American educational system. That is to say, everyone. There does not seem to be a way to get the most out of the resolution without this kind of holistic approach. Keep this in mind as we take this resolution apart.

Context and Impacts on Strategy

It’s tempting to think of standardized testing as somehow being birthed from No Child Left Behind. In some way, however, standardized testing has been a hallmark of the American public educational system, at some point in the K-12 spectrum, since nearly its inception.

Take, for instance, the SAT and ACT, both of which act as nearly mandatory hurdles on the road to college attendance. The SAT and ACT, as administered in their modern form, are relics of the 1940s. Today, the ACT and SAT are both essentially non-state entities.

Standardization, as implemented at the state and federal levels, also has a few generations of experience under its belt. States began large-scale testing in the 1970s, while the federal government dipped a toe into the water with the National Assessment of Education Progress, or NAEP, shortly thereafter.

America’s 21st century test apparatus owes its existence to No Child Left Behind—NCLB. Per a federal mandate, states are required to conduct annual standardized testing for reading and math from third through eighth grade. Additionally, testing must occur one more time between grades 10 and 12, inclusive. Science tests are also mandated, although these are administered more periodically.
Before going beyond NCLS, it is necessary to clear up a few misconceptions. The biggest one is that standardized testing suddenly became a pervasive phenomenon, cold turkey, as a result of federal legislation. In fact, many states already had independent testing protocols, albeit not as far-reaching one. NCLB was an expansion, not a beginning. In response to the new federal legislation, 23 states expanded the assessments they were already doing to meet federal criteria. Essentially, NCLB primarily introduced greater amounts of standardization, not testing.

So where did this resolution come from? A solid guess would be Common Core. The Obama presidency’s hallmark piece of education legislation has also spurred renewed interest in the “opt-out” movement—parents and educators removing their children from the claws of standardized testing. The idea that kids are tested too much, too frequently has become a pervasive one in American educational thinking. This resolution is a reasonable extrapolation of the current state of American standardized testing.

Just like NCLB, Common Core brings with it increased standardization, rather than increased testing per se. In response to the new legislation, two new consortia were formed: the Smarter Balanced Assessment Consortium and the Partnership for Assessment of Readiness for College and Careers. You’ve probably heard of the last one at some point—it’s PARCC. These two consortia are composed of member states and have been developing tests for multi-state use under Common Core.

Keep in mind that Common Core, in its current implementation, is not just another layer of the standardized testing cake sitting atop No Child Left Behind. In fact, the intent of the new testing regiment is to act as a replacement—a modernization—of NCLB. The first pilots of the curriculum were two years ago, and the program has been expanding since.

Increased standardization has come from tests hopping across state boundaries. Under NCLB, tests were typically restricted by state borders. NCLB simply provided a framework for each state to create and conduct standardized testing. So tests were standardized—as long as you looked only inside one state. With Common Core and its aligned consortia, the goal has been to develop tests which transcend state boundaries. It is a step towards truly national standardization, where data is comparable across any two state boundaries in the United States.

To recap: the current level of federally-mandated standardized testing amounts to a little over two tests annually per student, on average—reading and writing, with the potential for a science exam depending on the year. This seems like a reasonably low number.

However, this is just a federal requirement. So long as nothing is specified by a superior bureaucracy, every bureaucratic level of American government which oversees education—federal, state, and district—can implement standardized tests. And oftentimes, this is exactly what happens.

While every state administered NCLB exams once the law came into full effect, many also supplemented this
legislation with testing of their own. The current suite of tests includes exit, end-of-course, and diagnostic exams. As of 2012, half of American states administered exit exams, and fifteen used end-of-course exams, according to the Center for American Progress.

This filters down to the district level as well. While it’s easy to think of American public schools as limited and ailing (think about the proportion of media coverage concerning faltering budgets), the truth remains that many offer opportunities for students to specialize and excel. However, this requires identifying students for whom specialized services (gifted-education or vocational programs, to name just a couple) would be a good fit. Often, the burden for identification is passed down to districts from states, resulting in additional district-level testing. Indeed, some studies have shown that a plurality of a typical public school student’s test load comes from administration at a district level. We have, of course, included some of these evidence pieces in the body of this brief.

Districts also administer benchmark examinations of the same style as those given by states. However, they take on drastically different forms. For starters, the medium is different: tests at the district level are increasingly digital.

One other crucial difference is the timing. Federal- and state-mandated testing is often an end-of-year affair, with results potentially coming back to teachers four months later. District-led assessments, however, are used on an interim bases. That is, they are administered throughout a school year. Combined with their digital implementation, they are able to provide near-instant results to teachers, who can then use the results to modify instruction, review missed concepts, etc.

Because district exams are quickly administered and used throughout the year, they also get used as predictive tools. Often, superintendents can know their likely end-of-year state exam scores by looking at net results and trends in their district-level interim testing data.

This context will be important to how both sides framework their cases. Oftentimes, we think of American public policy as a top-down federal affair. Defense. Healthcare. Infrastructure. Most facets of American public services are easily linked to federal directives.

As far as power transfer to the federal government goes, however, education has lagged behind other public services. Fundamentally, data collection and decision-making still happens on a district level. This helps to explain one of the main calls against standardized testing: that it is redundant. Often, districts that find state or federal standards inadequate will implement their own versions of the same tests which ostensibly provide better, more actionable data.

There are other federal factors at play: Race to the Top and the Elementary and Secondary Education Act, to name a couple. For both teams to consider the resolution in a comprehensive manner, they will need to dissociate the notion of standardized testing from No Child Left Behind or other single pieces of federal legislation. To not do so risks muddling debate.
Framework 1: The Piecemeal Approach

“Resolved: On balance, standardized testing is beneficial to K-12 education in the United States.” We provide multiple takes on the definition of standardized testing in this brief, so I will leave this term up to the discretion of individual teams. As far as selection of a definition goes, however, teams should, in the process of preparation, single out case studies to ensure they can distinguish between what is and is not a standardized test in the heat of round. In particular, electronic testing methods require a nuanced approach. How teams decide to treat electronic testing—both typical tests and analytical learning software—will determine how they attack opponent advocacies concerning these topics.

Of greater interest, from a frameworking perspective, is the term “K-12 education.” What does it truly mean for something to be of benefit to “K-12 education”? The problem here is that this term encompasses just about anything a team wants it to. If the resolution had specified some component—say, student achievement—this resolution would be relatively straightforward. But for a team to simply launch into their contentions without specifying what exactly K-12 education entails sounds like a recipe for muddled debate.

The most functionally intuitive approach is to break K-12 education down into its constituent pieces and attack each of them separately. In doing so, teams will put together comprehensive advocacies. This will allow them to more concretely address the resolution while also easing the process of CX by enabling category-by-category comparisons between cases to look for logical and evidentiary deficiencies.

Fundamentally, the goal of K-12 education is to better the future of the country. When the educational system churns out individuals with a brighter future, the net potential for the nation as a whole goes up. But there are many more parties to the education system: parents, teachers, and administration, in particular. With that in mind, we can break “K-12 education” into four constituent components: students, teachers, parents, and administration. Of course, teams likely will not devote equal time to all of these, and some (likely the last two) may be cut out entirely. But in breaking K-12 down into smaller chunks, teams make it easier to win rounds by starting off with a clear purpose. I’ll be focusing mainly on the first two criteria.

Students

What does it mean to improve K-12 education for students? The most logical basis for discussion seems to be potential: what course of action maximizes a given student’s potential?

For Con teams, the path to arguing this side of the resolution is straightforward and has many routes. One approach is the simple “time and money” method: testing takes time and money. Time is in the form of all the hours spent filling in bubbles and short-answer questions. It adds up. Keep in mind the context of this debate: the federal, state, and district levels all prescribe testing regimens, and these all fall at the feet of children.

Con teams will likely have to stick with opportunity cost arguments on the monetary front. Currently, standardized testing accounts for roughly a quarter of a percent of total education costs. So relatively, this
number is small. In absolute terms, however, it is billions of dollars. As far as American public spending goes, billions of dollars is still not a crushing amount of money. The opportunity cost argument is a great lead-in for Con teams to get creative and proactive with their argumentation. While PF will never be Policy debate, the burden on Con teams, when arguing the opportunity costs, is to show where those opportunities are. If standardized tests provide even a shred of benefit and have no better alternative, they could arguably be worth the billions of dollars.

For Pro teams this resolution entails putting on the shoes of nearly every possible student. It may not be easy. I’m fortunate enough to have gone to a school with advanced placement classes in every possible subject and a well-funded debate team; we even had a computer lab attached to the debate room where I spent countless nights building cases and spitballing debate ideas. I was lucky enough to have stability and opportunity. Many students are not.

For many school districts, something as simple as keeping track of students is difficult. Especially in low-income districts, any given student may drop and re-enroll in a school multiple times per year as their family moves between rental properties. Oftentimes, teachers will return to classrooms in the fall to find that their students have substantially regressed over the summer. Again, this happens more often in low-income areas, where stability and opportunity don’t magically appear in the summer months.

In contributing to this brief, the most haunting, provocative piece I read was Meredith Broussard’s “Why Poor Schools Can't Win at Standardized Testing,” published in The Atlantic. Generally, this piece reads like a Con debater’s rhetorical cheat sheet. But I would encourage Pro teams to give it a read—and to read between the lines. Broussard investigated the Philadelphia public school system and found that for low-income schools, preparing for standardized tests was a near-impossibility.

The reason was textbooks: for students to do optimally on standardized tests, they needed the right books. But, of course, the district could not afford those books. More damningly, the district decided it could not afford any books, with the per-student book budget set to exactly $0.

More damningly, the system for tracking books as well was horribly broken. New books, which no one could confirm existed, were found in the backs of closets. The centralized computer system which tracked books didn’t actually have accurate data for whether those books existed and where they were.

Broussard’s prescription was to stop giving tests, or at least tests tied to specific literature. This is where Pro teams start gaining some footing.

For Pro teams, a great way to approach student outcomes is to mix comparative data with a strong advocacy. For instance, let’s take books. Many school districts cannot afford books due to budget shortfalls. Add up the budget shortfalls across all the school districts in America and we get a number far in excess of what is spent annually on standardized testing. So cutting standardized testing clearly is not the solution to America’s educational budget woes.
But what is standardized testing actually good at? Standardization, put simply. Keep in mind the context of this topic: many decisions and data collection schemes occur at the district level. For many low-income districts, there aren’t actually enough resources to make sure this happens in a comprehensive way. The result is exactly what Broussard found: things slip through the cracks. Resources go unaccounted for. More damningly, students go unaccounted for.

Pro teams can thus position standardized testing as a means to stop students from “falling through the cracks.” In context of a world where districts have neither the funding nor the resources to keep track of everything they need to in an accountable way, standardized testing presents a bastion of stability. Standardized testing produces massive quantities of data which, when handled by entities with adequate resources (e.g. at the state or federal level), does not get lost.

Crucially, Pro teams are not bound by this resolution to advocate standardized testing in its current form. There is no reason for Pro teams to delude themselves into thinking (and subsequently arguing) that America’s current implementation of standardized testing is optimal. For instance, Pro teams can advocate standardized testing based on general standards of proficiency, rather than particular materials. Pro teams can advocate for public development of test questions (similar to PARCC), which would prevent private publishers from tailoring their own materials and exams to each other (driving up prices in the process).

Fundamentally, the biggest asset for Pro teams is that, at its core, an advocacy of standardized testing can be an advocacy for more, better-managed data. And when it comes to the students who slip through the cracks—those who are constantly moving; those who are at risk; those who, due to an accident of geography, attend lesser-funded schools, do not have the resources they need to succeed—having data is a crucial first step to righting wrongs.

**Teachers**

We can break down the impacts of standardized testing on teachers into two subgroups: education and results. That is, standardized testing impacts how teachers educate their students—what curriculum is used, what kind of problems students are generally given, what gets disproportionate amounts of time—and how they are evaluated.

For Con teams, we have provided no shortage of data to address either of these facets. It would be naïve to think that how teachers are evaluated and what they teach are somehow not linked. For many districts, the implementation of standardized testing has also ushered in an era of predictive analytics—the use of data to predict and project an individual’s performance. Standardized test scores, crunched using “value-added” data analysis techniques, have increasingly been used by (typically urban) school districts as part of their evaluative teaching methodologies.

The shift in how teacher performance is measured has, by many accounts, been mirrored by teacher performance itself. A common notion is that teachers have begun “teaching to tests.” While this is not necessarily entirely true, a decent well of data supports this general theory.
One other interesting angle for Con teams will be to analyze the differences in impacts between low- and high-performing schools. There already exists a high correlation between performance on standardized tests and socioeconomic status—many academics have already pointed out that standardized testing, in its current form, tends to measure how much a student’s parents earn in income, more than anything else. That, in itself, is the foundation of the stock argument that standardized tests are discriminatory. A more intriguing approach would for Con teams to examine how curricula are impacted by standardized testing—are teachers in low-income schools, with higher pressures to improve scores, “teaching to the test” more?

This would lay the foundation for a holistic Con argument which truly encompasses discrimination in the American educational system. Between teachers teaching to tests and test material itself not being conducive to the development of truly applicable problem-solving skills, Con teams can demonstrate how testing systematically disadvantages students based on socioeconomic status.

For Pro teams, pragmatism will be the recipe of choice. The fundamental goal is for the best teachers to be teaching the best material. Pro teams, by again using comparative scrutiny, can demonstrate standardized testing to be the most beneficial option.

Let’s look initially at the first component listed above: the actual education being provided. Standardization, almost by definition, provides a template for teachers to instruction. In some way, yes, this is teaching to the test. But for teachers faced with a scarcity of teaching resources, standardization acts as a starting point from which to educate. Again, we see standardization as a tool to combat the phenomenon of falling through the cracks.

But the most promising arena is teacher evaluation. At first, it seems profoundly unfair that teachers are evaluated based on value-added criteria derived from standardized test scores. And maybe, to a certain extent, it is. But two factors can swing this argument to the Pro.

The first is an analysis of the current state of teacher evaluation. While teachers may be evaluated on multiple criteria, only two, in many cases, determine whether an individual is retained during budget cuts: degree attainment and seniority. The current system of teacher retention is essentially designed to reward past success—getting graduate degrees and teaching for a long time—instead of current success. As imperfect as it may be, standardized testing provides a much more meaningful analytical criterion for teacher retention.

More importantly, as mentioned in the previous section, Pro teams are not limited to advocating the current system. An advocacy of standardized testing in tandem with in-person evaluation would ensure that quality teachers are retained, and that the pressure to “teach to the test” is lessened (since other meaningful criteria are used as well).

Pro teams should sense a common theme at this point: by unshackling themselves from defending the status quo and taking up affirmative advocacies, Pro teams unburden themselves of tricky arguments and impossible justifications of ineffective policies. This seems to be the most effective ticket to a winning round on the Pro for this resolution, especially considering that people in general (and, by extension, lay judges) are on the whole
unreceptive to the idea of standardized testing as it is. Pro teams can win by proving that standardized testing has a relevant future, rather than a defensible past.

**Framework 2: Floors and Ceilings**

Thus far, I have only brought up a tiny fraction of the available arguments and contexts for this resolution. But let’s take a look at what’s been presented so far.

On the Con, the arguments are about limitations: testing limits the time and money used on students, while also hampering teachers’ ability to vary their curriculum. On the Pro, the theme is accountability: ensuring students and teachers don’t slip through the cracks. This sets up the floors and ceilings framework.

In this case, the floor is everyone’s minimum performance: what is the minimum level of student accomplishment and teacher competency. On the other hand of the spectrum is the ceiling: how can the best students and teachers maximize their potential?

I won’t be breaking this framework into constituent components or bringing up additional arguments tailored to them; this is more a way of thinking about the evidence you come across and mentally sorting it, potentially into a comprehensive case (if you’re a fan of this framework, of course).

For the Pro, after a few weeks of dealing with evidence pertaining to this resolution, the floors framework seems intuitive: standardized testing raises the collective floor of American education. Fundamentally, this framework is rooted in treating standardized testing, very simply, as a data collection mechanism. When it comes to the highest-performing students and teachers, we don’t necessarily need data. We know that they are generally not at risk of immediate failure. But what about low-performing students? What about students who are falling behind, or constantly moving, or simply regressing over each subsequent summer? What about the teachers who are no longer pulling their weight? It takes data to identify these individuals and track trends. And for Pro teams, consideration of a world where this data doesn’t exist can form the impetus of a strong advocacy and case.

On the Con, the opposite is generally true. Standardized testing creates a box that, as evidence can show, will limit the scope of both teacher and student potential. This can mean the best teachers aren’t teaching the best material This can mean the best students aren’t tackling the most intellectually provocative material. This is all fodder for Con teams under the “lowered ceiling” framework.

This kind of frameworking will likely impact the style of debate a team faces. In presenting standardized testing as impacting the potential of students—either raising that of the lowest achievers or lowering that of the highest achievers—teams risk turning the debate into one over the fundamental purpose of the American educational system and who it is truly meant to serve. This might be a debate you’re more comfortable with tackling than the minutiae of standardized testing’s particular impact. And if you’re looking to push potentially worse-prepared opponents into territory they’re unqualified to handle, this might just do the trick.
Wrap-up

When prepping for debates, I would try to hear as many opinions as possible before turning my preliminary evidence into a research path and subsequent case. For this resolution in particular, I’d encourage you to do the same—not to farm contentions, but to broaden your perspective (and win some rounds in the process). You’re surrounded by people—teachers, friends, parents, administrators—who are impacted by this resolution.

As always, best of luck this upcoming month.

- Daniel Tsvankin
Pro Evidence
Testing Allows Teachers to Help Students

*Standardized testing identifies areas where students need improvement. ABH*


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.*

*Standardized testing is necessary for teachers to objectively know which subjects their students are struggling with. Such testing assists teachers develop a curriculum to best suit their students’ needs.*

*Standardized testing helps teachers assess knowledge. LZ*


Tests and quizzes in the classroom are perhaps one of the most important ways in which teachers can formally assess the knowledge of their students, but of course homework can be used for this purpose, too. Testing is typically seen as an evaluation of what students have learned, and indeed this is true. Conscientious teachers will pay attention to how students perform on tests and use that knowledge to inform their teaching in the future. If many students fail a particular topic on the test, it may be a sign to spend more time covering that material next time or use a different approach to teaching the materials. Teachers can also learn how individual students perform and what the students’ respective strengths and weaknesses are. In turn, teachers can use that information to guide further instruction.
Non-test based solutions just aren’t accurate at measuring knowledge levels. LZ


Teachers often drastically overestimate what they believe their students to know (Kelly, 1999) and testing provides one way to improve a teacher’s estimation of their students’ knowledge. The problem of “the curse of knowledge” permeates education. That is, instructors (especially those just beginning) can fail to realize the state of knowledge of their students and pitch their presentations at too high a level. (Most readers can think of their first calculus or statistics course in this regard.) The general idea is that once we know something and understand it well, it is hard to imagine what it was like not to know it. For example, Newton (1990) conducted a study in which students sat across from each other separated by a screen. Each was given a list of 25 common tunes that most Americans know (Happy Birthday to You, the Star Spangled Banner, etc.). One student (the sender) was picked to tap out the tune with his or her knuckles on the table and give an estimate of the likelihood that the other student could name the tune. The other student (the receiver) tried to decipher the tune and name it. This is a classic situation similar to a teacher and student where one person knows the information (tune, in this case) and is trying to communicate it to the other person who does not know it. When the senders judged how well they did in communicating the tune to the other student, they thought they succeeded about 50% of the time! However, the students on the receiving end of the taps could recognize the tune only 3% of the time! When the sender was tapping out Happy Birthday, she was hearing all that music in her mind’s ear and tapping in time to it. What the receiver heard, however, was a series of erratic taps. This tale is an allegory of an expert in a subject matter trying to teach it to a novice, especially the first time. Again, it is hard to know what it is like not to know something you know well.
**Teachers Support Standardized Testing. ABH**


Interviews with 61 teachers and administrators in 4 Minnesota school districts suggest that, in their judgment, Minnesota’s state-mandated tests were well-aligned with curricular priorities and teachers’ instructional goals, emphasizing critical thinking as well as competencies needed to pass the Basic Standards exit exam, and avoiding the type of recall item that would require drill and memorization. This result, in combination with a survey showing that 85 percent of Minnesota teachers support the exit exam, suggests that Minnesota has been unusually successful in designing a high stakes testing system that has garnered teacher support. The success of Minnesota’s model suggests that unintended narrowing of the curriculum due to high stakes testing may be avoided if pressure on teachers to narrow the curriculum is reduced through well-designed, well-aligned exams.

**PRO can use this study to show that when executed properly, teachers generally support standardized testing.**
Testing Good

Standardized testing motivates students to study. LZ


Probably the most influential indirect benefit of testing is the one described in general terms at the beginning of the chapter: Having frequent quizzes, tests, or assignments motivates students to study. Every professor and every student knows that many students procrastinate and often do not study until the night before a test. Often university courses include only a midterm and a final exam, and it is no surprise that the episodes of studying occur primarily just before tests. Mawhinney, Bostow, Laws, Blumenfeld, and Hopkins (1971) documented this point in controlled circumstances, with tests given daily, weekly, or every three weeks. Studying was most copious and evenly spaced with daily testing. With less frequent testing, study behavior occurred only before the tests (see also Michael, 1991). In addition, in their survey of student behaviors described previously, Kornell and Bjork (2007) found that 59% of students, when choosing what to study, chose topics that were due soon or already overdue. More frequent testing across the semester would encourage students to study more and would space their studying over several weeks.

