B | Economic Studies at BROOKINGS

Evidence Speaks Reports, Vol 2, #53
May 24, 2018

What can NAEP tell us about how much US children are learning?

Matthew M. Chingos

Executive Summary

Scores from the National Assessment of Educational Progress (NAEP), dubbed the "nation's report card," are often used to compare student achievement across states. An important limitation of NAEP is that it does not track the performance of individual students over time, so inferences about how much students are learning must be made by comparing scores from tests given to different groups of students every two years.

This report presents the results of different exploratory analyses that take advantage of the fact that the same birth cohorts are tested four years apart on the 4th- and 8th-grade NAEP exams. For example, I compare 8th-grade scores from the 2017 NAEP to 4th-grade scores from the 2013 NAEP. I then contrast these measures of change over time to demographically adjusted 8th-grade scores published by the Urban Institute.

I find that states with similar 8th-grade performance vary widely in their 4th-to-8th-grade increases (and vice versa). Both measures provide potentially useful information, and neither is clearly better given that the increase measure ignores differences in educational quality through 4th grade whereas the 8th-grade score ignores unmeasured differences in student characteristics captured by the 4th-grade score.

I also find that states vary significantly in the extent to which educational progress that benefits their 4th-grade students continues to benefit the same cohorts of students by the end of middle school. Many states that see gains in 4th-grade scores do not see any gains for the same cohorts of when they are tested in 8th grade, raising concerns that some of the education reforms of the last 15 years have changed when students learn key skills but not whether they have learned them.

Introduction

The 2017 NAEP scores released last month revealed national test-score performance that was largely unchanged from 2015, when scores had dipped on three out of four tests.[1] The long-term trends in performance are still positive, but 4th-grade scores have now been stagnant for a decade while 8th-grade scores have posted small increases over the last 10 years.

These trends cry out for explanation—and many commentators are happy to oblige—but the truth is that NAEP scores can tell us how much students know but not why scores have increased, decreased, or remained the same.

A key limitation of NAEP is that, while it provides the only national snapshot of student performance in 4th and 8th grade, it does not track the performance of individual students over time. As a result, inferences about how much students are learning must be made by comparing scores from tests given to different groups of students every two years. Fourth-graders in 2017 are an entirely different group of children from fourth-graders in 2015, and policies enacted in 2016 could have potentially affected those tested in 2017 but not those tested in 2015.

This report presents new analyses of state-average NAEP data that attempt to address the limitation of changing samples of students by following cohorts of students from 4th grade in a given year to 8th grade four years later. NAEP selects new samples of students at every test administration, so it is unlikely that any individual student would be tested in both years. But both groups of students are selected to be representative of students in their state in that grade and year, so comparing the two scores provides a useful proxy for how much knowledge a cohort of students has gained over time.[2] This analysis should be regarded as exploratory given the limitations of comparing NAEP scores across grades.[3]

I compare these measures of change over time to demographically adjusted scores that my colleagues at the Urban Institute and I have calculated using the restricted-use, student-level NAEP data. These adjusted scores compare the average performance of students in each state compared to demographically similar students around the country.[4] These scores are a better way to compare performance across states than simply using the raw NAEP scores.

The increase from 4th to 8th grade is a useful measure in part because it controls for any family or state characteristics that are reflected in the 4th-grade score (such as income or how much families value education). But, as a result, the increase measures ignore any differences in state education policies that affect 4th-grade scores. For this reason, 8th-grade scores may be a

better summary measure of state performance.

