

Statement to the New York State Senate Standing Committee on Public Education

On the State of the State's Testing Program

Fred Smith

Administrative Analyst (Ret.)
New York City Public Schools

October 29, 2013

Good morning. Thank you for giving me the opportunity to address the Education Committee.

My name is, Fred Smith. I retired from the New York City public school system after 33 years where I worked on test research and development and as an administrative staff analyst. I hold a graduate degree in measurement and evaluation from Teachers College and am currently involved with Change the Stakes, a group of parents and educators concerned about the damage excessive testing is doing to education.

I offer you a testing person's perspective and findings on the testing program and where we stand today concerning New York's core-aligned assessments. I will present my analysis and findings about the program.

~~~~~

Let me start by saying that I attended a forum on testing last year that had a regent, a union president, a school principal, a professor and a local parent/activist as speakers. Each had a different slant on the testing program reflective of their important roles. To a person, each was critical of how the testing program was hurting children.

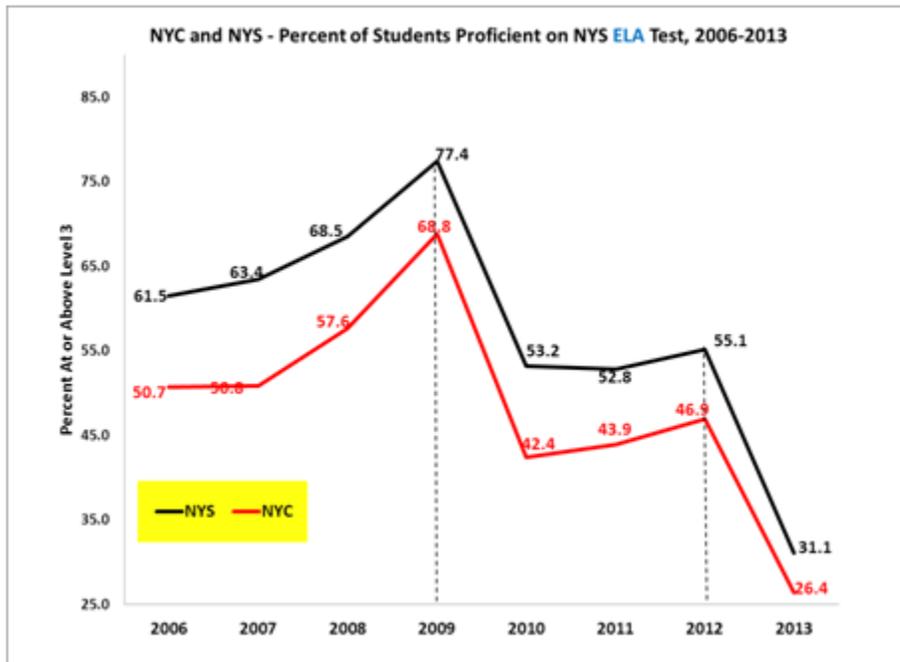
It made me think that testing has been a defective engine driving the train, but no one seems able to repair it or stop it from furiously racing ahead.

Today I will provide information about significant flaws in the program, particularly its field testing procedures; what makes the "core-aligned" 2013 exams a non-starter in efforts to establish a baseline; and the need for truth in testing and timely disclosure of test-related data in order to treat parents and the public with respect and restore a measure of trust in the integrity of testing.

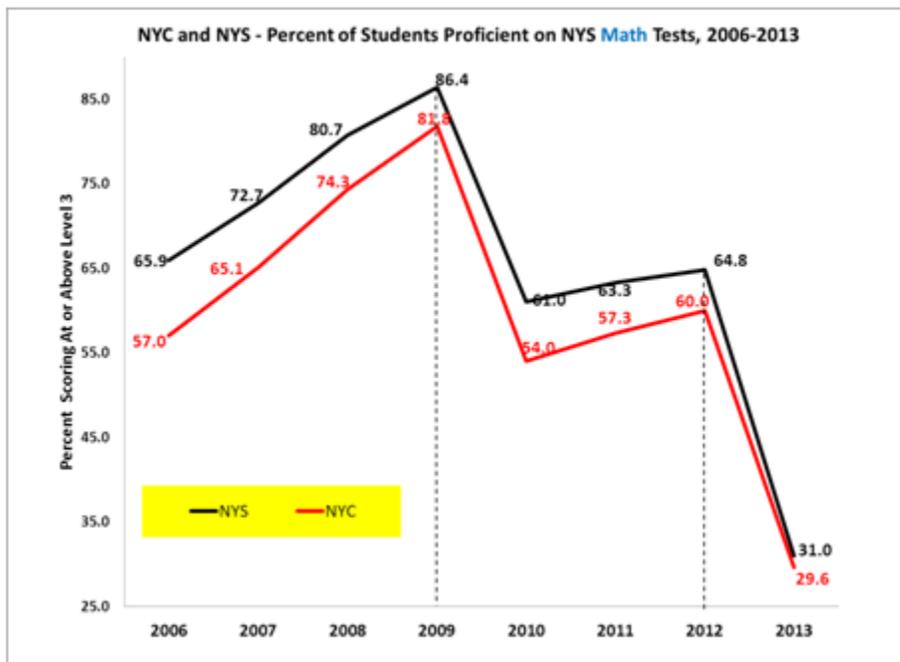
### **Recent History and Broad Trends in the New York State Testing Program**

Let's start with the big picture. These charts trace the New York State and New York City results on the English Language Arts (ELA) and Math tests between 2006 and 2013. About 1.2 million students take the exams every year in grades 3 through 8, including 420,000 in the City.

First, let's look at the percentage of students deemed to be proficient (i.e., reaching Performance Level 3) each year. Two things are noteworthy: Large changes appear from one year to the next. This pattern has been observed here and elsewhere following the testing mandates of the No Child Left Behind Act. But sudden movement up or down simply doesn't happen in education, especially when a huge population is involved. Learning proceeds gradually and incrementally even if we would like to speed its course. The line graphs raise questions about the scales being used to weigh student achievement and the veracity of the ELA and math results.