Testing doesn't produce rote learning and feedback prevents this problem. LZ


Second, critics sometimes argue that retrieval practice through testing produces “rote” learning of a superficial sort, as if the student can parrot back the information but not really understand it or know it in a deep fashion. Learning is said to become “inert” or “encapsulated” in little factoid bubbles. Perhaps this criticism is justified in some cases, but we think that good programs of quizzing with feedback usually prevent this problem. We reviewed evidence previously showing that retrieval (via testing) can lead to deep knowledge that can be used flexibly and transferred to other contexts (e.g., Butler, 2010). Again, the burden is on the critics to show that testing leads to problems rather than simply asserting that these problems might exist. The next two criticisms are based on data and must be taken more seriously.
Testing doesn’t necessarily trade off with other classroom activities. LZ


First, quizzing in class may take time away from other critical classroom activities, such as lectures, discussion, and demonstrations. Is that a problem? This point is true to an extent, but how does one know (in absence of proper studies) whether these activities are better than retrieval via quizzing? For example, Karpicke and Blunt (2011) showed that retrieval practice produced better retention later than did concept mapping, a widely used study technique. We expect that when other such studies are conducted, they may show that some quizzing is as beneficial as, or more beneficial than, an equal amount of time spent on lecturing (just as testing is better than restudying). In addition, as discussed above, having classroom quizzes may keep motivation up and provide the indirect benefit of having students study more. At any rate, we do not think this criticism holds water, but future research may change our opinion.
Standardized testing enhances retrieval processes. LZ


Testing in school is usually done for purposes of assessment, to assign students grades (from tests in classrooms) or rank them in terms of abilities (in standardized tests). Yet tests can serve other purposes in educational settings that greatly improve performance; this chapter reviews 10 other benefits of testing. Retrieval practice occurring during tests can greatly enhance retention of the retrieved information (relative to no testing or even to restudying). Furthermore, besides its durability, such repeated retrieval produces knowledge that can be retrieved flexibly and transferred to other situations. On open-ended assessments (such as essay tests), retrieval practice required by tests can help students organize information and form a coherent knowledge base. Retrieval of some information on a test can also lead to easier retrieval of related information, at least on delayed tests. Besides these direct effects of testing, there are also indirect effects that are quite positive. If students are quizzed frequently, they tend to study more and with more regularity. Quizzes also permit students to discover gaps in their knowledge and focus study efforts on difficult material; furthermore, when students study after taking a test, they learn more from the study episode than if they had not taken the test. Quizzing also enables better metacognitive monitoring for both students and teachers because it provides feedback as to how well learning is progressing. Greater learning would occur in educational settings if students used self-testing as a study strategy and were quizzed more frequently in class.
Wheeler and Roediger (1992) conducted an experiment in which a strong testing effect occurred, although the experiment was mostly about a different topic. We present selected conditions here from their experiment to make our points about testing. Their subjects saw 60 pictures while they listened to a story, with instructions that they would later be asked to recall the names of the pictures. The pictures were integrated into the story so that when an object was named in the story, the picture appeared on the screen. Subjects were told that paying attention to the story would help them retain the pictures (which was true). After hearing the story and seeing the pictures, subjects were given free recall tests in which they were given a blank sheet of paper and had to recall as many of the names of the 60 pictures as possible. After hearing the story, one group of subjects was told that they could leave and return a week later for a test. A second group was given a single test that lasted 7 min and then they were excused. The third group was given three successive 7-min tests after the learning phase; that is, they recalled the pictures once, were given a new blank sheet and recalled as many items as possible a second time, and then repeated the process a third time. The group that recalled pictures once recalled about 32 pictures and the group that recalled them three times recalled 32, 35, and 36 pictures (i.e., performance increased across tests, a phenomenon called hypermnesia; Erdelyi & Becker, 1974). For present purposes, the data of most interest are those on the final retention test 1 week later when the students returned to the lab for more testing. Students in all three groups had heard the story and seen the pictures once, so the only difference among the three groups was how many tests they had taken just after studying the materials (0, 1, or 3). How did this manipulation affect recall? The data to answer this question are shown in Figure 1, where it can be seen that those who had not been tested recalled 17.4 pictures, those who had been tested once recalled 23.3 pictures, and those who had previously been tested three times recalled 31.8 pictures. Thus, taking three tests improved recall by nearly 80% a week later relative to the condition with no tests. Another way to consider the data is by comparing the scores on the immediate test just after study to those a week later. Recall that on the first test after study, subjects produced about 32 items. We can assume that those subjects who were not tested immediately after study could have recalled 32 had they been tested, yet a week later they could recall only 17, showing 45% forgetting. However, the group that was tested three times immediately were still able to recall 32 items a week after study, thus giving three tests essentially eliminated forgetting after a week. This outcome shows the power of testing.
Testing helps identify gaps in knowledge. LZ


The testing effect represents a direct benefit of testing; the second benefit is indirect. Taking a test permits students to assess what they know and what they do not know, so that they can concentrate study efforts on areas in which their knowledge is deficient. Students may take a practice quiz, realize which questions or items they got wrong, and then spend more time studying the items they missed. For example, Amlund, Kardash, and Kulhavy (1986) found that subjects corrected errors on a second test if they had an intervening study session after the first test. Other research shows that when students receive opportunities to restudy material after a test, they spend longer on restudying items that were missed than those that were correctly retrieved (see Son & Kornell, 2008).

Testing facilitates further learning. LZ


Kornell and Bjork (2007) provided evidence from a laboratory experiment that students are typically unaware that learning can occur during testing. In one experiment, students learned a set of Indonesian–English vocabulary words by repeated trials. They had the option of studying the pairs or being tested on them (with feedback) on each occasion and could switch between the two modes at any point. Most students began in the study mode, although nearly everyone changed to the test mode after the first two trials. Kornell and Bjork interpreted this outcome as indicating that students wanted to achieve a basic level of knowledge before testing themselves. In addition, Kornell and Bjork also reported the results of a survey in which students were asked whether they quizzed themselves while studying (using a quiz at the end of a chapter, a practice quiz, flashcards, or something else); 68% of respondents replied that they quizzed themselves “to figure out how well I have learned the information I’m studying” (Kornell & Bjork, 2007, p. 222). Only 18% of respondents recognized that testing actually facilitated further learning.
Another benefit of retrieval practice is it can enhance learning during future study sessions. That is, when students take a test and then restudy material, they learn more from the presentation than they would if they restudied without taking a test. This outcome is called test-potentiated learning (Izawa, 1966). The benefits of test potentiation are distinctly different from the direct benefits of testing per se, although in many practical situations (e.g., receiving feedback after tests) the two are mixed together. Izawa (1966) was perhaps the first researcher to study the test potentiation effect and has contributed much to our understanding of test potentiation. Her initial forays into the area emerged after asking questions about whether learning could occur during a test. She proposed three specific hypotheses. First, neither learning nor forgetting occurred on test trials. Second, learning and forgetting (as well as learning of incorrect information) could occur on test trials. Finally, although learning and forgetting might not occur on a test session, taking a test might influence the amount of learning during a future study session. Izawa studied how different patterns of study, test, and neutral trials affected later performance. Across many experiments (e.g., Izawa, 1966, 1968, 1970), Izawa concluded that neither forgetting nor learning occurred on test trials, but taking a test could improve the amount of material learned on a subsequent study session. While this conclusion may appear to contradict the basic finding of the testing effect, the contradiction is resolved by examining how learning and forgetting are defined in Izawa’s basic paradigm. Izawa’s conclusion was that no learning or forgetting occurred during a test trial, but she made no assumptions about how learning or forgetting would be affected after the test trial; the testing effect can be interpreted as a slowing of forgetting after the test.
Test improves organization of knowledge. LZ


Another indirect benefit of retrieval practice is that it can improve the conceptual organization of practiced materials, especially on tests that are relatively open-ended (such as free recall in the lab or essay tests in the classroom). Gates (1917) postulated that one of the reasons retrieval practice leads to increased performance is that retrieval (or recitation, as he called it) causes students to organize information more than does reading. He suggested that as students actively recall material, they are more likely to notice important details and weave them into a cohesive structure. Masson and McDaniel (1981) showed that an additional testing session after study resulted in higher performance on delayed recall and recognition tests and, more important, that the additional test yielded higher organization on the final recall test. Their primary measure of organization was the adjusted ratio of clustering (ARC), which is a measure of how often words from the same category are recalled together in free recall with an adjustment for the overall level of recall. Scores range from 1 to 1, with 1 representing perfect organization or clustering and 0 representing chance clustering (Roenker, Thompson, & Brown, 1971). Masson and McDaniel’s results suggested that the test resulted in improved organization and higher recall on final tests. More recently, other research (Zaromb, 2010; Congleton & Rajaram, 2010) has explored the relationship between testing and organization. Experiments reported by Zaromb and Roediger (2010), for example, showed that retrieval practice during testing improves both the organization of materials and their recall. In fact, the increased organization from previous retrievals may provide an underlying mechanism of the testing effect, at least in free recall.
Testing doesn’t just shoehorn knowledge – it facilitates application to new situations. LZ


One criticism of retrieval practice or testing research is that students may be learning little factoids in a rote, verbatim way. Critics complain that testing is the old “kill and drill” procedure of education from 100 years ago that produces “inert knowledge” that cannot be transferred to new situations. However, proponents of testing argue that retrieval practice induces readily accessible information that can be flexibly used to solve new problems. This issue leads to the crucial question of whether knowledge acquired via retrieval practice (relative to other techniques) can be applied to new settings. Recent research shows that the mnemonic benefits of taking a test are not limited to the specific questions or facts that were tested; retrieval practice also improves transfer of knowledge to new contexts. Transfer may be defined as applying knowledge learned in one situation to a new situation. Researchers often categorize transfer as being near or far; near transfer occurs if the new situation is similar to the learning situation, whereas far transfer occurs if the new situation is very different from the learning situation. Barnett and Ceci (2002) proposed a taxonomy for transfer studies, arguing that transfer might be measured on many continuous dimensions (e.g., knowledge domain, physical context, temporal context, etc.).

Evidence suggests testing is even better than restudying. LZ


Some evidence suggests that repeated testing can facilitate transfer better than restudying. For example, Carpenter, Pashler, and Vul (2006) showed that testing with word–word paired associates (denoted by A–B here) improved performance on a later test relative to addi- tional study opportunities. When given A, subjects could recall B more often when they had previously been tested relative to only studying the pairs. More important, Carpenter et al. also tested subjects’ recall for the A member of the pair when they were given B, so they were tested on the member of the pair that was not directly retrieved during initial testing. Recall was improved for these A items when learning had occurred via testing relative to repeated studying. Repeatedly testing with one member of the pair transferred to higher performance in recalling the other member of the pair. This could be considered a case of near transfer.
Testing improves retention even on non-tested material. LZ


One potential limiting factor of implementing testing in a classroom setting is choosing which material to test. It is unrealistic for an instructor to test students on everything. Fortunately, research on testing suggests that retrieval practice does not simply enhance retention of the individual items retrieved during the initial test: taking a test can also produce retrieval-induced facilitation—a phenomenon that shows testing also improves retention of nontested but related material. Chan, McDermott, and Roediger (2006) were the first to coin the term retrieval-induced facilitation, providing evidence for the effect in three experiments. Students studied a prose passage and then completed two initial short answer tests, restudied the passage twice, or did nothing (the control condition). Those in the initial testing group answered questions related to a subset of information from the passage. More important, another subset from the passage was not tested during the initial test, but this material was related to the questions that had been answered on the initial test. In the restudy condition, students read the answers but did not receive a test. After 24 h, all the students returned to complete a final test covering the entire passage. Results of the final test revealed that retention of the nontested information was superior when students had taken a test relative to conditions in which they restudied the material or in which they had no further exposure after study. Chan et al. concluded that testing not only improves retention for information covered within a test, but also improves retention for nontested information, at least when that information is related to the tested information.
Testing improves metacognitive accuracy. LZ


Another benefit of testing is improvement of metacognitive accuracy relative to restudying (e.g., Roediger & Karpicke, 2006a; Shaughnessy & Zechmeister, 1992). This point is related to the second one discussed—testing informs students as to what they know and what they do not know. However, in this case, the focus is on students’ accurate predictions of their future performance. Testing permits students to have better calibration of their knowledge. If students only study material repeatedly, they may think that their familiarity with the material means that they know it and can retrieve it when needed. However, such familiarity can be misleading. These points have direct implications for educational settings—the better students are at differentiating what they do know and what they do not know well, the better they will be at acquiring new and more difficult material and studying efficiently (Thomas & McDaniel, 2007; Kornell & Son, 2009). Therefore, instead of simply restudying, teachers can administer quizzes and students can self-test to determine what material they know well and what material they do not know well. Students’ ability to accurately predict what they know and do not know is an important skill in education, but unfortunately students often make inaccurate predictions. When students reread material repeatedly, they are often overconfident in how well they know the material. Taking a test, however, can lead to students becoming less confident, a finding known as the underconfidence-with-practice effect (Koriat, Scheffer, & Ma’ayan, 2002; see also Finn & Metcalfe, 2007, 2008). Testing can help compensate for the tendency to be overly confident, which results in a more accurate assessment of learning.
Students do not self test themselves, so testing is a valuable classroom tool. LZ


Testing is a powerful way to improve retention, but when students are given control over their own learning, they do not often choose to test themselves or do not test themselves very frequently (Karpicke, 2009; Kornell & Bjork, 2007). During paired-associate learning, when students are given the opportunity to drop, restudy, or retest on items they have correctly retrieved, they often choose to drop items despite benefits that would accrue if they continued to test themselves. When given control early in the learning phase, students often choose to study pairs instead of testing themselves on them and receiving feedback. These decisions seem to be guided by their inflated judgments of learning, but they lead to poor learning strategies (Karpicke, 2009; Metcalfe & Finn, 2008). Students seem to lack a good theory about what study strategies are effective. As noted in a previous section, surveys have shown that university students do not realize the direct benefits of retrieval practice as a study strategy. Future research is needed to determine if students can be educated on this aspect. For example, if students experience the benefits of retrieval practice on learning in one context, will they then adopt this strategy for learning in a different context? While we must await the answer to this question, we can say that testing does cause students to become less overconfident in the judgments of learning (even to the point of underconfidence, as in the underconfidence- with-practice effect). Because tests generally improve metacognition, educators should encourage their students to self-test during learning and while studying.
Another indirect benefit of testing is that tests create a release from proactive interference. Proactive interference occurs when sets of materials are learned in succession; the previous material learned influences the retention of new materials in a negative manner. Thus, proactive interference refers to the poorer retention of material learned later, caused by prior learning (Underwood, 1957; see Crowder (1976) for a review). Elongated study sessions may therefore cause a buildup of proactive interference. However, research has shown that when tests are inserted between study episodes, they cause a release from proactive interference and enable new learning to be more successful. Szpunar, McDermott, and Roediger (2007) reported evidence of a release from proactive interference caused by testing in a paradigm in which subjects learned five lists of words. During learning, each list was separated from the next list by an immediate test or a short break of equivalent length. The group that took tests between each list performed better on a final test relative to the group that took short breaks. In addition, the tested group was able to recall a greater proportion of studied words from the most recent list relative to the no-test control group. Thus, taking tests after learning each list protected the subjects from proactive interference during learning.
Testing Closes Achievement Gaps

*Testing key to identify racial achievement gaps – first step to fixing them. ELT*

To civil rights groups, standardized tests are an important tool for pointing out — and hopefully fixing — those racial differences in achievement. No Child Left Behind helped reveal how black and Latino students, or students with disabilities, were performing on average, relative to the average performance of white students — and, perhaps more importantly, tried to require states to fix it.

Standardized tests "are the only available, consistent, and objective source of data about disparities in educational outcomes, even while vigilance is always required to ensure tests are not misused," a coalition of 12 national civil rights groups wrote in a May letter. "These data are used to advocate for greater resource equity in schools and more fair treatment for students of color, low-income students, students with disabilities, and English learners. "They concluded: "We cannot fix what we cannot measure."(1)
Testing ensures students with disabilities are not left behind. ELT

David Stephens and Jaclyn Long [University of Connecticut], 2010, “Standardized Assessments and Students with Disabilities”
http://epsy5121fall2010.wikispaces.com/Standardized+Tests+and+Students+with+Disabilities

Data gathered from alternate assessments is usually different from that gathered from the general assessment, however, if they are aligned well with the same academic standards, then the performance of students with disabilities on alternate assessments can show progress toward achieving the same necessary skills and knowledge expected of all students (Roach & Elliot, 2010). Based on the ability of states to test some students with disabilities using alternative assessments, the desire to hold schools and districts accountable for all student achievement is made attainable. The mandated system of accountability under NCLB represents a major change in how the achievement of students with special needs is measured. Prior to provisions to IDEA 1997 and NCLB, the primary source of accountability for students with special needs were the goals written in their Individual Educational Plans (IEPs) (McLaughlin & Thurlow, 2010). These goals are often not standards based, making direct peer-to-peer comparison less reliable and valid. Privacy provisions within the IEP prevented schools from reporting student achievement data for students with special needs. Measuring student achievement based on IEP goals did not hold anyone accountable; no consequences were established for failure to attain goals (McLaughlin & Thurlow, 2010). In essence, accountability in terms of students receiving special education services was determined based on if IEPs were filled out correctly for each student. Students were rarely given full access to the general education curriculum, and therefore not expected to participate in standardized tests with their non-disabled peers.

In an age where "teaching to the test" has become a common term, it is important to know just what is being sacrificed in order for students with disabilities to take and pass standardized assessments. Students with disabilities are often the neediest students in schools, and therefore require the most intensive supports and resources. The passing of NCLB in 2001 marked the first time students with disability's scores were reported publicly. The focus of achievement for students with disabilities has made a dramatic shift from meeting individual goals to performing on standardized assessments. With this shift, instruction has focused increasingly on content covered by state assessments. With this shift in instructional focus, something is being sacrificed for students with disabilities. Future research should focus on what instruction is being cut out of the curriculum for students with disabilities in order to fit in more time for teaching content specifically on the assessment. McLaughlin and Thurlow (2010) wonder, "how current accountability practices that require teachers to teach specific subject matter, result on kinds of educational outcomes that are valued by students with disabilities [such as employment and independence]?
Federal Standards Do Not Detract From Education

Most student testing time is spent on district-level exams.


The Center for American Progress is an independent nonpartisan educational institute

Despite the perception that federally mandated state testing is the root of the issue, districts require more tests than states. State tests alone are not to blame for testing fatigue. District-level tests play a role too. Students across all grade spans take more district-required exams than state tests. Students in K-2 are tested three times as much on district exams as state exams, and high school students are tested twice as much on district exams. But even students in grades that must be assessed per No Child Left Behind took between 1.6 and 1.7 times more district-level exams than state exams. Most of the district-level tests in use were interim benchmark exams that are taken two to four times throughout the year. Other district-wide exams included diagnostic tests and end-of-course exams for students taking certain required courses.

Students are tested as frequently as twice per month and an average of once per month. Testing can occur very frequently for some students. Students in grades in which federal law requires annual testing—grades 3-8—take the most tests. This means about 10 tests, on average, throughout the year. But in the Jefferson County school district in Kentucky, which includes Louisville, students in grades 6-8 were tested approximately 20 times throughout the year. Sixteen of these tests were district-level assessments. In the Sarasota County, Florida, school district, middle school students were tested 14 times on state and district tests throughout the year. These interruptions in instruction may likely be contributing to public sentiment regarding students being overtested. Students in grades K-2 and 9-12, who do not take or are less frequently tested using federally required state exams, take the fewest number of tests—approximately six tests in a year.

An intuitive advocacy stemming from this evidence is increasing standardization and decreasing testing—that is, reducing redundancy between levels of testing administered by bureaucratic institutions.
Federally-overseen standardized testing is a crucial enabler of parental choice. DAT
http://www.brookings.edu/research/papers/2015/01/08-chalkboard-annual-testing

Whitehurst is a senior fellow at the Brookings Institution.

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.

Inherently, standardized testing is the most uniform and applicable way for information to be made available nationally. The impact, then, is on parental choice and, consequently, student success.
State-level Standardization Improves Education

*States with well-implemented holistic standards boost student achievement.* DAT


**Burke is the Will Skillman fellow in education policy at The Heritage Foundation.**

Instead of signing on to common standards that will drive state curricula, state education leaders should strengthen state standards and tests. States should follow the example of models like Massachusetts or Virginia in creating solid standards and aligned assessments. State standards can also be strengthened by continually raising the bar on achievement. As students reach content proficiency, the proficiency bar should be raised to further challenge students to meet the demands of college coursework and competitive careers.

States with outstanding standards and tests have taken great pains to ensure proper and precise learning sequencing. This is appropriate at the state level, where teacher certification and other integrated factors of a quality education system are determined.

The Bay State requires teachers to be proficient in all aspects of the Massachusetts Comprehensive Assessment System (MCAS) and in all subject matter content, and it aligns teacher testing to state standards.[25] Mastery of general content knowledge and subject matter knowledge required by teachers helps to ensure that standards are aligned both horizontally, so that students learn content aligned by grade level, and vertically, to eliminate redundant content and verify subject mastery.[26] An initial criticism of the common core standards was that there were “grade-sequencing problems in some places…such as requiring a math skill in one grade level without prerequisite skills in the previous grade level.”[27]

**Most education policy happens at the state level. Some of the most successful states (e.g. Massachusetts) use standardization in tandem with state-based evaluative methods to create a holistic picture of student success. This allows them to analytically determine education policy without artificially capping teachers’ or students’ ceilings.**
While federal control is unnecessary, states are adept at comprehensive testing. DAT

http://www.brookings.edu/research/papers/2015/01/08-chalkboard-annual-testing

Whitehurst is a senior fellow at the Brookings Institution.

