



Class Size Matters
124 Waverly Place, NY, NY 10011
phone: 212- 674- 7320
www.classsizematters.org
email: info@classsizematters.org

How high-stakes testing is unreliable and undermines the quality of education in NYC Schools

Testimony to the NYC Council Education Committee

Leonie Haimson, Class Size Matters

November 25, 2013

Thank you for holding these hearings today, Chair Jackson. Over the last few years, the state and the city have increased their emphasis on high-stakes exams, just as the NY State Education Department has shown itself to be unable to produce a decent, reliable exam with a credible scoring system in at least ten years.

That's why there have been wild gyrations from year to year in the percent of students making the grade. For example, 77% of NYS students were at level 3 or 4 in English in 2009; this dropped to 53% in 2010 and 31% in 2013..

For nearly a decade, from at least 2003-2010, there was rampant test score inflation, with many of the same state and city officials who are now proclaiming that the scores are too low and our schools are failing claiming with equal conviction that the earlier, rising test scores showed that NYC schools were improving rapidly.

The state test score bubble allowed NYC Mayor Bloomberg renew mayoral control and win a 3rd term by maintaining that his high-stakes testing regime was working, when the reality was that, according to everyone who was paying attention, the exams had gotten overly predictable and the scoring far too easy over time.

At the same time as the state exams showed huge increases, scores on the more reliable national exams called the NAEPs showed little progress. In fact, [NYC made smaller gains](#) on the NAEPs than nearly any other large school district in the country during these years.¹

The last two years of exams created by Pearson have been especially disastrous; from the [multiple errors in questions and scoring](#) on the 2012 exams (including the [infamous Pineapple passage](#), which we wrote about first on the NYC parent blog) to the [epic fail of this year's tests](#)

¹Class Size Matters, "NYC Achievement Gains Compared to Other Large Cities since 2003," January 2012, <http://www.classsizematters.org/nyc-second-to-last-among-cities-in-student-progress-on-the-naeps-since-2003/>.

– which were too long, riddled with ambiguous questions and replete with commercial logos for products like Mug Root Beer.²

Top students were unable to finish these shoddy exams, and many left in tears and had anxiety attacks. To make things worse, the exams featured reading passages drawn straight from Pearson [textbooks](#) which were assigned to some students in the state and not to others.³

In fact, the cut scores for proficiency were set so high that parents of students in districts on Long Island and Westchester -- where more than 90% graduate from high school in four years, go onto four year college where they do very well -- were told that most of their children also failed the exams and would not be college and career ready. These parents have clearly said they don't believe these results, they don't trust SED and they don't want their kids discouraged by their results on exams that are so flawed and have so little credibility.

The truth is that the new cut scores that determine the different proficiency levels on the state exams – which decide how many kids “pass” or are at Level 3 and 4 -- are arbitrary and set by Commissioner King. He can set them to create the illusion that our schools are rapidly improving, as the previous Commissioner did, or he can set them to make it look that our public schools are failing, as Commissioner King now is doing, to bolster support for his other policies – including the Common Core, student personal data-collection and sharing with vendors, test-based teacher evaluation and charter expansion.

The primary evidence that Commissioner King now bases his overly-harsh cut scores upon is that the results are mirror the percent of students who test “proficient” or above on the NAEPs. Yet while the NAEPs are reliable to discern trends in test scores, because they remain relatively stable over time, the cut scores that determine the different achievement levels are VERY controversial. Even the National Academy of Sciences has questioned the setting of the NAEP proficiency levels and points out:

Just one third of American fourth graders were said to be proficient in reading by NAEP in the mid-1990s at the very time that international assessments of fourth-grade reading judged American students to rank Number Two in the world.⁴

Though Deputy Chancellor Suransky has claimed he has added funds for intervention services and support for struggling students the opposite has happened. Class sizes have gone up

² See: NYC Public School Parents, “NYS Educators agree: Flawed, confusing and misleading exams,” April 18, 2012, <http://nycpublicschoolparents.blogspot.com/2012/04/nys-educators-agree-flawed-confusing.html>. See also: Leonie Haimson, “The Lessons of Pineapplegate,” *WNYC Schoolbook*, June 6, 2012; NYC Public School Parents, “This year’s NYS/Pearson ELA exams: an Epic Fail,” April 17, 2013, <http://nycpublicschoolparents.blogspot.com/2013/04/looking-for-feedback-on-this-years-ela.html>.

³ Kathleen Porter-Magee, “The problem with Pearson-designed tests that threatens thousands of scores,” *Washington Post Answer Sheet*, May 6, 2013.

⁴ James Harvey, executive director of the National Superintendents Roundtable, “NAEP: A flawed benchmark producing the same old story,” *Washington Post Answer Sheet*, November 4, 2011, http://www.washingtonpost.com/blogs/answer-sheet/post/naep-a-flawed-benchmark-producing-the-same-old-story/2011/11/03/gIQAAbnonmM_blog.html.

sharply, for the sixth year in a row, and in the early grades are now the largest in 15 years.⁵ Schools are now spending one day per week of the 37 ½ minutes meant for small group instruction for struggling students instead on scoring tests and prepping for the teacher evaluation system. Some schools are even closing for half days to spend even more time on these purposes. None of this is helping kids learn.

