EDUCATION AND TRAINING

RESEARCH REPORT

Varsity Blues

Are High School Students Being Left Behind?

Kristin Blagg *Matthew M. Chingos*

May 2016

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is dedicated to elevating the debate on social and economic policy. For nearly five decades, Urban scholars have conducted research and offered evidence-based solutions that improve lives and strengthen communities across a rapidly urbanizing world. Their objective research helps expand opportunities for all, reduce hardship among the most vulnerable, and strengthen the effectiveness of the public sector.

Contents

Acknowledgments

This report was funded by an anonymous donor. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at www.urban.org/support.

The authors thank Tom Loveless and Martin West for helpful comments on drafts of this report.

Executive Summary

Efforts to improve the academic quality of US schools often leave high schools behind. Federal accountability policies require annual testing in math and reading for grades three through eight but only once in high school, where schools are also held accountable for increasing graduation rates. The relative lack of attention to secondary schools has coincided with disappointing national student achievement results for high school students, even as achievement among elementary and middle school students has risen significantly since the 1990s.

Has high school quality stagnated, or even deteriorated, as student achievement has increased in elementary and middle schools? Or are there other factors, such as changing demographics or declining student effort, which explain stagnant high school achievement? This report addresses these questions using student-level data from the National Assessment of Educational Progress (NAEP). We examine the results of nationally representative math and reading tests that have been administered since the early 1970s to better understand why the academic gains posted by elementary and middle school students have not persisted into high school.

The data strongly suggest that stagnant achievement among high school students is a real phenomenon. This result is consistent across different versions of NAEP and with other achievement tests and does not appear to result from changes in who is taking the test (e.g., as a result of rising high school graduation rates), flaws in test design and administration, or declining student effort.

Understanding why students are leaving high school with math and reading skills not much better than their parents awaits better data and additional research. We recommend several improvements to NAEP, including the regular assessment of high school students across the nation and in each state, as is done for younger students. There is also a critical need for researchers and policymakers to renew their focus on high schools and ensure that the academic gains that elementary and middle schools have produced are not squandered.

Varsity Blues: Are High School Students Being Left Behind?

Elementary and middle schools have long been the focus of education reform efforts. In Hamilton, Ohio, at the 2002 signing of the No Child Left Behind (NCLB) act, President George W. Bush stressed that states would be called upon to "design accountability systems to show parents and teachers whether or not children can read and write and add and subtract in grades three through eight."[1]

Bush was speaking to a group of high school students at the time, yet high schools have received significantly less attention in federal policy. NCLB promoted annual test-based accountability for elementary and middle schools, but subjected high schools to much looser testing requirements. States were required to assess students' mastery of math and reading standards at least once between 10th and 12th grade, a continuation of requirements from the previous reauthorization of the Elementary and Secondary Education Act in 1994. In lieu of additional annual testing, high school graduation rates were incorporated into the federal school accountability measures.

When Congress reauthorized NCLB as the Every Student Succeeds Act in late 2015, little changed in federal policy regarding high school. High schools are still required to test students once in math and reading using either a state-created or nationally recognized high school assessment test. In addition, states must include graduation-rate goals for high schools in their accountability systems.

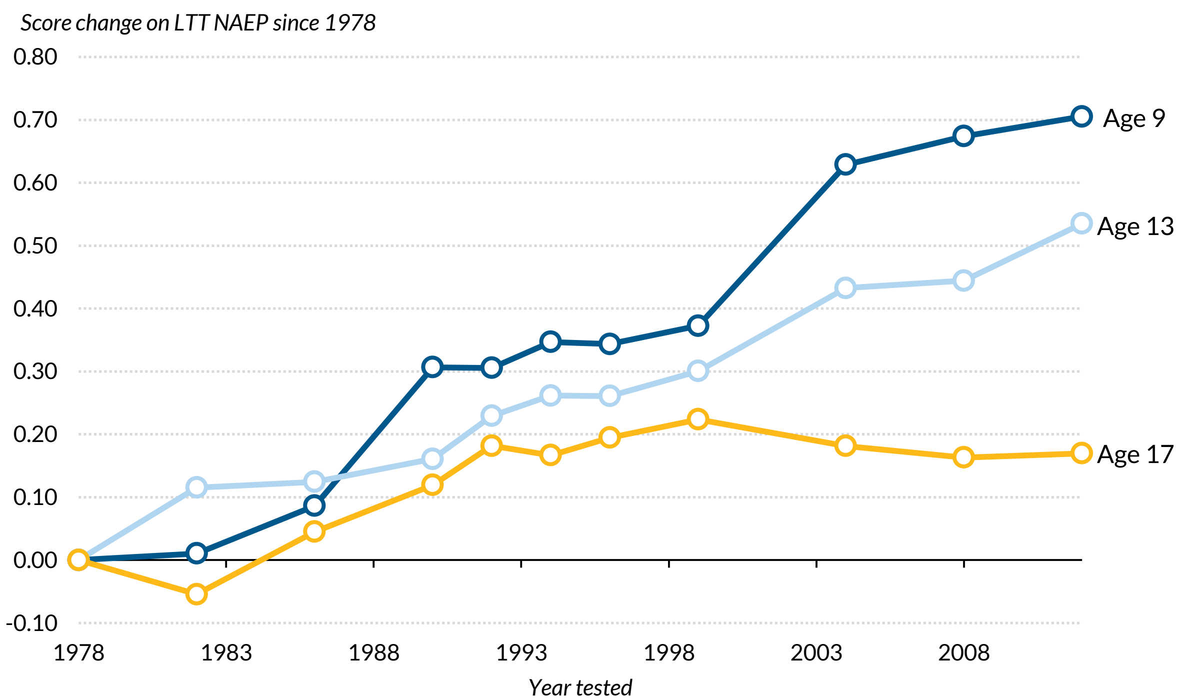
The lack of attention on high schools, relative to elementary and middle schools, has coincided with disappointing student achievement results for high school students, even during (and following) periods of substantial increases in achievement among elementary and middle school students. These trends are alarming, especially as many students are entering higher education unprepared for college-level work (Sparks and Malkus 2013). The goal of this report is to assemble the key data sources on high school achievement and empirically examine competing explanatory hypotheses for the stagnant trends.

We draw on the National Assessment of Educational Progress (NAEP), widely regarded as the Nation's Report Card. NAEP comprises a set of nationally representative assessments of student achievement in elementary, middle, and high schools. Since 1990, mathematics and reading assessments, known as main NAEP assessments, have been conducted about every two to four years for 4th- and 8th-grade students, and roughly every four years for 12th-grade students. In addition, a second set of tests, the long-term trend (LTT) NAEP assessments, have been administered in mathematics and reading to 9-, 13-, and 17-year-olds roughly every four years since the 1970s.[2]

On both the main and LTT assessments, elementary and middle school students have registered substantial achievement gains over time. However, national high school student achievement (12th-grade students and enrolled 17-year-olds, most of whom are in 11th grade) has remained largely stagnant. For example, while 9- and 13-year-olds saw increased performance on the LTT mathematics assessment by 0.71 and 0.54 standard deviations from 1978 to 2012, the performance of 17-year-olds inched upward only 0.17 standard deviations (figure 1). Even more worrying is that achievement for younger students has increased markedly since the early 1990s while high school achievement has been flat.

FIGURE 1

LTT NAEP Mathematics Score Changes, in 1978 Standard Deviations



This trend has puzzled education researchers and policymakers alike. Given the positive trends in student achievement among younger students, we should expect that at least some of those academic gains would persist into high school. Has high school quality stagnated, or even deteriorated, as student achievement has increased in elementary and middle schools? Or are there other factors, such as demographics or declining student effort on low-stakes exams, that explain stagnant high school achievement? We address these questions using student-level, restricted-use main and LTT NAEP data, which allow us to investigate the underlying trends in student populations that may affect overall student achievement.

The infrequency of NAEP administration makes it difficult to track student achievement at the high school level—an important problem in its own right. But the available data strongly suggest that stagnant achievement among high school students is a real phenomenon, and it cannot be explained away by rising high school graduation rates or problems with how the tests are designed and administered. The NAEP tests for high school are not perfect measures of academic achievement, but they are consistent across tests (main versus LTT) and with the results of other national achievement tests.