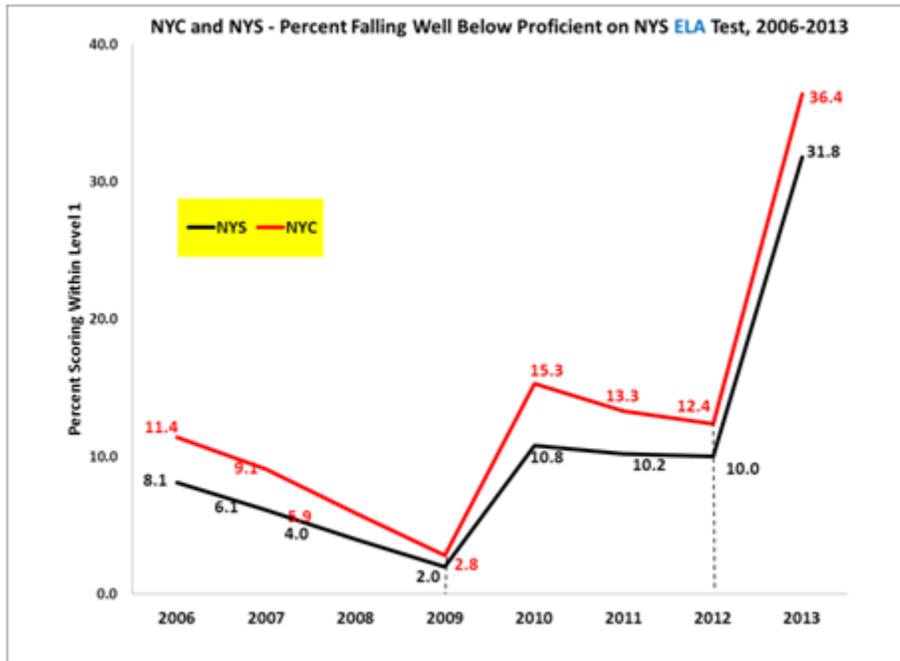


Secondly, precipitous turning points occurred in 2009 and this year. They punctuate our doubts about the testing program, the quality of the examinations and the meaning of the results.

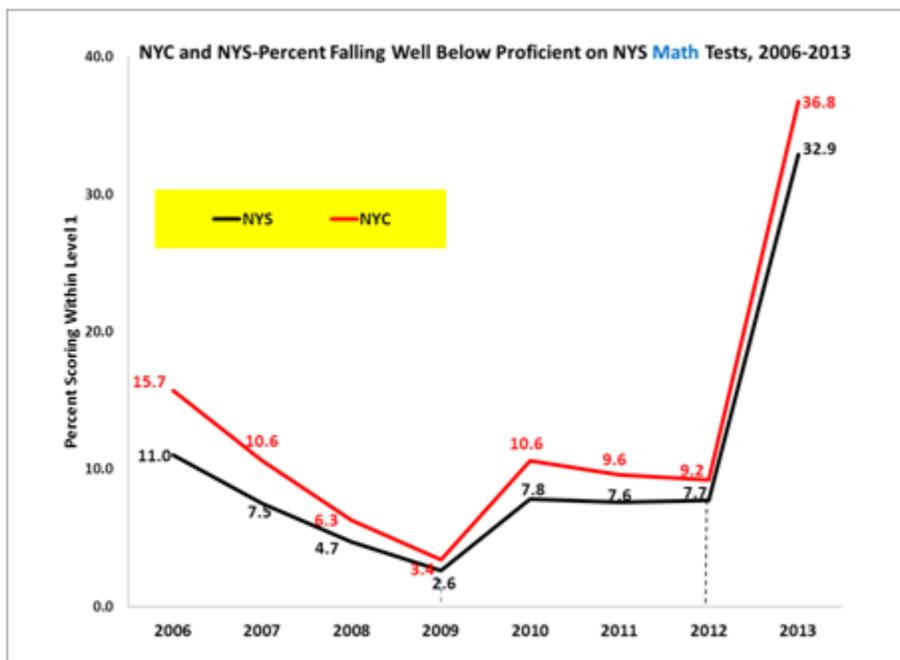


They tell us more about variations in the tests that were used and in the cut off scores that were set than they do about increases or decreases in student attainment and growth. It is typical that when new tests are introduced scores fall (2009) and the next year they level off or begin to rise until another test series is brought in (2013). Next year expect our results to hold steady or rise.

The flip side of this picture is seen in the percentage of children classified as well below proficient (Level 1) in ELA and math. At the bottom of the score range there are also shifts with extreme reversals recorded between 2009 and 2010 and from April 2012 to last April.



The changes at the low end of the distribution raise a red flag about the way New York State’s tests were developed under contract with the State Education Department by CTB/McGraw-Hill and Pearson, Inc. And they beg for honest answers to questions about how SED established the cutoff points each year that delineated the tests’ four performance levels, which determined how many students “met the learning standards” or required extra instructional support.



The incredibly high percentage of children reaching Level 3 and Level 4 in 2009 made this a pivotal year for the testing program. Regents Chancellor Merryl Tisch realized the results could not be sustained—77% proficiency on the ELA and 86% in math. She pledged to make the tests more rigorous and less predictable. No one, however, was held accountable for how the low scoring thresholds had been put in place that enabled these results.

While Ms. Tisch was discounting the 2009 scores, other politicians were celebrating the new heights in proficiency. In New York City the mayor was seizing reelection-year credit for education policies that led to the near disappearance of Level 1.

But an analysis of the cut scores revealed that children could reach Level 2 merely by guessing the low number of correct answers required to reach that point. The fact that the number of students in that category had shrunk so much was an artifact of how low the bar had been set.

The table shows the fluctuations in the number of Level 1 students from 2008 to 2013. It raises concerns that children who might have had serious academic problems and been eligible for instructional intervention were denied services in 2010 because they did better on the easy 2009 exams to reach Level 2.