Figures 1 and 2 compare, for math and reading respectively, the 4th-to-8th-grade score increases to the demographically adjusted 8th-grade scores in each state. In math, states that post larger increases between grades also tend to have higher 8th-grade scores but the correlation is not perfect. For example, Massachusetts and California both post above-average increases, but Massachusetts has much higher 8th-grade scores. The NAEP data do not reveal the extent to which this is due to unmeasured differences between students in the two states vs. education policies and practices that affect 4th-grade performance.
(See Figure 1)

Reading scores (Figure 2) tell a different story, in that there is little systematic relationship between the 4th-to-8th-grade increase and 8th-grade performance. There are thus even more examples of states that diverge in terms of their performance on the two measures. For example, California and Maryland have similar 8th-grade scores but wildly different gains between 4th and 8th grades. This could mean that Maryland's education system better supports reading skills through 4th grade, but that California students make up for the initial deficit in the years that follow.
(See Figure 2)

This example raises the question of whether educational progress has been exaggerated by students learning math and reading skills sooner than they used to (scores at younger ages rising) but not leaving school with greater knowledge (stagnant scores at older ages). NAEP scores over longer periods to time tend to show the largest increases for younger students and the smallest increases for older students (with especially dismal results for high-school students).[5]

I contribute evidence to this discussion by examining whether 10-year changes in demographically adjusted 4th-grade scores correspond to 10-year changes in 8th-grade scores for the same pairs of cohorts (4th graders in 2003 and 2013 and 8th graders in 2007 and 2017).[6] I report the results in Figures 3 and 4 for math and reading, respectively.

Figure 3 shows that every state saw an increase in 4th-grade math scores between 2003 and 2013. But only 30 states posted gains in 8th-grade math scores for the same cohorts over this period. There is a positive correlation between increases measured at 4th and 8th grades, but many states deviate from that general relationship.

For example, Arkansas, Kentucky, and Maryland all increased their 4th-grade scores by more than 10 points (more than a year of learning, as the average difference between 4th- and 8th-grade scores is about 40 points), but those gains evaporated by 8th grade. But several states, including Nevada and Hawaii, did see gains captured at both grades, although the gains measured in

8th grade were considerably smaller than those in 4th grade. On average across all states, the 10-year gain was 7.6 points in 4th grade but only 0.3 points in 8th grade.
(See Figure 3)

Reading scores (Figure 4) tell a similar story with some differences. Once again, gains measured at 4th and 8th grades are modestly correlated, but the average gains are more similar (3.3 points in 4th grade and 2.6 points in 8th grade). Florida and Nevada posted large reading gains that persisted in both grades, whereas a number of states posted modest gains at 4th grade that did not translate into an improvement in 8th grade.
(See Figure 4)

This analysis of state-average NAEP data reveals two key findings by comparing the achievement data of representative samples of the same birth cohorts taken at different points in time.

First, measuring states based on their 4th-to-8th-grade increases often produces different inferences than measuring them based on 8th-grade performance. It is not clear which measure is better given that the increase measure ignores differences in educational quality through 4th grade whereas the 8th-grade score ignores unmeasured differences in student characteristics captured by the 4th-grade score.

Second, states vary significantly in the extent to which educational progress that benefits their 4th-grade students continues to benefit the same cohorts of students by the end of middle school. The fade-out of improvements, especially in math, raises concerns that some of the education reforms of the last 15 years have changed when students learn key skills but not whether they have learned them by 8th grade.

This analysis speaks to the value of longitudinal data systems that can track students throughout their elementary and secondary schooling, so that progress over time can be tracked in a more comprehensive way. But state data systems are generally not well equipped for this purpose because they typically only begin testing students in 3rd grade and tests change every few years so that trends over longer periods of time cannot be accurately measured.

NAEP could play to its current strengths and mitigate its weaknesses by adding a longitudinal component that tracks a nationally representative sample of students over time, from well before 4th grade to well after 8th grade.

