The "Phantom" Collapse of Student Achievement in New York: Lessons for Educators as States Implement the Common Core

John Cronin, Ph.D.
Senior Director of Research at NWEA

Nate Jensen, Ph.D.
Research Scientist at NWEA

January 2014

On August 7th, 2013, the New York State Education Commissioner, John King, announced the initial results of the state's new assessment, which was designed to measure college and career readiness relative to the Common Core Learning Standards. Commissioner King noted that the proficiency rates on these assessments were significantly lower than proficiency rates on the prior year's assessment. In reading, the proportion of proficient students dropped from 55% to 31%, and in mathematics, the proficiency rate dropped from 65% to 31%. These changes in student test performance have caused some educators and policymakers in the state to question how these test results are used, including calls to delay high-stakes evaluations of student and teacher performance based on results from these new assessments.

The observed drops in proficiency rates reflect a change in the difficulty of the proficiency standard and not a decline in student scores or performance. That is, the cut scores on these tests—scores that denote whether a student was proficient—were raised, making it more difficult for students to meet the proficiency threshold on the new tests than on the test that was used in prior years. The Commissioner himself noted that the new standards were a break from past practices in his press release:

> "**These proficiency scores do not reflect a drop in performance, but rather a raising of standards to reflect college and career readiness in the 21st century.** I understand these scores are sobering for parents, teachers, and principals. It's frustrating to see our children struggle. But we can't allow ourselves to be paralyzed by frustration; we must be energized by this opportunity. The results we've announced today are not a critique of past efforts; they're a new starting point on a roadmap to future success."[1]

Unfortunately, the Commissioner's key distinction that student performance did not decline, but that students were held to a higher proficiency standard on the new tests, was not fully understood. For example, the New York Times led a story on the release of these test results with the following headline:

> "Test Scores Sink as New York Adopts Tougher Benchmarks"[2]

The New York Times correctly picked up on the fact that the new tests were aligned to a more rigorous set of standards, but the report that test scores sank is inaccurate. In fact, the Times only reported that the number of students passing these tests dropped dramatically, as the Commissioner noted, while failing to acknowledge that the Commissioner also said that these changes in proficiency do not indicate a drop in performance. This distinction, as we will show, is extremely important.

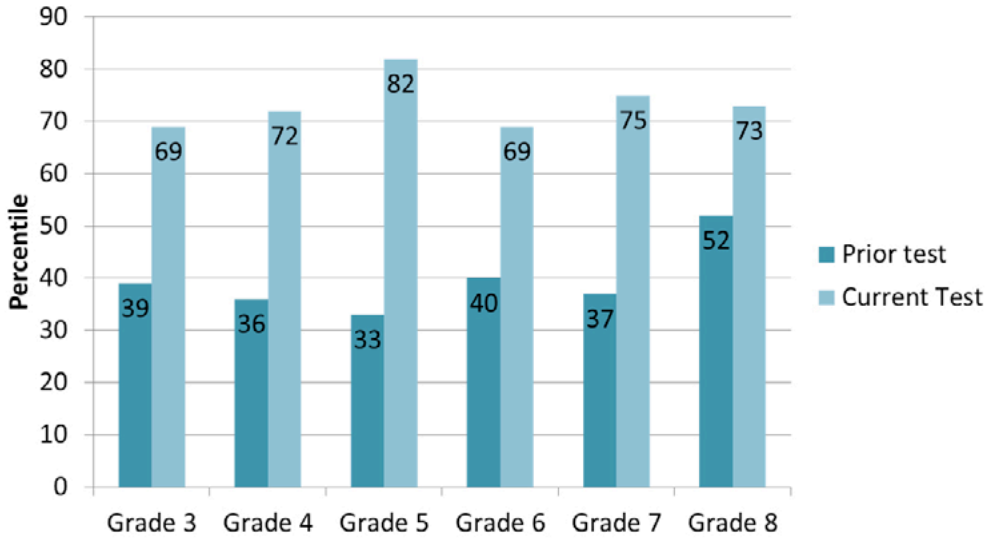
Think of the problem this way. Let's assume that we are testing the jumping ability of a group of 6th graders. We've decided that a proficient 6th grader should be able to high jump three feet, so we test all 6th graders against that standard and find that 75% are proficient because they can jump that high. Now let's assume that after the test we decided that this standard doesn't reflect the performance of an athlete "on track" for college, so we raise the bar to five feet. After we raise the bar, we find that only 20% of the group of 6th graders could clear this benchmark. Did the 6th graders' jumping ability decline? Of course not. The students could still jump just as high, but their jumping ability was held against a higher standard in the second test.

