

The phantom collapse of student achievement in New York

John Cronin and Nate Jensen
Phi Delta Kappan 2014 96: 60
DOI: 10.1177/0031721714553412

The online version of this article can be found at:
<http://pdk.sagepub.com/content/96/2/60>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Phi Delta Kappan* can be found at:

Email Alerts: <http://pdk.sagepub.com/cgi/alerts>

Subscriptions: <http://pdk.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Sep 22, 2014

[What is This?](#)

The phantom collapse of student achievement in New York



Common core

By John Cronin and Nate Jensen

Results from New York's first Common Core state tests appeared to show a big decline in student achievement. But a deeper look revealed a much different, brighter reality.

In August 2013, New York State Education Commissioner John King released initial results of the state's new assessment, which was designed to measure college- and career-readiness relative to the Common Core State Standards. King noted that proficiency rates on the assessments dropped significantly from the prior year's — from 55% to 31% in reading and from 65% to 31% in math. These changes in student test performance caused educators and policy makers to question how test results were used and included calls to delay evaluations of student and teacher performance based on the results.

But in reality the observed drops in proficiency rates reflect an increase in the proficiency standard and not a decrease in student scores or performance. That is, the state raised the cut scores on these tests, which denote whether a student was proficient. That act made it more difficult for students to meet the new proficiency threshold. In a press release, the commissioner said the new standards broke from past practices:

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 (New

JOHN CRONIN (john.cronin@nwea.org) is a senior director, and **NATE JENSEN** (nate.jensen@nwea.org) is a research scientist at Northwest Evaluation Association, Portland, Ore.



York State Education Department, 2013).

Unfortunately, the Commissioner's message that student performance did not decline, rather that students were held to a higher proficiency standard, was not fully understood. For example, a *New York Times* headline read, "Test scores sink as New York adopts tougher benchmarks" (Hernandez & Gebeloff, 2013). The *Times* correctly said the new tests were aligned to a more rigorous set of standards but inaccurately reported that test scores sank. In fact, the number of students passing these tests dropped dramatically, as the Commissioner noted, but the *Times* and other media failed to acknowledge that the changes in proficiency did not indicate a drop in performance. This distinction is extremely important.

Think of the problem this way. Let's assume that we're 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 stu-

This case illustrates one of the primary problems with state testing programs: They are not consistent.

dents 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. And it was difficult to know whether student test scores actually improved or declined from a year earlier because scores from the 2013 and 2012 tests were reported on different scales.

Nevertheless, one important question remains: Did student performance in New York actually decline between 2012 and 2013? One way to answer this question is to compare student performance across both years using the same measurement scale while holding the proficiency threshold constant. This would let us draw conclusions about whether student test performance actually changed since 2012 and, if so, in what way.

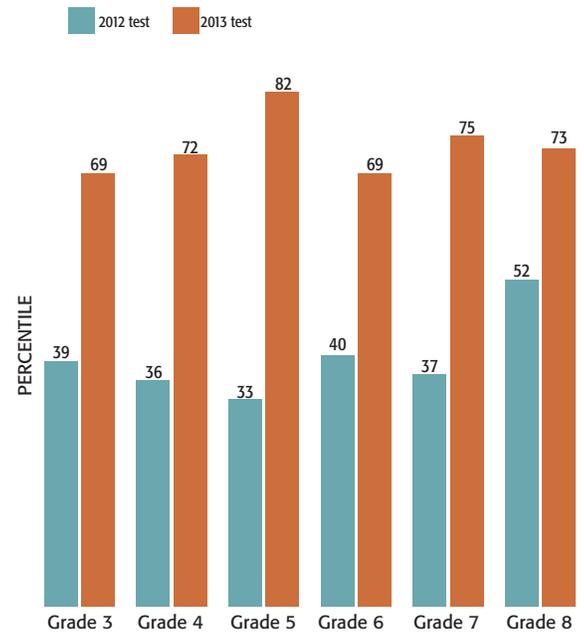
Northwest Evaluation Association (NWEA) works with many New York school systems that use the Measures of Academic Progress® (MAP®) assessment to measure student performance on the state's mathematics and reading standards. The assessment is a computer-adaptive test aligned to the state's curriculum standards 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 (Ryan & Brockmann, 2009).

Figure 1 shows the differences in estimated proficiency cut scores, expressed as a percentile rank relative to NWEA's nationally representative norming sample across the two years on the mathematics tests (Thum & Hauser, 2012). These national 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 to be considered proficient on the state test. 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.