Differences in the quality and rigor of standards for what students should learn are unlikely to have a substantial effect on academic achievement (as our colleague Tom Loveless has shown) and, in any case, do not require federal involvement. The Common Core, for example, was initiated by states operating through the National Governors Association. Since there is no imperative for federal involvement either from the perspective of constitutional responsibility or from the functional perspective of unmet need, there is no reason for requirements for state standards to be part of the ESEA reauthorization.

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.

Pro teams will face a wide variety of “stock” attacks. The key will be to prepare an advocacy that steps around them, which is surprisingly doable. One such attack is that the federal government’s one size fits all approach inherently doesn’t work for education. By advocating a state-based approach (as seen in this section), Pro teams can still position standardized testing as a “responsive” option.
Testing As Part of A Larger Educational Framework

Here, we present advocacy solutions for the Pro. It will be difficult for Pro teams to argue that standardized testing works in a vacuum. To help persuade skeptical lay judges, Pro teams will need to present a compelling advocacy that shows standardized testing fitting productively into a progressive educational framework.

*Teachers find standardized test information constructive*

http://www.wsj.com/articles/SB10001424052702304723304577366023832205042

[Thomas Kane, Harvard Graduate School of Education]: Across the country, school systems are reinventing the ways they evaluate and provide feedback to teachers. Although I don't believe student test scores should be the sole factor in teacher evaluation, I believe just as strongly that they have an important role to play.

Clear evidence for that conclusion comes from the Bill and Melinda Gates Foundation's Measures of Effective Teaching project, which I lead. The project has been working with 3,000 teacher-volunteers in six school districts to test different forms of feedback for teachers, including their students' gains on test scores. Ours and other recent studies confirm that achievement-gain measures provide valuable information and should not be ignored.

To clarify: We should focus on gains in test scores, not end-of-year scores. Any estimate of how much the student has improved while in the teacher's class must take into account the fact that students start at different points. We want to know how much a teacher contributes to student growth during the time students are in that teacher's classroom.
Testing provides sorely-needed data. DAT

http://www.wsj.com/articles/SB10001424052702304723304577366023832205042

[Thomas Kane, Harvard Graduate School of Education]: No information is perfect. But better information should lead to better decisions. Currently, high-stakes personnel decisions in K-12 education are primarily based on two factors: experience and graduate degrees. In the recent recession, thousands of teachers were terminated based simply on their seniority. As imperfect as the current measures of effective teaching are—and they must be improved—using multiple measures provides better information about a teacher's effectiveness than seniority or graduate credentials.

A high-quality system of performance feedback for teachers requires money, roughly 2% of teacher payroll costs. Given tight budgets, school systems will have to reallocate resources to cover the cost. But there is no other investment a school leader could make that would offer more bang for the buck.
Standardized test scores make it possible to predict student life outcomes. DAT

http://www.brookings.edu/research/papers/2015/01/08-chalkboard-annual-testing

Whitehurst is a senior fellow at the Brookings Institution.

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.

Pro teams don’t need to take the stance that standardized test are the end-all. By frameworking standardized tests as “another” predictive tool, Pro teams establish massive impacts: being able to gauge student performance across entire lifetimes, not just school years.
Standardized tests are a piece of modern evaluation systems. DAT


TNTP is an organization promoting education access and quality.

Smart retention hinges on the ability of school and district leaders to accurately identify Irreplaceables and low-performing teachers. States and school districts need to replace outdated teacher evaluation systems that rate nearly all teachers “satisfactory” and give them little useful feedback on their performance.

Research has shown that combining value-added data with the results of classroom observations and student surveys provides a more complete and accurate picture of a teacher’s success. Although using value-added data was the most practical way to conduct the research for this report, we strongly believe that teacher evaluations in the real world should use a “multiple measures” approach.

However, school and district leaders don’t need to wait for better evaluations to start focusing on smart retention. While working to build new evaluation systems, they can use existing information to better understand their teachers’ performance. For example, the Houston Independent School District developed a “staff review” process while it worked to build a comprehensive new evaluation system. As part of the process, principals gave each teacher an informal performance rating based on the results of standardized tests, classroom observations and all other available performance information. Research shows that principals can make these kinds of judgments accurately, especially when it comes to the highest and lowest performers.

The process helped the district support smarter retention decisions by requiring principals to discuss the retention of every high- and low-performing teacher with their managers. Principals needed to explain everything they had done to retain their Irreplaceables. If they were not working to dismiss or counsel out a low-performing teacher, they needed to make a compelling case for giving that teacher another year to improve...
Testing Data is Key To School Choice

Annual testing is key to prevent bad accountability systems from being implemented. ELT


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.

A majority of people support standardized testing. ELT


A majority of the public supports testing and accountability. A large majority — 71 percent — say they support annual mandatory testing as an indicator of how a school is performing (Public Agenda, 2003). A more recent poll reported that 57 percent of the general public thinks the amount of testing is currently about right or not enough (Phi Delta Kappan, 2005).

Because the public funds schools, use this evidence to support the idea that schools should be accountable to the people funding them.
Con Evidence
Hurts Learning

Fixation on standardized testing detracts from real teaching and learning. LZ


Students are frustrated. Parents are concerned. Educators are calling for change. Finally, our elected leaders are starting to listen. Finally, we have a chance to change the heavy emphasis on high-stakes standardized testing in our schools. The Obama administration recently said that the obsession with standardized tests has gone too far, and announced that it wants to revisit this policy and cap the amount of time public school students spend taking standardized tests. This is great news for our students and our schools. Today in our schools, some students will take 112 standardized tests between kindergarten and graduation. One hundred twelve. And why? The testing obsession in public education is based on the idea that we need to hold our kids, teachers and schools accountable for meeting high standards. You'll get no complaint from educators about making sure students learn what they need to know and are tested on how well they learn it. That's what educators do every day. But the fixation on standardized testing is actually making it harder to do it. The amount of time students spend taking tests is one problem. The amount of time teachers need to spend preparing kids for these tests is another. Together, testing and preparation for testing consume literally weeks of school time. If we didn't need all of that time to get kids ready for these standardized tests, what would we be doing? Teaching. And what would students be doing? Learning. But it's not just about time. It's also about focus. Standardized testing has directed too much attention to a few content areas that are easily tested, sacrificing lessons that build kids' natural curiosity, critical thinking and problem-solving skills. The high stakes associated with standardized tests have led to days of mindless test-prep and remediation. That's not real teaching, and it certainly isn't real learning.
Standardized testing is toxic to a culture of learning. LZ


But I do know that the overreliance on standardized testing distracts students and teachers and is utterly toxic to the culture of learning in our public schools. That's why I appreciate the president's change of heart on standardized testing, and that's why I'm grateful that Gov. Tom Wolf has secured a one-year delay from the federal government on the use of the new PSSA scores in Pennsylvania's School Performance Profiles. That's a good start, but it's just that — a start. We have advocated for a three-year moratorium on the use of PSSA results in the school performance profiles, because, if we're going to keep using these tests, the state should at least give schools the time to do it right and adjust their curriculum. I'm also for a full review of the policies and purpose behind standardized tests and a complete analysis of whether or not they actually make a positive difference for our students. Tests are supposed to help students learn. Right now, our toxic standardized tests miss that mark — by a long shot. That's why I'm reminding people what our public schools have been forced to sacrifice to comply with government orders to administer these tests. We've sacrificed teaching and learning. With less standardized testing, we can get back to what our public schools are supposed to do.

A school with high test scores loses the incentive to improve. LZ


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.

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. 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 must empty their walls and hallways for many weeks. Classes are under lockdown with limited access to restrooms, and some schools turn to daily announcements or even pep rallies to “prepare” students for testing.
Standardized testing hurts teacher creativity. LZ


You take what is happening in education. Right now, in recent years, there’s a strong tendency to require assessment of children and teachers so that you have to teach to tests. And the test determines what happens to the child and what happens to the teacher. That’s guaranteed to destroy any meaningful educational process. It means the teacher cannot be creative, imaginative, pay attention to individual students’ needs. The student can’t pursue things, maybe some kid is interested in something, can’t do it because you got to memorize something for this test tomorrow. And the teacher’s future depends on it, as well as the student. The people sitting in the offices, the bureaucrats designing this, they’re not evil people, but they’re working within a system of ideology and doctrines that turns what they’re doing into something extremely harmful.

Artificially ranking people with tests hurts younger children. LZ


First of all, you don’t have to assess people all the time… People don’t have to be ranked in terms of some artificial [standards]. The assessment itself is completely artificial. It’s not ranking teachers in accordance with their ability to help develop children who will reach their potential, explore their creative interests. Those things you’re not testing. So you are giving some kind of a rank, but it’s a rank that’s mostly meaningless. And the very ranking itself is harmful. It’s turning us into individuals who devote our lives to achieving a rank. Not into doing things that are valuable and important. It’s highly destructive at the lower grades. This is elementary education, so you are training kids this way. And it’s very harmful. I could see it with my own children. When my own kids were in elementary school, at a good quality suburban school, by the time they were in third grade they were dividing up their kids into dumb and smart. You’re dumb if you’re lower tracked, smart if you’re upper tracked. Think of what that does to the children. It doesn’t matter where they’re tracked, the children take it seriously… If you’re caught up in that it’s just extremely harmful. It has nothing to do with education.
Standardized testing hurts the creative potential of education – it turns us into machines that only maximize wealth. LZ

https://creativesystemsthinking.wordpress.com/2015/02/21/noam-chomsky-on-the-dangers-of-standardized-testing/

Education is developing your own potential and creativity. Maybe you’re not going to do well in school and you’ll do great in art. That’s fine. What’s wrong with that? It’s another way of living a fulfilling wonderful life, and one that is significant for other people as well as yourself. The whole idea [of ranking] is harmful in itself. It’s kind of a system of creating something called “economic man.” There’s a concept of economic man, which is in economics literature. Economic man is somebody who rationally calculates how to improve his own status (and status basically means wealth). So you rationally calculate what kinds of choices you should make to increase your wealth, and you don’t pay attention to anything else. Maximize the number of goods you have, cause that is what you can measure. If you do that properly, you are a rational person making informed judgments. You can improve your “human capital,” what you can sell on the market. What kind of human being is that? Is that the kind of human being you want to create? All of these mechanisms- testing, assessing, evaluating, measuring- they force people to develop those characteristics… These ideas and concepts have consequences…

This narrow, mechanistic approach to education does not correspond to the reality of child development and brain science, but the metrics and assessment train charges down the track nevertheless. So what's wrong with that, you might ask? Isn't school about teaching kids stuff and then testing them to see what they've learned? In a word, "No." It simply doesn't work, especially with young children. As Boston College Professor Peter Gray wrote in a recent Psychology Today article: Perhaps more tragic than the lack of long-term academic advantage of early academic instruction is evidence that such instruction can produce long-term harm, especially in the realms of social and emotional development. "Direct instruction" does increase scores on the tests the instruction is aimed toward, even with very young children. This self-fulfilling prophecy is not surprising. But multiple studies also show that the gains in performance are fleeting -- they completely wash out after 1-3 years when compared to children who had no such early direct instruction. "Wash out" is too kind. A comprehensive study of kindergartens in Germany revealed, as Gray writes: Despite the initial academic gains of direct instruction, by grade four the children from the direct-instruction kindergartens performed significantly worse than those from the play-based kindergartens on every measure that was used. In particular, they were less advanced in reading and mathematics and less well adjusted socially and emotionally.


In another extensive study of poor children in Ypsilanti, MI, young boys and girls who were in academic, instruction-based early education programs were, by age 23, more than twice as likely to have arrest records, less likely to be married and suffering from various types of emotional impairment compared to their peers who attended play-based preschool.
Measuring the wrong thing isn’t benign – it results in problems like gaming the system and high stress among others. LZ


These behaviors (pressing academic work on young children) are a direct result of measuring the wrong thing (test scores). If we measured the right things (social development, curiosity, empathy, imagination and confidence), we would engage in a whole different set of education behaviors (play, socialization, arts programs, open-ended discovery). After nearly 20 years of reading, observing, teaching and presiding over a school, I'm convinced that this simple statement -- "Measure the wrong things and you'll get the wrong behaviors" -- is at the root of what ails education, from cradle to grave. Measuring the wrong thing (standardized scores of 4th graders) drives the wrong behaviors (lots of test prep and dull direct instruction). In later school years, measuring the wrong thing (SAT and other standardized test scores, grade point averages, class rank) continues to invite the wrong behaviors (gaming the system, too much unnecessary homework, suppression of curiosity, risk-aversion, high stress).

We should measure things that aren’t as profitable but more beneficial like creativity. LZ


Measuring the right things is more complicated and less profitable. But if we measured, even if only in our hearts, the things that we should truly value (creativity, joy, physical and emotional health, self-confidence, humor, compassion, integrity, originality, skepticism, critical capacities), we would engage in a very different set of behaviors (reading for pleasure, boisterous discussions, group projects, painting, discovery, daydreaming, recess, music, cooperation rather than competition).
Standardized testing takes time away from other subjects. ABH


NCLB put test-based accountability into federal law, subsequently solidifying the state standardized test as the sole benchmark through which all schools are measured. The consequences of this cannot be underestimated. As the annual progress of schools is judged by single standardized tests in reading and mathematics, the panic created by such a policy has had a snowball effect of emphasizing passing the test over the general quality of the school experience: the more emphasis placed on test scores, the less emphasis placed on the general school experience. Once tests have such high stakes attached to them, instructional time is supplanted by test preparation resulting in a shortened and weakened classroom experience. In a thoughtful editorial from a few years back, curricularist Peter Hlebowitsh (2007) noted upon this damaging effect that high stakes standardized testing can have on the school curriculum: We have known for years that school experiences in high-stakes-testing environments generally reduce themselves to what is being tested. The effect is that art, music, and such skills sets as critical thinking, creativity, cooperative behavior, and many others get short shrift in the classroom, primarily because such matters typically have little or no place on the exams. (p. 28).

**Standardized testing causes schools to “teach to the test”, meaning that teachers will spend more time focusing on reading and mathematics, but not other critical subjects such as art and music, because those subjects aren’t found on standardized tests.**
Testing Prevents Students From Learning Life Skills. ABH


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.

Even if PRO Shows that standardized testing increases academic performance, CON can argue that it prevents students from growing as balanced individuals. Teachers are less likely to be able to teach social skills or help students develop emotionally, because they have to prepare students for a standardized exam.

The JRS Office of Policy Research compiles research and policy for both policymakers and teachers.

Research shows, for example, that GED recipients perform about as well as high school graduates on standardized tests but have much worse life outcomes because they often lack important qualities such as curiosity, conscientiousness, perseverance, and sociability. ELA teachers could help more students develop these so-called “soft skills” or non-cognitive abilities if they didn’t have to focus on drills for tests. Instead, ELA teachers have to, for instance, cut back on large-scale projects that require perseverance, reduce the number of literary texts that engender the empathy necessary to sociability, and limit opportunities for developing student curiosity. Student learning that could lead in positive directions is diminished when tests prevent teachers from helping students develop the noncognitive abilities that support better life outcomes.

Another limitation on student learning results from the negative perceptions standardized tests can give to students about themselves and their own abilities. Studies show that elementary school students can begin to lose their sense of themselves as capable, able to do well in school and graduate, when they see unknown adults as controlling the administration and consequences of the standardized tests they are required to take. Even the very best ELA teachers have difficulty fostering learning in students who do not believe in their own abilities.

Student learning is also limited by testing’s inflexible sorting of students into categories of proficient or not proficient. It can be very difficult for students designated as not-proficient to imagine themselves as effective readers and writers. This test-generated binary is troubling because it gives no space to the full range of features that comprise effective reading and writing. Students who have literacy abilities that extend beyond but do not fully encompass the narrow band of skills measured by standardized tests may not understand or appreciate their own capacities and become disengaged from school.

Even if Pro teams can demonstrate that standardized testing is improving education, the main quantitative metric to measure this is standardized tests themselves, which don’t measure a wide enough performance band to adequately give a snapshot of performance trends.
**Teachers focus less on social studies and science. ABH**


HOW DOES STANDARDIZED TESTING AFFECT CURRICULUM AND INSTRUCTION IN ELEMENTARY CLASSROOMS? Curricular and instructional changes have also occurred as a result of high stakes testing. More than 82% of the mentor teachers “to a thorough extent” make sure that test objectives are covered in their instruction as well as adjusting instructional plans based on students’ most recent test scores. Significant instructional time is spent regularly throughout the year on test preparation activities in 68% of the classrooms. In more than 60% of the classrooms, students are given worksheets that review expected test content, and students are given practice in the kinds of item formats that are on the test. 78% of the teachers indicate that test preparation activities have increased over the past three years while remaining the same in 22% of the classrooms. No teacher indicates that test preparation has decreased or indicates “don’t know.”

**Curriculum changes to improve student test scores has increased/occurred in 89% of the classrooms** and remains the same in 10%. In contrast are the significant decreases in the attention given to social studies and science? **More than 66% of the teachers indicate that attention to science has decreased and 60% indicate that social studies instruction has decreased.**

This is a good statistic CON can use to quantify how teachers are changing their focus, and “teaching to the test”.
Testing decreases students’ motivation to learn. ABH


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.

Standardized testing inherently hurts learning because students are focused on the extrinsic motivation of a test score. Adding importance to these scores takes away from intrinsic motivations to learn throughout life.
Effective assessment requires room for self-assessment. ABH


Evaluation is the process of using qualitative and quantitative data to arrive at a value judgment of a child’s abilities. Assessment refers to the means whereby we get to that judgment. Viewing assessment as a means to an end leads us to examine more closely the daily interactions and processes that children go through as they learn. **Teaching, learning, and assessment are intricately woven together in the classrooms where children grow and learn.** In order to provide quality instruction that is developmentally appropriate for children and leads to the furtherance of democratic values, teachers recognize that assessment must be an integral part of the curriculum. It is continuous and permeates every aspect of the curriculum, both for the teacher and students. **Quality assessment that informs instruction focuses on students’ strengths rather than pointing out their weaknesses.** It allows teachers to determine what students can do rather than what they cannot do; teachers thus build knowledge on a firm foundation of strengths. **Effective assessment involves self-assessment. When children are allowed and encouraged to self-assess, they begin to understand why they are doing what they are doing. They have a sense of their own success and growth, which leads to empowerment and greater risk taking—the very values we wish schools to foster.** Finally, effective assessment involves active collaboration among teachers, children, and parents. All work together for the good of the child.

**CON can use this to show that standardized tests are harmful because they do not allow for self-assessment. Self-assessment is key because that is how students are able to learn from their mistakes and see their development over time.**
Testing cuts out science from curriculums. ELT


But while Common Core is supposed to ensure all kids acquire a broad base of knowledge, in New York, it appears to be doing the opposite at some schools. Because students need significantly better math and reading skills to pass the new tests, even higher-income schools are turning to test prep. Many schools are focusing more on math and reading than ever before.

“I have usually two to three hours of math during my school day,” says Aidan Fryling, a 6th grader in public school on Long Island. “We barely ever do any science.”

Aidan says this is a change from the way things were before Common Core. And it’s not just what they’re learning, it’s how they’re learning it.

“I just remember doing a lot more projects before Common Core. Like, a lot more hands-on projects. And now it’s a lot more sheets, packets.”

He says he and his classmates take a lot of practice tests.
Standardized Testing decreases success in college. ELT

http://www.washingtonmonthly.com/college_guide/blog/do_standardized_test_make_us_l.php

Here’s an interesting finding. You know those bubble tests public school teachers hate? The ones that are supposed to measure learning and improve education quality? They may, effectively, do the opposite.

Standardized tests now widespread in America’s public elementary and secondary schools may be hindering students’ ability to succeed in college, according to a new paper. According to a report issued by the Gordon Commission on the Future of Assessment in Education and presented at the annual meeting of the National Academy of Education:

Over the past two decades, our country has been trying to build a standards-based accountability system as a foundation for a more equitable and higher-achieving education system. In practice, however, we have created a test-based accountability system that does not reflect the standards we aimed for at the beginning of the 1990s, much less today’s fewer, clearer, higher Common Core Standards.

Several studies, using several different methodologies, have shown that the state tests do not measure the higher-order thinking, problem solving, and creativity needed for students to succeed in the 21st century. These tests, with only a few exceptions, systematically overrepresent basic skills and knowledge and omit the complex knowledge and reasoning we are seeking for college and career readiness.

Standardized testing incentivizes bad education reform. ELT

http://www.washingtonmonthly.com/college_guide/blog/do_standardized_test_make_us_l.php

But it’s not just that the tests don’t measure or facilitate the acquisition of the skills needed for success after high school. As the report explains “the misrepresentation of standards by most current accountability tests has had negative effects on teaching and learning, especially for poor and minority students.” The tests drive changes in education, for sure, but they’re likely driving the wrong changes.
Ensures we lose the best teachers – good models support teacher autonomy. ELT


Testing already overtested children to collect some “hard” data to evaluate their teachers is yet another abominable idea that American politicians have picked out of the sky. While it has about as much chance as a passing cloud of improving teacher quality, it’s quite likely to worsen instruction for children and drive our best teachers out of the field.

