Grade 3-8 ELA and Math Testing in New York State 2006 - 2014

An Informational Presentation
February 3, 2015
Purpose

- To inform parents, educators and other interested citizens about changes to the grade 3-8 ELA and math tests, and testing program, in New York State
- To draw a connection between those changes and student performance
- To bring this information into the conversation about the link between teacher and principal annual professional performance evaluation (APPR) and student test results
Reliability . . .

- . . . means that the assessment you give delivers results consistently from administration to administration.

- My bathroom scale is not reliable (this is true). When I step on it first thing in the morning it says I weigh 163 pounds. I step off. I step back on and it says I weigh 165 pounds. I step off. I step back on and it says I weigh 162. Why, you might ask, do I keep this piece of junk? I’m just too lazy to throw it out and buy a new one.
Reliability Slide Two

- It doesn’t matter to anyone except me whether I buy a new bathroom scale.

- But what if my monthly mortgage payment, or rent, was tied to my weight?

- What if, for every pound I gained, my payments went up by $100? And for every pound I lost they went down by $100?

- Then it would matter a lot and I would want to know for sure that my scale was reliable.
Reliability Slide Three

- Now let’s talk about APPR, or Annual Professional Performance Review, in New York State.

- At the present time 20% of teacher evaluations in grades 4-8, ELA and math are tied to the ELA and math tests.

- Governor Cuomo and Chancellor Tisch have proposed raising this to 40% or 50%.

- Principal evaluations and school and district ratings are also tied to test results.

- Isn’t it important for NYS tests to deliver results that can be fairly and accurately compared from year to year?

- WHAT IF THEY DON’T?
Take a deep breath . . .

- . . . The next few slides show a handful of different graphs, charts and tables.

- They also provide information that will help you interpret the data, and understand what it shows about the history of tests and testing in New York State.

- Take your time; read the corresponding text, and you will learn important facts about how much the tests, and the testing program, have been changed over the past decade.
This graph shows the percentage of NYS students achieving proficiency (scoring at levels 3 or 4) on the grade 3-8 ELA and math tests from 2006 through 2014. It appears that student performance peaked in 2009, then declined unevenly since, reaching a low in 2013. Or has it???
Let’s take a look “under the hood.”
Factors to Consider

- When in the school year the tests are administered
- Which corporation developed the tests
- The duration, structure and content of the tests
- Which standards are assessed—2005 standards, or the New York State Common Core Learning Standards?
- Transparency—are teachers and principals able to study the tests and test results to learn about student performance and adjust instruction accordingly? Can parents obtain detailed information about their student’s performance?
- Where are the “cut scores” which determine the boundary between Level 2 (below proficient) and Level 3 (proficient)?
From 2006 through 2009 it appeared that, under Commissioner Richard P. Mills, students were making steady progress toward the No Child Left Behind (NCLB) requirement of 100% proficiency by 2014.
2006-9: Specifics

- Tests were developed by CTB/McGraw-Hill. In 2005 only students in grades 4 and 8 were tested. Students in all grades 3-8 were tested beginning in 2006.

- Tests were administrated consistently in January (ELA) and March (math), over either 2 or 3 days, depending on grade and subject.

- Length ranged from 75 minutes in grade 3 and 5 ELA, and 85 minutes in grade 3 math, to 160 minutes in grade 8 ELA and math.

- Test questions were aligned to the 2005 learning standards.

- Teacher and principals could study the tests and adjust instruction. Data specialists received detailed technical reports about specific questions.

- Between 2006 and 2009 the “cut scores” were lowered, particularly in math, making it easier for students to achieve Level 3, or “proficiency.”
Math 2006-2009 Score Inflation

How to read this table: In 2006, grade 3 students had to earn 64.0% of 39 points on the math test to achieve Level 3, or “proficiency.” In 2009 grade 3 students had to earn 54% of 39 points. There was a change of -10 percentage points required for proficiency (percentages rounded).