| <b>Changes in the Number* and Percent of Students Classified as Level 1 on the New York State Exams (Level 1: Student Shows Serious Academic Problems)</b> |                |                 |                |                   |               |                 |                |                   |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|-----------------|----------------|-------------------|---------------|-----------------|----------------|-------------------|
| <b>English Language Arts</b>                                                                                                                               |                |                 |                |                   |               |                 |                |                   |
| Year                                                                                                                                                       | New York State |                 |                |                   | New York City |                 |                |                   |
|                                                                                                                                                            | Number Tested  | Percent Level 1 | Number Level 1 | Number Difference | Number Tested | Percent Level 1 | Number Level 1 | Number Difference |
| 2008                                                                                                                                                       | 1,207,778      | 4.0%            | 48,311         |                   | 417,155       | 5.9%            | 24,482         |                   |
| 2009                                                                                                                                                       | 1,200,460      | 2.0%            | 24,009         | <b>-24,302</b>    | 415,579       | 2.8%            | 11,755         | <b>-12,727</b>    |
| 2010                                                                                                                                                       | 1,196,283      | 10.8%           | 129,199        | <b>105,189</b>    | 414,609       | 15.3%           | 63,399         | <b>51,644</b>     |
| 2011                                                                                                                                                       | 1,195,432      | 10.2%           | 121,934        | -7,265            | 417,574       | 13.3%           | 55,610         | -7,789            |
| 2012                                                                                                                                                       | 1,192,129      | 10.0%           | 119,213        | -2,721            | 416,918       | 12.4%           | 51,690         | -3,920            |
| 2013                                                                                                                                                       | 1,182,266      | 31.8%           | 375,961        | <b>256,748</b>    | 415,257       | 36.4%           | 151,196        | <b>99,506</b>     |
| <b>Mathematics</b>                                                                                                                                         |                |                 |                |                   |               |                 |                |                   |
| Year                                                                                                                                                       | New York State |                 |                |                   | New York City |                 |                |                   |
|                                                                                                                                                            | Number Tested  | Percent Level 1 | Number Level 1 | Number Difference | Number Tested | Percent Level 1 | Number Level 1 | Number Difference |
| 2008                                                                                                                                                       | 1,217,789      | 4.7%            | 57,236         |                   | 424,807       | 6.3%            | 26,960         |                   |
| 2009                                                                                                                                                       | 1,211,360      | 2.6%            | 31,495         | <b>-25,741</b>    | 423,392       | 3.4%            | 14,305         | <b>-12,655</b>    |
| 2010                                                                                                                                                       | 1,210,384      | 7.8%            | 94,410         | <b>62,915</b>     | 425,738       | 10.6%           | 45,055         | <b>30,750</b>     |
| 2011                                                                                                                                                       | 1,207,539      | 7.6%            | 91,773         | -2,637            | 426,173       | 9.6%            | 41,103         | -3,952            |
| 2012                                                                                                                                                       | 1,202,504      | 7.7%            | 92,593         | 820               | 424,626       | 9.2%            | 39,087         | -2,016            |
| 2013                                                                                                                                                       | 1,190,104      | 32.9%           | 391,544        | <b>298,951</b>    | 421,938       | 36.8%           | 155,087        | <b>116,000</b>    |

\* The NYC Numbers at Level 1 are exact as reported on the DOE web site. The NYS Numbers are approximate--based on multiplying the numbers tested by the reported percentages.

In 2009, there was a drop off of about 24,000 Level 1 students statewide on the ELA. This included 12,727 in New York City. Did they no longer have serious learning problems? With the imposition of the “tougher” exams last April, the number of students falling well below proficiency surged by **more than 250,000** across the state including close to **100,000** students in the City. If this result has educational relevance, what is SED doing to meet its obligation to these children so suddenly marked deficient?

And in 2010, what was done for the 100,000+ students statewide, of whom 51,644 were from the City, when it turned out they quadrupled the ranks of those who had worrisome academic problems? Perhaps we can learn something from the 2010 boom about what to do with 2013’s explosion of Level 1 children—assuming, of course, that the cut off scores and the boxes they fill with numbers hold any educational meaning.

### **The 2009 Crossroads: Cut Scores and a Concession about Field Testing**

Ms. Tisch’s announcement of the 2009 results brought news that was too good to be true. Three years of rising scores had reached an implausible apogee. She promised that SED would take testing in a better direction. At the same time she firmly rejected probing how the statewide exams had become indefensible. But reporters, who usually find technical data “too arcane” for their audiences needed an explanation about why the cut off scores had been set so low.

It came in muted comments from Howard Everson, SED’s chief testing adviser and chair of its Technical Advisory Group. It centered on field testing. Field testing is done to try out items on samples of students. Statistics are generated that let publishers see how easy or difficult these items are likely to be should they appear on future exams. The data also guide publishers and the state in judging the number of score points students will need to reach each performance level.

There are two ways to field test items: By embedding them in the body of actual (operational) exams, where they are taken under real testing conditions, though they don’t count in the scoring; or by putting the trial items in separate “stand-alone” test booklets that are given to students at a different time than the operational tests.

Embedded items are the preferred way to try items if their inclusion doesn’t unduly lengthen testing time and enough forms are created to yield a sufficient number of items to choose from.

Mr. Everson noted that SED’s practice of releasing operational tests after they had been given prevented the field testing of items in a real test environment. He meant it precluded publishers from embedding developmental items among the items that made up the operational exams because disclosure would have exposed these items and blocked further use. So stand-alone field testing was done leading up to 2009.

Beyond that, the efficacy of field testing rests on ensuring that the students taking the trial items constitute samples that are representative of the populations that eventually take the operational tests, which will consist of those items that survive the field tests. This doesn’t only mean that

the samples match the population demographically. To be representative also requires that students in the sample are motivated to make their best efforts to do well on the field tests.