How to Use NAEP Data

NAEP is too often treated like a public opinion poll, the results of which are tea leaves that can be read by analysts and commentators interested in how school quality varies across places and over time. But NAEP reflects much more than just the quality of schools. In particular, NAEP scores reflect the underlying population of students taking the test. Schools in Massachusetts serve a different student population than schools in Mississippi, and the student population in 2016 looks different from the student population in 1990 or 1980.[3]

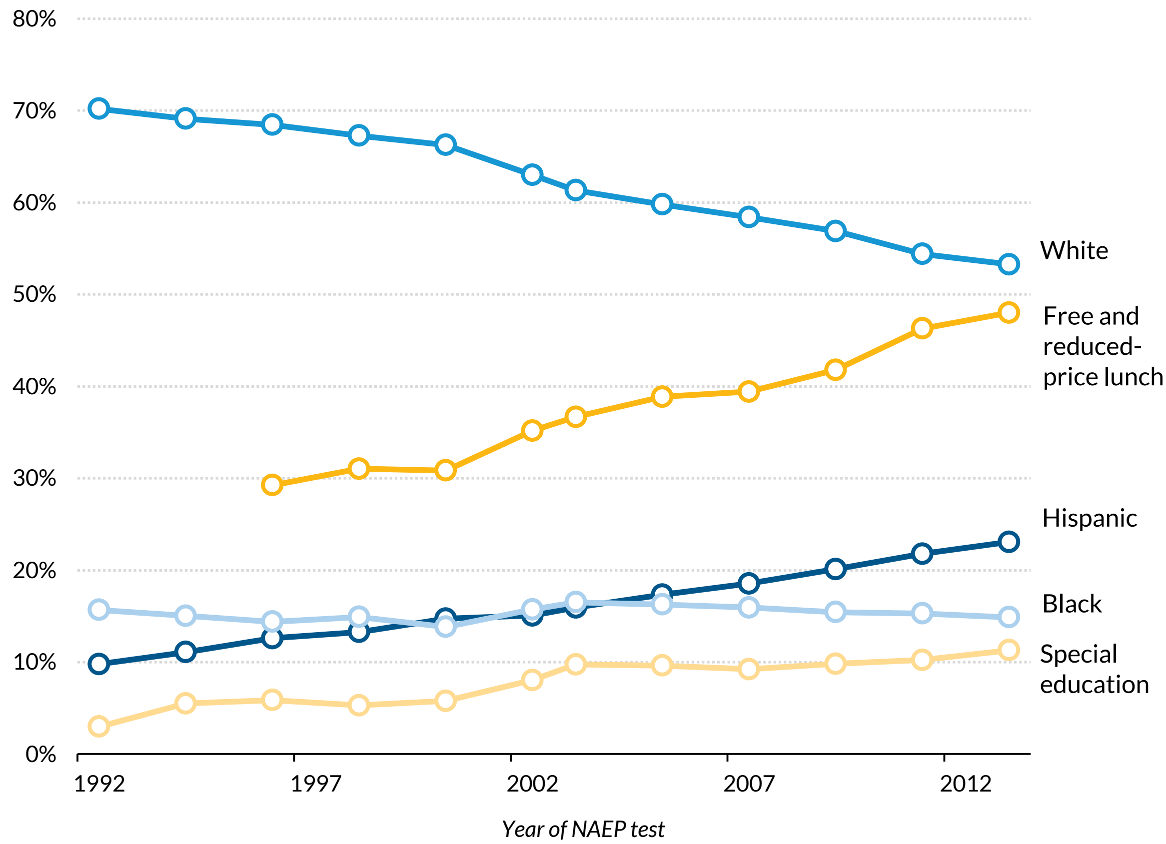
Given the changes in the nation's student population over time, an assessment of NAEP-score trends should also account for underlying demographic trends. Such an adjustment allows national trend scores to reflect how demographically similar students perform academically over time, assuming that the relationship between demographics and test performance is constant over the period analyzed. The adjusted trends do not isolate school quality, as they still reflect changes in non-education policies, culture, and unmeasured student characteristics, but they are a marked improvement over simply eyeballing the raw scores.

The US student population has changed substantially since the start of NAEP testing in the 1960s and 1970s, with most of these changes occurring since the start of main NAEP testing in the early 1990s (figure 2). In particular, the nation has seen an increasing Hispanic share of the population (and decreasing white share). In addition, the percentage of students identified as eligible for free and reduced-price lunch and as eligible for special education services has risen considerably over time. Appendix table A.1 shows that the increase in the Hispanic share was accompanied by increases in the percentage of students who do not speak English at home and who are identified as English language learners.

FIGURE 2

Demographics of Main NAEP Test Takers

Share of students over time



Note: Demographics averaged across 4th- and 8th-grade main mathematics and reading NAEP tests. We do not use 12th-grade data because 12th-grade students are assessed less frequently

Some of these trends are the result of national demographic shifts, but others may be at least partly driven by policy changes and school decisionmaking that affect how demographics are measured. For example, over the past decade, eligibility criteria for the National School Lunch and School Breakfast programs have undergone significant change. States now directly certify students from households who receive Supplemental Nutrition Assistance Program benefits, and, starting in 2010, districts began to adopt community eligibility, allowing schools to grant free lunch to all students if 40 percent or more would qualify through direct certification (Moore et al. 2013; Neuberger et al. 2015, note 2).

These policy changes may have contributed to the increase in the share of students identified as eligible for free and reduced-price lunch, which has risen steadily despite the fact that measures of childhood disadvantage usually correlate with the state of the economy and thus are cyclical (Child Trends Databank 2015). The policy decisions schools and districts make that affect an individual student's classification, but they also affect the overall measurement of free and reduced-price lunch recipients and other designations, such as special education students and limited English proficiency students. In other words, trends in these characteristics potentially reflect changes in both the student population and changes in federal, state, and local policy.

Including measures policy changes affect, such as free and reduced-price lunch, could overadjust the results, but excluding them could also underadjust. For this reason, we report results both with and without measures we view as correlated with policy decisions as well as demographic changes.[4]

We adjust NAEP scores for demographic shifts over time by estimating how scores would have changed given observed changes in the student population but holding constant the relationship between demographics and test scores (estimated using 2003 data). Our methodology is detailed in the technical appendix; the key assumption we make is that the relationship between demographics and NAEP scores is stable over time.[5] The fact that the US student population is increasingly composed of demographic groups who tend to score less well on NAEP means that adjusting for these shifts increases the NAEP-score gain from the mid-1990s to the present. Unadjusted NAEP scores also increased, which means that NAEP scores have increased in spite of demographic shifts that should tend to drag down the average.

Table 3 shows the unadjusted- and adjusted-score changes on the fourth- and eighth-grade math and reading NAEP tests from the mid-1990s through 2013 (the last year for which student-level data are available to researchers).[6] All changes are expressed in student standard deviation units. One way to interpret standard deviation units is to compare them with the average annual increase in test scores made by students at different grades. The average fourth-grade student sees an increase in student achievement of about half a standard deviation. For the typical eighth-grade student, the increase is about one-quarter of a standard deviation (Hill et al. 2008).

Unadjusted NAEP scores increased over this period, with especially large increases in math achievement. For example, the gain for fourth-grade students, averaged across reading and math tests, was 0.38 standard deviations, or about 75 percent of a year of learning. In other words, the average fourth-grade student in 2013 was roughly 75 percent of a grade level ahead of the average student in the mid-1990s.

Adjusting for national student population changes in age, gender, race, and language spoken at home—controls that school or district policies largely do not affect—increases the score gain by an average of 0.09 standard deviations across the four tests. When we add additional controls for special education, limited English proficiency, and free and reduced-price lunch eligibility, we find that the national adjusted scores increase by another 0.11 standard deviations, on average.

TABLE 1

Unadjusted and Adjusted Changes in Main NAEP Scores, 1996/98–2013

	No controls	Nonpolicy controls	All controls
4th-grade math	0.63	0.71	0.82
4th-grade reading	0.12	0.21	0.32
8th-grade math	0.35	0.43	0.55
8th-grade reading	0.11	0.19	0.31
Average	**0.30**	**0.39**	**0.50**

Notes: Changes are for 1996–2013 for math and 1998–2013 for reading. Changes are given in 2003 standard deviations.

Another way of interpreting these changes is to calculate how the average student from 2013 would compare with students who took the NAEP exam in the mid-1990s. An average (50th percentile) eighth-grade student on the math exam from 2013 would have scored in the 64th percentile in 1996. When controlling for non-policy demographic changes in the 2013 student population, this eighth-grade student would move up to the 67th percentile, and when controlling for all demographics, she would score in the 71st percentile of 1996 students.

These results show that accounting for demographics is important, but the size of the adjustment depends on exactly what demographic measures are included. We also report the results of a number of variations on our methodology in table A.1, which are largely consistent with our main results. However, we do find that adjusting for student-reported parental education, which we can only do for eighth-grade students, reduces the size of the demographic adjustment.

Demographic shifts are important, but are they the reason that 12th-grade scores have stagnated? Are other factors at play? Before providing new evidence on these questions, we first review the data limitations that make these questions challenging to answer.

Data on High School Achievement Are Limited

The cumulative growth in elementary and middle school NAEP scores is clear in both the main and LTT data. The trend in high school scores is much less clear, however, in large part because of data limitations. The 12th-grade main NAEP, and the 17-year-old LTT NAEP, are given less frequently than the 4th- and 8th-grade main NAEP tests, leading to less certainty about the overall trend of high school student achievement.

Main NAEP tests are given to 4th- and 8th-grade students on a biannual basis, whereas 12th-grade students typically take the test every four years. Moreover, the 4th- and 8th-grade tests assess a larger sample of students to produce valid average scores for individual states, whereas state participation on the 12th-grade exam is voluntary (13 states participated in the 2013 assessment, and the 2015 test did not include state sampling). Since 2003, when the 4th- and 8th- grade main NAEP became mandatory for all states, these tests have been given a total of seven times, but the 12th-grade main NAEP assessment has only been conducted four times (in 2005, 2009, 2013, and 2015).