<table>
<thead>
<tr>
<th>Math</th>
<th>2006</th>
<th>2009</th>
<th>Out Of</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>gr 3</td>
<td>64.0%</td>
<td>54.0%</td>
<td>39</td>
<td>-10</td>
</tr>
<tr>
<td>gr 4</td>
<td>57.0%</td>
<td>53.0%</td>
<td>70</td>
<td>-4</td>
</tr>
<tr>
<td>gr 5</td>
<td>59.0%</td>
<td>50.0%</td>
<td>46</td>
<td>-9</td>
</tr>
<tr>
<td>gr 6</td>
<td>57.0%</td>
<td>49.0%</td>
<td>49</td>
<td>-8</td>
</tr>
<tr>
<td>gr 7</td>
<td>60.0%</td>
<td>44.0%</td>
<td>50</td>
<td>-16</td>
</tr>
<tr>
<td>gr 8</td>
<td>56.0%</td>
<td>51.0%</td>
<td>69</td>
<td>-5</td>
</tr>
</tbody>
</table>
In 2010, under Commissioner David Steiner, tests were still based on the 2005 learning standards. Teachers could still study the tests and adjust instruction. Data specialists still had access to detailed reports. However, the ELA and math tests were administered later in the year. They were made more difficult and there were changes to the “cut scores,” or the level of achievement that defines “proficiency.”
This bar graph shows the change in the percentage of points required to achieve different levels of performance on the grade 4 ELA exams in 2009 and 2010.
In 2009 fourth grade students had to earn more than 35% of the available points on the ELA exam to move from Level 1 to Level 2.
They had to earn more than 61% of the available points (35 plus 26) to move from Level 2 to Level 3 on the ELA.
In 2010, fourth grade students had to earn more than 74% of the points available on the exam in order to score at Level 3.
In 2009 fourth-graders had to earn more than 51% of the grade 4 Math exam points to achieve Level 3. In 2010 they had to earn more than 71% of the available points. Cut scores on the other 2010 math and ELA tests were adjusted similarly.
Beginning in 2011, under Commissioner John B. King, the tests were made “secure.” Teachers could no longer review them to learn how students were being assessed, or where students had difficulty with specific questions. Although NY adopted Common Core standards beginning in 2010, the tests continued to assess the 2005 learning standards. The infamous “Pineapple and the Hare” question appeared on the 2012 grade 8 ELA test in the first round of assessments developed by the Pearson Corporation.
2011-12 Specifics

- Both the ELA and math were administered in May.
- 2011 tests were still developed by CTB/McGraw-Hill. Beginning in 2012 tests were developed by the Pearson Corporation.
- In both years tests still assessed student mastery of the 2005 learning standards.
- In 2011 tests ranged from 100 minutes over 2 days (gr 3 math) to 160 minutes over 3 days (grade 4 ELA).
- In 2012 all tests were 270 minutes long, over three days.
- Data specialists no longer had access to detailed technical information about specific questions.
- The “cut scores” defining proficiency remained relatively stable.
In 2013 and 2014 NYS ELA and math tests were aligned with the Common Core. Tests were shorter but the content and the types of questions were more difficult. In 2014 high achieving grade 8 students were exempted from the math test, if they took the algebra Regents exam. Between 2012 and 2013 the cut scores defining proficiency were lowered in most grades in ELA, but raised in most grades in math.
ELA 2012-2013: Change in Percent of Raw Score Points Required for Proficiency

<table>
<thead>
<tr>
<th>ELA</th>
<th>2012</th>
<th>2013</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>gr 3</td>
<td>75.0%</td>
<td>64.0%</td>
<td>-11</td>
</tr>
<tr>
<td>gr 4</td>
<td>71.0%</td>
<td>69.0%</td>
<td>-2</td>
</tr>
<tr>
<td>gr 5</td>
<td>73.0%</td>
<td>68.0%</td>
<td>-5</td>
</tr>
<tr>
<td>gr 6</td>
<td>71.0%</td>
<td>72.0%</td>
<td>+1</td>
</tr>
<tr>
<td>gr 7</td>
<td>78.0%</td>
<td>73.0%</td>
<td>-5</td>
</tr>
<tr>
<td>gr 8</td>
<td>79.0%</td>
<td>73.0%</td>
<td>-6</td>
</tr>
</tbody>
</table>

How to read this table: In 2012, grade 4 students had to earn 71% of the points on the ELA exam in order to achieve Level 3, or proficiency. In 2013 grade 4 students had to earn 69% of the points. This was a change of -2 percentage points from 2012 to 2013.
### Math 2012-2013 Change in Percent of Raw Score

#### Points Required for Proficiency

<table>
<thead>
<tr>
<th>Math</th>
<th>2012</th>
<th>2013</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>gr 3</td>
<td>75.0%</td>
<td>73.0%</td>
<td>-2</td>
</tr>
<tr>
<td>gr 4</td>
<td>59.0%</td>
<td>65.0%</td>
<td>+6</td>
</tr>
<tr>
<td>gr 5</td>
<td>59.0%</td>
<td>67.0%</td>
<td>+8</td>
</tr>
<tr>
<td>gr 6</td>
<td>57.0%</td>
<td>69.0%</td>
<td>+12</td>
</tr>
<tr>
<td>gr 7</td>
<td>54.0%</td>
<td>64.0%</td>
<td>+10</td>
</tr>
<tr>
<td>gr 8</td>
<td>54.0%</td>
<td>68.0%</td>
<td>+14</td>
</tr>
</tbody>
</table>

**How to read this table:** In 2012, grade 4 students had to earn 59% of the points on the math exam to achieve Level 3, or proficiency. In 2013 grade 4 students had to earn 65% of the available points. This was a change of +6 percentage points from 2012 to 2013.
Why did student performance on the ELA drop in 2013, when the bar for “proficiency” was lowered, and not raised?
Here are a few answers:

- The listening section, on which students generally did well in the past, was included in 2012 but eliminated in 2013.