High-performing nations -- like Finland, Japan and Singapore -- have more cultural differences than commonalities. When you consider those commonalities you have to wonder: Is the United States still located on planet Earth? In the nations that rank highest on the Programme for International Student Assessment, teachers are: respected professionals; trusted by school administrators; given the autonomy to make curricular decisions; planning and assessing in blocks of time built into their weekly school schedule; and encouraged, supported and expected to collaborate to improve instruction.
Doesn’t Measure

*Common core standards don’t actually measure learning. LZ*


And what are these standardized tests actually testing anyway? In Pennsylvania, policymakers adopted a variation of the Common Core standards, changed the PSSA tests in math and English language arts, and then changed the scores that show whether a student is "proficient." This all happened before schools and teachers had time to adjust their curriculum. This headlong rush to change tests is why fewer Pennsylvania students scored "proficient" on the PSSAs this year. Those declines have absolutely nothing to do with how well our kids and educators are doing. So, if they don't really show how well our kids are doing, what is the point of these tests? I'm still not sure, and neither are many of the educators and parents with whom I've talked.
Because of the limitations of standardized testing, it makes a poor measurement of actual learning. LZ


A Standardized Test's Assessment Mission The folks who create standardized achievement tests are terrifically talented. What they are trying to do is to create assessment tools that permit someone to make a valid inference about the knowledge and/or skills that a given student possesses in a particular content area. More precisely, that inference is to be norm-referenced so that a student's relative knowledge and/or skills can be compared with those possessed by a national sample of students of the same age or grade level. Such relative inferences about a student's status with respect to the mastery of knowledge and/or skills in a particular subject area can be quite informative to parents and educators. For example, think about the parents who discover that their 4th grade child is performing really well in language arts (94th percentile) and mathematics (89th percentile), but rather poorly in science (39th percentile) and social studies (26th percentile). Such information, because it illuminates a child's strengths and weaknesses, can be helpful not only in dealing with their child's teacher, but also in determining at-home assistance. Similarly, if teachers know how their students compare with other students nationwide, they can use this information to devise appropriate classroom instruction. But there's an enormous amount of knowledge and/or skills that children at any grade level are likely to know. The substantial size of the content domain that a standardized achievement test is supposed to represent poses genuine difficulties for the developers of such tests. If a test actually covered all the knowledge and skills in the domain, it would be far too long. So standardized achievement tests often need to accomplish their measurement mission with a much smaller collection of test items than might otherwise be employed if testing time were not an issue. The way out of this assessment bind is for standardized achievement tests to sample the knowledge and/or skills in the content domain. Frequently, such tests try to do their assessment job with only 40 to 50 items in a subject field—sometimes fewer. Accurate Differentiation As a Deity The task for those developing standardized achievement tests is to create an assessment instrument that, with a handful of items, yields valid norm-referenced interpretations of a student's status regarding a substantial chunk of content. Items that do the best job of discriminating among students are those answered correctly by roughly half the students. Developers avoid items that are answered correctly by too many or by too few students. As a consequence of carefully sampling content and concentrating on items that discriminate optimally among students, these test creators have produced assessment tools that do a great job of providing relative comparisons of a student's content mastery with that of students nationwide. Assuming that the national norm group is genuinely representative of the nation at large, then educators and parents can make useful inferences about students. One of the most useful of those inferences typically deals with students' relative strengths and weaknesses across subject areas, such as when parents find that their daughter sparkles in mathematics but sinks in science. It's also possible to identify students' relative strengths and weaknesses within a given subject area if there are enough test items to
do so. For instance, if a 45-item standardized test in mathematics allocates 15 items to basic computation, 15 items to geometry, and 15 items to algebra, it might be possible to get a rough idea of a student's relative strengths and weaknesses in those three realms of mathematics. More often than not, however, these tests contain too few items to allow meaningful within-subject comparisons of students' strengths and weaknesses. A second kind of useful inference that can be based on standardized achievement tests involves a student's growth over time in different subject areas. For example, let's say that a child is given a standardized achievement test every third year. We see that the child's percentile performances in most subjects are relatively similar at each testing, but that the child's percentiles in mathematics appear to drop dramatically at each subsequent testing. That's useful information. Unfortunately, both parents and educators often ascribe far too much precision and accuracy to students' scores on standardized achievement tests. Several factors might cause scores to flop about. Merely because these test scores are reported in numbers (sometimes even with decimals!) should not incline anyone to attribute unwarranted precision to them. Standardized achievement test scores should be regarded as rough approximations of a student's status with respect to the content domain represented by the test. To sum up, standardized achievement tests do a wonderful job of supplying the evidence needed to make norm-referenced interpretations of students' knowledge and/or skills in relationship to those of students nationally. The educational usefulness of those interpretations is considerable. Given the size of the content domains to be represented and the limited number of items that the test developers have at their disposal, standardized achievement tests are really quite remarkable. They do what they are supposed to do. But standardized achievement tests should not be used to evaluate the quality of education. That's not what they are supposed to do.
Trying to measure quality of education with standardized testing is like measuring temperature with a tablespoon. LZ


For several important reasons, standardized achievement tests should not be used to judge the quality of education. The overarching reason that students' scores on these tests do not provide an accurate index of educational effectiveness is that any inference about educational quality made on the basis of students' standardized achievement test performances is apt to be invalid. Employing standardized achievement tests to ascertain educational quality is like measuring temperature with a tablespoon. Tablespoons have a different measurement mission than indicating how hot or cold something is. Standardized achievement tests have a different measurement mission than indicating how good or bad a school is. Standardized achievement tests should be used to make the comparative interpretations that they were intended to provide. They should not be used to judge educational quality. Let's look at three significant reasons that it is thoroughly invalid to base inferences about the caliber of education on standardized achievement test scores.

Because large corporations own standardized tests, there is no real standardization among different educational objectives. LZ


The companies that create and sell standardized achievement tests are all owned by large corporations. Like all for-profit businesses, these corporations attempt to produce revenue for their shareholders. Recognizing the substantial pressure to sell standardized achievement tests, those who market such tests encounter a difficult dilemma that arises from the considerable curricular diversity in the United States. Because different states often choose somewhat different educational objectives (or, to be fashionable, different content standards), the need exists to build standardized achievement tests that are properly aligned with educators' meaningfully different curricular preferences. The problem becomes even more exacerbated in states where different counties or school districts can exercise more localized curricular decision making.
One size fits all just doesn’t work – it can’t align with what is being emphasized in a particular classroom. LZ


At a very general level, the goals that educators pursue in different settings are reasonably similar. For instance, you can be sure that all schools will give attention to language arts, mathematics, and so on. But that's at a general level. At the level where it really makes a difference to instruction—in the classroom—there are significant differences in the educational objectives being sought. And that presents a problem to those who must sell standardized achievement tests. In view of the nation's substantial curricular diversity, test developers are obliged to create a series of one-size-fits-all assessments. But, as most of us know from attempting to wear one-size-fits-all garments, sometimes one size really can't fit all. The designers of these tests do the best job they can in selecting test items that are likely to measure all of a content area's knowledge and skills that the nation's educators regard as important. But the test developers can't really pull it off. Thus, standardized achievement tests will always contain many items that are not aligned with what's emphasized instructionally in a particular setting.

Robust studies prove – tests do not cover what is actually taught. LZ


To illustrate the seriousness of the mismatch that can occur between what's taught locally and what's tested through standardized achievement tests, educators ought to know about an important study at Michigan State University reported in 1983 by Freeman and his colleagues. These researchers selected five nationally standardized achievement tests in mathematics and studied their content for grades 4–6. Then, operating on the very reasonable assumption that what goes on instructionally in classrooms is often influenced by what's contained in the textbooks that children use, they also studied four widely used textbooks for grades 4-6. Employing rigorous review procedures, the researchers identified the items in the standardized achievement test that had not received meaningful instructional attention in the textbooks. They concluded that between 50 and 80 percent of what was measured on the tests was not suitably addressed in the textbooks. As the Michigan State researchers put it, "The proportion of topics presented on a standardized test that received more than cursory treatment in each textbook was never higher than 50 percent"
Mismatches are very rarely recognized, but they definitely exist. LZ


Well, if the content of standardized tests is not satisfactorily addressed in widely used textbooks, isn't it likely that in a particular educational setting, topics will be covered on the test that aren't addressed instructionally in that setting? Unfortunately, because most educators are not genuinely familiar with the ingredients of standardized achievement tests, they often assume that if a standardized achievement test asserts that it is assessing "children's reading comprehension capabilities," then it's likely that the test meshes with the way reading is being taught locally. More often than not, the assumed match between what's tested and what's taught is not warranted. If you spend much time with the descriptive materials presented in the manuals accompanying standardized achievement tests, you'll find that the descriptors for what's tested are often fairly general. Those descriptors need to be general to make the tests acceptable to a nation of educators whose curricular preferences vary. But such general descriptions of what's tested often permit assumptions of teaching-testing alignments that are way off the mark. And such mismatches, recognized or not, will often lead to spurious conclusions about the effectiveness of education in a given setting if students' scores on standardized achievement tests are used as the indicator of educational effectiveness. And that's the first reason that standardized achievement tests should not be used to determine the effectiveness of a state, a district, a school, or a teacher. There's almost certain to be a significant mismatch between what's taught and what's tested.
Standardized tests cannot allow for meaningful comparisons among students because of limited collections of comparisons. LZ


A second reason that standardized achievement tests should not be used to evaluate educational quality arises directly from the requirement that these tests permit meaningful comparisons among students from only a small collection of items. A test item that does the best job in spreading out students' total-test scores is a test item that's answered correctly by about half the students. Items that are answered correctly by 40 to 60 percent of the students do a solid job in spreading out the total scores of test-takers. Items that are answered correctly by very large numbers of students, in contrast, do not make a suitable contribution to spreading out students' test scores. A test item answered correctly by 90 percent of the test-takers is, from the perspective of a test's efficiency in providing comparative interpretations, being answered correctly by too many students. Test items answered correctly by 80 percent or more of the test takers, therefore, usually don't make it past the final cut when a standardized achievement test is first developed, and such items will most likely be jettisoned when the test is revised. As a result, the vast majority of the items on standardized achievement tests are "middle difficulty" items. As a consequence of the quest for score variance in a standardized achievement test, items on which students perform well are often excluded. However, items on which students perform well often cover the content that, because of its importance, teachers stress. Thus, the better the job that teachers do in teaching important knowledge and/or skills, the less likely it is that there will be items on a standardized achievement test measuring such knowledge and/or skills. To evaluate teachers' instructional effectiveness by using assessment tools that deliberately avoid important content is fundamentally foolish.
Standardized tests suffer from confounding factors. LZ


The third reason that students' performances on these tests should not be used to evaluate educational quality is the most compelling. Because student performances on standardized achievement tests are heavily influenced by three causative factors, only one of which is linked to instructional quality, asserting that low or high test scores are caused by the quality of instruction is illogical. To understand this confounded-causation problem clearly, let's look at the kinds of test items that appear on standardized achievement tests. Remember, students' test scores are based on how well students do on the test's items. To get a really solid idea of what's in standardized tests, you need to grub around with the items themselves. The three illustrative items presented here are mildly massaged versions of actual test items in current standardized achievement tests. I've modified the items' content slightly, without altering the essence of what the items are trying to measure. The problem of confounded causation involves three factors that contribute to students' scores on standardized achievement tests: (1) what's taught in school, (2) a student's native intellectual ability, and (3) a student's out-of-school learning.
What should educators do instead? They should learn more about standardized tests, carry out an educational campaign against standardized tests, and arrange better assessments. LZ


I've described a situation that, from the perspective of an educator, looks pretty bleak. What, if anything, can be done? I suggest a three-pronged attack on the problem. First, I think that you need to learn more about the viscera of standardized achievement tests. Second, I think that you need to carry out an effective educational campaign so that your educational colleagues, parents of children in school, and educational policymakers understand what the evaluative shortcomings of standardized achievement tests really are. Finally, I think that you need to arrange a more appropriate form of assessment-based evidence. Learning about standardized achievement tests. Far too many educators haven't really studied the items on standardized achievement tests since the time that they were, as students, obliged to respond to those items. But the inferences made on the basis of students' test performances rest on nothing more than an aggregated sum of students' item-by-item responses. What educators need to do is to spend some quality time with standardized achievement tests—scrutinizing the test's items one at a time to see what they are really measuring. Spreading the word. Most educators, and almost all parents and school board members, think that schools should be rated on the basis of their students' scores on standardized achievement tests. Those people need to be educated. It is the responsibility of all educators to do that educating. If you do try to explain to the public, to parents, or to policymakers why standardized test scores will probably provide a misleading picture of educational quality, be sure to indicate that you're not running away from the need to be held accountable. No, you must be willing to identify other, more credible evidence of student achievement. Coming up with other evidence. If you're going to argue against standardized achievement tests as a source of educational evidence for determining school quality, and you still are willing to be held educationally accountable, then you'll need to ante up some other form of evidence to show the world that you really are doing a good educational job. I recommend that you attempt to assess students' mastery of genuinely significant cognitive skills, such as their ability to write effective compositions, their ability to use lessons from history to make cogent analyses of current problems, and their ability to solve high-level mathematical problems. If the skills selected measure really important cognitive outcomes, are seen by parents and policymakers to be genuinely significant, and can be addressed instructionally by competent teachers, then the assembly of a set of pre-test-to-post-test evidence showing substantial student growth in such skills can be truly persuasive. What teachers need are assessment instruments that measure worthwhile skills or significant bodies of knowledge. Then teachers need to show the world that they can instruct children so that those children make striking pre-instruction to post-instruction progress. The fundamental point is this: If educators accept the position that standardized achievement test scores should not be used to measure the quality of schooling, then they must provide other, credible evidence that can be used to ascertain the quality of schooling. Carefully collected, nonpartisan evidence regarding teachers' pre-test-to-post-test promotion of undeniably important skills or knowledge just might do the trick.
Standardized testing just can't measure learning – especially true when public education is a market driven activity. LZ


Measure the wrong things and you'll get the wrong behaviors. This simple statement succinctly characterizes why the American education system continues beating its head against the wall. Education reformers and so-called policy "experts" are constantly collecting and analyzing data. Many of these experts are, not surprisingly, economists. It's not for nothing that economics is sometimes called "the dismal science." The hostile takeover of education by non-educators is filled with intelligent sounding phrases: "evidence-based," "data driven," "metrics and accountability." At every level of schooling, mountains of data are collected to inform "best practices" based on the alleged cause and effect implications of data-based instruction and the feedback gleaned from tests. It is not coincidental that the education policy and reform business is highly profitable. Public education is estimated to be a $600-700 billion market. Those who drive the measuring and testing industry are first in line at the trough. Pearson Publishing, for example, has its greedy tentacles in nearly every school district in America. All the iterations of reform -- No Child Left Behind, Race to the Top and, more recently, the Common Core -- are driven by (and driving) the collection and interpretation of data. Throughout education, an increasingly rigid, closed loop of assessment is systematically making schools worse: Define things children should know or be able to do at a certain age; design a curriculum to instruct them in what you've decided they should know; set benchmarks; develop tests to see if they have learned what you initially defined; rinse and repeat.
There are systemic errors within standardized testing. ABH


Finally, all of these concerns should be viewed in the context of the testing industry today. Lines (2000) observed that errors are more likely in testing programs with greater degrees of centralization and commercialization, where increased profits can only be realized by increasing market share, “The few producers cannot compete on price, because any price fall will be instantly matched by others …. What competition there is comes through marketing” (p. 1). In Minnesota, Judge Oleisky (Kurvers et al. v. NCS, Inc., 2002) observed that Basic Skills Test errors were caused by NCS’ drive to cut costs and raise profits by delivering substandard Errors in Standardized Tests: A Systemic Problem NBETPP monographs 29 service—demonstrating that profits may be increased through methods other than marketing. With the recent passage of President Bush’s No Child Left Behind Act (NCLB)31, the testing industry in the US will become increasingly more centralized and more commercialized. An amplified demand for testing services without an appreciable increase in the number of service providers in the short term will intensify time pressures already experienced by the contractors. At the same time NCLB will heighten the reliance of state DOEs on the few contractors available, creating a situation whereby those who pay for a service become increasingly dependent on one that is more prone to error. Coupling these conditions with the lack of industry oversight creates conditions for a future that is ripe for the proliferation of undetected human error in educational testing.

The limited number of testing contractors coupled with the massive demand for testing will only increase the likelihood of oversight and human error that cause grading mistakes.
Standardized Testing errors have severe negative consequences. ABH


This paper contains a sizable collection of testing errors made in the last twenty-five years. It thus offers testimony to counter the implausible demands of educational policy makers for a single, error-free, accurate, and valid test used with large groups of children for purposes of sorting, selection, and trend-tracking. No company can offer flawless products. Even highly reputable testing contractors that offer customers high-quality products and services produce tests that are susceptible to error. But while a patient dissatisfied with a diagnosis or treatment may seek a second or third opinion, for a child in a New York City school (and in dozens of other states and hundreds of other cities and towns), there is only one opinion that counts – a single test score. If that is in error, a long time may elapse before the mistake is brought to light – if it ever is. This paper has shown that human error can be, and often is, present in all phases of the testing process. Error can creep into the development of items. It can be made in the setting of a passing score. It can occur in the establishment of norming groups, and it is sometimes found in the scoring of questions.

Standardized testing puts so much emphasis on a single test score, that any grading errors will have major consequences for the educational future of those involved. Even if the amount of errors are minimal, their magnitude cannot be overlooked.
Testing has negligible effect on student achievement – comparisons with other countries fail.

ELT


Conclusion 1: Test-based incentive programs, as designed and implemented in the programs that have been carefully studied, have not increased student achievement enough to bring the United States close to the levels of the highest achieving countries. When evaluated using relevant low-stakes tests, which are less likely to be inflated by the incentives themselves, the overall effects on achievement tend to be small and are effectively zero for a number of programs. Even when evaluated using the tests attached to the incentives, a number of programs show only small effects. Programs in foreign countries that show larger effects are not clearly applicable in the U.S. context. School-level incentives like those of the No Child Left Behind Act produce some of the larger estimates of achievement effects, with effect sizes around 0.08 standard deviations, but the measured effects to date tend to be concentrated in elementary grade mathematics and the effects are small compared to the improvements the nation hopes to achieve. (pg 85)
Teaching to the test makes results useless. ELT


To this point, we have discussed problems with tests as accountability measures even when best practices are followed. In addition, now that tests are being widely used for high-stakes accountability, inappropriate forms of test preparation are becoming more widespread and problematic (Hamilton et al., 2007). Test results may become increasingly misleading as measures of achievement in a domain when instruction is focused too narrowly on the specific knowledge, skills, and test question formats that are likely to appear on the test. Overly narrow instruction might include such practices as drilling students on practice questions that were released from prior years’ tests, focusing on the limited subset of skills, knowledge and question formats that are most likely to be tested, teaching test-taking tricks (such as the process of elimination for multiple-choice items or memorizing the two “common Pythagorean ratios” rather than learning the Pythagorean theorem), or teaching students to answer open-ended questions according to the test’s scoring rubric. When scores increase on a test for which students have been “prepared” in these ways, it indicates only that students have learned to correctly answer the specific kinds of questions that are included on that particular test. It does not indicate that that students have also attained greater mastery of the broader domain that the test is intended to represent (Koretz, 2002).
Standardized tests measure relative achievement but not actual learning. ELT


Have you noticed that the first year a standardized test is administered the lowest results are gained? Then, each successive year the scores are higher because the teachers and administrators re-align their lessons and curricula to better match the test objectives. This situation appears to illustrate that the policymakers’s reform was effective. Unfortunately, the scores usually plateau, though that won’t necessarily be the policymakers’s concern because it is likely that he or she will no longer be serving in public office. Are the students better prepared through this reform? Probably not. W. James Popham (2002), a professor emeritus in the UCLA Graduate School of Education, expresses his opinion that “it is wrong to evaluate a school by how well its students perform on standardized achievement tests” (¶ 1). He contends that standardized tests are constructed from the Army Alpha test, which was used during World War I to identify possible officer-training candidates. The Alpha test was designed to measure the aptitude of the candidates and to identify those candidates who are superior intellectually within the group of test-takers. The Alpha test rank-ordered the test-takers based on their aptitude. Currently, tests in the public schools are intended to measure student achievement, not aptitude. However, because the current standardized tests are based on the Alpha test model, they actually measure achievement as a student’s relative standing within the test group and not the acquired knowledge or actual preparation of a student.
Standardized testing decreases the quality of education over the year. ELT


Several studies, using several different methodologies, have shown that the state tests do not measure the higher-order thinking, problem solving, and creativity needed for students to succeed in the 21st century. These tests, with only a few exceptions, systematically overrepresent basic skills and knowledge and omit the complex knowledge and reasoning we are seeking for college and career readiness. The misrepresentation of standards by most current accountability tests has had negative effects on teaching and learning, especially for poor and minority students. The tests carry consequences, and many educators serving poor students aim to raise test scores in the most direct — in some cases, the only — way they know: They provide practice on exercises that substantially match the format and content of their state’s end-of-year accountability tests. These exercises often depart substantially from best instructional practice. Some studies have documented a systematic decline from fall to spring in the quality of instruction. In reading, for example, the complexity of texts that students engage with is lower — in the same classrooms, with the same children — in March than in October. And there is less discussion of text and word meaning as teachers direct children through workbook exercises that mimic state test items (Anagnostopoulos, 2003; Koretz & Hamilton, 2006; McNeill, 2002). Principals and district administrators encourage this practice. They introduce interim assessments that largely mirror the end-of-year tests rather than model the kinds of performance intended by the standards. They do this because the tests count, and they are afraid that without practice, students will not do well enough to meet adequate yearly progress (AYP) requirements.
Unfair

Standardized testing is racially unfair. LZ


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.