This is akin to what occurred in New York: student test performance, and subsequently what students learned, may not have changed at all—in fact, it may have improved—but students had to clear a higher proficiency threshold with the new test to be considered college and career ready, which contributed to the decline in student proficiency rates. Unfortunately, it was difficult to know whether student test scores actually improved or declined since last year, as scores from New York's prior and current tests were reported on different scales, which made comparing past and current scores challenging.

Nevertheless, one important question remains: Did student performance in New York actually decline between 2012 and 2013, or was it a *phantom decline* that was reported in the media? One way to address this question is to compare student performance across both years using the same measurement scale while holding the proficiency threshold constant. This would permit actual comparisons of student performance between 2012 and 2013, and would allow us to draw conclusions about whether student test performance actually changed since 2012, and if so, in what way. Northwest Evaluation Association™ (NWEA™) works with many New York school systems that use the Measures of Academic Progress® (MAP®) assessment to measure student performance on the state's mathematics and reading standards. The assessment is a computer-adaptive test that is aligned to the state's curriculum standards and reported on an equal interval scale. MAP is strongly correlated with both the prior version and current version of the New York state assessment, and as a result, we are able to estimate scores on our scale that correspond to the prior proficiency standards for New York as well as the new, more difficult, proficiency standards.[3]

In Figure 1 we show the differences in estimated proficiency cut scores, expressed as a percentile rank relative to the NWEA nationally representative norming sample[4], across the two years on the mathematics tests. These percentile ranks indicate that the level of performance required to demonstrate proficiency on the new assessment was considerably higher than what was required in 2012. For example, in 4th grade mathematics, students in 2012 under the prior standards needed to score at or above the 36th percentile on the State test in order to be considered proficient. In 2013, under the new college and career readiness standards, 4th grade students needed to score at or above the 72nd percentile to receive a proficient rating. These large differences in proficiency cut scores can be observed across all grade levels, and are present in reading as well.

Figure 1 – Difference between 2012 and 2013 New York Benchmark Proficiency Scores, Mathematics



Because the difficulty of the cut scores relative to the NWEA scale is known, we can use student MAP results to estimate what a school system's 2013 proficiency rate would have been had the proficiency cut scores not changed. To illustrate this, we selected six New York school systems with total enrollments of at least 3,000 students that used NWEA tests in at least 2012 and 2013, and tested nearly all of their students on both MAP and the required state assessment. These districts were not selected to be representative of all New York schools, nor does their performance necessarily reflect that of the state as a whole. We simply used these school systems to illustrate how changes in proficiency cut scores can impact the perception of a district's performance.

Table 1 shows the mean MAP scale scores in 4th grade mathematics for students in the six school systems from the spring 2012 and spring 2013 test administrations. The data show that, in these particular school systems, student performance in 4th grade mathematics actually improved between 2012 and 2013, and for some districts (such as District 3), that improvement was substantial. So, the perception that student test scores declined between 2012 and 2013 is a misperception, at least based on the test results from these six school systems. In fact, student performance in mathematics in these districts *improved for all grades tested*, with the exception of one district's 8th grade mathematics scores.

Table 1 – Mean MAP Scale Scores from Spring 2012 and Spring 2013, 4th Grade Mathematics

School System	Spring 2012	Spring 2013	Difference
District 1	218.1	219.6	+1.5
District 2	219.4	221.8	+2.4
District 3	210.9	224.8	+13.9
District 4	215.3	219.8	+4.5
District 5	218.1	221.3	+3.2
District 6	201.4	204.6	+3.2

But, given that proficiency rates are the summary statistic most often reported, it makes sense to look at how the change in standards impacted proficiency rates for this same group of 4th grade students over the same time period. In other words, if we applied the 2012 proficiency cut scores to the 2012 results for these students, and the higher 2013 proficiency cut scores to the 2013 results, what would be the subsequent impact on estimated proficiency rates in these six districts based on results from the MAP assessment? In this way, we can present results on our assessment in the same manner that proficiency results from the New York State assessments were originally reported to the public. In Table 2, we show estimated proficiency rates in our six school districts based on 2012 and 2013 MAP results, applying the proficiency standards that were in place at the time of testing.

