



Making the Grade in America's Cities

Assessing Student Achievement in Urban Districts

Kristin Blagg

June 2016

Many US education reform efforts have focused on the performance of students in large, urban school districts. Compared with their suburban and rural counterparts, urban school districts enroll larger proportions of students of color, and more of their students are eligible for free and reduced-price lunch (Sable, Plotts, and Mitchell 2010). Moreover, the achievement gap is larger within large city districts than for public school districts nationally. For example, on the National Assessment of Educational Progress (NAEP) in 2015, the average gap between black and white student scores was 20 percent larger in large city districts, and the gap between Hispanic students and white students was nearly 25 percent larger.

Recognizing the importance of understanding student achievement in large cities, the National Center for Education Statistics (NCES) has conducted biennial assessments of fourth- and eighth-grade reading and mathematics, known as the Trial Urban District Assessment (TUDA) program, since 2002. The TUDA, which assessed 21 school districts in 2015, is an extension of the NAEP assessment program the "Nation's Report Card," which also provides national and state data on student achievement.

The TUDA program has given researchers and policymakers a window into urban school district performance, providing the opportunity to track and compare student achievement in the public schools of cities such as New York City, Atlanta, Houston, Chicago, and Los Angeles. However, comparing scores across the TUDA districts is complicated by the existence of differences in student demographics among these cities. For example, in cities such as Boston, Dallas, and Cleveland, more than 90 percent of 2015 NAEP test takers were eligible for free and reduced-price lunch, whereas schools in Austin and Hillsborough County, Florida, had a free and reduced-price lunch percentage comparable with the national mean. And roughly 40 to 50 percent of fourth graders in Houston, Dallas, and San Diego were classified as English language learners in 2015, while less than 5 percent of fourth-grade students in Baltimore and Atlanta had English language learner status.

In this report, I examine the NAEP score trends of students in large cities overall, and provide new analysis of TUDA results by generating scores that are adjusted for a rich set of demographic controls. I examine the relative performance of the full set of 2013 TUDA participants (the most recent test administration for which student-level data are available to researchers), as well as performance changes among the subset of districts that participated in both 2005 and 2013.