MAP and a changed view

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 if the state had not changed the proficiency cut scores. To illustrate this, we selected six New York school systems with total en-

FIG. 1.
Difference between 2012 and 2013 New York proficiency cut scores, mathematics



rollments 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 affect the perception of a district's performance.

TABLE 1.
Mean MAP 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

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 ad-

TABLE 2.

Estimated proficiency rates on NWEA’s 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%



ministrations. 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 assertion that student test scores declined between 2012 and 2013 is incorrect — 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.

But, given that proficiency rates are the summary statistic most often reported, it makes sense to look at how the change in standards affected 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 effect 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 in place at the time of testing.

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 affect perceptions of student test performance.

Lessons learned

As other states transition to the new Common Core assessments, 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.

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.

Reports of declines in proficiency rates created the erroneous impression of a collapse in student achievement in New York. This was a phantom collapse.

TABLE 3.

Estimated proficiency rates on NWEA’s 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%

Of course, nothing is 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 (Cronin, Dahlin, Kingsbury, & Adkins, 2007). The problem thus was not with the change in standards; rather, the problem was the perceptions 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.

Educators must understand these changes and be prepared to address misperceptions that will arise when proficiency rates inevitably drop as the Common Core’s higher standards 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 into place to produce reasonable comparisons of current test results to prior test results. 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, for example, a school system tries 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).

The New York State Education Department released 2014 results in mid-August, showing improvement for students across the state in both math and reading. New York educators will now be able to compare student performance across multiple years on Common Core assessments, providing all stakeholders with valuable information. As other states move to implement the new assessments, it is important to consider what steps are being taken to make certain that schools do not experience the same “break” in testing data that occurred in New York.

Need for data literacy

Further, in this instance, the break in student testing data may mask the effect of important New York



Present a session to aspiring teachers
at the 2015 Future Educators Association® National Conference!

FEA National Conference

APRIL 24-25, 2015

Washington, D.C.

SESSION PROCESS OPENS SEPTEMBER 15.

<http://app.wizehive.com/apps/2015FEAproposal>



initiatives that could have had a significant influence on teaching and learning. The 2012-13 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 that program's effect on student learning statewide. The break in testing programs and particularly the failure to create a way to compare prior scores to current scores makes it much more difficult for researchers, the media, and the public to ascertain this effort's effect, if any, on student learning.

Finally, the New York narrative illustrates the need for educators to become data literate and to be able to coach the public when student achievement information is misrepresented. Proficiency rates will certainly decline if student performance declines, but they can also decline if the proficiency cut score is raised. 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 (Pew Research Center, 2012).

The fact that only 31% of New York students are proficient under the current standard 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

an inaccurate and cynical portrayal of the problem and would overlook what largely drove these declines in proficiency rates: 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. 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 in 2014 and beyond, this discussion should be based on sound and consistent testing data. If student achievement declines, educators should take appropriate steps to rectify the reason for it. However, if student proficiency goes down, this does not necessarily mean 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. ■

References

- Cronin, J., Dahlin, M., Kingsbury, G.G., & Adkins, D. (2007). *The proficiency illusion*. Washington, DC: Thomas B. Fordham Institute.
- Hernandez, J. & Gebeloff, G. (2013, August 7). Test scores sink as New York adopts tougher benchmarks. *New York Times*. www.nytimes.com/2013/08/08/nyregion/under-new-standards-students-see-sharp-decline-in-test-scores.html
- New York State Education Department. (2013, August 7). *State Education Department releases grades 3-8 assessment results*. New York, NY: Author. www.oms.nysed.gov/press/grades-3-8-assessment-results-2013.html
- New York State Education Department. (2014, August 14). *State Education Department releases grades 3-8 assessment results*. New York, NY: Author. www.nysed.gov/news/2014/state-education-department-releases-grades-3-8-assessment-results
- Pew Research Center. (2012, February 27). *Most parents expect their children to attend college*. Washington, DC: Author. www.pewresearch.org/daily-number/most-parents-expect-their-children-to-attend-college/
- Ryan, J. & Brockmann, F. (2009). *A practitioner's introduction to equating with primers on classical test and item response theory*. Washington, DC: Council of Chief State School Officers.
- Thum, Y.M. & Hauser, C. (2012). *RIT scale norms: For use with Measures of Academic Progress (MAP®) and MAP® for Primary Grades*. Portland, OR: Northwest Evaluation Association.



"My attorney, Mr. Hays, would like to discuss my algebra grade with you."