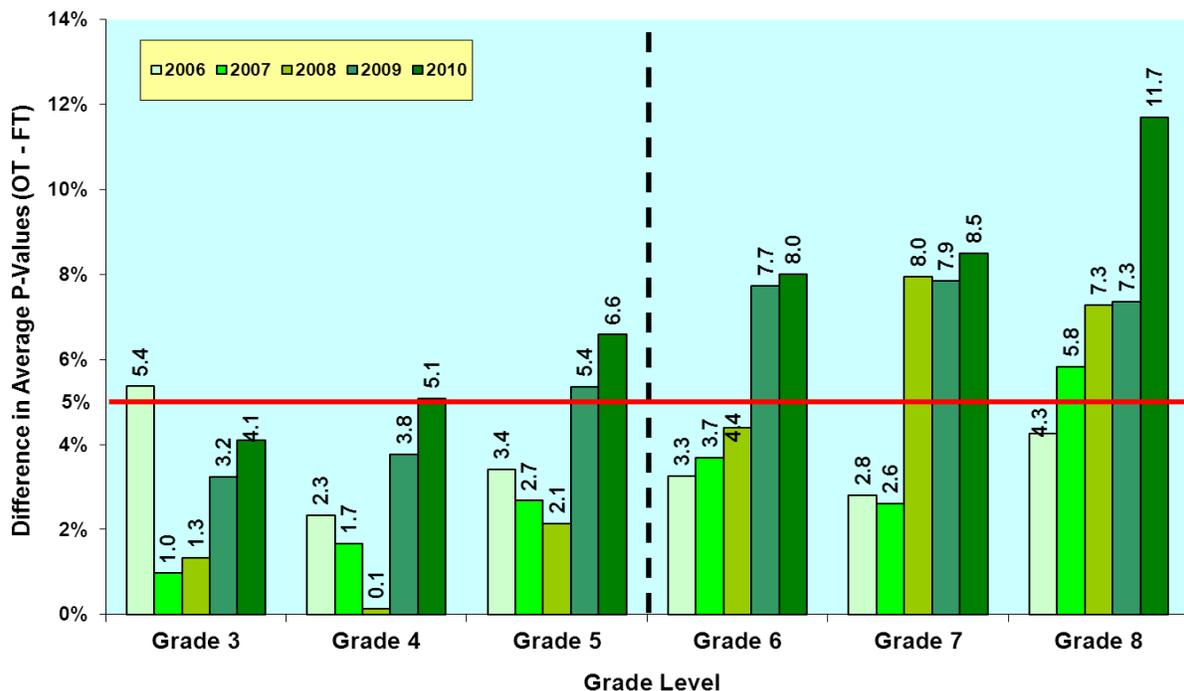
And the rub with the stand-alone field tests is their artificiality which dampens motivation. The state’s reliance on this approach yielded unreliable information about the items that appeared on the 2009 ELA and math exams. Students who took these field tests knew that performance on the items didn’t count for or against them. In short, they were inconsequential no-stakes exams and student indifference made the items seem harder than they later proved to be.

As reported by GothamSchools, SED’s chief testing adviser allowed that students who know they aren’t being graded on a test could be less motivated to do their best. “As a result, field tests often suggest that questions are more difficult than they actually are. And because they’re used to set the scale by which the real tests are graded, the end result is an easier state exam.”

The last statement acknowledges that the cut scores were set lower in 2009 because the stand-alone field test stats made the items appear to be relatively difficult. When the items proved to be easier on the operational test than had been projected, the results went off the charts, a combined effect of having easier items than expected and lower thresholds to reach.

I was able to analyze field test data that SED provided to me as a result of a FOIL request. The information enabled me to compare stand-alone field test item difficulty levels (aka p-values) on the tests (multiple-choice items combined with constructed response questions) by grade level with the corresponding statistics when these items appeared on the operational tests. The chart shows differences in average difficulty levels registered on the ELA from 2006 through 2010 for the same items based on the stand-alone try-outs vs the real exams. Higher p-values signal easier items.

**Extent to which ELA Operational Test Exceeded Field Test P-Values. 2006-10**



In each case, the average p-values for the items when the results counted exceeded the averages on the stand-alone field tests. The disparity appears to get larger over time and the gap is wider in 2010 compared to 2006. It is greater at the middle school than the elementary school level.

The findings are consistent with the conclusion that children do not put forth optimum effort on the field tests. Thus, the statistics that are generated underestimate how well students in the test population will do when the items appear on tests that matter. The fact that the field test data are not good predictors of operational test performance is problematic in terms of test development and to the extent that these data are used to inform decisions about where to set the cut scores.

[Note: In a separate analysis, the gaps between average field and operational test difficulties are even greater for the constructed response questions and underlie these findings.]

### **2009: Prelude to the Failure of New York State Core-Aligned Tests in 2013**

Having been made aware of the shortcomings of stand-alone field testing what has SED done since 2010? These have been the transitional years leading to a new assessment regimen brought forth in April 2013's ELA and math tests aligned to the Common Core Learning Standards.

The Contract with Pearson. In 2011 SED signed a five-year \$32 million contract with Pearson, Inc. to develop new tests that will be used to make high-stakes decisions about students, teachers and schools and say where they stand in relation to the Standards.

The Office of State Comptroller registered the agreement as a *consultant contract*, which is let in cases where “an individual or firm is hired to provide professional knowledge, skills or expertise. The award is usually made based upon an evaluation of both cost and technical expertise.” Such contracts cover services such as architecture, engineering, technology consulting, medical services, etc.” The development of state assessments is considered an etcetera.

So Pearson, a specialist and publisher in the field of testing, entered into an agreement with SED as both a consultant and vendor—guiding its client-agency in contract implementation, while selling SED goods and services in fulfillment of the deal.

The contract stipulated that embedded field testing procedures would be followed to produce multiple-choice items for the core-aligned tests for the duration of the agreement. This seemed like an improvement in reaction to what had happened when stand-alone field testing was done during CTB/McGraw-Hill's years as test vendor (and cited in the 2009 fiasco). But it proved immediately inadequate and problematic under Pearson and SED's new partnership agreement.

SED called for the embedding of too few items. As specified, only four sets of trial items per grade could be planted in the operational test booklets, inevitably resulting in an insufficient supply of items for potential use on subsequent operational tests.

For example, the four forms of the 2012 Grade 3 ELA test would contain its operational reading passages and associated multiple-choice items—material common to the forms. In addition, each form might contain a reading passage and six items being tried out for consideration for possible inclusion on the 2013 Grade 3 operational exam.

The decision on whether to select the material would be based on a review of the passages and the field test item statistics. But, at best, this arrangement meant that only four forms would be tried out, leaving a paucity of items to be field tested. To embed too many items on so few forms would lengthen the operational tests beyond practicable limits.

Coming into the contract, Pearson had to know that embedding try-out items on just four forms would not generate a large enough crop of items to pick from in constructing the next exams. It's not clear what Pearson, the vendor, advised SED to do. It is clear, however, that Pearson, the consultant, knew four forms alone couldn't work. The company held contracts in other states where many more embedded forms and try-out items were called for.

The upshot of this was that Pearson wound up doing a vast amount of stand-alone field testing in June 2012 to make up for the lack of embedded items so as to spin off enough multiple-choice items from which to assemble the April 2013 core-aligned ELA and math tests. The price tag for this substantial amount of extra work, which was not anticipated in the contract, remains unknown.

It's important to bring SED/Pearson's resort to stand-alone field testing in June 2012 into sharp relief. These field tests became the primary basis for constructing this year's first-ever core-aligned tests. Up to 20 stand-alone forms per grade and subject (ELA or math) were administered to schools throughout the state. Altogether the sampling plan gave field test assignments to nearly **1,900 schools** across New York State and targeted almost **400,000** students to take them.