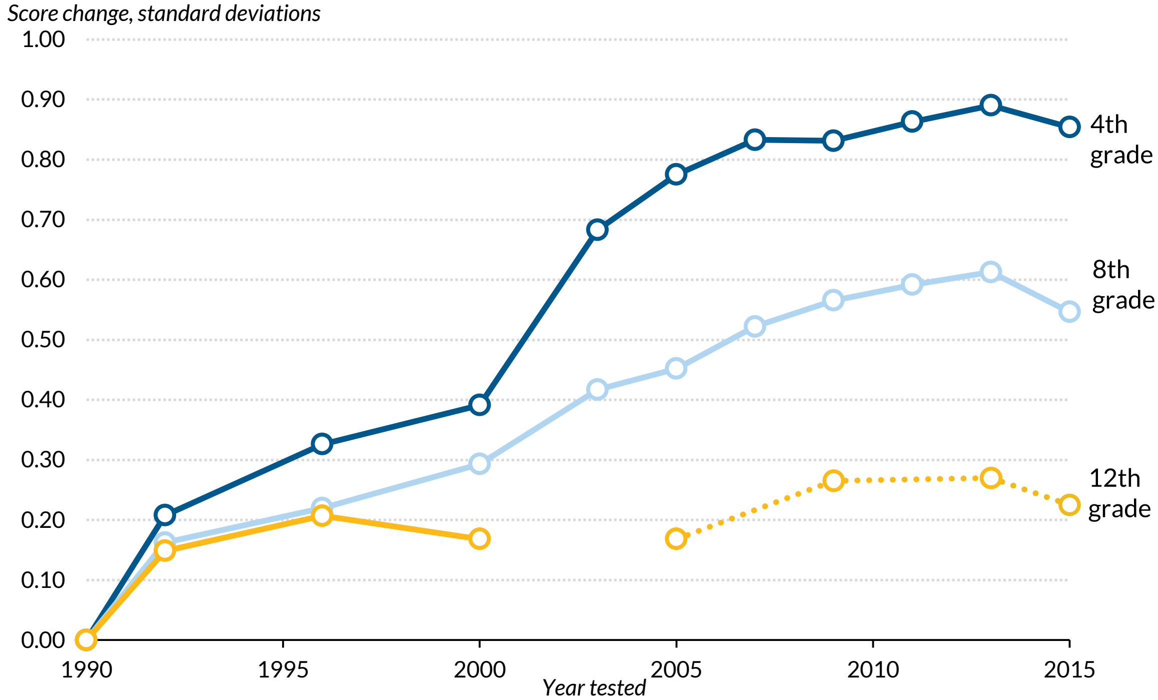
The LTT NAEP reading and math tests, though given less frequently, were consistently administered to all three age groups (9-, 13-, and 17-year-olds) every two to four years until 2012. However, budget cuts have led to the cancellation of the 2016 and 2020 administrations, so new LTT data will not be available until 2024 at the earliest, 12 years after the previous assessment.[7]

Trends on the main mathematics NAEP (figure 3), from 1990 to 2015, follow a similar pattern to the LTT mathematics NAEP (figure 1). Students in the elementary and middle school grades posted substantial gains—0.85 and 0.55 standard deviations, respectively—from 1990 to 2015.

Tracking 12th-grade mathematics performance is complicated by the introduction of a new NAEP math test in 2005 aimed at better reflecting high school curriculum.[8] However, performance on both 12th-grade math tests has stayed relatively flat over time, with scores growing by just 0.17 standard deviations between 1990 and 2000, and by 0.10 standard deviations from 2005 to 2013, a total change of 0.27 standard deviations (assuming no change between the last administration of the old test in 2000 and the first administration of the new test in 2005).

FIGURE 3

Main NAEP Math Score Changes, Expressed in 1990 Standard Deviations



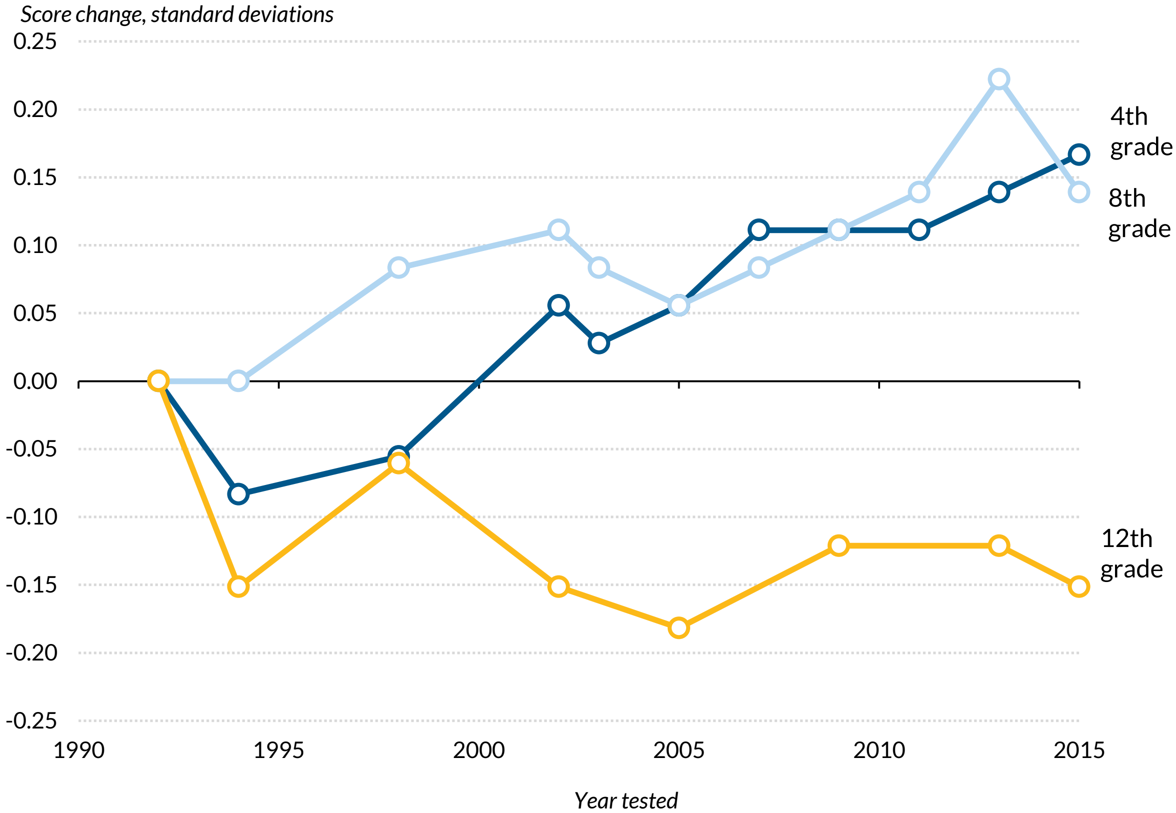
Notes: Testing accommodations were not given on the 1990 or 1992 exams. The 12th-grade math framework was adjusted in 2005. We restarted the score at the 2000 level, used the 2005 standard deviation and indicated this new test with a stippled line.

Trends on the main and LTT reading NAEP tests are noisier than the mathematics NAEP test trends, but they also indicate rising achievement in the 4th and 8th grades (and for 9- and 13-year-olds), with stagnant achievement in 12th grade (17-year-olds). Figure 4 shows the score trends for the main reading NAEP from 1992 to 2015. The nation's 4th-grade students improved their reading scores by 0.17 standard deviations over their 1992 performance, while 8th-grade students improved their scores by 0.14 standard deviations. In contrast, 12th-grade students in 2015 scored 0.12 standard deviations lower than their 1992 counterparts, and their performance has remained relatively flat since 2002.

Performance of 9-, 13-, and 17-year-olds on the LTT reading NAEP demonstrate similar trends over time, particularly between 1990 and 2012 (the most recent LTT NAEP test), when the nation's 9-year-olds posted growth of 0.28 standard deviations, and 13-year-olds gained 0.17 standard deviations. Over the same time period, 17-year-olds saw a decrease of 0.07 standard deviations (see figure A.1).

FIGURE 4

Main NAEP Reading Score Changes, in 1992 Standard Deviations



Note: Testing accommodations were not given on the 1990 or 1992 exams.

Hypotheses for Fade Out of NAEP Achievement Gains in High School

In both mathematics and reading, high school students lag behind their elementary and middle school contemporaries in performance on the main and LTT NAEP tests. Why are these achievement gains fading out by the end of high school? We examine four hypotheses to explain this puzzle and provide new empirical evidence wherever possible:

- Cohort adjustment: Not enough time has passed for gains of younger students to appear when those same cohorts are tested as high school students.
- Marginal graduate: Increasing high school persistence and graduation rates are dragging down scores of 12th-grade students (and 17-year-olds).

- Senioritis: More recent cohorts of high school students are taking NAEP less seriously, leading to lower scores.
- Measurement: NAEP has become less aligned to the subject matter taught in high schools over time.

Cohort-Adjustment Hypothesis

A potential explanation for stagnant high schools scores is that not enough time has passed for the gains from earlier grades to show up in the test scores of students in later grades. Underlying this hypothesis is the idea that, rather than looking at achievement level changes by test date, we should be tracking the performance of cohorts of students as they progress through school.