Standardized testing creates self-fulfilling prophecies that disadvantage racial minorities. LZ


This is not the only instance where race has become an important factor for how standardized tests are used in public education. Just last month public schools in both Virginia and Washington D.C. announced targets for how many students in each racial group must pass for schools to remain in good standing. For example, in Virginia only 45% of black students in each school must pass standardized math tests while 68% of whites, and 82% of Asians must do the same. Officials say that these plans are not discriminatory because students who are the farthest behind must progress the most, but critics reason that if one expects less from some students, those lower educational expectations will become a self-fulfilling prophecy for school districts and those students will fall even farther behind.
Research clearly demonstrates that tests unfairly advantage certain racial groups – only ending reliance on standardized testing solves. LZ


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.

The standardized testing gap is huge between the poor and the wealthy. LZ


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.
The history of standardized testing is racially biased. LZ


One thing all this research has shown us is that the issue lies with how we use these tests, not with the kids who take them. Just consider the history of standardized tests which — according to Columbia University Professor Nicholas Lemann’s history of the Educational Testing Service, The Big Test: The Secret History of America’s Meritocracy — were first developed in the 1940s as a way to exclude Jewish students from Ivy League campuses. Interestingly, Stanley Kaplan, today one of the largest test preparation organizations, got its start when Mr. Kaplan resolved to come up with test-taking techniques to “beat the test” and ensure that such students did well.

Colleges are already acknowledging the harms of testing, but K-12 still uses them and they are pervasive in the system. LZ


Today, as an acknowledgment of the inherent racial and economic inequity of standardized achievement tests, hundreds of colleges have already stopped requiring the SAT for college admission decisions. However, the same cannot be said for K-12, where scores on achievement tests are in part used for everything from admitting students to prestigious public schools to placing students in gifted or remedial programs, allocating federal funding, and even evaluating teachers. A growing number of parents, school boards, teachers and civil rights organizations are beginning to question the fairness of our overreliance on standardized tests and recently over 300 groups, including the NAACP Legal Defense Fund signed a petition to ask congress to ban the use of such tests. Given the recent developments in Washington D.C. and Virginia, it would seem that it’s about time.
Standardized testing is definitely racially biased. LZ


A boycott of the Northwest Evaluation Association's (NWEA) Measures of Academic Progress (MAP) test by teachers and students at Garfield High School in Seattle has reignited a movement against standardized testing in U.S. schools and debate over tests' inherent cultural bias. In an appeal of the Seattle Public School Board's 2010 decision to renew its contract with the NWEA, members of a parent group alleged "that the MAP test disadvantages non-English speaking students, special education students, minority and low-income children." Fair education reform advocates have long cited a litany of concerns about standardized testing, many of which address racial bias and discrimination, as author and researcher Harold Berlak explains in the journal Rethinking Education: Standardized testing perpetuates institutionalized racism and contributes to the achievement gap between whites and minorities. For instance, the deeply embedded stereotype that African Americans perform poorly on standardized tests hinders many African Americans' testing ability. Also, research has shown that minorities statistically have lower standardized test scores than whites because of existing, hidden biases in the development and administration of standardized tests and interpretation of their scores. Therefore, the achievement gap will not begin to close until current standards and assessment tests are significantly reformed.

The schools affected by standardized testing are mostly minority schools. LZ


In a month where (some of) America celebrates black history and the work of individuals who fought tirelessly for equality, the current uproar over standardized testing highlights just how much we have yet to overcome. Yesterday, Chicago Public Schools announced a preliminary hit list of 129 schools that are in danger of being closed by the start of the 2013-2014 school year. For some of these schools, closings are tied to test scores, and the majority of them are in African-American and Latino neighborhoods on the South and West sides of Chicago -- neighborhoods already besieged by crime, poverty and foreclosure, and subjected to significant population loss in the wake of the destruction of public housing; neighborhoods where children often lack health care and proper nutrition, and do not have access to test prep, books or other informal learning at home; and neighborhoods where parents have limited education and families are constantly stretched economically.

Education equity should be the norm, but from the makeup of standardized tests to the circumstances surrounding the lives of the students taking them, this equity remains elusive. Fiction-wise, it didn't exist on "Good Times" in 1974 or when Diff'rent Strokes presented the same theme four years later, and it isn't a reality for some minority students today. If schools are to test -- and reasoning for that alone is debatable -- then districts cannot expect fair evaluation when circumstances are different for each child. For many children, the burden of life alone is often so great that their primary goal is not a quality education, but day-to-day survival. And culturally, when academic outcomes are averaged across race and class, the achievement gap grows even greater. Society must first address these circumstances, and only then can there be real education reform. As we learned later in "The IQ Test," "It's hard to get the right answers when you don't understand the questions."
Standardized tests are designed for white, middle class kids. LZ


There was an interesting op-ed in The New York Times yesterday about how to fix standardized testing. The writer argues that the problem isn’t that the tests are given. The problem is the questions that are asked. “The problem is that the reading passages used in these tests are random,” says the writer, E. D. Hirsch Jr. “Children are asked to read and then answer multiple-choice questions about such topics as taking a hike in the Appalachians even though they’ve never left the sidewalks of New York, nor studied the Appalachians in school.” Hirsch goes on to explain that, according to a study, reading comprehension tests don’t so much evaluate reading comprehension skills. Rather, they are a measure of the student’s pre-existing knowledge of the subject matter. “The key to comprehension is familiarity with the relevant subject. For a student with a basic ability to decode print, a reading-comprehension test is not chiefly a test of formal techniques but a test of background knowledge,” is how Hirsch puts it. The suits in their offices who run the standardized tests will swear up and down that the reading comprehension texts are selected at random. As a veteran of public schools in New York, Connecticut, and California, I say, “Uhm, are you kidding me?” Reading comprehension sections tend to be about things like history, faraway places, or, yes, the Appalachian Mountains. Are these things our inner-city students would naturally know about? Of course not. Yes, things like history, geography, and the Appalachian Mountains are important. But that’s why we have sections that test this knowledge. The texts in reading comprehension should either be truly random across cultures, class, religions, and backgrounds, or they should be community specific. In New York, for example, this would mean test texts that focus on things like reggeton, Chris Brown, the Yankees, what’s the hottest show on BET, the cooking of popular Puerto Rican meals, what happens in Prospect Park at night, and the elements of MySpace. In other words, if what we’re trying to do is test reading comprehension, we can’t also be confusing our students about subjects they might never have encountered before. Right off the bat, reading the long, strange name of some random mountain range tells our students, “No, you don’t know about this.” And, “Yes, this test was not made with you in mind.” I say that these tests are racist because anyone who has administered a standardized test knows that the subject matter tends to be very educated-middle-class. And anything educated-middle-class focused is going to necessarily be biased toward whites. It’s simply how it is. Just out of curiosity, I looked up this past January’s Regents ‘Comprehension Examination.’ The topic of the two readings? Advice from a dietitian, and the ecological viability of using straw bales as an alternative building material. (Find this and other past tests here.) Now, if you don’t think a white kid from the suburbs is about one hundred times more likely to have talked about things like this in his home than the child of a Dominican immigrant in the city, you’re fooling yourself. These tests are racially biased, whether they mean to be or not.
Factors outside of the classroom impact test scores. ABH


Corporate reform groups such as Advance Illinois and Stand for Children claim teachers are the main factor in student academic achievement. Recent research shows otherwise: as much as 90\% of variation in student growth is explained by factors outside the control of teachers. Children who do not have access to health care, who are hungry, who are exhausted from night-time symptoms of asthma, who are fearful of violence in their communities, who do not have books or access to other informal learning at home, whose parents have limited education, whose families are constantly stressed by economic problems, and who do not go to libraries and museums in their free time are at an academic disadvantage. These factors are highly related not only to testing outcomes, academic achievement, future education and socio-economic success, but also to the racial, ethnic and class origins of individuals. The inequitable history of American society, politics, institutions and economic relations are at the root of these relations. As a result, when academic outcomes are average across subgroups such as race and class, glaring gaps appear.

Primarily economic factors can contribute to poor test grades, which in turn can limit opportunities for underprivileged students. Standardized testing reveals inequalities among different economic and racial groups, but wrongly assumes that such issues can be solved within a classroom.
Testing disproportionately targets urban schools. DAT


The Center for American Progress is an independent nonpartisan educational institute

District-level testing occurs more frequently and takes up more learning time in urban districts than in suburban districts. We examined how district-level testing compared across urban and suburban districts and found some substantial differences. Urban high school students, in particular, spend more time taking district-level exams than suburban high school students. Urban high school students take three times as many district-level tests and spend up to 266 percent more time taking them compared with suburban high school students.

A few districts, such as the suburban South-Western City School District in Ohio, do not have any required assessments for high school students. At the other end of the spectrum, high school students in the urban district of Jefferson County, Kentucky, are tested approximately, on average, 13 times throughout the year, and high school students in Denver Public Schools spend an average of nearly 17 hours taking district-level tests.

In grades 3-5 and 6-8, urban-district students spend approximately 80 percent and 73 percent more time, respectively, taking district-level exams than their suburban peers. In grades K-2, urban students spend about 52 percent more time on district tests.

Regardless of the harms of standardized testing, one big problem is that the harms themselves are not actually standardized—instead, urban students (more likely to be minority affiliated than suburban ones) bear the brunt of the costs. Even if the Con cannot demonstrate substantial, crippling harms, the fact that they'
Health greatly impacts test scores. ABH


We recognize that schools must prioritize academic achievement and that in the current school funding climate, health is often perceived as secondary, at best. However, results from this study and others indicate that creative approaches that integrate curricular and noncurricular school-wide efforts to promote healthy behaviors among all students are worth the investment. Examining the odds of achieving goal or above on all 3 standardized tests for each of the individual health assets, it appears that not having a television in the bedroom, being at a healthy weight and physically fit, being food secure, and eating at fast-food restaurants 1 time or less per week are the most important predictors of academic achievement in this study. Further, children who drink less soda and other sweetened drinks, are emotionally healthy, have quality sleep, feel safe in their neighborhoods, and are also significantly more likely to achieve goal on standardized tests. But beyond each individual health asset, it appears that any and all additional health-promoting effort cumulatively impacts academic achievement. Individual targeted initiatives may be insufficient to promote change; therefore, we must advocate against diffusion of responsibility (eg, just taking soda machines out of schools won't impact health or grades, so why bother) and for a more comprehensive approach.

This is yet another piece of evidence that shows that factors such as a student’s health (which is adversely affected by a poor socioeconomic status) can impact standardized test scores, meaning that they are not representative of a student’s academic ability.
Students’ socioeconomic status influences test scores. ABH


While this was the first time students took the computerized Common Core test, the scores released last week matched an all-too-familiar historical pattern: They followed the money. They reflected entrenched economic and racial patterns despite decades spent trying to counter them and level one of the country’s most important playing fields. In Piedmont, students are mostly white and Asian, a home costs $2 million on average and a typical family earns $207,000. In East Palo Alto, predominantly Hispanic and to a lesser extent black, the average house costs $600,000 and families make $50,000. More than 70 percentage points separates the Bay Area districts with the highest scores from the districts with the lowest. The achievement gap, the results show, is not only alive and well but possibly growing.

Standardized testing is harmful because the scores are greatly skewed by factors such as race and income. Also, this example is from September 2015, in case PRO tries arguing that this issue is improving.

Standardized tests means students with disabilities miss out on life skills. ELT


In an age where "teaching to the test" has become a common term, it is important to know just what is being sacrificed in order for students with disabilities to take and pass standardized assessments. Students with disabilities are often the neediest students in schools, and therefore require the most intensive supports and resources. The passing of NCLB in 2001 marked the first time students with disability's scores were reported publicly. The focus of achievement for students with disabilities has made a dramatic shift from meeting individual goals to performing on standardized assessments. With this shift, instruction has focused increasingly on content covered by state assessments. With this shift in instructional focus, something is being sacrificed for students with disabilities. Future research should focus on what instruction is being cut out of the curriculum for students with disabilities in order to fit in more time for teaching content specifically on the assessment. McLaughlin and Thurlow (2010) wonder, "how current accountability practices that require teachers to teach specific subject matter, result on kinds of educational outcomes that are valued by students with disabilities [such as employment and independence]?
Testing fails to capture achievements of students with learning disabilities. ELT

Disability Rights Advocates [legal counsel], 2001, “Do No Harm”

Standardized tests assume that each student taking the test will read in the same manner. However, a student with a learning disability cannot process words like other students. Every time a learning disabled student reads a word, it is as if he or she is encountering that word for the first time. Reading is thus a slow and tiring process. On these tests, the student is struggling with reading the test, not answering the questions. The students’ disability, not his or her ability, is being assessed. One reason that high-stakes assessments have a discriminatory impact on students with learning disabilities is because often when the tests were developed, little no attention was given to how the tests would impact learning disabled test takers. The sample population that is used by test developers to set the average scores for the tests usually does not include students with disabilities. When disabled students are included in the sample population, it is often unintentional, and the performance of these individuals is not separately tracked. Most testing publishers also do not give students with disabilities accommodations they need when testing a sample population, thus leading to a dearth of information and research about the true effect of an accommodation on a testing situation.

The way tests are written guarantees students with disabilities will fail. ELT

Disability Rights Advocates [legal counsel], 2001, “Do No Harm”

Handwritten Tests Can Unfairly Disadvantage Students With Learning Disabilities

Several states have also required tests to be handwritten, and learning disabled students are not allowed to use computers or typewriters. Not only is it hard to understand any rationale for this requirement in this modern age of technology and computers, but many students with learning disabilities require a computer or wordprocessor as a writing tool because their disability impairs the physical act of writing. Moreover, many of these students have been provided with such technology during their education as part of their legally mandated Individualized Educational Program or Section 504 Plan, and they will be inadequately prepared if required to suddenly handwrite their examinations on high-stakes assessments.

Many Multiple Choice Tests Unfairly Impact Dyslexic Students Because they Do Not Provide Sufficient Context

Most assessment systems have multiple choice sections. Multiple choice tests generally do not provide sufficient context for learning disabled students, and dyslexic students rely heavily on context as a compensating mechanism to identify words. The disadvantage of multiple choice tests could be eliminated for many learning disabled students through use of audio tests or readers.
Standardized Testing Hurts Teachers

*Teachers lose control of the curriculum. ABH*


Exploring the relationship between teachers and state-mandated standardized testing, and using the Michigan Evaluation Assessment Program (MEAP) and the discourses surrounding it as a case study of consequences, this study both corroborates findings from previous studies on teachers and testing and adds to them. Like McNeil’s (2000) study, it points out how state-mandated standardized testing is not simply an evaluator of a curriculum but its creator. Ordering particular forms and norms of teaching and learning while limiting others, the MEAP and its discourses provide conceptual order to teachers’ perceptions, investments, and actions, marking the boundaries of the permissible and the possible in social studies education. This study also illustrates that standardized testing impacts teaching in mostly negative ways, reducing teaching to low levels of intellectual engagement and teachers to implementers of externally designed curricula and pre-packaged materials intended to help them teach to the test (Cimbricz, 2002; Haney, 2000; Kohn, 2000; McNeil, 2000). As a result, teachers speak of feeling inadequate, anxious, frustrated, alienated, and angry (Smith, 1991).

Standardized testing forces teachers to teach to a test instead of being allowed to facilitate learning in their preferred ways. This leads to teachers feeling inadequate and struggling to teach a curriculum they didn’t create.
Testing robs teachers of resources. DAT


Kathleen Jasper left her post as an assistant principal of a Florida high school in early 2014 because, she says, of her mounting frustrations with testing. "I was being forced to implement bad education policy, especially with respect to testing," she said.

Florida is one of at least 36 states, by NPR Ed's count, that require or plan to require high school end-of-course exams in an array of subjects, as a condition of graduation.

If you want a high school diploma in the Sunshine State, you must pass tests in algebra, geometry, civics and U.S. history. That's on top of the state standardized tests (the FCAT) in math and reading, and every other test on the list.

These end-of-course tests are given throughout 10th, 11th and 12th grade, and each year there is time set aside for retakes. Schools, naturally, want to give students as many chances as possible to pass the tests, because the students need them to graduate.

The result? "I watched tests take up 40 to 50 percent of the year," says Jasper, who now maintains a blog and podcast about education. "Media centers were closed for the entire month of January. Laptops, every resource was sucked up into testing."

Debbie Brockett reports the same scenario unfolding on the other side of the country. She is the principal of Las Vegas High School, a 3,000-student, predominantly Hispanic and low-income school.

Unlike college professors, who simply assign books and leave it to the students to buy them, K–12 teachers have to provide students with books. But it’s not a simple matter of ordering one book per student per subject. Based on the schools I visited and the teachers I interviewed, each student needs at least one textbook and one workbook per class, plus a bunch of worksheets and projects the teacher pulls from assorted websites (not to mention binder clips and construction paper and scissors and other project-based materials). Books can be reused year to year, but only if the state standards haven’t changed—which they have every year for at least the past decade.
Teachers lose control of both content and scope. DAT


The JRS Office of Policy Research compiles research and policy for both policymakers and teachers.

Standardized tests narrow the entire curriculum in many schools, often squeezing out subjects such as music, art, foreign languages, and, especially in elementary grades, social studies, because they are not included in tests. For ELA teachers, these tests also lead to subject-specific narrowing. ELA teachers are required to focus their instruction on the literacy skills measured on standardized tests. Since reading is more prominent than writing in most tests, teachers spend more time on reading rather than writing, usually focusing on comprehension, not higher-order critical reading skills. Even when English language arts teachers deliberately teach beyond the test-based curriculum, important aspects of writing such as revision do not get attention, so students read a narrow range of texts and have limited opportunity to learn strategies for and the value of revising, rather than just proofreading, their writing.

Standardized tests also limit the type of writing students do. Many tests of writing include a significant portion of indirect measures such as multiple-choice and short-answer items that do not require students to write extended prose. These tests encourage teachers to emphasize a test-based approach that focuses on the application of a fixed set of skills, which means that students learn little about processes of composing and rhetorical dimensions such as audiences and purposes for writing. This limitation is exacerbated by the increasing reliance of standardized tests on machine scoring. Most prominently, tests developed by the Smarter Balanced (SB) consortium and Partnership for Assessment of Readiness for College and Careers (PARCC) use machine scoring to assess students’ levels of achievement in writing on the Common Core State Standards.

For Con teams, this card plays into the floors vs. ceilings argument: that standardized testing significantly lowers the ceiling on achievement as a cost of raising the floor only slightly. This is a case study in English & Language Arts (ELA) which aptly demonstrates these impacts.
Standardized test data is too inaccurate to be a reliable teacher performance gauge. DAT

http://www.wsj.com/articles/SB10001424052702304723304577366023832205042

[Linda Darling-Hammond, Stanford Center for Opportunity Policy in Education]: Proponents of using test scores concede that such measures are imperfect but argue that they still are useful in the same way batting averages are—as an approximate indicator of performance. But at best, teachers' value-added ratings in one year predict only 25% of the variance in ratings in the next year, leaving 75% or more to be explained by factors such as who is assigned to a teacher's class and what conditions he or she teaches under.

The National Research Council and the Educational Testing Service, among other research organizations, have concluded that ratings of teacher effectiveness based on student test scores are too unreliable—and measure too many things other than the teacher—to be used to make high-stakes decisions. Test-score gains can reflect a student's health, home life and attendance; schools' class sizes and curriculum materials; and the influence of parents, other teachers and tutors. Because these factors are not weighed, individual teachers' scores do not accurately reveal their ability to teach.

Nonetheless, New York City's value-added ratings will soon be used to determine continuation and dismissal of teachers there. And a recently passed state law will extend the practice to all public-school teachers in New York state, not just those teaching reading and math, requiring a dramatic increase in the amount of testing for children.

One-third of the state's principals have signed a letter protesting the new system because they believe it will mismeasure teachers, undermine collaboration and create disincentives for teaching the neediest students. Further, the principals worry that greater focus on teaching to multiple-choice tests will reduce the time for the research, writing and complex problem-solving students need to succeed in today's society.
Standardized tests disproportionately harm teachers of low-income students. DAT

http://www.wsj.com/articles/SB10001424052702304723304577366023832205042

[Linda Darling-Hammond, Stanford Center for Opportunity Policy in Education]: Proponents of test scores often rightly favor an evaluation method that combines measures of teachers' classroom practice with evidence of student learning, including tests. But for this to work, the test-score measures must be appropriate for the particular students and the curriculum being taught. Unfortunately, federally imposed teacher-evaluation policies insist on using state tests that do not measure growth, are poor measures of higher-order thinking skills and penalize teachers of the neediest students.

Among other things, the tests administered each spring ignore the differences in summer learning between more- and less-advantaged students. More affluent students have enriched summer experiences, so when they return each autumn, they start school further ahead of where they were in June. Poor students, by contrast, have few opportunities for summer learning. Most actually lose ground between June and September. Value-added measures wrongly attribute this loss to their teachers, further distorting the teacher-evaluation process.
The use of standardized tests to evaluate teachers has perverse impacts on needy districts.


The EPI is a nonpartisan public policy thinktank.

Besides concerns about statistical methodology, other practical and policy considerations weigh against heavy reliance on student test scores to evaluate teachers. Research shows that an excessive focus on basic math and reading scores can lead to narrowing and over-simplifying the curriculum to only the subjects and formats that are tested, reducing the attention to science, history, the arts, civics, and foreign language, as well as to writing, research, and more complex problem-solving tasks.

Tying teacher evaluation and sanctions to test score results can discourage teachers from wanting to work in schools with the neediest students, while the large, unpredictable variation in the results and their perceived unfairness can undermine teacher morale. Surveys have found that teacher attrition and demoralization have been associated with test-based accountability efforts, particularly in high-need schools.