| <b>Summary - June 2012, NYS and NYC Public Schools</b>                                                                                                                                                                |                    |                  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|------------------|
| <b>Assigned to Give ELA and Math Stand-Alone Field Tests</b>                                                                                                                                                          |                    |                  |
| <b>Estimated Number* of Students in Grades Targeted for Tests</b>                                                                                                                                                     |                    |                  |
| <b>ELA 2012</b>                                                                                                                                                                                                       | All Public Schools |                  |
|                                                                                                                                                                                                                       | April 2012         | June 2012        |
| Counties                                                                                                                                                                                                              | N-Tested           | Assigned Schools |
| New York State                                                                                                                                                                                                        | 227,401            | 1,026            |
| New York City                                                                                                                                                                                                         | 83,017             | 543              |
| Outside of New York City                                                                                                                                                                                              | 144,384            | 483              |
| <b>Math 2012</b>                                                                                                                                                                                                      | All Public Schools |                  |
| New York State                                                                                                                                                                                                        | 170,157            | 872              |
| New York City                                                                                                                                                                                                         | 62,233             | 455              |
| Outside of New York City                                                                                                                                                                                              | 107,924            | 417              |
| <b>ELA and Math</b>                                                                                                                                                                                                   | All Public Schools |                  |
| New York State                                                                                                                                                                                                        | <b>397,558</b>     | <b>1,898</b>     |
| New York City                                                                                                                                                                                                         | <b>145,250</b>     | <b>998</b>       |
| Outside of New York City                                                                                                                                                                                              | 252,308            | 900              |
| *Estimated number of students targeted for June field testing is based upon NYSED data on the number of students who took the April ELA and math tests in the schools and grades matching the field test assignments. |                    |                  |

In New York City about **1,000** schools had assignments seeking an estimated **145,000** students to try out items that would be used to create the April 2013 operational tests and inaugurate Common Core testing.

Given the sheer magnitude of this effort one would think it made sense. But SED knew openly from 2009 that stand-alone procedures had grave pitfalls. The field testing that was conducted back then had at least been planned.

In June 2012, the field tests had all the disadvantages inherent in the stand-alone approach. And the contract didn't call for them. In fact, field testing in June is arguably the worst time of year to try out items. At such a late point in the year the last thing students in grades 3 through 8 want to see is another test—students who have spent a year taking test after test, being asked to take another one that they know doesn't count.

Where did Regent Tisch stand on this matter? Where was Pearson when it came to consulting about these factors? What happened to the technical advisory group? It seems that common sense had to be abandoned in order to make the Common Core an inevitability.

But the questionable stand-alone field testing process became the launching pad for April's first core-aligned assessments and making the Common Core Standards a fait accompli in New York.

Last May the New York City Council passed a resolution calling upon the New York State Department of Education to immediately stop all stand-alone field testing of students. It came shortly after the 2013 exams and spoke in favor of embedded field testing and against the impending stand-alone field tests that were once again scheduled in June.

Two of the supporting clauses are noteworthy: *“Stand-alone tests are neither mandated nor necessary and their validity is questionable because if students know the tests do not carry any consequences they may not be motivated to perform well; and whereas, if using the field questions that were embedded in the April exams is insufficient (for the 2014 operational exams), the State should explore alternatives such as conducting field tests outside of school time while compensating the test takers, as is done for adults.”* [My statement to the Council in support of its action is appended.]

Though strongly worded, the resolution had no binding effect on SED. June 2013's field test assignments proceeded school by school. Pearson did stand-alone testing of multiple-choice items once more to make up for too few of embedded items. Statewide over **2,000** schools were invited to participate with **345,000** students targeted to help Pearson develop items. New York City accounted for more than **900** of the schools and **127,000** of the students. The cost for the additional work is a mystery. It was a service not mentioned in the contract.

A number of troubling aspects of the testing program emerge from this:

- It is a given that faulty field testing leads to the development of poor operational tests. In this case, the stand-alone approach was the foundation for the first core-aligned exams in April. Therefore, the new wave of measurement is off to a poor start.

- Yet, SED insists that this year’s ELA and math tests will provide a **baseline** against which to measure the achievement and growth of students in meeting the new standards. How solid can the baseline be if it was largely established on the grounds of unreliable June 2012 field test data?
- When the operational tests ensued in April, parents and teachers reported that children didn’t have time to complete them; that the test content was too difficult for students and inappropriate to their age range; that younger children were frustrated, upset and dispirited by the material. These observations were too widespread to be dismissed as anecdotal.
- SED’s and the City Department of Education’s response has been that it was necessary to make the tests tougher in order to have them reflect standards being put into place that emphasized critical thinking, a deeper understanding of math concepts, the reading of more complex texts, and a higher level of writing—all aims of the common core.
- Well-conceived field testing irons out issues of test administration, structure and content, including matters of test length and proper time limits. The timing and methodology of the June 2012 field tests made them a failure. The negative experiences students had with the 2013 operational exams and the dismal results sealed the case.
- Less time was allotted on the 2013 versus the 2012 multiple-choice items. These items weigh most in the test results. This certainly isn’t a way to test whether students possess the deeper qualities of thought that the common core seeks to develop.