Teacher evaluation based on value-added test scores is just as unreliable – even if one believes that test scores were the best evidence for a good teacher (which I don't)--with huge volatility from one year to the next.⁶

The school progress reports also swing widely from year to year as they are based in one year's changes in test scores – which research shows is 40-80% random. In fact James Liebman, who first devised the school grading system for DOE, promised that the grades would soon be based on three years of test scores but reneged on his promise soon after.⁷

DOE officials are very aware these metrics are unreliable, but like the teacher evaluation system, it allows them to brand more schools as failing or ineffective – to more easily close them. .

So why are NYSED and DOE officials determined to prove that more than 69% of the students throughout New York State are failing? This is the [Shock Doctrine](#) at work.⁸ Naomi Klein has observed that when you scare people enough, it is easier to persuade them to allow you to make whatever radical changes you want, since the status quo will be perceived as so disastrous.

In the case of SED and DOE, they want to convince parents that their corporate agenda, will somehow improve the quality of education in our state, even though there is little or no evidence to back up any of these policies.

Rick Hess, the conservative commentator at Education Week, revealed the motives behind the promoters of these exams in a column called the "[Common Core Kool-aid](#)":

First, politicians will actually embrace the Common Core assessments and then will use them to set cut scores that suggest huge numbers of suburban schools are failing. Then, parents and community members who previously liked their schools are going to believe the assessment results rather than their own lying eyes...

⁵ Class Size Matters, *Class Sizes Increase for the Sixth Year in a Row; Including Data Analysis and Charts*, November 18, 2013 at <http://www.classsizematters.org/class-sizes-increase-for-the-sixth-year-in-a-row-class-size-report/>; see also Rachel Monahan, "New York City class sizes average 25 children after growing six years", *NY Daily News*, Nov. 18, 2013.

⁶ Gary Rubinstein, "Analyzing Released NYC Value-Added Data Part 1-6," Feb. 26, 2012 <http://garyrubinstein.teachforus.org/2012/02/26/analyzing-released-nyc-value-added-data-part-1/>

⁷ See Beth Fertig, *Why cant U teach me 2 read? Three Students and a Mayor Put Our Schools to the Test*, Macmillan, September 15, 2009, p. 121 .

⁸ Naomi Klein, *The Shock Doctrine: The Rise of Disaster Capitalism*, New York: Metropolitan Books/Henry Holt, 2007.

Finally, newly convinced that their schools stink, parents and voters will embrace "reform." However, most of today's proffered remedies--including test-based teacher evaluation, efforts to move "effective" teachers to low-income schools, charter schooling, and school turnarounds--don't have a lot of fans in the suburbs or speak to the things that suburban parents are most concerned about....Common Core advocates now evince an eerie confidence that they can scare these voters into embracing the "reform" agenda.⁹

The Common Core is similarly flawed; the two “architects” who were hired by the Gates Foundation, David Coleman and Jason Zimba, neither one of whom ever taught K12 a day in their lives. There was little or no input from working classroom teachers or parents but plenty from the testing companies.

The philosophy behind the Common Core which is the same standardization for all – is bound to fail; and punish those kids, teachers and students most who are most at risk.

Joel Klein, who wrote an oped for Rupert Murdoch’s NY Post in support of the new exams, appropriately entitled the [The Good News in Lower Test Scores](#), now heads Amplify, Rupert Murdoch’s online learning division, which is the largest contractor for inBloom.¹⁰

For Klein and Murdoch, the drastic fall in state test scores is indeed good news, because it will make districts desperate to improve results and help them market their computer tablets, data systems, and software products to make more profit.

In the case of Pearson, the world’s largest educational corporation, more schools will now be convinced to buy their textbooks, workbooks, and test prep materials, as [900 NYC schools have now done](#) – in hope that their students may do better on the Pearson-state exams, that may even include the same reading passages as happened this year.¹¹

To achieve their ideological ends, politicians, corporations, and educrats are not only willing to define your children in terms of their test scores, but also to redefine them as failures – to help them implement their mechanistic, reductionist, and ultimately depersonalized vision of education.

⁹ Rick Hess, “The Common Core Kool-Aid,” *Education Week*, November 30, 2012, http://blogs.edweek.org/edweek/rick_hess_straight_up/2012/11/the_common_core_kool-aid.html.

¹⁰ Joel Klein, “The good news in lower test scores,” *New York Post*, August 7, 2013, <http://nypost.com/2013/08/07/the-good-news-in-lower-test-scores/>.

¹¹ “State Plans Record-Keeping Changes for Schools,” *NY1 News*, November 24, 2013, <http://www.ny1.com/content/news/education/199276/state-plans-record-keeping-changes-for-schools>.