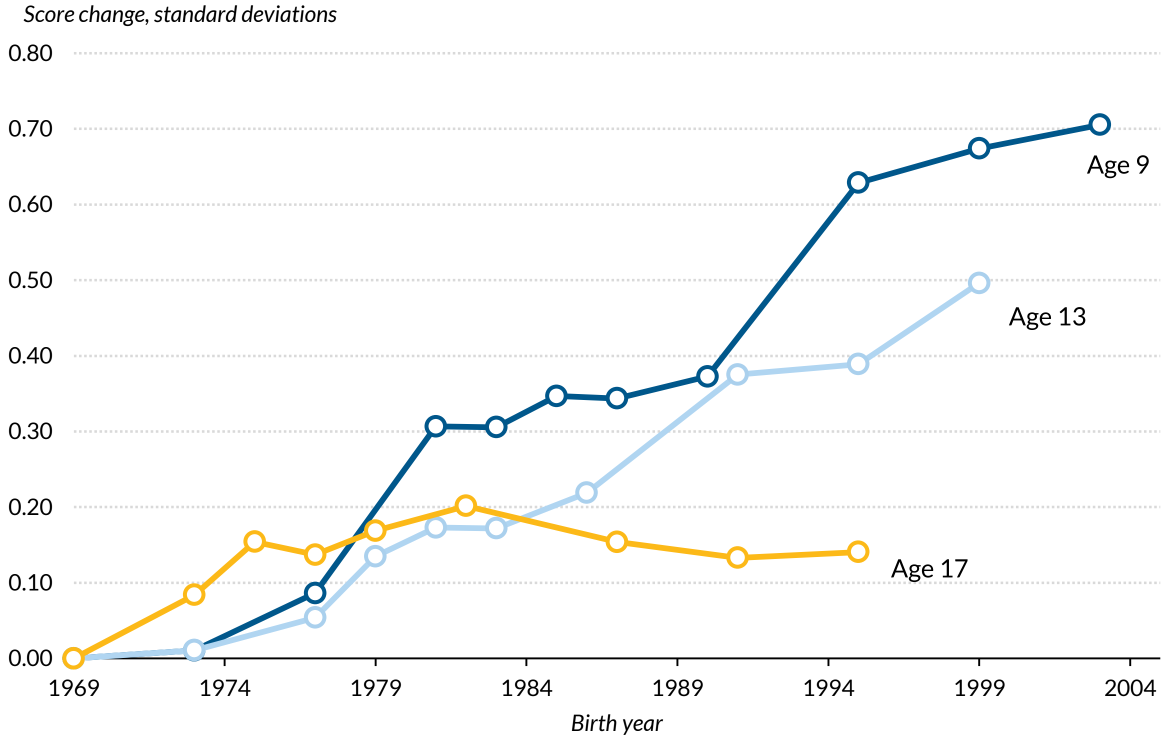
To evaluate this hypothesis, we examine LTT NAEP scores by birth year, rather than by year tested. In effect, this shifts test score data for 13-year-olds backward 4 years relative to data for 9-year-olds, and data for 17-year-olds backward 8 years.

Figure 5 illustrates the cohort analysis for the LTT mathematics NAEP results. We can see that there are periods, particularly for students born after 1985, when scores for 9- and 13-year-olds increased over previous cohorts. However, by the time those students were assessed at age 17, their scores remained stubbornly similar to those of past cohorts. A cohort analysis of the LTT reading assessment produces similar results (see figure A.2).

Although this analysis casts doubt on the cohort hypothesis, we cannot completely discount this theory. It is possible that future cohorts (those born later than 1995) will retain some of the gains that they posted when tested at age 9 and 13. Future data on these most recent cohorts would provide a better understanding of the magnitude of achievement fade-out over time, but unfortunately the next LTT data collection is planned for 2024 (birth year 2007).

FIGURE 5

LTT NAEP Mathematics Score Changes, by Cohort Birth Year



Notes: Scores are adjusted using the first year of cohort assessment (1978 standard deviation for 9 year olds, 1982 for 13 year olds, and 1986 for 17 years olds). The 2008 and 2012 assessment year scores are from the revised format test.

Marginal-Graduate Hypothesis

A leading hypothesis for stagnant high school NAEP scores posits that increases in high school persistence and graduation rates have increased the number of academically marginal students who remain enrolled in school and are therefore included in the 12th-grade (and 17-year-old) testing pool. In the words of the National Center for Education Statistics official who oversees NAEP administration, "What's happening is that students who would normally drop out of school are staying in.... Students who would normally not be taking our assessment, they're in there now at larger proportions."[9]

We test this hypothesis in two ways. First, we test whether implementing demographic adjustments has a larger effect on the test-score trends of high school students than it does for younger students (as we might expect to result from increasing high school persistence rates). Second, we examine whether the divergence in the scores of older and younger students coincides with a rise in high school persistence rates.

We performed demographic adjustments on the LTT reading and mathematics NAEP tests, similar to the adjustments that we showed for fourth- and eighth-grade students on main NAEP (table 2). To employ consistent controls across time, we used student-level data starting from 1982 (for mathematics) and 1980 (for reading) through 2008 (the last year for which student-level data are available to researchers).[10]

Employing controls for age (birth month), race and ethnicity, and gender, we see that the shift in student populations from the early 1980s to 2008 generates a relatively consistent upward adjustment of 0.09–0.14 standard deviations for all age groups. Performing an adjustment using just the data from 1990 to 2008 shows a similarly consistent adjustment (an increase ranging from 0.06 to 0.11 standard deviations).

TABLE 2

LTT Reading and Mathematics Achievement, Adjusted for Demographics

	Student Standard Deviations (1982/1980)			Student Standard Deviations (1990)		
	No controls	Race, gender, age	Race, gender, age, parent education	No controls	Race, gender, age	Race, gender, age, parent education
Math						
9-year-olds	0.68	0.80		0.40	0.51	
13-year-olds	0.39	0.49	0.40	0.36	0.43	0.39
17-year-olds	0.24	0.33	0.21	0.05	0.11	0.06
Reading						
9-year-olds	0.13	0.27		0.24	0.33	
13-year-olds	0.03	0.16	0.08	0.08	0.17	0.10
17-year-olds	0.01	0.15	0.07	-0.10	-0.03	-0.07

We also calculate an adjustment that takes into account student- reported parental education for 13- and 17-year-olds. This additional measure depresses the size of the adjustment, since students have reported relatively higher levels of parent education over time. This is consistent with our analysis of eighth-grade main NAEP scores (see table A.1).

The marginal-graduate hypothesis predicts substantially larger increases in the adjusted scores of 17-year-olds, compared with 9- and 13-year-olds. Although we do see that the demographic adjustment raises scores for 17-year-olds, we do not see a shift that is dramatically different from the shift for 9- and 13-year-olds. If anything, from 1990 to 2008, the magnitude of the score adjustment tends to be smaller for older students.

The demographics captured in the NAEP data are limited and thus are likely to miss characteristics such as academic ability and motivation that are correlated with both NAEP scores and high school persistence rates. But if rising high school persistence rates are partly behind stagnant NAEP scores, then we might expect rising persistence to coincide with a divergence between the NAEP scores of student cohorts when they are tested at age 9 and at age 17. When marginal students persist beyond compulsory age at a higher rate and are therefore assessed at age 17, we should expect that their relatively lower NAEP scores not only would depress performance relative to previous high school cohorts but would also decrease NAEP performance relative to the same cohort of students on earlier assessments.

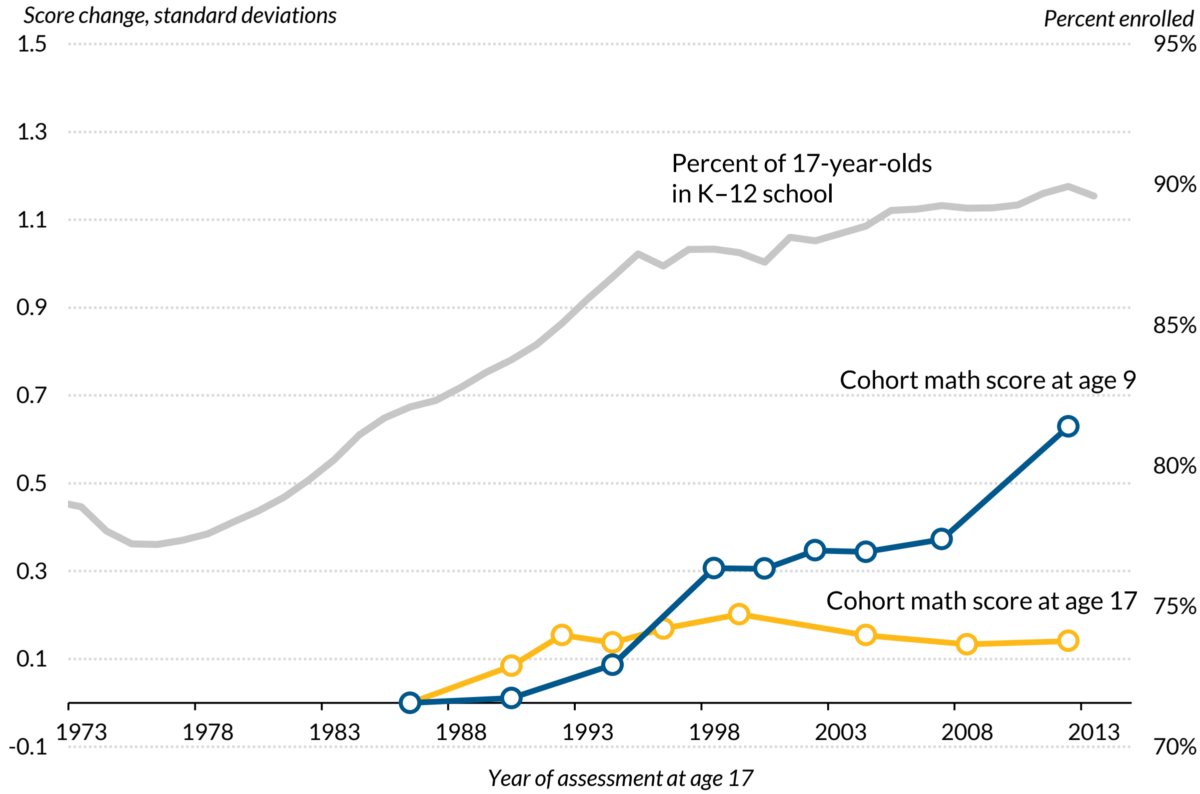
A comparison of NAEP scores and school enrollment rates of 17-year-olds for the same birth cohorts puts a significant dent in the marginal-graduate hypothesis. High school graduation rates have hit record highs for four straight years, reaching 82 percent for the class of 2014.[11] However, this recent progress is overshadowed by longer-term trends in 17-year-old enrollment in K–12 schools, which show that the largest increase in persistence among this age group took place much earlier.