| <b>ELA - Average Time Allocated to Multiple-Choice Items (Test Booklet 1), 2013 vs 2012</b>           |                        |                           |                           |                                   |                        |                           |                           |                           |
|-------------------------------------------------------------------------------------------------------|------------------------|---------------------------|---------------------------|-----------------------------------|------------------------|---------------------------|---------------------------|---------------------------|
| <b>2013 English Language Arts</b>                                                                     |                        |                           |                           | <b>2012 English Language Arts</b> |                        |                           |                           | <b>2013 v 2012</b>        |
| <b>Grade</b>                                                                                          | <b>Number of Items</b> | <b>Allotted Test Time</b> | <b>Avg. Time per Item</b> | <b>Grade</b>                      | <b>Number of Items</b> | <b>Allotted Test Time</b> | <b>Avg. Time per Item</b> | <b>Percent Difference</b> |
| 3                                                                                                     | 30                     | 70                        | 2.33                      | 3                                 | 36                     | 90                        | 2.50                      | <b>-6.7%</b>              |
| 4                                                                                                     | 30                     | 70                        | 2.33                      | 4                                 | 37                     | 90                        | 2.43                      | <b>-4.1%</b>              |
| 5                                                                                                     | 42                     | 90                        | 2.14                      | 5                                 | 39                     | 90                        | 2.31                      | <b>-7.1%</b>              |
| 6                                                                                                     | 42                     | 90                        | 2.14                      | 6                                 | 39                     | 90                        | 2.31                      | <b>-7.1%</b>              |
| 7                                                                                                     | 42                     | 90                        | 2.14                      | 7                                 | 39                     | 90                        | 2.31                      | <b>-7.1%</b>              |
| 8                                                                                                     | 42                     | 90                        | 2.14                      | 8                                 | 39                     | 90                        | 2.31                      | <b>-7.1%</b>              |
| <b>3 - 8</b>                                                                                          | <b>228</b>             | <b>500</b>                | <b>2.19</b>               | <b>3 - 8</b>                      | <b>229</b>             | <b>540</b>                | <b>2.36</b>               | <b>-7.0%</b>              |
| <b>Math - Average Time Allocated to Multiple-Choice Items (Test Booklets 1 &amp; 2), 2013 vs 2012</b> |                        |                           |                           |                                   |                        |                           |                           |                           |
| <b>2013 Mathematics</b>                                                                               |                        |                           |                           | <b>2012 Mathematics</b>           |                        |                           |                           | <b>2013 v 2012</b>        |
| <b>Grade</b>                                                                                          | <b>Number of Items</b> | <b>Allotted Test Time</b> | <b>Avg. Time per Item</b> | <b>Grade</b>                      | <b>Number of Items</b> | <b>Allotted Test Time</b> | <b>Avg. Time per Item</b> | <b>Percent Difference</b> |
| 3                                                                                                     | 61                     | 140                       | 2.30                      | 3                                 | 58                     | 180                       | 3.10                      | <b>-26.0%</b>             |
| 4                                                                                                     | 62                     | 140                       | 2.26                      | 4                                 | 62                     | 180                       | 2.90                      | <b>-22.2%</b>             |
| 5                                                                                                     | 62                     | 180                       | 2.90                      | 5                                 | 60                     | 180                       | 3.00                      | <b>-3.2%</b>              |
| 6                                                                                                     | 68                     | 180                       | 2.65                      | 6                                 | 60                     | 180                       | 3.00                      | <b>-11.8%</b>             |
| 7                                                                                                     | 68                     | 180                       | 2.65                      | 7                                 | 62                     | 180                       | 2.90                      | <b>-8.8%</b>              |
| 8                                                                                                     | 68                     | 180                       | 2.65                      | 8                                 | 62                     | 180                       | 2.90                      | <b>-8.8%</b>              |
| <b>3 - 8</b>                                                                                          | <b>389</b>             | <b>1000</b>               | <b>2.57</b>               | <b>3 - 8</b>                      | <b>364</b>             | <b>1080</b>               | <b>2.97</b>               | <b>-13.4%</b>             |

- My analysis of the data shows that the average time per item decreased by 7.0% on the core-aligned ELA and by 13.4% on the math items. The decrease was sharpest in math in grades three and four where average time went down by 26.0% and 22.2%.
- In addition, over the full course of the three-day testing period, the 3<sup>rd</sup> and 4<sup>th</sup> grades tests in ELA and Math are the only ones where the testing time decreased—by 22% in both grades for the ELA and by 22% and 15% for math in the two grades.
- Allowing less time to complete the tests contradicts the values and purposes espoused by the Common Core—richer cognitive development and understanding big ideas—turning the exams into speeded rather than power tests. Providing insufficient time to finish more challenging material nullifies whatever intrinsic logic the Common Core professes to have.
- The omission rate question is not a quibble. Giving children too little time to finish a test is a mark of poor planning and, once again, poor field testing.
- In the matter of children not having time to finish the items, SED Associate Commissioner for Assessment Ken Wagner reported that a routine analysis showed an average omission rate on the test items of 3% to 5%. He said this was typical. It served to refute what parents had been saying about an inordinate number of children being unable to complete the tests. It also implied that the testing parameters SED and Pearson had imposed were justifiable.
- My study of omission rates from 2006 to 2012, based on item analyses data presented in the publishers' Technical Reports, found a much lower rate than the level he cited for 2013—meaning that the new tests evinced higher than usual levels of incompleteness.
- On its face then, if the 3% to 5% averages are accurate, Mr. Wagner's study confirms what parents and teachers have said, although his purpose seems to have been to blunt their critical observations that the tests did not allow time to finish or forced students to rush their answers in a way that was detrimental to performance.
- Here is a deeper question to address than resolving differences in the instance of omissions: When officials go forward with findings, shouldn't they be obligated to produce the data on which they base their conclusions?

### **Sunshine, Transparency and Truth-in-Testing**

The lack of data on omissions is symptomatic of a pervasive problem that hovers over the State and City management of education and the handling of information, parents and the public.

It appears that executives can make claims without having to submit supporting data and analyses. This creates an imbalance. It gives them latitude to make assertions, often self-serving, with little chance that the conclusions they reach and the policies/programs that follow can be examined. It encourages increasingly bold decisions and actions to be taken in a high-handed manner. It is the opposite of good government and accountability.

Since the data are ultimately paid for by tax dollars, a mechanism must be created to ensure that whenever an "official" claim is made, especially one justifying an important action—the data upon which it relies must be made available to the public expeditiously.

Unfortunately, the FOIL process does not meet that need. FOILs have become a device agencies have used in a cumbersome, seemingly arbitrary way to deny or delay access to requested information. This thwarts timely meaningful debate that could preclude or alter an unwise action.

In the case of the omissions rate, the impression Mr. Wagner created is that 2013's rate was normal. This could be an honest interpretation of the data he has, or a mistaken understanding of the information, or a statement calculated to publicly defend a testing process against legitimate doubts raised by parents and educators—in order to rush ahead to next year before facts can get in the way of implementation. We won't know until he is moved to let us see for ourselves.

In brief, here are some discouraging developments I have observed in the two years since Pearson won its 5-year contract with SED. They depend on a one-two punch—disclosing less data or making them more difficult to obtain than in the past combined with SED's habit of keeping parents and the public in the dark.

#### Less Data and More Difficult to Obtain.

- Through 2010, ELA and math tests were posted on SED's Office of State Assessment web pages within weeks after the exams were given. Anyone could see the operational tests and the answer keys. SED has stopped providing online access to the test material—citing cost factors and possible re-use of the operational items (even though they had been exposed); the need to discourage more time from being devoted to test preparation than release of the items would promote; and the need to protect embedded field test items from being seen.