Figures

Figure 1. 8th-grade math scores vs. average change since 4th grade, by state (correlation=0.51)

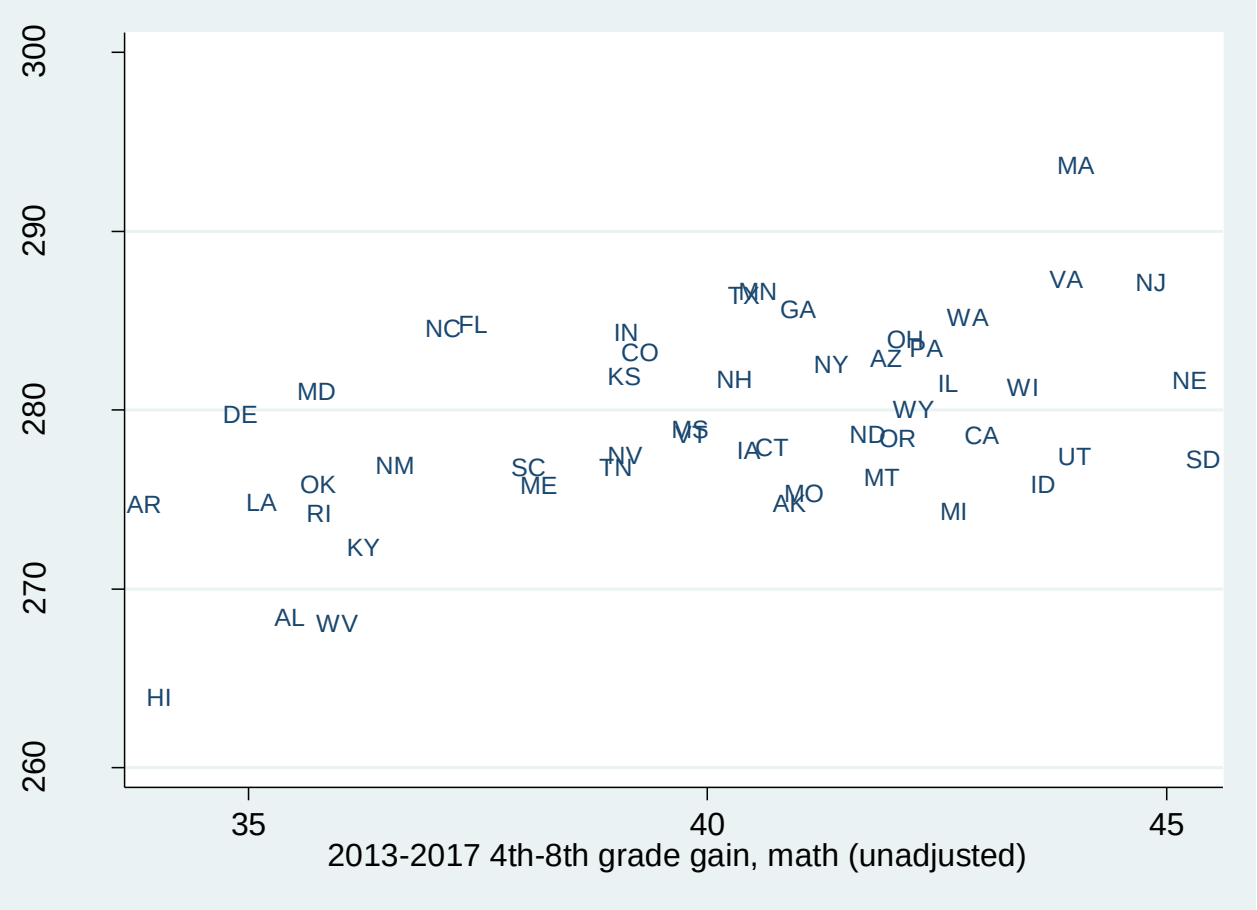


Figure 2. 8th-grade reading scores vs. average change since 4th grade, by state (correlation=-0.03)

