

Statement by the Council of the Great City Schools
On 2013 DC-CAS Scores

Michael Casserly, Executive Director
Ray Hart, Director of Research¹

The Council of the Great City Schools, the nation's primary coalition of large urban school districts, has been asked by a number of individuals over the last several weeks for its analysis of and opinions on the DC-CAS scores that the Office of the State Superintendent of Education (OSSE) released earlier this year and re-released this week.

The majority of questions we have been asked relate to the appropriateness of OSSE's use of prior year cut scores to report performance levels and whether this reporting method inflated the appearance of student progress.

Prior to the release of its 2013 test performance on the DC-CAS, OSSE decided to use past performance-level cut scores to report citywide student performance rather than newly developed cut scores based on the more rigorous Common Core State Standards (CCSS). In the opinion of the Council, which is familiar with the technical methodology used by OSSE, either decision was appropriate.

OSSE should have been more transparent in describing to the public what the scores meant and what they didn't mean, but its decision to report the data in a way that would allow the public to see how performance had changed over time was entirely sound. In releasing its new results this week, OSSE has correctly distinguished between old and new results and how they can be compared and not compared.

OSSE's Options

If OSSE had chosen to base its 2013 test results on common core standards, then it would have been unable to compare its data to any previous year's proficiency levels using the new scale. The ability to answer the question, *Did student reading and math proficiency improve?* would have been impossible with a methodology that attempted to compare results across two scales. What makes OSSE's new release appropriate was that it reanalyzed all prior year data on the new scale and it placed all current year data on the old scale.

The data clearly demonstrate that the public schools in D.C. have made substantial progress.

States across the country that have modified their state assessments to reflect the more difficult CCSS have faced the same decision that OSSE faced. *Do we use historic trend lines or create a new baseline grounded in the higher expectations?* The following statement from New York, which chose to release proficiency results based on common core assessment standards, illustrates this tradeoff and some of the pitfalls of employing this approach:

New York - I want to make it very clear that the change in test scores (including, possibly, one in your child's score) does not mean that students are learning less or that teachers and schools are performing worse than last year. Proficiency rates – the percentage of students meeting or exceeding the standards – on the new Common Core

¹ Ray Hart has a Ph.D. in education with emphasis in measurement, statistics and evaluation.

assessments cannot be compared with last year's proficiency results since the old scores are from an old test based on the former standards. This is a new beginning and starting point that will provide better, clearer information to parents, teachers, and principals about what our children know and are able to do. The results from these assessments will help you and your school directly address the learning needs of your child so that he or she gets and/or stays on track for college and career success.

To reiterate, two reports are possible with the 2013 exam data:

1. One can measure student proficiency using the new scaled score standards and place all historical scores on the new scale, or abandon the ability to compare historical results and changes in district performance over time, or
2. One can measure student proficiency using the scaled score standards of the old DC- CAS exam and maintain the ability to measure changes in student proficiency levels. Using the old cut scores or scaled scores will allow a direct comparison to historical performance data, but it does not permit comparison to common core rigor unless different scaling is used.

To ensure that the two approaches were clear, CTB McGraw-Hill (CTB), the vendor for the DC-CAS, scaled the new assessment such that scaled scores on the easier (old D.C. CAS) and more difficult (new D.C. CAS) tests were equated. A scaled score of 550 on the old exam was equal to a scaled score of 550 on the new exam. This was clear from the following quote in an email from CTB to OSSE dated June 18, 2013:

“Equating between the current year and last year is independent of cut score setting procedure. Purpose of equating is to place the current scale in the same scale as the last year, through common lining anchor items, and equating was done before setting final cut scores. After equating, a scale score 550 in this year equals to last year scale score 550, and student's scale scores are not affected by different cut scores....This standard setting new cuts is to assure the test changes are reflected fairly.”

In the same email, CTB goes on to state—

“If new cuts [Common Core cut scores] were applied to this year, we could not compare this year percentage of each performance level to last year directly because of cut score changes or test standard changes.”

In addition, one should keep in mind that the District of Columbia will receive state data on NAEP in October of this calendar year and the District of Columbia Public Schools (DCPS) will receive its NAEP results as a school district in December. The state and DCPS results will show a comparable trend line and the results on both will reflect rigor similar to the common core. In other words, the public will get from NAEP the best of what OSSE was trying to show using the DC-CAS—trend lines and common core-like rigor.