- I also had been able to FOIL for and acquire field test item statistics for every item selected for operational use. This allowed comparison of the field test items with corresponding operational items to see how well the two data sets matched—a window into test research procedures and the quality of the statewide exams. In 2011 only partial information was returned. In response to phone calls for the 2012 and 2013 data I was told that the stats weren't being released for fear the field test items would be seen—a contrived fear to justify SED reluctance to provide the stats—knowing the items didn't need to be revealed..
- By contract, Pearson is obligated to produce two reports each year. It is responsible for delivering a Technical Report that includes an analysis of all items—their difficulty levels and how well they functioned, including omission rates. The report is due in December. The 2012 Technical report was not posted until July 2013 (although it bore a 2012 date). This prevented scrutiny of 2012's operational tests until after April 2013's core-aligned exams had been given. Whatever knowledge might have been gained from the report pertinent to construction of the 2013 exams was rendered useless. This is consistent with SED's effort to write off the 2012 exams as being transitional and not comparable to 2013.
- The contract also calls for a Field Test Technical report due in December. The 2012 Field Test report has not yet appeared—ten months beyond the delivery data. It would shed light on how many schools and students actually participated in the stand-alone field tests (by grade, separately for ELA and math), underscoring the fact that the tests are not mandated. And the report would provide field test items statistics—hopefully linking stats for the items that were selected for inclusion on the operational tests with their operational item stats.

- Prior to 2011 preliminary operational item stats were posted online. The preliminary data hardly varied at all from the data in the final report. Therefore, I believe that preliminary item analysis information is available (particularly in light of Mr. Wagner's 3% to 5% findings about the average percentage of omissions). There is no excuse at this time for keeping the 2013 operational item data a secret. The information is known to SED. Its disclosure would not entail revelation of the items per se. Nor would presentation of the data entail increased costs of future exams.
- In the same hidden way, there has never been an explanation about how much it cost to have Pearson provide two extra rounds of stand-alone field tests in June 2012 and 2013 or what the cost was in terms of teacher hours or the value of classroom time they took up.
- An item on the State Comptroller's Open Book web site indicates that an amendment was approved on September 25<sup>th</sup> paying Pearson an additional \$2.8 million for a change in the scope of the contract. What were the changes and why were they necessary?
- Before we are steamrolled into next April's operational tests, I am asking the Board of Regents to demand that the embedded and stand-alone field test data gathered in 2012 and the corresponding 2013 operational test data be released so that a proper assessment can be made of the 2013 exams—the initial core-aligned ELA and math tests that were just given to serve as the baseline against which next April's results will be measured. They should also order a public accounting of how much it has cost to fulfill the terms of the contract and to pay for modifications.
- I respectfully ask the Senate Committee on Education to use its considerable influence to bring these data to light.

Keeping Parents and the Public in the Dark. I know that in recent weeks, this committee has heard heartfelt testimony from parents about the advent of the Common Core and the disquiet it has caused them. I want to add a few test-related points regarding the top-down conduct of the program and the way the largest educational constituency, parents and their children, have been treated in order to advance the Common Core.

- First, Change the Stakes is a group of parents working on testing and other school issues. We have developed material for parents and, for the record, I am appending one of our handouts to this statement. It calls for Sunshine, Transparency and Truth-in-Testing.
- Truth-in-testing is a good place to start since so much time and worry revolve around testing children each year. The testing program invites distrust because SED and DOE have not been required to demonstrate to parents and educators that the results are valid for reaching high-stakes decisions about students, teachers and schools.
- The Standards for Educational Testing prepared by the American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education has provided professional guidelines about test construction. SED and Pearson have disregarded the well-respected principles and good practices that the Standards embody. This is flagrantly wrong when it comes to providing evidence of validity and data that supports the purposes and decisions its tests are intended to serve.
- Change the Stakes wants to help bring the need for Truth-in-Testing to the attention of legislators and the media. Action is needed this year that will allow parents to review test material, check for scoring accuracy and see that the scores are used properly, fairly and

within their limits. A process must be put in place for test-takers to challenge errors and appeal decisions—many of which appear to be ad hoc and arbitrary.

- We applaud the work of Senator Lavallo over the last 30 years in championing the cause of Truth-in-Testing. We support and urge him to revise the law and put forth strong winning legislation in the coming session.
- Parents (and teachers) must be kept informed in a timely way about educational programs and parental consent must be obtained for children to participate in them.
- Stand-alone field testing has been an area in which SED has gone out of its way to make parents unaware of the tests. A copy of the 2011 Field Test School Administrator’s Manual says that “parents should be informed of the dates of testing and the purpose of the tests” under the heading “Preparation of Students for Field Testing.” A year later, under the same heading, this sentence was deleted from the Manual.
- Over the last three years, SED has been able to create a vacuum of knowledge about field testing by issuing memos to superintendents and principals throughout New York vaguely describing testing that would be coming along and offering more information later in the year. The recipients of these memos have not been advised to inform parents the field tests in any systematic way.
- In 2012 and 2013 parents began protesting the field tests and objecting to the fact that they had not been informed. Some boycotting took place because the tests are not mandated. (In fact, the RFP leading to the contract indicated that bidders should plan field testing to oversample students because there would likely be a 30% rate of non-participation.)
- In addition to being opposed to excessive testing, the argument against the field test rests on the belief that parents and guardians should be informed of school-based activities their children may be involved in; that field testing is part of a commercial research and development process; that offering children and classroom time to the service of test publishers without parental knowledge and consent is a form of exploitation; that subjects who agree to participate in such projects are usually compensated; and that this action and the covert way it has been taken is otherwise arrogant and disdainful of parents.
- Evidently SED thinks it has an unqualified right to continue to operate in this one-way manner. A startling memo to superintendents and principals was posted on its web site dated August 2013—from Ken Wagner. It announces an ambitious plan to conduct **not one but two** separate stand-alone field test projects next spring. They would come in the middle of the operational ELA and math tests. Here’s the link:  
<http://www.p12.nysed.gov/assessment/fieldtest/updateparccmemo>.
- No doubt they will be on the same scale that targeted approximately **750,000** students in grades 3 through 8 over the last two Junes, including **275,000** in the City. The same stand-alone approach will be replicated—the one that failed in 2009, 2012 and 2013.
- It seems as if SED sees its mission as running interference for Pearson and the Common Core. It operates imperiously. It is not being forthright with parents and the public.