Figure 6 shows Current Population Survey data on the percentage of surveyed 17-years-olds who were enrolled in an elementary or secondary grade.[12] The marginal-graduate hypothesis would lead us to expect the most divergence between the LTT NAEP scores of 17-year-olds and younger students for the birth cohorts who experienced the largest increase in high school enrollment rates. But if anything we find the opposite. The largest jump in 17-year-old enrollment occurred between roughly 1980 and 1995, with an enrollment rate increase of approximately 10 percentage points. NAEP scores did not begin to diverge significantly by age until roughly 2000, a period during which school enrollment rates were comparatively flat.

The marginal-graduate hypothesis has obvious intuitive appeal, but both of our empirical tests suggest that rising high school persistence rates have played little role in stagnant high school achievement.

FIGURE 6

17-Year-Old Enrollment and LTT NAEP Results



Source: Current Population Survey.
Note: Scores are adjusted using the first year of cohort assessment (1978 standard deviation for 9-year-olds and 1986 for 17-years-olds). Trend for percent of 17-year-olds in a K–12 school is calculated as a five-year moving average of annual weighted Current Population Survey percentages.

Senioritis Hypothesis

Another hypothesis with intuitive appeal posits that today's 12th-grade students (and 17-year-olds) take the NAEP tests less seriously than previous high school students. Perhaps students are jaded by the increasing number of standardized tests that they take in their school career and do not make an effort on the test. If this is true, it would depress scores relative to the true skills of today's high school students to a greater degree than previous cohorts.

There is some evidence that students perform better on the NAEP when they have an incentive to do so. When 12th-grade students were offered a financial incentive for taking the NAEP test, with the amount of the incentive contingent on correct answers, they performed, on average, about 0.14

standard deviations higher (Braun, Kirsch, and Yamamoto 2011). This study and others like it show that NAEP tests may somewhat underestimate students' academic ability because of the low-stakes nature of the assessment, but they do not tell us whether the motivation of high school students has changed over time.

Data limitations make it difficult to assess how student effort on NAEP tests has evolved over time, but neither of two brief analyses we conducted indicate declining effort. First, we examine the average proportion of test items skipped by students on the LTT NAEP (either by leaving them blank, marking "I don't know," or failing to reach the item during the allotted time). We find that the percentage of unanswered questions stays relatively constant over time, at an average of roughly 6 to 8 percent from 1978 to 2004.[13]

Second, we examined student self-reports of the amount of effort they put into the main 12th-grade NAEP. We found that the level of students who report trying "much harder," "harder," or "about as hard" on the assessment as on a school test stayed constant at roughly 50-60 percent of students from 1992 to 2009.

More research is needed to better understand possible changes in student effort on low-stakes tests such as the NAEP, but the available evidence provides no reason to believe that effort has declined.

Measurement Hypothesis

The final hypothesis we examine is that the main and LTT NAEP tests fail to accurately capture changes in high school student achievement over time because of issues related to assessment design and administration. For example, perhaps there are evolving problems with how high school students are sampled, or perhaps the test content is becoming less aligned with what students learn in high school.

Participation rates on the NAEP do decline as students grow older, with the weighted percent of students participating hovering around 90 percent for 9- and 13-year-old LTT participants compared with 75–85 percent for 17-year-olds (Perie et al. 2005).[14] However, participation rates on the LTT NAEP do not exhibit any consistent trend over time (see table A.3).

In contrast, 12th-grade participation rates on main NAEP increased starting in 2005, likely bolstered by the inclusion of a pilot state assessment program in 2009, in which 11 states participated (National Center for Education Statistics 2010). Studies of NAEP nonparticipation rates show that, all else equal, increasing participation has the potential to lower scores. However, the effect of student and school nonparticipation is fairly small—one study found that a change of 20 percentage points in the school nonparticipation rate would likely bias the estimate by less than one NAEP scale score point (about 0.03 standard deviations) (Grissmer 2007).

The main NAEP assessment is designed to keep pace with the content and assessment formats that students are exposed to in their schools. NAEP frameworks are occasionally revised (as in the 2005 12th-grade mathematics framework) to better align with what students are expected to learn. In addition, the National Assessment Governing Board, which oversees the NAEP subject frameworks, has argued that content on the 12th-grade main NAEP exams align closely with the current versions of the SAT and ACT.[15]

The LTT NAEP assessment does not undergo framework content changes so that it remains a constant measure of student performance over time. However, if the content of the LTT NAEP exam is substantially different from the content that high school students are exposed to, we might expect to see the trends in main NAEP scores diverge from those of the LTT NAEP. Instead, both LTT and main NAEP exams demonstrate similar trends.

Further evidence that the stagnation of high school achievement (but not earlier grades) is not confined to any particular test or set of tests comes from international assessments. The Trends in International Mathematics and Science Study (TIMSS) shows that the performance of US fourth- and eighth-grade students has improved significantly over time, mirroring NAEP results. Fourth-grade math scores rose by 0.27 standard deviations from 1995 to 2011 and eighth-grade scores rose by 0.21 standard deviations over the same period.[16] The Progress in International Reading Literacy Study (PIRLS), which assesses reading skills in fourth grade, also showed significant gains for US students from 2006 to 2011.[17]

In contrast, the Program for International Student Assessment (PISA), which assesses the performance of 15-year-olds (typically enrolled in 9th or 10th grade), has shown flat achievement for the United States. American students' average score on reading has not changed significantly between 2000 and 2012, and their average math score has not changed from 2003 to 2012.[18] In contrast, countries like Germany, Indonesia, Mexico, Poland, and Portugal posted significant gains in both math and reading over similar time periods.

The High School Conundrum

The empirical evidence we have assembled rejects most of the hypotheses put forward for stagnant high school NAEP scores, especially those that try to explain away this trend by pointing to measurement-related issues. Why does it seem that high schools do not sustain the achievement gains realized at the elementary and middle school levels? We briefly discuss a few competing (but not mutually exclusive) theories, all of which are fertile ground for future research.

One possibility is that average high school quality has deteriorated, as state and federal policy efforts have focused more on elementary and middle schools. For example, perhaps schools and districts responded to incentives NCLB and other policies created by focusing more attention and resources on elementary and middle schools, partly at the expense of high schools. As a result, today's students are entering high school better prepared than previous generations, but those gains are erased by the diminished academic quality of high schools.

A second possibility is that high schools are not getting worse but function in a way that leads to fade-out of overall test achievement gains from earlier grades. This could be the case if high schools have not responded to the increasing achievement levels of their ninth-grade students, instead producing the same "outputs" with better "inputs." This could result from the fact that high school is more of an assortment of subject-specific courses, rather than a coherent academic program that builds on achievement in prior grades. This could also be a function of the increased academic sorting that occurs in high schools, where students may opt into less-challenging courses even if they have the preparation to pursue more challenging academics.

The upshot of both of these possibilities is that our school *system* is teaching students key academic skills in math and reading *earlier*, as evidenced by 4th- and 8th-grade scores, but not increasing *overall* achievement in these subjects, as evidenced by 12th-grade scores. This does not mean that gains achieved earlier will not translate into benefits later in life, such as improved educational attainment and income, as studies of many educational interventions make clear (Chetty, Friedman, and Rockoff 2014; Deming et al. 2015; Dynarski, Hyman, and Schanzenbach 2013). But it is certainly cause for concern, especially in light of evidence that educational achievement as measured by test scores is a much stronger predictor of economic growth than educational attainment as measured by years of schooling (Hanushek et al. 2008).

The evidence shows that high schools are not sustaining academic achievement growth created in the earlier grades, but persuasive explanations for why this is the case and what to do about it awaits better research. To produce that research, we need better data on high school student achievement. The final section of this report provides a starting point for that necessary discussion.

Conclusion and Recommendations

The evidence presented above is not conclusive, but it suggests that researchers and policymakers need to take a hard look at high schools and undertake efforts to collect more and better data on high school achievement. We make the following recommendations:

Invest in routine high school NAEP: NCLB left high schools behind by requiring state-level NAEP participation in 4th and 8th grade but not 12th grade. Participation in the 12th-grade main NAEP assessment should become mandatory for all states and should be carried out on the same schedule as the earlier grades (i.e., every two years instead of every two to four years). This would provide researchers and the public with a better understanding of how high school achievement varies across the nation and could produce insights as to what policies increase student achievement in high school.

Do not abandon long-term trend NAEP: This report demonstrates that LTT NAEP is an important and valuable source of data on the achievement of students in elementary, middle, and especially high school. The return on the significant public investment in LTT NAEP will be highest if this test continues to be administered at four-year intervals, rather than going on a 12-year hiatus from 2012 to 2024, as is planned.

Link NAEP to other data: Some traditional student demographic measures, such as a student's free and reduced-price lunch status, lack the consistency and granularity needed to accurately track changes in the population of NAEP test takers. The National Center for Education Statistics has acknowledged the growing difficulty of determining socioeconomic status by commissioning a study on alternate ways to assess income differences in the NAEP student population (Cowan et al 2012). We believe that NAEP could be markedly improved by linking student data to other administrative datasets, such as Internal Revenue Service data on parental income. Publishing aggregate data on student performance by family income would protect families' privacy while providing a more nuanced picture of academic performance at all grade levels.