- The 2013 exams had more reading passages than the 2012 exams. Reading passages were more likely to be dense, informational (non-fiction) text at high levels.

- The 2013 exams had more multiple choice questions, and the questions were more likely to have “better” and “best” answers, rather than “right” or “wrong” answers. It took longer for students to consider options and make choices.

- Many 2013 questions required students to return to reading passages repeatedly, in order to interpret specific words and lines. This was very challenging, distracting and time consuming especially for young children, students with disabilities and English Language Learners.

- Anecdotal reports suggest that many students did not complete the 2013 exams in the time allowed.
The Bottom Line?

- Since 2006 the NYS Education Department has switched test development companies, assessed new standards, altered the content of tests, adjusted the testing calendar, lengthened and shortened the tests, raised and lowered the cut scores, and eliminated transparency in the testing program.

- There are many reasons why test-based evaluation of students, teachers, principals and schools is bad science, bad policy, and bad for children.

- However, as long as State Ed regularly alters the measure used for professional performance evaluation and accountability ratings,

- Attempting to draw sound conclusions about student, teacher, principal or school performance based on test results is absurd.
Because they have been changed in so many ways, and so frequently, NYS tests are a “Bad Scale” by which to measure student learning, evaluate teachers, and rate schools.
Addendum
Given how frequently the NYS tests have been changed, is there a different measure that provides consistent information about student achievement?

- Yes, the NAEP, which stands for “National Assessment of Educational Progress.”

- NAEP tests are administered every other year, on a sampling basis, to students in grades 4 and 8, in every state, in reading, math, and science.

- NYS student results on the NAEP have been relatively stable over time:
NYS NAEP Results 2005 – 2013

% Proficient NYS NAEP 2005-2013

<table>
<thead>
<tr>
<th></th>
<th>2005</th>
<th>2007</th>
<th>2009</th>
<th>2011</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading 4</td>
<td>33</td>
<td>36</td>
<td>36</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>Reading 8</td>
<td>33</td>
<td>32</td>
<td>33</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Math 4</td>
<td>36</td>
<td>43</td>
<td>40</td>
<td>36</td>
<td>40</td>
</tr>
<tr>
<td>Math 8</td>
<td>31</td>
<td>30</td>
<td>34</td>
<td>30</td>
<td>32</td>
</tr>
</tbody>
</table>
In contrast to the NYS NAEP results, NYS ELA and math results have fluctuated widely. Scores clustered tightly in 2013 when State Ed made the policy decision to define “proficiency” on the ELA and math tests at a level consistent with the NAEP.

Note—in 2005 only students in grades 4 and 8 were tested in ELA and math. This graph shows the average of those results for students state-wide.
Finally,

- The next few slides illustrate the impact of these changes in testing on a single cohort of students—New York State students who were in third grade in 2009, fourth grade in 2010, and so forth:
In 2009, 75.8% of New York State’s third grade students achieved proficiency (Level 3) on the ELA exam. 92.9% of NY’s third graders achieved proficiency on the math exam. In 2010, when it appears that the same group of students suddenly did much worse in fourth grade, the tests had been made longer and more difficult. In addition, the bar for “proficiency” had been raised by 14 percentage points on the ELA, and 20 percentage points in math.
In 2011 and 2012 these students advanced through the fifth and sixth grades. The standards assessed, the structure of the exams, and the cut scores remained relatively constant. Testing times were increased.

In 2013, tests were aligned with the Common Core. The level of difficulty on the ELA was increased, and the math tests introduced new content and new types of questions. The minimum scores required for “proficiency” were lowered in ELA but raised in math.
Between 2013 and 2014 the NYS testing program remained relatively stable. Some assessments were shorter, but the content and format of the exams, and the level of performance required for proficiency, remained largely constant. In 2014 8th graders who took the algebra Regents exam were exempted from the NYS grade 8 math test.
Sources: NYSED.gov

Assessment section: Test administration manuals

Information and Reporting Services section: Raw-to-Scale Score charts, Scale-to-Performance level charts, technical manuals, annual assessment result reports