Individual teacher rewards based on comparative student test results can also create disincentives for teacher collaboration. Better schools are collaborative institutions where teachers work across classroom and grade-level boundaries toward the common goal of educating all children to their maximum potential. A school will be more effective if its teachers are more knowledgeable about all students and can coordinate efforts to meet students’ needs.

Con teams should look at the knock-on effects of impacts as they come across them in searching for the evidence. For instance, while standardized testing hurts teachers (wasted time, less autonomy, inaccurate measurement, etc.), the impacts passed to students are more alarming. Impacts on teachers play into a broader Con argument concerning institutional inequality.
Test-based teacher evaluations pervert teachers’ incentives. DAT


The EPI is a nonpartisan public policy thinktank.

In one recent study, economists found that peer learning among small groups of teachers was the most powerful predictor of improved student achievement over time. \{53\} Another recent study found that students achieve more in mathematics and reading when they attend schools characterized by higher levels of teacher collaboration for school improvement.\{54\} To the extent that teachers are given incentives to pursue individual monetary rewards by posting greater test score gains than their peers, teachers may also have incentives to cease collaborating. Their interest becomes self-interest, not the interest of students, and their instructional strategies may distort and undermine their school’s broader goals.\{55\}

To enhance productive collaboration among all of a school’s staff for the purpose of raising overall student scores, group (school-wide) incentives are preferred to incentives that attempt to distinguish among teachers.

Individual incentives, even if they could be based on accurate signals from student test scores, would be unlikely to have a positive impact on overall student achievement for another reason. Except at the very bottom of the teacher quality distribution where test-based evaluation could result in termination, individual incentives will have little impact on teachers who are aware they are less effective (and who therefore expect they will have little chance of getting a bonus) or teachers who are aware they are stronger (and who therefore expect to get a bonus without additional effort). Studies in fields outside education have also documented that when incentive systems require employees to compete with one another for a fixed pot of monetary reward, collaboration declines and client outcomes suffer.\{56\} On the other hand, with group incentives, everyone has a stronger incentive to be productive and to help others to be productive as well.

Test-based evaluations turn teaching into a rat race. Con teams should be clear that despite the immediate harms being to teachers, the overall harms end up on students.
Teachers feel pressured by the media. ABH


HOW DO NEW ACCOUNTABILITY STRUCTURES INFLUENCE THE STRESS AND/OR PRESSURE ON TEACHERS IN ELEMENTARY CLASSROOMS? All of the teachers indicate that they feel increased pressure to improve their students’ standardized test scores with the greatest pressures being felt from the media, school boards, and their principals (see Table 1). Specific instrument questions can be located in the section of the instrument, “Teachers.” Approximately 95% of the teachers indicate “testing creates a lot of tension for teachers and/or students.” A factor analysis of pressure variables indicates that three major types of pressures exist for these teachers: administrative pressure, media pressure, and pressure from other involved parties (see Table 2). Administrative pressure is evaluative pressure as the personnel involved have an actual or perceived role of evaluation and includes the principal, school board and “other,” which is comprised of school superintendents and district office personnel. In essence, these people are the teachers’ “bosses.” The factor of pressure from other interested parties constituent is a nonevaluative role but comprises people who have a vested interest in the school. This factor includes other teachers, parents, and community members. The media pressure factor stands alone. It is worth noting that the respondents perceive media pressure to be the greatest pressure.

CON should use this card to show another warrant of why teachers feel more stressed because of standardized testing, and then demonstrate the negative consequences of having teachers that feel stressed.
The Opportunity Costs of Standardized Testing

As always, the premise of the opportunity cost argument is simple: if there is an alternative to standardized testing which cuts existing harms without introducing any new ones, it is harmful not to pursue that alternative. Consequently, the presence of easily-implementable alternatives which achieve the same results, or evidence indicating that testing is a waste of time/money, indicates that standardized testing is not beneficial to K-12 education.

Better data can be collected “on the fly”. DAT

http://www.npr.org/sections/ed/2015/01/06/371659141/what-schools-could-use-instead-of-standardized-tests

Stealth assessment. Similar math and reading data, but collected differently.

The major textbook publishers, plus companies like Dreambox, Scholastic and the nonprofit Khan Academy, all sell software for students to practice math and English. These programs register every single answer a student gives.

The companies that develop this software argue that it presents the opportunity to eliminate the time, cost and anxiety of "stop and test" in favor of passively collecting data on students' knowledge over a semester, year or entire school career. Valerie Shute, a professor at Florida State University and former principal research scientist at ETS, coined the term "stealth assessment" to describe this approach.

Stealth assessment doesn't just show which skills a student has mastered at a given moment. The pattern of answers potentially offers insights into how quickly students learn, how diligent they are and other big-picture factors.

"Invisible, integrated assessment, to me, is the future," Kimberly O'Malley, the senior vice president of school research at Pearson Education, told me. "We can monitor students' learning day to day in a digital scenario. Ultimately, if we're successful, the need for, and the activity of, stopping and testing will go away in many cases."
Standardized testing creates time-wasting redundancies. DAT


The Center for American Progress is an independent nonpartisan educational institute

There is a culture of testing and test preparation in schools that does not put students first. Based on our analysis, test-taking time does not appear in itself to be problematic. But the culture of testing, particularly in urban districts, may play a more prominent role in the schooling experience. The frequency at which testing interrupts the school calendar in some districts and the fact that the testing burden seems to disproportionately affect urban schools are important aspects of today’s testing culture that warrant additional consideration. Our research also finds that some districts and states may be administering tests that are duplicative or unnecessary, including the use of practice tests.

More difficult to capture but an integral part of the testing conundrum nonetheless is how much time schools spend on test-preparation activities, training teachers on how to administer assessments, and analysis of test results. District-level staff in the Rochester City School District, for example, acknowledged that the time it took to administer and analyze pre- and post-tests, which they were using for teacher-evaluation purposes, was a motivating reason to consider alternatives strategies and eliminate some district-wide tests.
Non-standardized testing provides a better assessment of future performance. DAT


http://www.npr.org/sections/ed/2015/01/06/371659141/what-schools-could-use-instead-of-standardized-tests

Research shows that at least half of long-term chances of success are determined by nonacademic qualities like grit, perseverance and curiosity. As states expand access to pre-K, they are including social and emotional measures in their definitions of "high quality" preschool. As one component of a multiple-measures system, all schools could be held accountable for cultivating this half of the picture.

The Montgomery County Public Schools in Maryland survey both students and teachers on social and emotional factors and use the results to guide internal decision-making. The district uses the Gallup student poll, a 20-question survey that seeks to measure levels of hope, engagement and well-being.

"Engagement" is basically a measure of how excited students are to be in the building. Last year, 875,000 students took the Gallup poll nationwide, in grades 5-12. **According to one study, student hope scores on this poll do a better job of predicting college persistence and GPA than do high school GPA, SATs or ACT scores.**
The time costs of standardized testing. DAT


Learning Lab reports on innovation and reform in education.

The tests might not accurately address the learning that students do, either.

A recent Washington University research paper shows that high-stakes standardized tests may be a better indicator of socioeconomic status than anything else, thus exacerbating existing achievement gaps along class lines. It found that children from wealthier communities outscored poorer children by as much as 60 percent on standardized tests.

“The plethora of state-high-stakes-tests has created a new dilemma: achieving social justice in the public schools,” said Donald C. Orlich, the paper’s author. “The poor, disfranchised, minority and disabled children have fallen into education’s ‘achievement gap.’”

Other concerns have been raised by the amount of time testing can take away from learning time.

A report by the American Federation of Teachers shows that the time that students take tests can range from 20 to 50 hours per year. In addition, students can spend 60 to more than 110 hours per year in test prep in high-stakes testing grades.

Perversely, when testing time eats into educational time, this will typically hurt minority and poor students disproportionately.
Comprehensive information already exists. DAT


Burke is the Will Skillman fellow in education policy at The Heritage Foundation.

What kind of information do parents need about their children’s educational performance? First, they need to know whether their children are mastering the curriculum content. State criterion-referenced tests, which measure a student’s mastery of the content outlined by state standards, currently supply this kind of information. Parents also need to know that when the state test determines that, for example, a child has mastered third-grade content, the child is keeping pace with third-grade students across the country. In other words, parents need to know how rigorous their state standards and tests are. To provide this information, some states also offer norm-referenced tests, which measure student achievement compared to other students nationally.

Another tool that can provide comparative information is the National Assessment of Educational Progress (NAEP), which is administered to a sample of students in each state. In this way NAEP provides an external “audit” and common gauge on the quality of state standards and tests.

The meaningful information that parents and other taxpayers need is already available. The tools already exist to supply straightforward information on student, teacher, and school performance—sometimes referred to as report cards on the school system. All states are currently required by the No Child Left Behind Act (NCLB) to create such report cards. Some states, such as Florida and Massachusetts, supply more detailed reporting and straightforward information than others. What has been missing in some other cases is transparency about that information. If access to information has been inadequate, that does not justify a national standards and testing regime. Rather, policies should insist on clear reporting of the essential data to parents and other taxpayers.
National standardization strikes down more effective state initiatives. DAT


Burke is the Will Skillman fellow in education policy at The Heritage Foundation.

The variation in state standards is one of the most frequently cited reasons for adopting national standards and tests.[19] But the same pressures that detract from the quality of many state standards are likely to plague national standards as well. As a result, the rigor and content of national standards will tend to align with the mean among states, undercutting states with higher quality standards.[20]

For example, the Obama Administration’s proposal would force Massachusetts to abandon its highly regarded state standards and sign on instead to a set of national standards that are well beneath the rigor and content of the current state standards.[21] If it fails to do so, Massachusetts would stand to lose $275 million a year in federal funding for Title I.[22] For states like Massachusetts, the Obama Administration’s plan means facing the prospect of losing out on federal funding if they refuse to water down their standards.

Secretary of Education Arne Duncan refers to the varying quality of state standards as “50 different goal posts.”[23] That is a catchy phrase, but it begs the question of whether the national standards movement is more concerned with uniformity than it is with excellence. Uniform minimum-competency standards on a national level would provide a one-size-fits-all approach that would likely lead to decreased emphasis on advanced work and a generally dumbed-down curriculum.

Centralized standards and testing would eliminate the possibility of competitive pressure for increasing standards of excellence.

With respect to testing in education, standardization seeks to create a uniform “floor”, which is laudable. Con teams should be paying attention instead to the ceiling—the ways in which standardization acts to lower students’ ability to act on their potential.
Visualized: time spent on standardized assessments. DAT


Senior Associate Director of Education Issues for AFT.
Figure 3B - Estimated Annual Test Preparation Time, Eastern District

<table>
<thead>
<tr>
<th>Hours Per Test-Taker</th>
<th>K</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>10</td>
<td>22</td>
<td>22</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>50</td>
</tr>
<tr>
<td>Benchmark</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>State-Mandated</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>35</td>
<td>35</td>
<td>30</td>
<td>30</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>70</td>
<td>40</td>
</tr>
</tbody>
</table>
Financial Costs of Testing

Spending per student varies wildly across the nation, and is cumulatively massive. DAT

Ujifasa, Andrew. Standardized Testing Costs States $1.7 Billion a Year, Study Says.” 24
http://www.edweek.org/ew/articles/2012/11/29/13testcosts.h32.html

Standardized-testing regimens cost states some $1.7 billion a year overall, or a quarter of 1 percent of total K-12 spending in the United States, according to a new report on assessment finances.

The report released Nov. 29 by the Washington-based Brown Center on Education Policy, at the Brookings Institution, calculates that the test spending by 44 states and the District of Columbia amounted to $65 per student on average in grades 3-9 based on the most recent test-cost data the researchers could gather. (The Brown Center report was not able to gather that data from Connecticut, Iowa, Oklahoma, South Carolina, West Virginia, and Wyoming.)

It also says that the District of Columbia spends the most on its assessments per student—$114—of the 45 jurisdictions Brookings measured, followed by Hawaii, Alaska, Delaware, North Dakota, and Massachusetts. New York, where test scoring is a local responsibility, spent the least—$7 per student—followed by Kansas, North Carolina, Oregon, and Utah.

Despite the relatively small amount states spend on tests overall, compared with total education spending nationally, the report, written by Brown Center fellow Matthew M. Chingos, warns that the testing costs take on growing importance during difficult budget periods for states.

While the two consortia developing tests for the Common Core State Standards in English/language arts and mathematics (adopted by 46 and 45 states, respectively) may help reduce overall costs for states, the report says, “it is not yet clear whether larger consortia ... are a better choice than smaller ones formed more organically” from a cost standpoint.

While testing spending isn’t high relative to total spending, the issue is that for many states, testing is essentially a fixed cost: states will slash education budgets but keep this spending, which is not directly generating benefits, intact.
**Funds are essentially flowing from teacher salaries to private testing corporations. DAT**


http://www.edweek.org/ew/articles/2012/11/29/13testcosts.h32.html

If the money for standardized assessments was instead put toward teacher raises, the report estimates that each teacher in the country would receive, on average, a raise of $550, or 1 percent, based on data about teacher salaries and other factors from the 2012 Digest of Education Statistics.

The report also includes information on the major contractors that provide services for the states’ primary assessment contracts, although they don’t represent all state spending on tests.

It found that six vendors overall accounted for the bulk of the states’ $669 million of annual spending for tests required under the No Child Left Behind Act in grades 3-8 and once in high school. That spending amounted to $27 per student on average. Of all the contractors, the report says that New York City-based Pearson Education received the most money (39 percent), followed by McGraw-Hill Education, also in New York (14 percent), and the Maple Grove, Minn.-based Data Recognition Corp. (13 percent).

Senior Associate Director of Education Issues for AFT.
Big Data and the Obsolescence of Standardized Testing

Rather than standardized testing, educators can implement software programs which simultaneously guide and track students. Data is thus collected and analyzed in real time. It costs less than standardized testing while providing much more relevant, actionable results and augments, rather than detracts from, learning time.

Case study: adaptive programs simultaneously teach and track students. DAT

Fine-grained student data can be structured into meaningful chunks to provide evidence of student problem-solving sequences, knowledge state, and strategy. An example of this use of fine-grained data that is in wide-scale use is the ASSISTments tutoring system, currently used by more than 20,000 students in the New England area. Designed by researchers at Worcester Polytechnic Institute and Carnegie Mellon University, ASSISTments combines online learning assistance and assessment activities. ASSISTments tutors students on concepts while they practice on problems, and provides educators with a detailed assessment of students’ developing skills. While ASSISTments is widely used in fourth to 10th grade mathematics and science, it is also finding use in English and social studies. This wider adoption across subjects is due in part to teachers’ ability to write their own questions.

When students respond to ASSISTments problems, they receive hints and tutoring to the extent they need them. At the same time, ASSISTments uses information on how individual students respond to the problems and how much support they need from the system to generate correct responses as assessment information. Each week, when students work on ASSISTments, it learns more about their abilities and, thus, can provide increasingly appropriate tutoring for each student and can generate increasingly accurate predictions of how well the students will do on the end-of-year standardized tests. In fact, the ASSISTments system, taking into account information on the quantity and quality of help that students request, has been found to be more accurate at predicting students’ performance on the state examinations than the number of items students get correct on benchmark assessments (Feng, Heffernan, and Koedinger 2009).
Adaptive software can focus on improving low-scoring students’ performance. DAT


User behavior modeling in education often characterizes student actions as on- or off-task and can be used as a proxy for student engagement. It relies on the same kinds of learning data used in predicting user knowledge plus other measures, such as how much time a student has spent online (or on the system), whether a student has completed a course, documented changes in the classroom or school context, attendance, tardiness, and sometimes a student’s level of knowledge as inferred from his or her work with the learning system or from other such data sources as standardized test scores. Baker and colleagues have conducted a series of studies on detecting and adapting to students’ off-task behaviors (called gaming the system) in adaptive learning systems that teach algebra (Baker et al. 2004, 2006). They found that gaming behaviors (such as clicking until the system provides a correct answer and advancing within the curriculum by systematically taking advantage of regularities in the software’s feedback and help) were strongly associated with less learning for students with below-average academic achievement levels. In response, they modified the system to detect and respond to these students and provide them with supplementary exercises, which led to considerably better learning. Similar research has been done in unscripted environments that are more open-ended than the well-defined domain of mathematics. For instance, Blikstein (2011) has presented an automated technique and a case study to assess, analyze, and visualize behaviors of students learning computer programming.

Online learning systems log student data that can be mined to detect student behaviors that correlate with learning. Macfayden and Dawson (2010) analyzed learning management system tracking data from a Blackboard Vista-supported course and found variables that correlated with student final grade. Fewer than five variables were found to account for 30 percent of the variation in student final grades, and their model could correctly pick 81 percent of students who failed the course.
Testing Creates Cycles of Failure

*Students don’t get the feedback they need to succeed.* DAT


Nevada is another state that requires end-of-course exams, two each in reading and math.

"Thirty-seven percent of the month of October was taken up with testing," Brockett said. "And the same is true in March. January is another heavy testing month. But the test prep may kill us even more." She estimates one day entirely devoted to prep for every day of testing.

The average pass rate for an end-of-course exam at Brockett's school is 33 percent. That means most students have at least one retake, which are given several times a year. They may retake as many times as needed to pass, even as the material covered on the test fades farther and farther behind them.

"The kids who retake are the ones who need more instruction, but the more they retake, the less instructional time they get," she said.

These tests are not graded quickly. In some cases, a student who fails a test may have just a few days before the next retake — not enough time to work on what he or she got wrong.

Both Brockett and Jasper said test days disrupt an entire school. Even students who aren't sitting for a specific test may find themselves moved all over the building, or they may end up marking time watching movies for several days.
Standardized testing does not increase student performance. DAT


Burke is the Will Skillman fellow in education policy at The Heritage Foundation.

Proponents argue that national standards will make American students more competitive with their international peers. They point to international evaluation measures such as the Trends in International Mathematics and Science Study (TIMSS), in which American students rank in the middle of the performance distribution. Proponents note that countries that outperform the United States have national standards and that the U.S. needs national standards to move up in the ranking.[11]

But the relationship between existence of standards and strong educational outcomes is not clear. While the countries that outperform the United States on international tests have national standards, so do most of those countries that score lower than the U.S.[12] In further defiance of the hypothetical rule, Canada handily outscores the United States on international exams but has no national standards.[13] Even the relationship between the quality of state standards in the U.S. and academic performance is weak and inconsistent across subject areas.[14]

More careful attention is needed to understand the role that national standards play in other countries before asserting that national standards would add the same value in the United States. Alternatively, state standards and tests might be a closer analogy to standards and assessment systems in countries with populations the size of American states. There are limits to international comparisons in education given the size, diversity, and federal system of the United States.

The hypothetical benefit of testing is having data to help guide student achievement. The macro-level results don’t seem to bear this out.
Testing Hurts Early Learning

*Standardized testing in early elementary grades has increased. ABH*


We document striking increases in the use of standardized tests in early elementary classrooms. In 2010, roughly 30 percent of kindergarten teachers reported using standardized tests at least once a month. This is 2.6 times more often than the rate reported by first grade teachers in 1999. Further, there was a 20 percentage point increase in both kindergarten and first grade teachers indicating that they consider children’s performance relative to state or local standards very important or essential. Overall, our results suggest that kindergarten classrooms became increasingly similar in structure and focus to typical first grade classrooms of the late nineties, but that first grade classrooms have also shifted away from art, music and science instruction and increased their emphasis on assessment.

*Standardized testing has become quite common in grades as early as kindergarten and first grade. CON can use this information to attribute some of the lapses in early education to testing.*

*NCLB increased pressure on kids in kindergarten. ABH*


There is some empirical evidence supporting the claim that NCLB, or accountability pressures more broadly, impacted the learning experiences of young children (Russell, 2007). In a qualitative case study of a Texas elementary school, Booher-Jennings (2005) described the intense pressure teachers in the untested early grades (K-2) felt to prepare their students for third grade assessments and the reduction of recess to 15 minutes per week, despite the concerns of early childhood teachers. In a larger-scaled investigation, Jacob (2005) showed that high-stakes accountability led children in early, untested grades to be “preemptively retained” so that they would not be included in standardized testing. Similarly, several recent studies demonstrate that low-performing teachers in high stakes grades are disproportionately reassigned to untested elementary classrooms, and that this harms children’s learning (Fuller & Ladd, 2013; Grissom, Kalogrides, & Loeb, 2014).

*A review of the literature discussing the impact of high stakes standardized testing on early learning indicates that teachers feel pressured by these exams. They tend to move that pressure on to students, because of the consequences of reporting low-test scores.*
Students get less exposure to art and science. ABH


Given the substantial shift towards full-day kindergarten over the period investigated, it was conceivable that we would find increases in time devoted to all subjects. However, this was not the case. We document substantial reductions in time spent on art, music and science (though not social studies, physical education or recess). Further, teachers in 2010 were far less likely to indicate that their classroom included various activity centers including art areas, dramatic play areas, science areas, or water/sand tables. These trends are consistent with the possibility that a heightened focus on literacy and math instruction crowded out coverage of other subjects. Taken together with the drops we document in “childselected activities” and the increases in “teacher-directed instruction” as well as the heightened use of textbooks, workbooks and worksheets, our results may also suggest important shifts in the pedagogical approaches to kindergarten instruction.

Teachers feel pressured to prepare kindergarten students for standardized tests, so they focus less on subjects that aren’t tested such as art, music, and science.

Introducing tests too early damages children’s perception of school. ABH


These concerns may be heightened for very young children. The notion of holding young children to academic standards is new, and preschool-aged children are not accustomed to the formal instruction that current standards are likely to foster. Moreover, their perceptions of school and of themselves as learners are just forming; formal instruction introduced too early, or that is too hard or uninteresting, could turn young children off to schooling before they even get started and thus handicap their academic development throughout their school years. And, as mentioned above, social-emotional development and nonacademic skills at school entry predict school performance. Even if academic success in school was the only concern, previous research strongly suggests the importance of attending to many dimensions of children’s development in preschool.