Consider a longitudinal NAEP: An important limitation of NAEP is that it only captures the performance of individual students at a single point in time. The National Center for Education Statistics should investigate the feasibility of piloting a longitudinal NAEP assessment that tracks cohorts of students over time. For example, a subset of the 4th-grade students in the 2017 NAEP could be resampled for follow-up during the 2021 8th-grade NAEP and the 2025 12th-grade NAEP. This type of longitudinal data collection could be embedded within the existing NAEP assessments and would enable more definitive analyses of how student achievement evolves as students progress through school.

Focus on high school achievement: The accountability movement in education produced a substantial amount of data on students, teachers, schools, and districts. However, because most achievement data are produced for students in grades three through eight, many researchers have focused their attention on elementary and middle school. Researchers and policymakers should renew their focus on high schools, centering not only on issues of high school completion, but also on issues of academic achievement. Evidence is needed to understand whether high schools should test students more frequently, as is required for earlier grades by the annual testing provision of the Every Student Succeeds Act, and to grapple with the optimal role of grade-specific and course-specific testing in high school accountability measures. More and better data are needed, but there is likely much that could be done with existing state data on end-of-course tests and high school exit exams.

The NAEP assessments are fundamental for researchers and the public to understand the academic performance of the nation's students. For too long, the academic performance of the nation's high school students has been overlooked or explained away. All of the available data provide a wake-up call for researchers and policymakers to renew their commitment to these students and ensure that the academic gains that elementary and middle schools have produced are not squandered.

Appendix A. Data and Methodology

This report draws on restricted-use, student-level data on long-term trend (LTT) and main NAEP reading and mathematics tests. We have access to all available datasets (up to 2008) from the LTT NAEP, which has been administered to 9-, 13-, and 17-year-olds since 1978 (mathematics) and 1971 (reading). The last administration of the LTT NAEP was in 2012, but student-level data from that administration are not yet available to researchers.

We also have access to all available datasets (through 2013) from the main NAEP, which has been administered to 4th-, 8th-, and 12th-grade students since 1990 (mathematics) and 1992 (reading). The most recent administration of main NAEP was in 2015, but student-level data from that administration are not yet available to researchers.

LTT NAEP Variables

We use data on students in the national reporting sample (R2SAMP = 1 in pre-1984 assessments, COHORT =1, 2, or 3 for assessments from 1984 onward). To minimize the time that students sit for the tests, no student takes an entire test. For the analysis, we use a statistical estimate of what each student's score on the test would have been had he or she taken the test in its entirety. For the LTT NAEP, this estimate is based on a set of five plausible test-score values and between 46 and 74 replicate weights (depending on test administration year; 62 replicate weights have been consistently used for both math and reading since the early 1990s).

For demographic adjustments of the data over time, we use a set of four student-level control variables that were available in the data from 1982 (mathematics) and 1980 (reading) onward. The variables used and their coding are as follows:

- SEX: gender (male or female)
- SRACE: race and ethnicity, from school records (white, black, Hispanic, Asian, American Indian, or other)
- PARED: Parent education, as reported by student (no high school, high school graduate, some college, or college graduate)

    - This variable was only reported for 13- and 17-year-olds. In 1980 and 1982, this variable was coded as PARED5

- BIRTHMO/MOB/BMONTH and BIRTHYR/YOB4/YOB/NEWYOB: Age in October of testing year using date of birth estimated as 15th day of birth month in birth year
    - All ages were confirmed to be in the age range of the given assessment (i.e., between 9.00–9.99 for 9-year-olds)

LTT NAEP Adjustment Methodology

All analyses are run separately by age and subject and are weighted to be nationally representative (using weight variable WEIGHT in pre-1990 assessments, and ORIGWT in assessments given from 1990 onward).

We perform an adjustment across time by estimating regression coefficients using the student-level data from 1992, roughly halfway between the available time periods for mathematics (1982–2008) and reading (1980–2008). Specifically, we regress the test score of each student in 1992 on the set of control variables described above. Control variables are included in the regression using dummy variables identifying each of the groups of students for each construct with the exception of the one arbitrarily chosen group that is the omitted category (except for age, which is included as a continuous variable).

Using this regression, we estimate a residual for each student across all assessment years, which is the difference between their actual score and a predicted score based on the relationship between the predictors and test scores in 1992. We calculate the adjusted change in the national average score between any two years as the change in the weighted average of the student-level residuals between those years.

Main NAEP Variables

We use data on students in the national reporting sample (RPTSAMP= 1). To minimize the time that students sit for the tests, no student takes an entire test. For the analysis, we use a statistical estimate of what each student's score on the test would have been had he or she taken the test in its entirety. For the 2013 main NAEP, using the procedures described in the documentation provided with the restricted-use data, this estimate is based on 20 plausible test-score values and 62 replicate weights (in prior years, including 2003, there were 5 plausible values instead of 20).

For demographic adjustments of the data over time, we use a set of student-level control variables which were available in the data from 1996 (mathematics) and 1998 (reading) onward. The variables used for the report adjustment and their codes are as follows:

- SEX: gender (male or female)
- SDRACE: race and ethnicity variable used by NCES to report trends (white, black, Hispanic, Asian, American Indian, unclassified/two or more races)
  - » This variable is school-reported, though supplemented at times with student-reported data and is coded as DRACE in the 1990, 1992, 1996, 1998, and 2000 assessments.
- SLUNCH: eligibility for federal free and reduced-price lunch program (not eligible, eligible for reduced-price lunch, eligible for free lunch, or other or missing)
- BMONTH and BYEAR: Age on February 1 of testing year, using date of birth estimated as 15th day of birth month in birth year , with ages more than two years from the mean weighted national age recoded to the mean
- LEP: student classified as an English language learner (yes or no)
- IEP: student classified as having a disability (yes or no)
- B018201: language other than English spoken at home (never, once in a while, about half of the time, or all or most of the time)
  - » This variable is coded as B003201 in the 1998 and 2000 assessments and as B003201A in the 1990, 1992, 1992, and 1996 assessments (never, sometimes, or always); we recoded B018201 responses of "once in a while" or "about half of the time" as "sometimes", and "all or most of the time" as "always" to make the variable consistent over time.

In addition, we conducted two robustness checks with additional variables. These variables were coded as follows:

- BA21101–BE21101: Student-reported Hispanic or Latino background, with options for selecting more than once choice (not Hispanic or Latino, Mexican/Mexican American/Chicano, Puerto Rican/Puerto Rican American, Cuban/Cuban American, and/or other Hispanic or Latino)

- About 7 percent of eighth-grade students and 5 percent of fourth-grade students who identified as Hispanic selected more than one background category in 2013. This variable is coded as B003101, allowing only one option, in the 1996, 1998, 2000, and 2002 assessments.

- PARED: Student-reported parent education (did not finish high school, high school graduate, some college, college graduate, or parent education unknown/missing)
  - This variable is not available for fourth-grade students.

Main NAEP Adjustment Methodology

All analyses were run separately by grade and subject and are weighted to be nationally representative (using weight variable ORIGWT).

We performed an adjustment across time by estimating regression coefficients using the student-level data from 2003, roughly halfway between the available time periods for mathematics (1996–2013) and reading (1998–2013). Specifically, we regress the test score of each student in 2003 on the set of control variables described above. Control variables are included in the regression using dummy variables identifying each of the groups of students for each construct, with the exception of the one arbitrarily chosen group that is the omitted category (except for age, which is included as a continuous variable).

Using this regression, we estimate a residual for each student across all assessment years, which is the difference between their actual score and a predicted score based on the relationship between the predictors and test scores in 2003. We calculate the adjusted change in the national average score between any two years as the change in the weighted average of the student-level residuals between those years.

Main NAEP Robustness Analysis

To test the robustness of these main NAEP estimates, we conducted four additional adjustments. Two of these adjustments were for changing the base year of the adjustment from 2003 to the first available assessment year with complete control variables (1996 for math, 1998 for reading) or the last available assessment year (2013). Two other adjustments were made using Hispanic and Latino background or parent education as additional demographic (nonpolicy) factors. The results of these adjustments are largely consistent with the adjustments we provide in the body of the report and are available in table A.2.