Table 2 - Estimated Proficiency Rates on NWEA MAP Assessments from Spring 2012 and Spring 2013, 4th Grade Mathematics

School System	2012 Proficiency Rate Relative to the 2012 Proficiency Cut Score	2013 Proficiency Rate Relative to the 2013 Proficiency Cut Score	Difference
District 1	89.1%	54.9%	-34.2%
District 2	87.9%	56.1%	-31.8%
District 3	95.5%	65.1%	-30.4%
District 4	82.6%	53.0%	-29.6%
District 5	85.6%	58.9%	-26.7%
District 6	36.8%	13.5%	-23.3%

These results reflect the scenario that was widely reported in New York—each district's proficiency rate declined substantially, creating the illusion that student achievement collapsed. But in these six districts, student performance in grade 4 on the MAP assessment actually improved from 2012 to 2013 (as we showed in Table 1). So what would student test results have looked like in these six districts if we evaluated the 2012 and 2013 results using just the 2013 proficiency cut score?

In Table 3, we show 4th grade mathematics proficiency rates from both 2012 and 2013, using only the 2013 cut scores to estimate these results. When the cut score is held constant across both years, we found that proficiency rates actually improved, which is what we would expect given that mean student achievement also improved in each school system. The results shown in Tables 2 and 3 provide a straightforward illustration of how simply changing proficiency cut scores can impact perceptions of student test performance.

Table 3 - Estimated Proficiency Rates on NWEA MAP Assessments from Spring 2012 to Spring 2013 Holding the 2013 Proficiency Cut Score Constant, 4th Grade Mathematics

School System	Spring 2012 Proficiency Rate Relative to the 2013 Proficiency Cut Score	Spring 2013 Proficiency Rate Relative to the 2013 Proficiency Cut Score	Difference
District 1	45.5%	54.9%	+9.4%
District 2	46.6%	56.1%	+9.5%
District 3	53.2%	65.1%	+11.9%
District 4	32.6%	53.0%	+20.4%
District 5	46.3%	58.9%	+12.6%
District 6	5.5%	13.5%	+8.0%

Lessons Learned

As other states transition to the new Common Core assessments, we anticipate that the New York narrative is likely to be repeated. Because cut scores on new Common Core assessments are intended to reflect "college and career readiness," they are likely to be more challenging than cut scores on nearly every states' prior NCLB test. Cut scores from previous versions of state accountability assessments were set in a context in which every student was expected to demonstrate proficient performance by 2014, and schools were sanctioned if proficiency rates

weren't improving rapidly enough to eventually meet this requirement. Given this environment, it was perfectly reasonable for states to set low proficiency standards, as the consequences of not doing so would have been that virtually every school in every state would have been under some form of sanction.

Of course, there is nothing intrinsically wrong with raising expectations for student performance. In fact, a "college and career ready" level of performance is more consistent with aspirations of parents and students than the prior standards, which were inconsistent and based on an amorphous concept of proficiency.[5] The problem thus was not with the change in standards; rather, the problem was the misperceptions that were created because the past scale used for the New York test could not be compared to the present scale. Because of this, the state could not report whether student achievement improved or declined, it could only report that proficiency rates had dropped dramatically.

It is critical that educators understand these changes and are prepared to address misperceptions that will arise when proficiency rates inevitably drop as the higher standards associated with the Common Core are implemented. In New York, Commissioner King presented this change accurately—the proficiency standards increased in difficulty, and as a result, proficiency rates dropped, but this did not mean that student performance collapsed. Unfortunately, reports of declines in proficiency rates (rather than actual declines in scores) created the erroneous impression of a collapse in student achievement. This was a **phantom collapse**, and as illustrated in our six district example, schools with apparent declines in proficiency rates actually showed improvements in student achievement between 2012 and 2013.

While educating the public about the actual meaning of the changes in proficiency standards is essential, the New York narrative also illustrates the importance of maintaining consistent, longitudinal achievement data over time. This case illustrates one of the primary problems with state testing programs—they are not consistent. The 2013 New York State test was a complete break from the prior assessment, and unfortunately, no mechanism was put in place to produce reasonable comparisons of current test results to prior test results. This is unfortunate, as this disconnect renders a school system's prior test results largely useless, not only because 2012 data cannot be compared to the current results, but because it makes it impossible to connect the current and future data to achievement trends that were established in the years before 2013. This creates challenges when a school system tries, for instance, to evaluate a reading program that began a five-year cycle of implementation in 2011 with state data collected from two distinct state tests that cannot be compared. This makes it especially important for school systems to maintain their own measures of student achievement to ensure that they can track student performance over time. In New York, school systems that maintained their own student achievement measures had data that allowed them to see whether

student test scores had actually declined, or if students had made improvements from year to year in math and reading (as was the case in our six example districts).