CON should use this to show that the simple action of administering tests too early could give young children, who are very impressionable, a bad perception of school in general. This could lead to them straying away from academics as they get older.
Children need to build up self-confidence at a young age. ABH


There is no reason why teaching academic skills needs to undermine young children’s positive dispositions for learning. Children’s beliefs about their competencies, their expectations for success, and their enjoyment of learning depend on the nature and difficulty of the tasks they are asked to complete and the nature of evaluation and the feedback they receive (see Stipek, 2002b). Children’s self-confidence is maintained by working on tasks that are difficult enough to give them a sense of accomplishment and satisfaction but are not too difficult for them to complete. The variability in children’s skill levels is why rigidly paced instruction is inappropriate; if all children are asked to do the same task it will invariably be too easy for some students and too difficult for others and thus either boring or frustrating for many children. By implementing what is known about tasks and instruction that are motivating for young children, teachers can avoid negative effects of teaching basic academic skills.

Standardized tests will always be too easy for certain children and too challenging for others. This hurts their self-confidence, as children are either engaging in tasks that are too easy, or tasks that are too difficult for them. The idea of a standardized test prevents teachers from being able to vary the curriculum based on the needs of their students.
Rigid standardized testing does not accurately measure the ability of young students. ABH


The institutionalization of standardized assessments for young kids threatens to turn preschool into an academic environment that is too regimented for youngsters. I know a girl who, when given an aptitude test at age 4, refused to answer the questions because she just didn’t feel like it that day. That’s the way 4-year-olds act sometimes. She was scored as essentially having the aptitude of a monkey. That’s the way standardized assessments are, and that’s no way to judge a 4-year-old. The Race to the Top criteria say that any assessments developed should conform with the recommendations of the National Research Council’s reports on early childhood. One such report, published in 2008, entitled Early Childhood Assessment: Why, What, and How,” makes clear how difficult it is to create valid assessments for young children: “Assessments of any type must be selected and implemented with care, but special attention is needed when using direct assessments with young children. It requires greater attention to establishing a relationship with the child, to ascertaining whether the task is familiar and comprehensible to him or her, to limiting length of the session and the child’s discomfort, to recognizing the role of conditions like hunger or fatigue, and to recognizing the possibility of bias if the tester is a caregiver or otherwise connected to the child. Instruments that have the most user-appeal often do not have the best psychometric properties. For example, portfolios of children’s artistic productions contain rich information but are hard to rate reliably. In the experience of committee members, selection of instruments is often more influenced by cost, by ease of administration, and by use in other equivalent programs than by the criteria proposed here.”

Standardized testing is inherently problematic for younger students. So many factors can impact how a child performs on an exam on one given day including the child’s attitude that day, and if the child is hungry or tired.
Tests Cause a Harmful Level of Stress

*Students have harmful physical reactions to stress. ABH*


New tests said to be aligned to the Common Core State Standards and designed by Pearson were given to New York’s students for the first time this past spring even as teachers were still absorbing the new standards and learning how to teach to the Core. State education officials warned parents that the standards and tests were more difficult than students were accustomed to, and they even predicted that overall test scores would drop 30 percent. When the scores were released in August, it turned out that the scores plummeted exactly 30 percent (raising questions about how the officials knew the exact percentage drop). The principals’ letter on the new exams lists a number of problems with the exams and said many children reacted “viscerally” to the tests: We know that many children cried during or after testing, and others vomited or lost control of their bowels or bladders. Others simply gave up. One teacher reported that a student kept banging his head on the desk, and wrote, “This is too hard,” and “I can’t do this,” throughout his test booklet. It urges parents to help children who scored poorly understand that it isn’t their fault.

This evidence shows that standardized tests are stressing students out so much that they’re having adverse physical reactions.
Failure on standardized tests has severe emotional consequences. ABH


Emotional impact. We anticipated that it would be difficult to gauge the full emotional impact of test failure on students because they were surveyed almost 2 years after being told that they failed the test and because they had since learned that they actually passed the test. Although the intensity of their feelings might well have dissipated, our focus groups and anecdotal observations persuaded us that most of these students had vivid recollections of their emotional reactions to being informed that they failed the test. The survey results indicated that more than three fourths of the students reported adverse emotional reactions. For example, half of the students were willing to admit that they felt stupid. More than half reported that they felt depressed, and more than half reported that they felt embarrassed. Most students reported multiple negative reactions. As expected, girls reported greater emotional impact than boys did. We do not know if the impact was truly greater on girls because research on gender differences yields extensive evidence that females generally disclose more about their emotions than males do (Dindia & Allen, 1992). There is also evidence that females may have greater awareness of their emotional states and a capacity to describe their experiences in more detail (Barrett et al., 2000). Nevertheless, both boys and girls reported adverse reactions to test failure, and both genders demonstrated increasing impact with grade level. Anecdotally, we learned of students who felt shocked and humiliated to the point of feeling physically ill, with symptoms such as nausea and vomiting. Some students felt sufficiently depressed and discouraged to seek professional counseling. Other students reported intense feelings of anxiety and panic as they contemplated taking the exam again. The literature on test anxiety (Spielberger & Vagg, 1995) makes it clear that such symptoms are not rare, but there appears to be far more research on anticipatory test anxiety than the aftereffects of test failure. The increasing prevalence of highstakes testing in the United States signals a need to expand the scope of test-anxiety research.

Failure on these high-stakes standardized exams causes a range of psychological and emotional issues for students. The most common one being that they feel “stupid” or “not good enough”. CON should use this evidence to show the long term harms of testing on students.
Stress from testing has long term consequences. ABH


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).

This piece of evidence says two important things. First, stress has harmful health impacts throughout life, and second, standardized testing especially increases stress because it makes children feel helpless.
Popular Complaints Are Historically Unfounded

*Studies suggest that the positive impacts of testing outweigh the commonly discussed harms.*

ABH


One hundred years’ evidence suggests that testing increases achievement. Effects are moderately to strongly positive for quantitative studies and are very strongly positive for survey and qualitative studies. The overwhelmingly positive results of the qualitative research review, in particular, may surprise some readers. The results should be considered reliable because the base of evidence is so large—245 studies conducted over the course of a century in more than 30 countries. Qualitative studies have been held up by testing opponents as a higher standard for studies of educational impact. Labeling quantitative studies as narrow, limited, or biased, they have pointed toward qualitative studies for an allegedly broader view. But, the qualitative studies they cite have not investigated the effect of tests on student achievement. Rather, they have focused on popular complaints about tests, such as “teaching to the test,” “narrowing the curriculum,” and the like. Indeed, many of those studies have not considered any possible positive effects and looked only for effects that would be considered negative.

This meta-study looked and hundreds of studies done regarding testing and concluded that the overwhelming majority of both qualitative and quantitative studies show that testing is beneficial. PRO can use this to show that many of the claims against testing are not backed up by actual data.
Taking tests does not suck up disproportionate amounts of time. DAT


The Center for American Progress is an independent nonpartisan educational institute

Actual test administration takes up a small fraction of learning time. Students spend, on average, 1.6 percent or less of instructional time taking tests. This corresponds to findings from other similar examinations of testing time.81 On average, students in grades 3-5 and 6-8 spend 15 and 16 hours, respectively, on district and state exams. In contrast to the average total hours of instructional time, the amount of time spent on test-taking is comparatively small.82 These students did spend more time on state tests than district tests—nearly three more hours, on average. Students in grades K-2 and 9-12, who take the fewest number of tests—approximately six tests in a year—spent the least amount of time taking tests in the year at approximately four and nine hours, respectively. The fact that these students do not take or are less frequently tested using federally required state exams is a contributing factor.
Testing takes up only enough time to produce meaningful results. DAT

http://www.brookings.edu/research/papers/2015/01/08-chalkboard-annual-testing

Whitehurst is a senior fellow at the Brookings Institution.  

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.

While the frequency of testing is a common complaint point, it is impossible to do any kind of evaluation (whether by standardized testing or some other means) on a less-than-annual basis. Pro teams should be working from the framework that some kind of evaluative measures are inherently necessary to successful educational programs. Standardized testing cannot be evaluated in a vacuum during rounds.
The Financial Cost of Testing is Negligible

*The minimal monetary cost of standardized testing is worth the results. ABH*


At the same time, states must not be penny wise and pound foolish. Results of standardized tests are being used for high-stakes decisions, ranging from which schools to close to which students to hold back a grade to which teachers to fire. All of the Common Core assessments discussed above cost less than the price of a textbook, as do most of the assessments currently in use through the country. The most recent data indicate that U.S. schools spend about $10,500 per student each year. In that context, it seems almost silly to worry about whether it costs $15 or $30 or $45 to properly measure how much students are learning.

The overall cost of standardized testing is very small compared to how much is already spent on education. It is financially reasonable to administer such tests to ensure that the other more expensive academic resources are being allocated properly.

*Standardized tests don’t need to be pervasive or costly. DAT*


Sampling. A simple approach. The same tests, just fewer of ’em. Accountability could be achieved at the district level by administering traditional standardized tests to a statistically representative sampling of students, rather than to every student every year.

That's how the "Nation's Report Card" works. Formally known as the National Assessment of Educational Progress, or NAEP, it's one of the longest-running and most trusted tests in the U.S. education arsenal, even though it's not attached to high stakes. It's given to a different sample of students each year, in grades 4, 8 and 12. The widely respected international test PISA is given to a sample of students too.

The solution is essentially to treat standardized tests like a rigorously-conducted poll. This ensures data validity while minimizing costs.
Cost savings of eliminating standardized tests are negligible. DAT


We estimate that states nationwide spend upwards of roughly $1.7 billion on assessments each year, after adjusting the $669 million figure to (1) account for the fact that six percent of students are located in states for which we were unable to obtain data, (2) reflect spending on assessments not included in states’ primary assessment contracts, and (3) include state-level spending on assessment-related activities that are not contracted out. This seemingly large number amounts to only one-quarter of one percent of annual K-12 education spending. Were all statewide assessment activities to cease and the funding used to hire new teachers, the pupil-teacher ratio would only fall by 0.1 students. If instead the costs were devoted to an across-the-board pay increase for teachers, the average teacher would see her salary increase by one percent, or about $550.

The relative impacts of assessment spending are relatively small, on the surface. This card also helps refute the opportunity cost argument—the money spent on standardized testing would likely have few impacts, if any, elsewhere.
Districts can minimize costs even further. DAT


This relatively low level of spending on assessment, combined with concerns that the quality of tests in many states is not high enough to use them for highstakes purposes such as teacher evaluation, strongly suggests that states should seek efficiencies in order to absorb budgets cuts without compromising test quality or to free up resources that could be reinvested in upgrades to assessment systems. A clear strategy for cost savings suggested by our data is for states to collaborate on assessments so as to share the fixed costs of test and item development over larger numbers of students. Our cost model predicts substantial savings from collaborating on assessments. For example, a state with 100,000 students that joins a consortium of states containing one million students saves an estimated 37 percent, or $1.4 million per year; a state of 500,000 students saves 25 percent, or $3.9 million, by joining the same consortium.

Collaborating to form assessment consortia is not a new idea, and is in fact the strategy being pursued by nearly all of the states that have adopted the Common Core standards. Our model cannot be used to estimate the cost of the tests being developed by the Common Core consortia because they include innovative features not part of most existing systems and because they are substantially larger (in terms of students covered) than any existing state assessment system. But our model does suggest that these consortia will create opportunities to realize significant cost savings, all else equal, compared to the current model of most states going it alone.

This card provides advocacies for Pro teams to counter the perception that standardized tests are inefficient (regardless of these claims’ actual truth).
Gains in teacher quality outweigh costs. ELT


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 — the single most important school-based factor for fostering student learning — 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.
Teachers’ Performance is Measured Accurately

*Standardized testing is a surprisingly good indicator of teacher performance. DAT*

http://www.wsj.com/articles/SB10001424052702304723304577366023832205042

[Thomas Kane, Harvard Graduate School of Education]: Third, although the current state tests focus too heavily on easy-to-measure skills and need to be improved, it's not true that the teachers with larger gains on such tests are simply coaching students for the state tests. In our Gates Foundation study, the students with the largest gains on the state tests also tended to have larger gains on other tests which probed students' conceptual understanding in math and their writing skills. These students also were more likely to report high levels of effort and enjoyment in class.

Moreover, Dr. Chetty and his colleagues found that students whose teachers had high achievement gains on state tests had higher earnings as adults. According to the study, taking aggregate results and comparing classes of similar size, an elementary school class with a teacher in the top 5% of achievement gains is estimated to earn $250,000 more in the students' lifetimes than a class led by a teacher with average achievement gains.

This card makes two impactful points which together combine for the basis of a syllogistic argument: 1) Teacher quality alone can impact student performance short- and long-term
Value-added testing is the best way to evaluate teachers when combined with other measures.


There are competing views on the best way to approach incorporating student test scores into teacher evaluations. The model that’s currently accepted as the most accurate is known as the “value-added model” because it’s supposed to measure exactly how much of a student’s progress made over the course of the year is thanks to her teacher. It compares each student’s test scores to their own past scores and to other students’ scores. “It’s not perfect, but there is no other measure as good as [the value-added model],” Hull said. “But that’s not to say that other measures shouldn’t be used as well. You really want to have multiple measures of student outcomes.”

The Measures of Effective Teaching (MET) Project, the magnum opus of teacher-evaluation studies — every expert I spoke with on this topic pointed me to it — confirms the idea that student growth, often in the form of test scores, should be included in evaluations, along with other metrics. The project, which was funded by the Bill & Melinda Gates Foundation, was a three-year research partnership between 3,000 teacher volunteers and dozens of independent research teams who together developed and tested multiple measures of teacher evaluations to help better identify effective teaching. It didn’t recommend one specific model, given that every district has its own needs, but it did highlight the fact that “heavily weighting a single measure may incentivize teachers to focus too narrowly on a single aspect of effective teaching and neglect its other important aspects.”
Testing Produces Crucial Data

*Accounting for all other variables, test scores predict students’ future incomes.* DAT


http://www.brookings.edu/research/papers/2015/01/08-chalkboard-annual-testing

Whitehurst is a senior fellow at the Brookings Institution.

Association between student test scores in grades 3-8 and earnings at age 28 (with earnings adjusted for student and school characteristics)

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.

Standardized testing must be thorough and annual to produce useful results. DAT

http://www.brookings.edu/research/papers/2015/01/08-chalkboard-annual-testing

Whitehurst is a senior fellow at the Brookings Institution.

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.
Opting Out Bad

*Opting out affects everyone. ELT*


Standardized tests, as ‘high stakes tests,’ have been misused over time to deny opportunity and undermine the educational purpose of schools, actions we have never supported and will never condone. But the anti-testing efforts that appear to be growing in states across the nation, like in Colorado and New York, would sabotage important data and rob us of the right to know how our students are faring. When parents ‘opt out’ of tests—even when out of protest for legitimate concerns—they’re not only making a choice for their own child, they’re inadvertently making a choice to undermine efforts to improve schools for every child.
AT: Test too frequently

Frequent testing is key to good data. ELT

http://www.brookings.edu/research/papers/2014/01/22-more-testing-whitehurst

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.
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.

This evidence responds to two arguments: that testing every year is bad and that tests are not reliable. When the Con presents evidence that tests are unreliable, you should always be asking what kind of tests they are measuring, and then explaining why the fault is specific to that kind of test.
Frequent testing solves teaching to the test. ELT


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.
AT: Teaching to the Test

Test scores are higher when teachers do not focus on test prep. ELT


What instruction produces higher test scores? ► Curriculum is aligned to state standards & tests ► Student achievement data informs instruction & supports ► Test preparation is embedded and does not take up a lot of valuable class time ► Lessons emphasize student learning & higher order thinking
Con Counters
AT: Unaccountable

The system is already gamed and a simple report would give the same information as a standardized test. LZ


People will say: "That's crazy! Schools will fudge results. Grade data means nothing because teachers apply different standards with different values. Let's give them all one reliable test. And won't this proposal create a whole new bureaucracy?" 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.
Grades differ anyways and publishing grades gives the same level of accountability. LZ


We have the grade problem at my high school. In the same course or department, a B in one classroom might be an A, or even a C, in another. It's a problem for us, and, likely, a problem in most schools. "To sum up, we don't learn much from standardized accountability, and we have lost a great deal by giving it so much prominence." But it has also been an opportunity. Recognizing our grading differences, we opted to create a common conception of achievement, our graduate profile, and department learning outcomes with rubrics. Our standards now align closely with the Common Core State Standards. Second, we created common performance tasks that measure these standards and formative assessments that scaffold to them. Third, we look together at student work. Fourth, we have begun to grade each other's students on these common tasks. We could publish the results of these performance tasks, and the public would have a good idea of what we're good at and what we're not. For example, our students effectively employ reading strategies to comprehend a text, but are often stymied by a lack of vocabulary or complex syntax. We've also learned most of our students can coherently develop a claim, citing the appropriate evidence to support it when choosing from a restricted universe of data. They aren't as good when the universe of data is broadened. They are mediocre at analysis, counter-arguments, rebuttals, and evaluation of sources, though they have recently gotten better at evaluating sources as we have improved our instruction and formative assessments. A small percentage of our students do not show even basic competency in reading and writing. That's better information than we've ever received from standardized testing. What's also started to happen is that teachers who use the same standards and rubrics, assign the same performance tasks, and grade each other's work are finding their letter grades starting to align. And, this approach has led to a lot of frank discussions. For example, why are grades different? Where we have looked, different conceptions of achievement and rigor seem most important. So we have to talk about it. The more we do, the more aligned we will become, and the more honest picture of achievement we can create. It has been fantastic professional development—done without external mandates. We have a long way to go, but we can understand the value of our efforts and see improvement in student work. I would not advocate publishing individual teachers' grades because it would cause the same problems as publishing individual teachers' standardized-test results, but grades by subject, grade level, and demographic categories could be fair game externally. Internally, those breakdowns should stimulate hard conversations and necessary professional development. Of course, this proposal would have to be negotiated and modified locally to avoid the punishment/reward cycle of other accountability measures that force people to conform and tempt them to cheat. The goal is to spur the collaboration and conversation necessary for improvement. Well, that's you district, some might say. It's got a unique collaborative culture and a better sense of achievement than most. You can't do that across the nation. Why not? With the common core, a definition of achievement exists. And teachers are more likely to respond to professional development and accountability more concretely connected to their daily work. They are more likely to improve.
AT: Closes Achievement Gaps

*Curriculum narrowing from tests hurts disadvantaged students most. ELT*


Research suggests that in response to the high-stakes standardized tests in No Child Left Behind, school districts narrowed their curriculums to focus more strongly on reading and math, the most-tested subjects. And those effects were strongest at low-performing schools that tend to enroll a disproportionate number of poor and minority students.

So some civil rights activists argue that despite its good intentions, the law actually harmed the students it was meant to help.

Read this with evidence that talks about why narrowing the curriculum to focus on just math and reading is bad. This piece of evidence serves to magnify those impacts in response to a pro contention about how testing closes achievement gaps.
Accountability Doesn’t Mandate Standardization

http://www.npr.org/sections/ed/2015/01/06/371659141/what-schools-could-use-instead-of-standardized-tests

Scotland is a place where you can see many of the approaches above in action. Unlike the rest of the U.K., it has no specifically government-mandated school tests. Schools do administer a sampling survey of math and literacy, and there is a series of high-school-exit/college-entrance exams that are high stakes for students. But national education policy emphasizes a wide range of approaches to assessment, including presentations, performances and reports. These are designed to measure higher-order skills like creativity, students' well-being and technological literacy as well as traditional academics. Schools and teachers have a lot of control over the methods of evaluation.

At the school level, Scotland maintains accountability through a system of government inspections that has been in place in the U.K. since 1833. Inspectors observe lessons, look at student work and interview both students and staff members.

This card is mostly meant to give some ideas on what a Con advocacy could look like. Con teams need to subvert a pervasive notion that a lack of standardization is paralleled by a lack of accountability.
Decentralization will actually increase accountability. DAT


Burke is the Will Skillman fellow in education policy at The Heritage Foundation.

On the other hand, national standards and testing are unlikely to overcome these deficiencies. These problems are too deeply ingrained in the power and incentive structure of the public education system. A national standards debate threatens to distract from these fundamental issues. Centralized standard-setting would force parents and other taxpayers to relinquish one of their most powerful tools for school improvement: control of the academic content, standards, and testing through their state and local policymakers. Moreover, it is unclear that national standards would establish a target of excellence rather than standardization, a uniform tendency toward mediocrity and information that is more useful to bureaucrats who distribute funding than it is to parents who are seeking to direct their children’s education.

Common national standards and testing will not deliver on proponents’ promises. Rather than addressing the misalignment of power and incentives from which many public education problems arise, national standards and testing would further complicate these same problems. An effort by the Clinton Administration to produce national standards and tests during the 1990s was roundly rejected because of strong opposition among Members of Congress, state leaders, and others.[1] This renewed push for common national standards and assessments should be similarly resisted.

Instead, federal policy can improve the alignment of power and incentives in public education by enhancing transparency of existing accountability tools and providing flexibility in program funding for states to do the same. State policy should advance systemic reforms that better align power and incentives with educational outcomes, including enhanced accountability and parental empowerment through educational choice. By pursuing this combination of reforms, Americans can better address the core issues that continue to inhibit meaningful education reform.