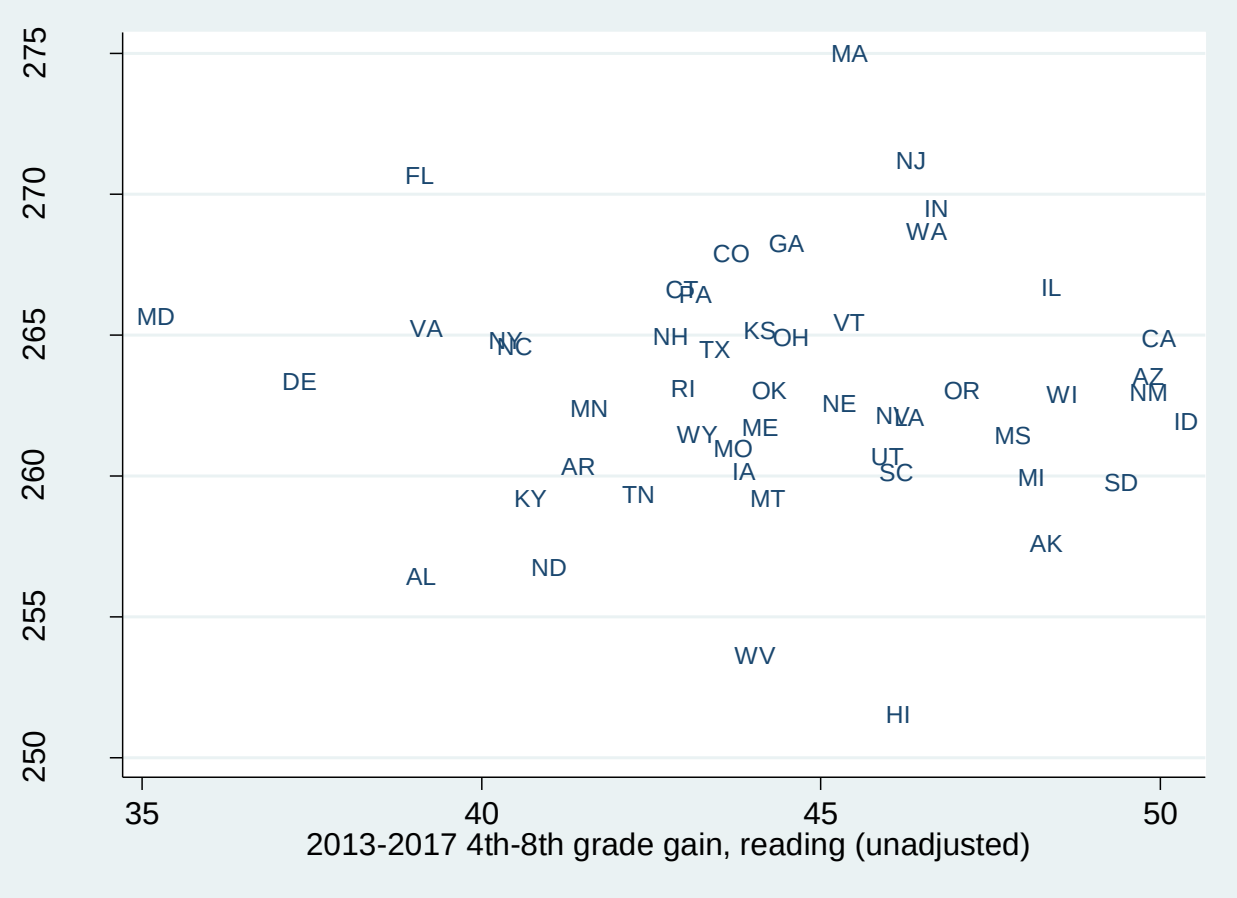


Figure 3. Change over 10 years in math scores of 2003 4th-grade cohort, measured in 4th and 8th grades, by state (correlation=0.50)