TABLE A.1

LTT NAEP (9 and 13 year old) and Main NAEP (4th and 8th grade) Demographics over Time

	Race/Ethnicity					Lunch Status		Non-English Language at Home		Learning Needs		Parent Education (8th Grade / 13 year old)			
Year	White	Black	Hispanic	Asian	American Indian	Free lunch	Reduced-price lunch	Sometimes, once in a while, or half the time	Always/all or most of time	Limited English proficiency	Special education	Did not graduate high school	High school graduate	Attended some college	College graduate
LTT NAEP demographics															
1971	84.3%	14.0%	1.7%												
1975	80.4%	13.0%	0.9%												
1978	79.8%	13.5%	5.6%	0.8%	0.1%										
1980	79.3%	13.7%	5.6%	0.8%	0.5%							10.2%	30.7%	16.5%	32.6%
1982	78.9%	14.0%	5.2%	1.5%	0.3%							10.7%	34.4%	14.1%	32.1%
1984	75.8%	15.0%	7.0%	1.7%	0.2%							8.6%	36.0%	9.4%	35.6%
1986	76.6%	14.6%	6.4%	1.5%	0.6%							7.8%	30.5%	15.3%	37.0%
1988	75.8%	15.5%	6.1%	1.8%	0.7%							7.9%	30.9%	10.0%	41.7%
Main NAEP demographics															
1992	70.2%	15.7%	9.8%	2.6%	1.4%			26.4%	6.7%	0.9%	3.0%	8.4%	24.1%	18.5%	40.8%
1994	69.1%	15.0%	11.1%	2.9%	1.4%			25.6%	7.2%	1.9%	5.5%	6.9%	21.4%	19.5%	42.8%
1996	68.5%	14.4%	12.6%	3.0%	1.4%	23.5%	5.8%	24.9%	7.5%	2.0%	5.9%	7.6%	23.6%	18.1%	41.7%
1998	67.3%	14.9%	13.3%	3.1%	1.3%	25.5%	5.6%	27.3%	8.6%	2.3%	5.3%	6.9%	21.9%	17.5%	44.3%
2000	66.3%	13.8%	14.7%	3.2%	1.8%	25.2%	5.7%	28.0%	8.3%	3.0%	5.8%	6.9%	21.8%	18.5%	41.7%
2002	63.0%	15.7%	15.1%	4.2%	1.2%	28.1%	7.1%	28.6%	12.9%	5.5%	8.0%	6.3%	16.8%	18.7%	45.9%
2003	61.3%	16.5%	15.9%	4.3%	1.2%	29.3%	7.3%	29.7%	14.8%	6.5%	9.8%	6.6%	18.4%	17.4%	46.2%
2005	59.8%	16.3%	17.3%	4.5%	1.2%	32.0%	6.9%	29.4%	15.6%	6.7%	9.6%	6.8%	17.0%	17.2%	47.3%
2007	58.4%	15.9%	18.5%	4.7%	1.2%	33.4%	6.0%	29.7%	16.5%	7.2%	9.2%	7.1%	16.8%	17.0%	47.5%
2009	56.9%	15.4%	20.1%	5.0%	1.1%	35.6%	6.2%	29.2%	17.2%	6.9%	9.8%	7.3%	16.4%	16.4%	48.3%
2011	54.4%	15.3%	21.8%	5.4%	1.1%	40.9%	5.4%	29.0%	18.2%	7.7%	10.2%	7.4%	15.9%	15.9%	49.1%
2013	53.3%	14.9%	23.1%	5.3%	1.1%	42.9%	5.1%	28.5%	18.3%	7.3%	11.3%	7.3%	15.5%	15.2%	49.9%

TABLE A.2

Unadjusted and Adjusted Changes in Main NAEP Scores, 1996/98–2013

		Student Standard Deviations (2003)		Student Standard Deviations (96/98)		Student Standard Deviations (2013)		Student Standard Deviations (2003, Hispanic Subgroups)		Student Standard Deviations (2003, Parent Education)	
	No controls	Nonpolicy controls	All controls	Nonpolicy controls	All controls	Nonpolicy controls	All controls	Nonpolicy controls	All controls	Nonpolicy controls	All controls
4th-grade math	0.63	0.71	0.82	0.79	0.97	0.70	0.82	0.73	0.83		
4th-grade reading	0.12	0.21	0.32	0.25	0.38	0.19	0.33	0.23	0.34		
8th-grade math	0.35	0.43	0.55	0.45	0.58	0.40	0.54	0.45	0.56	0.38	0.49
8th-grade reading	0.11	0.19	0.31	0.18	0.30	0.16	0.31	0.20	0.32	0.15	0.26
Average	0.30	0.39	0.50	0.41	0.56	0.36	0.50	0.40	0.51		

Notes: Changes are for 1996–2013 for math and 1998–2013 for reading. Changes are reported in 2003 standard deviations for all adjustments. The second and third columns provide an adjustment using 1996/1998 and 2013 as the base regression years, respectively. The fourth and fifth columns introduce the control variables of Hispanic subgroups and parent education, using 2003 as the base regression year.

TABLE A.3

LTT Reading NAEP Participation Rates over Time

	Weighted percent of schools participating before substitution			Weighted percent of students participating			Overall participation		
	Age 9	Age 13	Age 17	Age 9	Age 13	Age 17	Age 9	Age 13	Age 17
1971	92.5%	92.0%	90.5%	90.9%	84.2%	73.5%	84.1%	77.5%	66.5%
1975	93.9%	92.8%	91.0%	87.2%	85.2%	73.2%	81.9%	79.1%	66.6%
1980	94.5%	93.2%	90.5%	90.5%	85.5%	74.2%	85.5%	79.7%	67.2%
1984	88.6%	90.3%	83.9%	92.9%	89.2%	78.9%	82.3%	80.5%	66.2%
1988	87.2%	92.7%	78.1%	92.5%	90.2%	82.1%	80.7%	83.6%	64.1%
1990	87.0%	89.0%	79.0%	92.5%	90.2%	82.1%	80.5%	80.3%	64.9%
1992	87.0%	85.3%	80.9%	93.8%	90.8%	83.3%	81.6%	77.5%	67.4%
1994	86.7%	79.7%	80.1%	94.1%	91.8%	84.2%	81.6%	73.2%	67.4%
1996	83.5%	82.0%	81.7%	95.6%	92.2%	83.8%	79.9%	75.6%	68.5%
1999	84.9%	80.8%	74.0%	94.4%	92.1%	80.2%	80.2%	74.4%	59.4%
2004	85.3%	83.2%	73.4%	94.5%	92.4%	75.5%	80.7%	76.9%	55.4%
2008	95.9%	95.4%	90.3%	94.9%	93.8%	87.7%	92.2%	90.6%	81.1%
2012	82.0%	81.0%	80.0%	95.0%	93.0%	88.0%	-	-	-

FIGURE A.1

LTT NAEP Reading Score Changes, in 1971 Standard Deviations

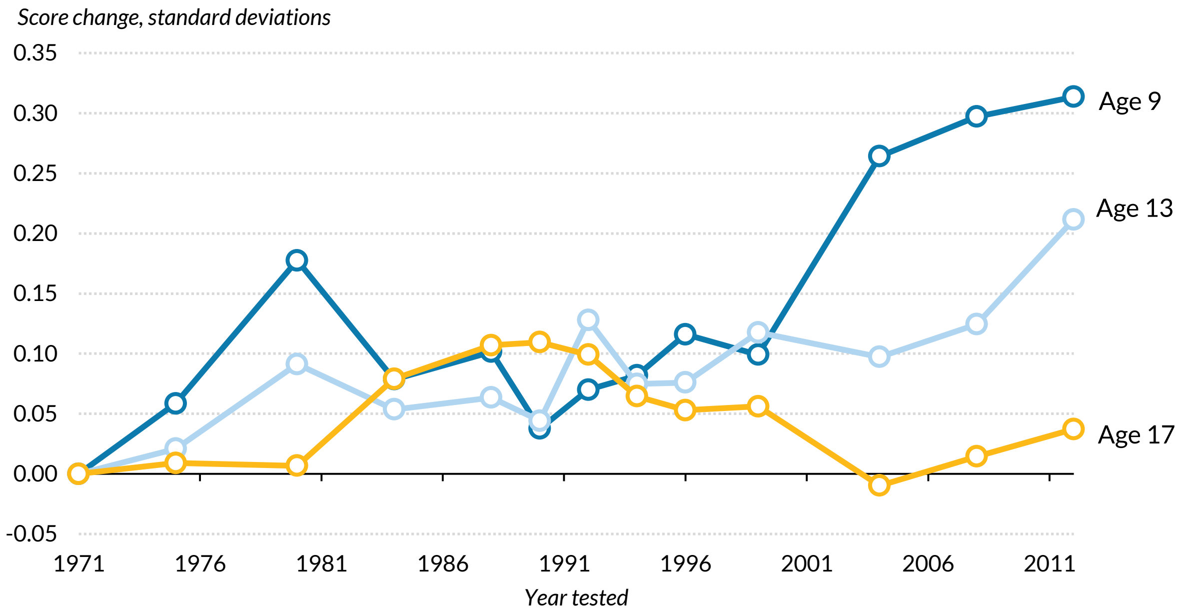
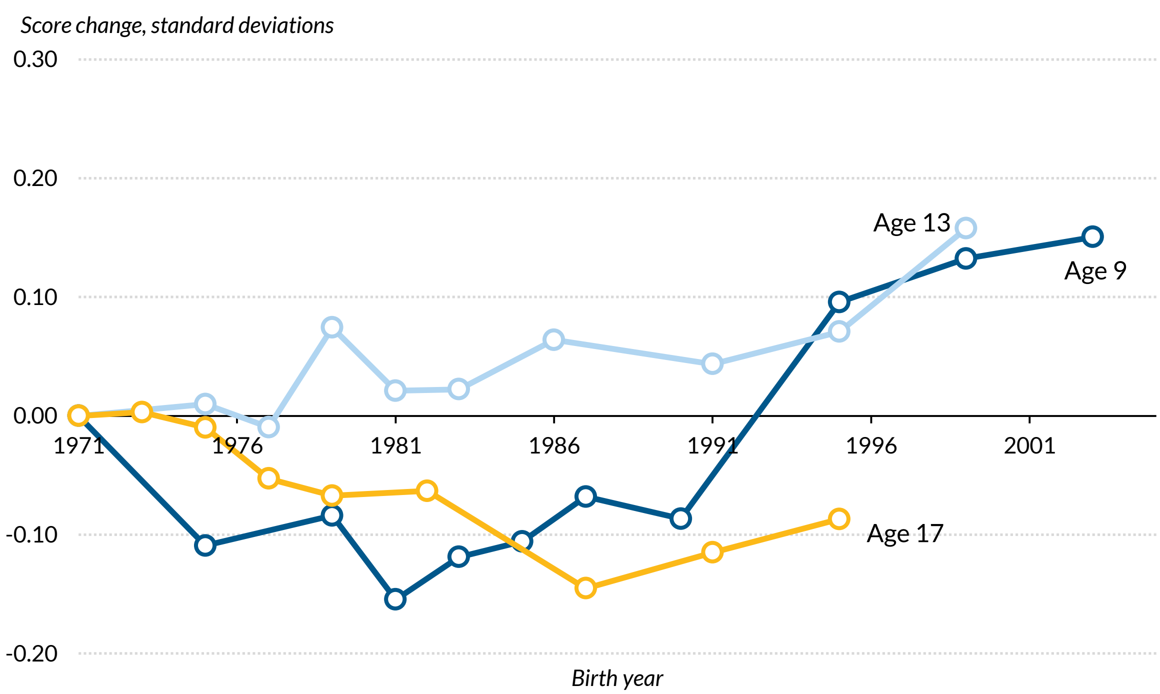