While federal standards can create accountability of some sorts, it’s more likely to be accountability to bureaucratic goals and educational mediocrity than anything else. Con teams should diverge accountability from quality in their discussion of the issue.
Standardization Gives Publishers Monopolies

*Publishers make billions selling standardized test material that poor districts can’t afford.*

*DAT*

*Poor students and districts essentially can’t afford the solutions to standardized tests.* DAT


When a problem exists in Philadelphia schools, it generally exists in other large urban schools across the nation. One of those problems—shared by districts in New York, D.C., Chicago, Los Angeles, and other major cities—is that many schools don’t have enough money to buy books. The School District of Philadelphia recently tweeted a photo of Mayor Michael Nutter handing out 200,000 donated books to K-3 students. Unfortunately, introducing children to classic works of literature won’t raise their abysmal test scores.

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.

Standardization almost inherently guarantees monopolization, which prices out the students who need to advance most. It’s a perversion of the educational ideal from which standardized testing initially sprung.
Test Data Is Not Informative

Value-added modeling of test data is wildly inaccurate. DAT


The EPI is a nonpartisan public policy thinktank.

Nonetheless, there is broad agreement among statisticians, psychometricians, and economists that student test scores alone are not sufficiently reliable and valid indicators of teacher effectiveness to be used in high-stakes personnel decisions, even when the most sophisticated statistical applications such as value-added modeling are employed.

For a variety of reasons, analyses of VAM results have led researchers to doubt whether the methodology can accurately identify more and less effective teachers. VAM estimates have proven to be unstable across statistical models, years, and classes that teachers teach. One study found that across five large urban districts, among teachers who were ranked in the top 20% of effectiveness in the first year, fewer than a third were in that top group the next year, and another third moved all the way down to the bottom 40%. Another found that teachers’ effectiveness ratings in one year could only predict from 4% to 16% of the variation in such ratings in the following year. Thus, a teacher who appears to be very ineffective in one year might have a dramatically different result the following year. The same dramatic fluctuations were found for teachers ranked at the bottom in the first year of analysis. This runs counter to most people’s notions that the true quality of a teacher is likely to change very little over time and raises questions about whether what is measured is largely a “teacher effect” or the effect of a wide variety of other factors.

The “big data” movement in education is only positively impactful if the data is meaningful in real-world terms. This so far does not appear to be the case with standardized test score data.
There are too many distortions for test scores to accurately gauge teacher performance. DAT


The EPI is a nonpartisan public policy thinktank.

Teachers’ value-added evaluations in low-income communities can be further distorted by the summer learning loss their students experience between the time they are tested in the spring and the time they return to school in the fall. Research shows that summer gains and losses are quite substantial. A research summary concludes that while students overall lose an average of about one month in reading achievement over the summer, lower-income students lose significantly more, and middle-income students may actually gain in reading proficiency over the summer, creating a widening achievement gap. Indeed, researchers have found that three-fourths of schools identified as being in the bottom 20% of all schools, based on the scores of students during the school year, would not be so identified if differences in learning outside of school were taken into account. Similar conclusions apply to the bottom 5% of all schools.

For these and other reasons, even when methods are used to adjust statistically for student demographic factors and school differences, teachers have been found to receive lower “effectiveness” scores when they teach new English learners, special education students, and low-income students than when they teach more affluent and educationally advantaged students. The nonrandom assignment of students to classrooms and schools—and the wide variation in students’ experiences at home and at school—mean that teachers cannot be accurately judged against one another by their students’ test scores, even when efforts are made to control for student characteristics in statistical models.

Recognizing the technical and practical limitations of what test scores can accurately reflect, we conclude that changes in test scores should be used only as a modest part of a broader set of evidence about teacher practice.
Test data merely mirrors socioeconomic status. DAT


St. Francis Xavier University Faculty of Education.

There is a considerable amount of literature on how social inequalities along racial, social class, and gendered lines are reproduced and sustained through school practices (Fine & Weis, 2001; Kozol, 1992; Thomson, 2002). Arguably, standardized testing (re)produces these inequalities (Lipman, 2004; McNeil, 2000). For example, educational researchers have consistently identified economic factors as having a positive or negative impact on student achievement (Brownell et al., 2006; Kohn, 2000). As confirmed in a recent report from the Manitoba Centre for Health and Policy, when it comes to testing and educational outcomes, “children from families of lower socio-economic status (SES) don’t do as well in school as those from wealthier families” (Roos et al., 2006, p. 1). This trend holds true for most students; specifically, “each step up the socioeconomic ladder is associated with better outcomes” (Roos et al., 2006, p. 1). The Manitoba Centre for Health and Policy also confirmed that by the time students reach secondary school, those with lower SES are more likely to leave school early, fail standardized tests or not write them, or be behind in school levels (as cited in Roos et al., 2006). Ungerleider (2006) says “in Canada approximately 70 percent of the variation in student learning is not attributable to school factors but to student, family and community characteristics” (p.877).

Test data doesn’t have an impact beyond mirroring socioeconomic data, rendering it redundant.
Too many factors affect test performance for the data to be reliable. DAT


The authors of the paper studied test results for the Bagrut, a series of tests similar to the SAT that Israelis take when they finish high school. They compared those results to data on the students' future earnings and to air pollution measurements on the day they took the test. They found that even a moderate increase in the level of air pollution on the days of the test reduced students' incomes by about 2 percent by the time they became adults.

That's right: Air pollution had a small effect on test scores, but the ramifications of those small differences throughout students' lives affected their careers and their incomes.

Israeli universities rely more heavily on Bagrut scores than do American schools on the SAT, so it is not clear whether random factors like air pollution during an SAT administration could have such a dramatic effect on American students' future careers.

That said, the Obama administration effectively required that teachers be evaluated on the basis of students' test scores until recently. In districts where those evaluations have been used to make decisions about teachers' pay, standardized tests can have real consequences for educators, if not for their students.

This is where Con teams can point out the Catch-22: testing only has high impacts if it is high-stakes, almost by definition, but high-stakes testing has too much variance to be a credible determinant of anything significant.
Some issues, like systemic inequalities, are obvious without test data. DAT


So researchers have been going back to the drawing board, trying to prove that, no matter which measuring stick you use, the worst teachers usually end up teaching the most disadvantaged kids. Last month, one of the top researchers in this field, labor economist Dan Goldhaber, published a new study with some of the most convincing evidence yet.

His study, written with two colleagues, one at the American Institutes for Research, where Goldhaber is a vice president, and one at Macalester College, measured teacher quality in three different ways for every teacher in the state of Washington. They looked not only at student test score gains, but also at years of teacher experience and teacher licensing exam scores. Their study, "Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students," was published online in the journal Educational Researcher on June 29, 2015.

And no matter which of these three measures of teacher quality they used, guess what? They got the same result. Disadvantaged students across the state's elementary, middle and high schools ended up with the worst teachers – the ones who not only produced the smallest test score gains, but also had the fewest years of experience and the lowest licensure exam scores.
Tests measure only the least important learning goals. ELT


As a result of these necessary decisions about how to focus the content and the types of questions, the resulting test will measure only a subset of the domain being tested. Some material in the domain will be reflected in the test and other material in the domain will not. If one imagines the full range of material that might be appropriate to test for a particular subject—such as eighth grade mathematics as it is taught in a particular state—then the resulting test might include questions that reflect, for example, only three-quarters of that material. The rest of the material—in this hypothetical example, the remaining quarter of the subject that is excluded—would simply not be measured by the test, and this missing segment would typically be the portion of the curriculum that deals with higher levels of cognitive functioning and application of knowledge and skills.
Successful Alternatives Exist

Finland doesn’t use standardized tests. ABH


In fact, in Table 2, we see evidence that newspapers already suggest that cheating is inevitable. Instead of asking what conditions are responsible for the wave of cheating that we see, most newspapers apparently have accepted cheating as a fact of life. Of course this need not be the case. **High-stakes tests are not the only way to evaluate schools and students.** It is worth noting that Finland, the highest achieving country in the world in reading, mathematics and science, apparently have no standardized tests that resemble ours whatsoever, though they use teacher made tests in their classroom and school accountability system. Their system uses high standards for allowing teachers into the profession, awards high pay and bestows high status to those that enter teaching, provides rigorous and extensive professional development for the teachers, and depends on trusting relationships to improve academic achievement.112 Clearly there are highly successful models of how to build a national school system that we should study before assuming that our corrupting high-stakes accountability system is the only one that will work.

If PRO argues that there is no better alternative than standardized testing, CON could bring up the example of Finland as a country that doesn’t utilize standardized testing, but has been ranked highly for its education.
Moving away from standardized testing was effective in Israel. ABH


Matriculation 2000 was a 5-year project aimed at moving from the nationwide traditional examination system in Israel to a school-based alternative embedded assessment. Encompassing 22 high schools from various communities in the country, the Project aimed at fostering deep understanding, higher-order thinking skills, and students’ engagement in learning through alternative teaching and embedded assessment methods. This article describes research conducted during the fifth year of the Project at 2 experimental and 2 control schools. The research objective was to investigate students’ learning outcomes in chemistry and biology in the Matriculation 2000 Project. The assumption was that alternative embedded assessment has some effect on students’ performance. The experimental students scored significantly higher than their control group peers on low-level assignments and more so on assignments that required higher-order thinking skills. The findings indicate that given adequate support and teachers’ consent and collaboration, schools can transfer from nationwide or statewide standardized testing to school-based alternative embedded assessment.

Research on Israeli schools found that using school-based assessment can be more effective than using nationwide standardized testing if enough support is given to teachers.
Teacher’s evaluations are more accurate than test scores. ABH

Quay, Lorene., Steele, Donald. [Lorene Quay was a professor at Georgia State]

Throughout the period of this study, teachers’ judgements of children’s academic development, expressed by means of a simple device, the DRS, were superior to the scores of a standardized test, the DPII, for predicting the future achievement of children. Although this study does not provide an explanation for this superiority, one explanation might be that the DRS allows the teacher to make a global assessment of a child’s academic development based on her observations of some combination of characteristics, whereas standardized tests require the teacher to observe and record only the presence or absence of a number of specific skills defined by the test developer. Thus, the DRS might provide a vehicle for teachers to express judgements based on their schemas about the way various child skills and characteristics combine or interact to produce academic outcomes. This study provides evidence that teachers can make accurate judgements early in children’s schooling about their later academic development and achievement.

**CON** can suggest that instead of using one dimensional standardized test scores, schools can allow teachers to evaluate students based on a variety of factors. This study found that judgments made by teachers were more predictive of a child’s future than standardized test scores.
Students could be assessed based on portfolios. ABH


A portfolio is an organized, purposeful, integrated collection of student work that exhibits process, progress, achievement, and effort over time (Garan, 2004; Salvia & Ysseldyke, 2001; Schipper & Rossi, 1997). Learning is perceived as evolving and changing and includes shared authority and meaningful integrated instruction. Within the portfolio process, assessment and instruction are viewed as recurring processes that inform each other. Self-assessment is at the heart of portfolios and allows children to critically examine the experiences and process of learning. Meaningful and purposeful assessment occurs through the ongoing use of portfolios. The student and the teacher work collaboratively to establish goals for learning and standards for selected work. Students are given a choice about what is selected to show their growth, thus creating ownership for their learning. It is through this ownership that motivation to learn increases. Knowledge and learning is no longer perceived as the ability to correctly select an answer on a multiple-choice test, but rather is seen as occurring in many contexts. Portfolios are a place to view process, which allows students and teachers to effectively evaluate the actual learning taking place. They enable students to see that learning is a dynamic, interactive, ongoing process. Portfolios are valuable for all children, especially those in the younger grades. Because they are intricately connected to instruction and curriculum, portfolios provide a foundation upon which future learning can be built. They allow children to practice analytical and critical thinking, both vital to the pursuit of knowledge. They force children to take ownership for their own learning; thereby, children grow in confidence and self-esteem. Unlike standardized tests that only evaluate an end product, portfolios allow young children to see themselves as learners and in control of that learning from the beginning.

One alternative to standardized testing is portfolio-based assessments. These assessments allow students to have greater control over their assessments, submit multiple pieces of work to be assessed (instead of one exam), and focus on work relating to the curriculum.
**Performance based assessments empirically more successful. ELT**


We call on the civil rights community to support the work of educators around the nation who are working to develop authentic forms of assessment that can be used to help support students to develop critical thinking. **Innovative programs like New York City Consortium Schools** have a waiver from state standardized tests and instead use **performance based assessments** that have produced dramatically better outcomes for all students, even though they have more special needs students than the general population—and have **demonstrated higher graduation rates, better college attendance rates, and smaller racial divides in achievement** than the rest of New York’s public schools.
Standardized testing decreases graduation rates – incentives better. ELT


Conclusion 2: The evidence we have reviewed suggests that high school exit exam programs, as currently implemented in the United States, decrease the rate of high school graduation without increasing achievement. The best available estimate suggests a decrease of 2 percentage points when averaged over the population. In contrast, several experiments with providing incentives for graduation in the form of rewards, while keeping graduation standards constant, suggest that such incentives might be used to increase high school completion.
Pro Case

Introduction:

Imagine you’re boarding a plane. As you pass by the cockpit, you notice that the windows have been covered up and all the sensors are gone. Altimeters, fuel gauges, engine sensors—all gone. That’s not a plane you’d want to fly on.

Likewise, it would be unreasonable for the American public school system to function without any data. This data comes from standardized testing. This data is what allows parents to choose schools for their children, teachers to know what level of students they are attempting to teach, and policymakers to effectively direct their increasingly marginal resources.

My partner and I are not here to convince you that standardized testing is an optimal, all-encompassing solution to America’s education problems. Standardized testing works in tandem with other data collection and policymaking methodologies, and it is only through a holistic approach that includes standardized testing that K-12 education in America improves. It is because standardized testing greases the wheels of public education that my partner and I affirm the resolution Resolved: On balance, standardized testing is beneficial to K-12 education in the United States.

Contention One: Testing Improves Teaching

We begin with the albatross in the room: students themselves. We do not contest the notion that in pure mathematical terms, standardized testing takes time away from strictly “learning” activities. But we also bring up a substantially more pervasive premise on which to base this discussion: students benefit from having good teachers. The difference between a good teacher and a bad teacher is the difference between a success and failure for many marginal students. And without data, the American educational system effectively biases its students toward failure. As Thomas Kane of the Harvard Graduate School of Education explains:

Currently, high-stakes personnel decisions in K-12 education are primarily based on two factors: experience and graduate degrees. In the recent recession, thousands of teachers were terminated based simply on their seniority. As imperfect as the current measures of effective teaching are—and they must be improved—using multiple measures provides better information about a teacher's effectiveness than seniority or graduate credentials.

The variables Kane ticks off—seniority and graduate degrees—are empirical and objective, relatively. But neither measures the actual quality of education provided. With standardized testing—a relatively objective measurement mechanism—school districts can target terminations based on merit. And the difference between terminating thousands of young teachers and thousands of less-adept teachers is tens of thousands of student outcomes.
Contention Two: Standardized Testing Gives Students Control

As we age, the consequences of failure get larger and larger. Students can fail—homework assignments, exams, classes—and be no worse off in the long run. But failure only has utility if it is spotted and diagnosed. This is the fundamental role of standardized testing in a K-12 environment. The Brookings Institution’s Grove Whitehurst explains:

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.

If standardized tests are truly predictive of life outcomes, which Whitehurst’s data confirms, the American educational system would be negligent in not administering them. Standardized tests, inherently, provide a comparative snapshot of a student’s likely future trajectory. Without the data—without confirmation that they need to maybe spend more time on certain subjects or change their study habits—students are flying blind.

Contention Three: Testing Improves Retention

My partner and I have already shown the benefits of standardized testing with respect to teacher quality and student self-determination. The final component of student success in the American educational system is, of course, the education itself. It’s tempting to consider standardized testing as “lost” time, seeing as it’s not spent on instruction. However, students can be learning even in a test environment. In fact, they can be learning more. Henry L. Roediger III of Washington University in St. Louis discusses the impacts of standardized testing on student information retention:

Chan et al. concluded that testing not only improves retention for information covered within a test, but also improves retention for nontested information, at least when that information is related to the tested information.

Essentially, compared to simply re-studying information, students retain what they’ve learned better when they’re tested on it. But education is not simply about plugging as much information as possible into a student. Fundamentally, education is the development of problem-solving skills. Roediger covers these as well:

Recent research shows that the mnemonic benefits of taking a test are not limited to the specific questions or facts that were tested; retrieval practice also improves transfer of knowledge to new contexts. Transfer may be defined as applying knowledge learned in one situation to a new situation.

As far as general educational goals go, testing in general appears to improve both the particular (information retention) and big-picture (problem-solving) goals built into the American educational system.
Taking an even wider look at the educational system, it is necessary to remember a fundamental purpose of K-12 education: to give every child a chance at a successful life. This only happens with effective development of problem solving skills; when students are aware of their strengths and weaknesses; and when the best teachers, rather than the most senior ones, are teaching them.
Con Case

Introduction:

Last year, districts in Philadelphia had a $0 budget for books. Not a penny. At the same time, a quarter of a percent of America’s education budget— that still billions of dollars— was spent on standardized testing. My partner and I find the idea of spending limited money on tests which students don’t have the resources to take perversely dystopian, to say the least. Fundamentally, standardized testing has encouraged test-taking at the expense of problem solving, without solving any problems itself. It is for this reason that my partner and I negate today’s resolution, Resolved: On balance, standardized testing is beneficial to K-12 education in the United States.

Contention One: Standardized testing doesn’t measure anything

One of the primary goals of standardization is data collection. By administering the same test to a massive sample on a regular basis (typically annually), administrators can measure student progress in a uniform way. Unfortunately, standardized testing falls short of its promise. As explained by W. James Popham of UCLA’s Graduate School of Education:

> The designers of these tests do the best job they can in selecting test items that are likely to measure all of a content area's knowledge and skills that the nation's educators regard as important. But the test developers can't really pull it off. Thus, standardized achievement tests will always contain many items that are not aligned with what's emphasized instructionally in a particular setting.

Essentially, the beauty of American education is that it’s not standardized. It’s impossible for a standardized test to adequately measure both a dominantly white, high-income school and a dominant minority, low-income one using the same set of questions because those students likely should not be learning exactly the same material at exactly the same pace. A world in which a standardized test is measuring every student equally and adequately is, in a cruel Catch-22, a world in which every student is not getting an equally adequate education.

Contention Two: Standardized testing harms student outcomes

Standardized tests do not exist in a vacuum. Fundamentally, students must use their own resources— time spent taking tests and money to acquire the requisite textbooks for those courses— to enable a standardized testing regimen. This is fine if the benefits outweigh the costs. But as it stands, precious resources are used to create testing regimes which harm students even further. The University of Virginia’s Daphna Bassok elaborates further:
Overall, our results suggest that kindergarten classrooms became increasingly similar in structure and focus to typical first grade classrooms of the late nineties, but that first grade classrooms have also shifted away from art, music and science instruction and increased their emphasis on assessment.

My partner and I present a reasonable goal for the American educational system: that the goal is to produce competent people. This is far, far removed from the notion of a competent test taker. When we think of what goes into a competent human being (and competent member of the workforce)—work ethic, creativity, problem-solving, teamwork—these are all marginalized institutionally as a result of standardized testing at any early age. So what happens when students are told that they haven’t done well on these standardized tests? UVA’s Dewey Cornell presents some harrowing results:

The survey results indicated that more than three fourths of the students reported adverse emotional reactions. For example, half of the students were willing to admit that they felt stupid. More than half reported that they felt depressed, and more than half reported that they felt embarrassed. Most students reported multiple negative reactions.

The issue isn’t simply that testing cripples students’ self-esteem and subsequent passion toward their own education. The issue is that these harms are the results of testing which doesn’t even measure real-world skills or qualities. This is unacceptable.

**Contention Three: Standardized testing harms teaching**

In an ideal scenario, teachers can make up for the harms created by standardized testing by independently improving student outcomes. Unfortunately, standardized testing also handicuffs teachers while lowering overall instructional quality. Michigan State University’s Avner Segall, for instance, found that “standardized testing impacts teaching in mostly negative ways, reducing teaching to low levels of intellectual engagement and teachers to implementers of externally designed curricula and pre-packaged materials intended to help them teach to the test.” So teachers are already hamstrung in terms of what they can teach. But even within this small box, the most effective teachers will likely produce the best results.

But the cruelest Catch-22 of standardized testing is that it risks identifying the most effective teachers as the least effective ones. According to Linda Darling Hammond of Stanford’s Center for Opportunity Policy in Education:

The National Research Council and the Educational Testing Service, among other research organizations, have concluded that ratings of teacher effectiveness based on student test scores are too unreliable—and measure too many things other than the teacher—to be used to make high-stakes decisions.

Nonetheless, New York City's value-added ratings will soon be used to determine continuation and dismissal of teachers there. And a recently passed state law will extend the practice to all public-school teachers in New York state, not just those teaching reading and math, requiring a dramatic increase in the amount of testing for children.
By masquerading as an effective means of measuring students, standardized tests are actually both them and teachers by putting competent teachers’ jobs at risk as a function of factors entirely outside their control.

It should be no surprise, then, that while states spend billions on standardized testing the U.S. continues to lag behind much of the developed world. Standardized testing has essentially turned billions of taxpayer dollars into a fount of useless data, crushed morale, and poor teacher retention decisions. Instead of focusing on true, real-world-applicable outcomes, standardized testing has ensured that students have a narrower range of educational background presented to them by teachers who have both little control of their curriculum and how their evaluative data will shake out with respect to their perceived competence. And as students and teachers cede control to test publishers and district administrators, they also cede control over their own darkening future.