Responding to OSSE's Critics

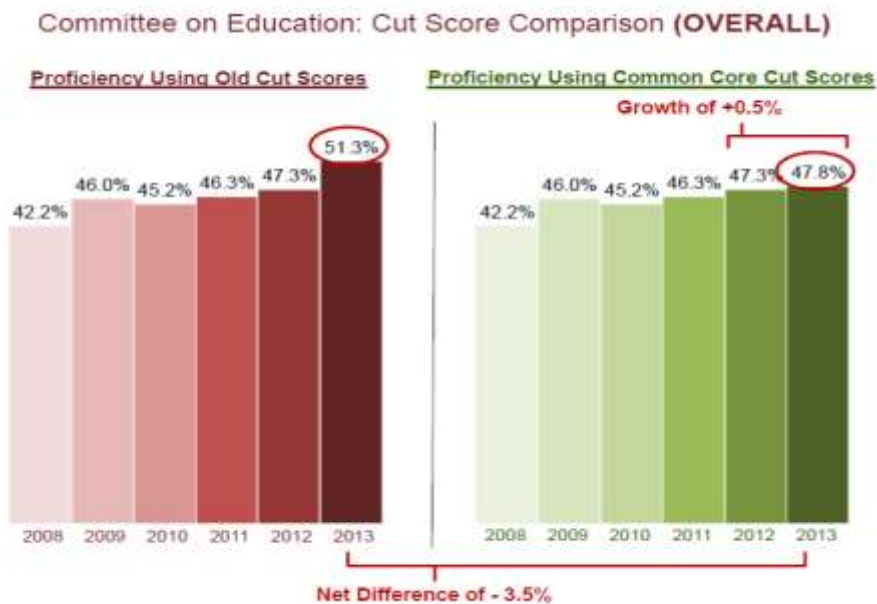
Given this background on the data and methodology employed by OSSE, the public should know that statements made by Mr. David Catania on how to interpret the DC-CAS scores are incorrect. First,

comparing the results of the new cut scores to the proficiency rates of the old cut scores is inappropriate. The chart to the right in Figure 1 below, which is found on Mr. Catania’s website (http://www.davidcatania.com/dccas_2013), should not include any proficiency rates from prior years because the scores use a new baseline for 2013 and are not comparable to prior years.

Consequently, the website’s assertion, “Fact: Applying the appropriate cut scores would have resulted in little overall growth in student proficiency.” is simply false. Also, drawing a line between the 2013 scores on the left of the graph using the old cut scores and the 2013 scores on the right is inappropriate because the two sets of results are based on different standards.

In other words, Mr. Catania has juxtaposed scores that are not comparable in order to assert—incorrectly—that there has been “little overall growth in student achievement.” If what OSSE did could be characterized as cheating, then so could this comparison because it misleads the public about the district’s progress.

Figure 1.



Second, the same website also contends, “Fact: In 2013, students answered fewer questions correctly than ever before.” Here, the statement is accurate on its face but highly misleading. The truth is that students answered fewer questions correctly on the new common core-based assessment, but the items on the new assessment were more difficult. Perhaps an analogy would be helpful in describing what test developers do when reporting and interpreting assessment results. Suppose we gave two children of equal ability two different assessments. John is administered the easier assessment, test A, and Allan is administered the more difficult assessment, test B. Because the two students have the same overall math ability, we would expect John to get a higher percentage of the items correct on the easier test and Allan to get fewer items right on the harder test. We would not want to make the mistake of concluding that because Allan took a more difficult assessment that his overall math ability is somehow diminished, as that would not be true.

Equating assessments of differing difficulty is a common practice for test developers. By looking at the items on the two exams, test developers can determine that John and Allan answered the same types of items correctly and missed the same types of items (despite the fact that Allan had fewer easy items on his test). They can then equate the two exams so that John's scaled score and Allan's scaled score are equal, as they should be. The process does not reward John for taking an easy test nor penalize Allan for taking a harder test, and their scores are directly comparable. And this was what CTB did with the DC-CAS when it placed the new test on the old scale.

As the analogy illustrates, fewer items correct on a more difficult exam is expected and supports the fact that the new assessment is indeed more difficult. In fact, changes in the mean raw scores, as reported on Mr. Catania's website, do not indicate lower student performance levels at all. Student performance can increase as mean raw scores decline if items on the exam are more difficult.

Third, the motives behind OSSE's decision to use trend lines rather than a new baseline to report DC-CAS data have been characterized as questionable. In fact, this decision was appropriate, well within OSSE's administrative discretion, and maybe even prudent. Had OSSE decided to move to the new baseline, it might have undermined the city's new personnel and school accountability systems. And it would run the risk of having to upend the system again in 2015 when the new PARCC assessments are administered.

Finally, Mr. Catania has repeatedly asserted that statements by the Council of the Great City Schools or its Executive Director Michael Casserly should be dismissed because the chancellor of the District of Columbia Public Schools, Kaya Henderson, sits on the organization's board of directors. We are proud of the fact that Ms. Henderson does indeed sit on the Council's board, but the claim of partiality ignores the fact that extremely critical reports on the D.C. schools were written when Paul Vance and Clifford Janey led the school system—and both sat on the Council's board. In fact, the organization has written hundreds of scathing reports on its own members in order to spur improvements in student achievement and system operations. Our concern here, and always, is to ensure that the best interests of students are considered and that their performance is accurately interpreted.