FIGURE A.2

LTT NAEP Reading Score Changes, by Cohort Birth Year



Notes: Scores are adjusted using the first year of cohort assessment (1980 standard deviation for 9-year-olds, 1984 for 13-year-olds, and 1988 for 17-years-olds). 2008 and 2012 scores are from the revised format test.

Notes

1. "President Signs Landmark No Child Left Behind Education Bill," White House (news release), January 8, 2002, http://georgewbush-whitehouse.archives.gov/news/releases/2002/01/20020108-1.html.
2. LTT assessments previously included writing and science assessments, which have been discontinued. See "How Were the NAEP Long-Term Trend Assessments Developed?" National Center for Education Statistics, last modified June 25, 2013, https://nces.ed.gov/nationsreportcard/ltt/howdevelop.aspx.
3. For additional discussion of these issues, see Chingos (2015) and a forthcoming data visualization tool.
4. We report the results of a series of additional specifics in table A.1 and discuss the variation in results in appendix A.
5. Table A.1 shows that we obtain similar results if we adjust the main NAEP scores using 1996, 1998, or 2013 data instead of the 2003 data.
6. We use 1996/1998 as the starting year for the analysis in table 3 because these were the first years that free and reduced price lunch eligibility was measured in the NAEP data. We also do not examine 2015 scores in our analysis because the student-level data needed to make the demographic adjustments has not yet been made available to researchers.
7. Tom Loveless, "Common Core's major political challenges for the remainder of 2016," *Brown Center Chalkboard* (blog), Brookings Institution, March 30, 2016, http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2016/03/30-common-core-political-challenges-loveless.
8. "Mathematics Framework Changes," National Center for Education Statistics, last modified October 23, 2013, https://nces.ed.gov/nationsreportcard/mathematics/frameworkcomparison.asp
9. Joy Resmovits, "American Students Perform Worse As They Reach Higher Grades," *Huffington Post*, May 7, 2014, http://www.huffingtonpost.com/2014/05/07/naep-2013-high-school_n_5276767.html.
10. The 2012 student-level LTT NAEP data will be made available to researchers in the winter of 2016–17.
11. Anya Kamenetz, "U.S. High School Graduation Rate Hits Record High," *NPR*, December 15, 2015, http://www.npr.org/sections/ed/2015/12/15/459821708/u-s-high-school-graduation-rate-hits-new-record-high.
12. These data are from the October supplement of the Current Population Survey. The percentage of enrolled 17-year-old students is calculated as the weighted proportion of 17-year-olds enrolled in a K–12 grade (thus eligible to be sampled for the LTT NAEP).
13. A new version of the NAEP test, given in 2004 and 2008, omitted "I don't know" as an option for multiple choice questions. This did depress the number of nonresponses we observed in the student level data.
14. "NAEP Technical Documentation: Participation, Exclusion, and Accommodation Rates for Age 17 LTT Mathematics in 2008," NAEP, last modified August 2, 2010, http://www.nationsreportcard.gov/ltt_2012/participation.aspx#0-1;"About the Assessment: Participation Rates," NAEP, accessed April 12, 2016, https://nces.ed.gov/nationsreportcard/tdw/weighting/2008/ltt_response_exclusion_and_accommodation_rates_for_age_17_in_2008.aspx.
15. "Content Comparisons and Alignment: Is the Content of NAEP Similar to Other Relevant Tests?" National Assessment Governing Board, accessed April 13, 2016, https://www.nagb.org/what-we-do/preparedness-research/types-of-research/content-alignment.html.

16. "Trends in International Mathematics and Science Study," 2011 results, figures 1 and 3, National Center for Education Statistics, accessed April 13, 2016, https://nces.ed.gov/timss/tables11.asp.
17. Jack Buckley, "NCES Statement on PIRLS 2011 and TIMSS 2011," National Center for Education Statistics, December 11, 2012, https://nces.ed.gov/whatsnew/commissioner/remarks2012/12_11_2012.asp.
18. "Program for International Student Assessment (PISA): Trends in Student Performance—Trends in U.S. Performance," National Center for Education Statistics, accessed April 13, 2016, http://nces.ed.gov/surveys/pisa/pisa2012/pisa2012highlights_6.asp.

References

Braun, Henry, Irwin Kirsch, and Kentaro Yamamoto. 2011. *An Experimental Study of the Effects of Monetary Incentives on Performance on the 12th-Grade NAEP Reading Assessment*. New York: Teachers College Record. http://www.tcrecord.org/Content.asp?ContentId=16008.

Child Trends Databank. 2015. *Children in Poverty*. Bethesda, MD: Child Trends. http://www.childtrends.org/?indicators=children-in-poverty.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79.

Chingos, Matthew. 2015. *Breaking the Curve: Promises and Pitfalls of Using NAEP Data to Assess the State Role in Student Achievement.* Washington, DC: Urban Institute: http://www.urban.org/research/publication/breaking-curve-promises-and-pitfalls-using-naep-data-assess-state-role-student-achievement/view/full_report.

Cowan, Charles D., Robert M. Hauser, Robert A. Kominski, Henry M. Levin, Samuel R. Lucas, Stephen L. Morgan. Margaret Beale Spencer, and Chris Chapman. 2012. *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation.* Washington, DC: National Center for Education Statistics. http://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf.

Deming, David J., Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. forthcoming. "School Accountability, Postsecondary Attainment and Earnings." *Review of Economics and Statistics*. http://scholar.harvard.edu/files/ddeming/files/deming_txaccountability_dec2015.pdf?m=1453408176.

Dynarski, Susan, Joshua Hyman, and Diane Whitmore Schanzenbach. 2013. "Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion." *Journal of Policy Analysis and Management* 32 (4): 692–717.

Grissmer, David. 2007. *Estimating Effects of Non-Participation on State NAEP Scores Using Empirical Methods*. Washington, DC: American Institutes for Research and National Center for Education Statistics. http://www.air.org/sites/default/files/downloads/report/AIR_NVS_Grissmer_revised_0.pdf.

Hanushek, Eric A., Dean T. Jamison, Eliot A. Jamison, and Ludger Woessmann. 2008. "Education and Economic Growth: It's Not Just Going to School, but Learning." *Education Next* 8 (2): 62–70.

Hill, Carolyn J., Howard S. Bloom, Alison R. Black, and Mark W. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives* 2 (3): 172–7.

Moore, Quinn, Kevin Conway, Brandon Kyler, and Andrew Gothro. 2013. "Direct Certification in the National School Lunch Program: State Implementation Progress, School Year 2012–2013—Summary." Washington, DC: US Department of Agriculture. http://www.fns.usda.gov/sites/default/files/NSLPDirectCertification2013_Summary.pdf.

National Center for Education Statistics. 2010. *The Nation's Report Card: Grade 12 Reading and Mathematics 2009 National and Pilot State Results*. NCES 2011-455. Washington, DC: Institute of Education Sciences, US Department of Education.

Neuberger, Zoë, Becca Segal, Caitlin Nchako, and Kathleen Masterson. 2015. *Take Up of Community Eligibility This School Year*. Washington, DC: Center on Budget and Policy Priorities. http://www.cbpp.org/research/take-up-of-community-eligibility-this-school-year.

Perie, Marianne, Rebecca Moran, Anthony D. Lutkus, and William Tirre. 2005. *NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics*. Washington, DC: National Center for Education Statistics. http://nces.ed.gov/nationsreportcard/pdf/main2005/2005464.pdf.

Sparks, Dinah, and Nat Malkus. 2013. "First-Year Undergraduate Remedial Coursetaking: 1999–2000, 2003–04, 2007–08." Washington, DC: National Center for Education Statistics. http://nces.ed.gov/pubs2013/2013013.pdf.

About the Authors

Kristin Blagg is a research associate in the Income and Benefits Policy Center at the Urban Institute, focusing on education policy. Before joining Urban, she spent four years as a math teacher in New Orleans and New York City. She has also been a human capital consultant for federal government clients.

Matthew M. Chingos is a senior fellow at the Urban Institute, where he studies education-related topics at both the K-12 and postsecondary levels. Chingos's areas of expertise include class-size reduction, standardized testing, teacher quality, student loan debt, and college graduation rates.

Statement of Independence

URBAN INSTITUTE · ELEVATE · THE · DEBATE

2100 M Street NW
Washington, DC 20037

www.urban.org