Further, in this instance, the break in student testing data may mask the impact of important New York initiatives that potentially had a significant impact on teaching and learning. The 2012-2013 year was the first year in which the state implemented a new, high stakes, teacher evaluation program. Given the stakes, it seems critical to evaluate the impact that program is having on student learning statewide. The break in testing programs, and particularly, the failure to create any means to compare prior scores to current scores, makes it much more difficult for researchers, and the media or public, to ascertain what impact (if any) this effort has had on student learning.

Finally, the New York narrative illustrates the need for educators to become data literate, and be able to coach the public when student achievement information is misrepresented, whether that occurs in the media or elsewhere. Proficiency rates will certainly decline if student performance declines, but they can also decline if the proficiency cut score becomes more difficult. **That distinction is incredibly important.** New York (and other states) recognized the need to raise standards because the prior proficiency standards did not reflect a level of performance that aligned to the aspirations of students and their parents (who almost universally embrace college attendance as their goal).[6] The fact that only 31% of New York students are proficient under the current standard[7] means that challenge is perhaps greater than what would have been recognized from reports based on student performance relative to the prior set of proficiency standards. But any implication that this represented deterioration in the performance of schools would reflect a cynical portrayal of the problem, and would overlook what largely drove these declines in proficiency rates—that the proficiency standards were more difficult in 2013 than in 2012.

The phantom collapse of student achievement in New York reflects a misguided narrative of supposed school failure that does little more than feed distrust about public education, and comes at a time when educators are working to raise expectations for student learning to better prepare them to be successful throughout high school and beyond. As the Common Core is implemented, schools will face the challenge of responding to higher standards. And as we evaluate the performance of these schools, this discussion should be based on sound and consistent testing data, rather than negatively opining about the failure of these schools to stack up to an ever-changing set of proficiency standards. *If student achievement goes down*, appropriate steps should be taken to rectify the reason for this decline. However, *if student proficiency goes down*, then it is important to remember that this does not necessarily mean that student achievement has declined, and the potential reasons behind these drops in proficiency—such as the implementation of a higher proficiency standard—should be clearly and accurately articulated to parents, teachers, and the public as a whole.

Learn more about NWEA Research by calling 866-654-3246 or explore more at **NWEA.org/Research**.

Founded by educators nearly 40 years ago, Northwest Evaluation Association (NWEA) is a global not–for–profit educational services organization known for our flagship interim assessment, Measures of Academic Progress (MAP). More than 7,400 partners in U.S. schools, school districts, education agencies, and international schools trust us to offer pre-kindergarten through grade 12 assessments that accurately measure student growth and learning needs, professional development that fosters educators' abilities to accelerate student learning, and research that supports assessment validity and informed policy. To better inform instruction and maximize every learner's academic growth, educators currently use NWEA assessments with nearly 8 million students.

References:

1. New York Department of Education (August 7, 2013). *State Education Department Releases Grades 3-8 Assessment Results.* Retrieved on December 3, 2013 from http://www.oms.nysed.gov/press/grades-3-8-assessment-results-2013.html
2. Hernandez, J. and Gebeloff, G. (August 7, 2013). Test Scores Sink as New York Adopts Tougher Benchmarks. *New York Times.* Retrieved on December 3, 2013 from http://www.nytimes.com/2013/08/08/nyregion/under-new-standards-students-see-sharp-decline-in-test-scores.html
3. NWEA uses an equipercentile methodology described in Ryan, J., & Brockmann, F. (2009). A practitioner's introduction to equating with primers on classical test theory and item response theory. CCSSO: Washington, DC.
4. Northwest Evaluation Association (2011). RIT Scale Norms: For Use with MAP and MAP for Primary Grades. Portland, OR: Author.
5. Cronin, J., Dahlin, M., Kingsbury, G.G., and Adkins, D. (2007). *The Proficiency Illusion.* Washington D.C. Thomas B. Fordham Institute.
6. Pew Research Center (February 27, 2012). Most Parents Expect their Children to Attend College. Retrieved on December 17, 2013 from http://www.pewresearch.org/daily-number/most-parents-expect-their-children-to-attend-college/
7. New York State Education Department (August 7, 2013). State Education Department Releases Grades 3-8 Assessment Results. Retrieved on December 12, 2013 from http://www.oms.nysed.gov/press/grades-3-8-assessment-results-2013.html

Partnering to Help All Kids Learn® | NWEA.org | 503.624.1951 | 121 NW Everett St., Portland, OR 97209

March 2015