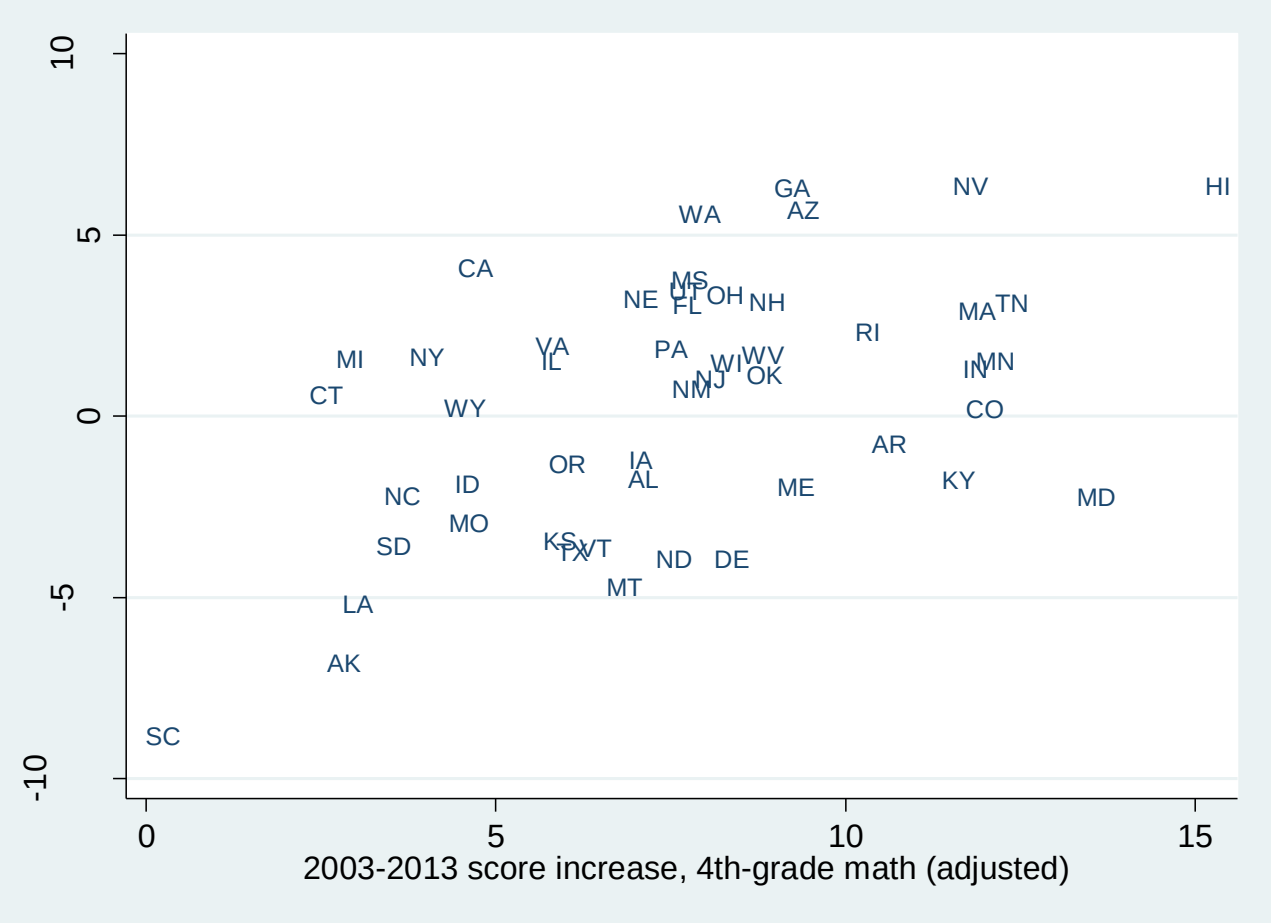
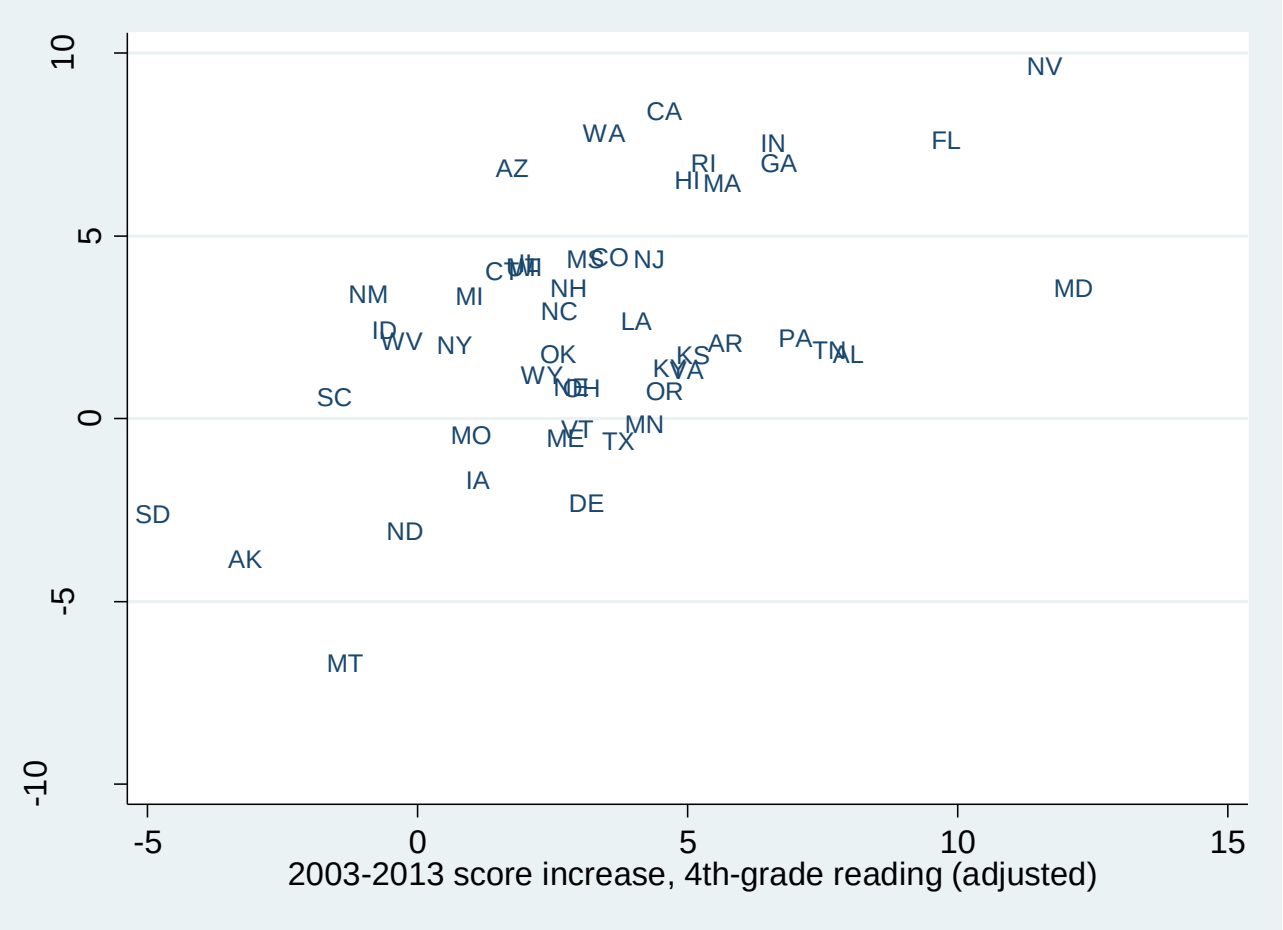


Figure 4. Change over 10 years in reading scores of 2003 4th-grade cohort, measured in 4th and 8th grades, by state (correlation=0.55)



[1] https://www.nationsreportcard.gov/reading_math_2017_highlights/

[2] The cohort can change over this period due to migration into and out of the state, but such changes over relatively short periods of time are likely to be small. I do not use the demographically adjusted scores discussed below for this part of the analysis because they are not designed to be comparable across grades.

[3] https://files.eric.ed.gov/fulltext/ED528992.pdf

[4] http://apps.urban.org/features/naep/

[5] https://www.urban.org/sites/default/files/publication/80251/2000773-varsity-blues-are-high-school-students-being-left-behind_2.pdf

[6] I use demographically adjusted scores that are re-normed each year so that, nationally, the adjusted mean score is the same as the unadjusted mean score. As a result, the scores are scaled such that national trends are not adjusted for national changes in demographics.