In dealing with a program such as the Common Core, imposed on such a grand scale, wisdom requires that the public have the right to a full account of what’s happening—the truth, if you will. Truth-in-Testing must stand high on our agenda as a pillar of principle, making sunshine and transparency top priorities in governance and educational reform.

Finally, in the many voices this committee has listened to across New York State in the last six weeks there is one message: The Common Core is an express engine that can’t take children to the top of the mountain. Parents and teachers are the steady little engine that can.

## **Appendix**

**May 21, 2013-Statement by Fred Smith, City Council Hearing on Resolution Calling for New York State Department of Education to Immediately Stop All Stand-Alone Field Testing for Students**

I strongly support the resolution under consideration concerning stand-alone field testing.

Stoppage of the June field tests will end the charade that this procedure is needed to ensure the development of good instruments for next April's exams.

Field testing in the public schools allows publishers like Pearson to do market research for free on the products it will sell us in the future. In this case, the New York State Education Department (SED) is letting Pearson try out test items in June to see how samples of students react to them.

Are the items too hard or too easy? Are they confusing or un-discriminating? Publishers gather statistics on how each item performs—not to measure student achievement—but to evaluate their potential for use on subsequent operational exams, like the ones that will be given next year.

The fatal flaw inherent in the stand-alone approach is that the information obtained on each item is unreliable. And items are the bricks used to build tests. Constructing exams based on weak data makes them meaningless. The problem is glaring when tests are used to reach high-stakes decisions, in such matters as student promotion, teacher effectiveness and school ratings.

Why are the data unreliable? Because students taking the field tests know the results have no consequences for them. The results don't count, it's June and students have been battered by tests all year. They are not motivated to do well on the stand-alone field tests. The data they generate on each item are not predictive of how students will respond to the same items when they appear on tests that count.

This is not a theory. The English Language Arts (ELA) and math tests that students took just four weeks ago were developed from last June's (2012) stand-alone field tests. Reports from teachers and parents indicated that students couldn't finish these exams; that the items were extremely hard and beyond the grade level of the students taking them; and less time was allotted to each item than the year before. If students had taken the June 2012 stand-alone field tests seriously, these shortcomings would have been identified and the problems would have been addressed.

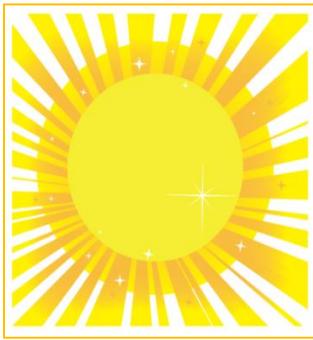
Now, SED wants us to embark on the same excursion we took last year—replicating stand-alone field test procedures that just failed us miserably, while the evidence of that failure is still warm.

It has tapped no fewer than 1,282 New York City public schools for field testing. **1,282!!** That will involve spending more than just one hour taking the field tests. The effort and disruption involved in these assignments add up to sacrificing at least a full school day—to say nothing of the vast costs in lost teacher salary and learning opportunities.

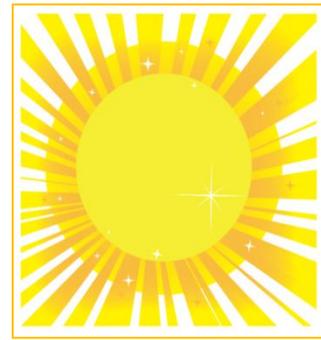
And beyond perpetuating the same kind of damaging high-stakes testing (thereby precluding better assessment techniques from emerging), continuing the stand-alone field tests imposes ethical questions that SED ignores. Why aren't children treated as subjects in a commercial research and development project and paid for their time? And why are parents kept in the dark about the field tests—keeping them uninformed about procedures they might otherwise object to?

In sum, stand-alone field tests yield shaky data, rendering tests that rely upon the information invalid and supporting the conclusion that this approach to field testing is a non-starter. Without doubt, halting the June stand-alone field tests is the right action for the City Council to take.

*Fred Smith, a testing specialist and consultant, retired as an administrative analyst for the New York City public schools. He is a member of Change the Stakes, a parent advocacy group.*



**SUNSHINE**  
**TRANSPARENCY**  
**TRUTH -IN- TESTING**



As parents, we all know that sunlight helps every living thing stay healthy and grow (unless you are a mushroom). We at Change the Stakes know daylight, warmth and fresh air not only nurture our children, but we want to be sure their schools also let the sun shine in.

Change the Stakes has spoken with parents who agree. They feel that they have been kept in the dark for too long by the New York State Education Department [SED] and the New York City Department of Education [DOE]. Many believe that policies and decisions that have an impact on their children come crashing down on them with little respect or attention given to their concerns. We want to ensure that We, the Public, are not left out of public education.

For that to happen, the top education officials in Albany and those following orders from City Hall must take actions in ways that are transparent to all individuals interested in quality schools for all children. These include policy-making and budget-making decisions that are equitable to everyone. They also include keeping parents (and teachers) informed in a timely way about new programs and seeking parental consent for children to participate in them.

Transparency means administrators need to work in ways that are obvious and free of deceit; that are readily understood and open to review and input from parents. They must act in the light of day—providing direct access to all evidence needed to evaluate their plans before and after they are approved and implemented.

Transparency comes before accountability. There is no real accountability without it. And *transparent* doesn't exist without each *parent* having a chance to be engaged in vital school-related matters.

Since so much (too much) time and worry revolve around testing children each year, perhaps, the best starting point to bring sunshine and transparency into the schools is with the state's ELA and math testing program. One of the reasons this program has been so harmful is because SED and DOE have not been required to give parents and educators information about how the tests were developed and that prove the results are valid for reaching high-stakes decisions about students, teachers and schools.

Change the Stakes is reaching out to parents and teachers to help us gain strength and bring the need for Truth-in-Testing to the attention of legislators and the media. Action is needed this year so we can review test material, check for scoring accuracy and see that the scores are used properly, fairly and within their limits. SED and DOE must be held accountable for disclosing the contents of exams, as well as data about their quality. A process must be put in place for test-takers to challenge errors and appeal decisions—many of which appear to be arbitrary.

Please join with us in our sunshine campaign: to insist on transparency in the way schools are governed and operate and to ensure that truth-in-testing becomes the norm for all aspects of test administration. Visit our website at [www.changethestakes.org](http://www.changethestakes.org) Thank you.