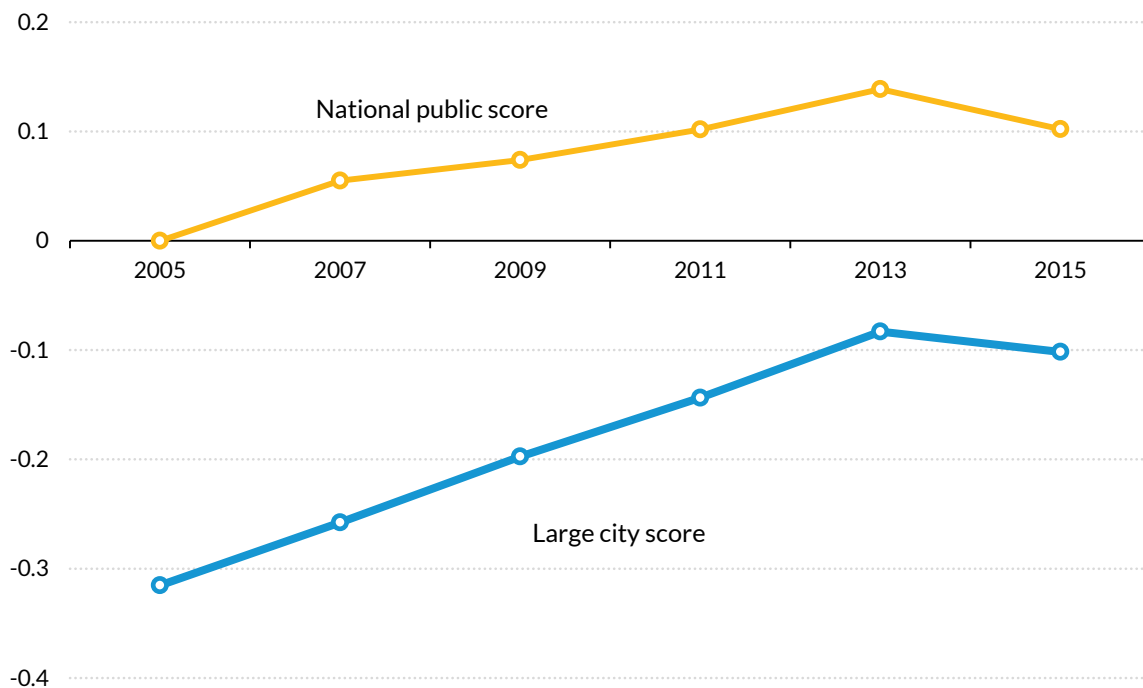
Rising Achievement in Large Cities

NAEP results for the nation's public school students have generally trended upward over the past several years, with a gain of roughly one-tenth of standard deviation from 2005 to 2015, averaged across the fourth- and eighth-grade tests.¹ However, public school students who live in large cities (defined as a large central city with a population of 250,000 or more) have doubled that score gain, posting an average improvement of 0.21 standard deviations over the same period. In fact, although public school students in large cities tend to score lower than public school students overall, they have closed about a third of the gap with national scores over the past decade (figure 1).

FIGURE 1

Main NAEP Score Change, in 2005 Standard Deviations

Averaged across 4th and 8th grade mathematics and reading tests



Source: Urban Institute analysis of NAEP data.

Note: NAEP = National Assessment of Educational Progress.

Though the NCES stresses that the large city designation is not equivalent to what might be thought of as “inner city,” public school students in large cities are more likely to be low-income, from minority racial or ethnic groups, or to be English language learners.² In 2015, 72 percent of large city students were classified as eligible for free and reduced-price lunch (compared with 53 percent of public school students overall), and 81 percent were students of color (compared with 50 percent overall). The rapid growth in NAEP achievement for students who live in large cities, inclusive of TUDA districts, motivates the work to understand and compare performance among these large cities.

Measuring District Performance

Like NAEP state scores, TUDA scores have been used to tout the efficacy of charter schools in Los Angeles, to assess the effect of mayoral control in New York City, and to bolster support for increased school funding in Cleveland.³ Assertions such as these overlook the fact that the student populations in these school districts vary substantially across cities and over time. The Los Angeles Unified School District serves a different population than Detroit Public Schools, and student demographics in cities such as Atlanta and Washington, DC, have shifted substantially over the past decade.

To compare the performance of different urban districts, we must first account for the student demographic differences between districts. I do this by adjusting the scores using demographic variables from the restricted-use student-level NAEP data from 2013 (the most recent available year). I adjust using the variables of gender, race and ethnicity, eligibility for free and reduced-price lunch, limited English proficiency, special education status, age, whether the student was given a testing accommodation, the amount of English spoken at the student’s home, and the student’s family structure (e.g., two-parent, single-parent, and foster).⁴ I conduct this statistical adjustment using only the students who were sampled from large cities, which is inclusive of the TUDA district samples. This adjustment includes roughly 160,000 observations across the four tests, or approximately 25 percent of students tested in 2013 (largely because of the oversampling necessary to generate TUDA scores, the large cities sample accounts for 15 to 17 percent of the weighted national score).

Figure 2 shows the unadjusted scores of all 21 TUDA districts, and the city of Washington, DC (which includes both district and charter schools), averaged across the four main NAEP tests, as well as the scores adjusted for the student characteristics listed above. (The source data for figure 2 appear later in the brief in table 1.) Even when accounting for a rich set of demographic differences among students in large cities, performance still varies substantially across TUDA districts, with a difference of 28 scale score points (slightly less than one standard deviation) separating the highest adjusted-score district (Boston) from the lowest (Detroit). This demographic adjustment did produce a 20 percent reduction in the overall range in unadjusted scores among TUDA districts, which is an average of 35 scale score points across the four tests (the difference between Charlotte–Mecklenburg and Detroit).

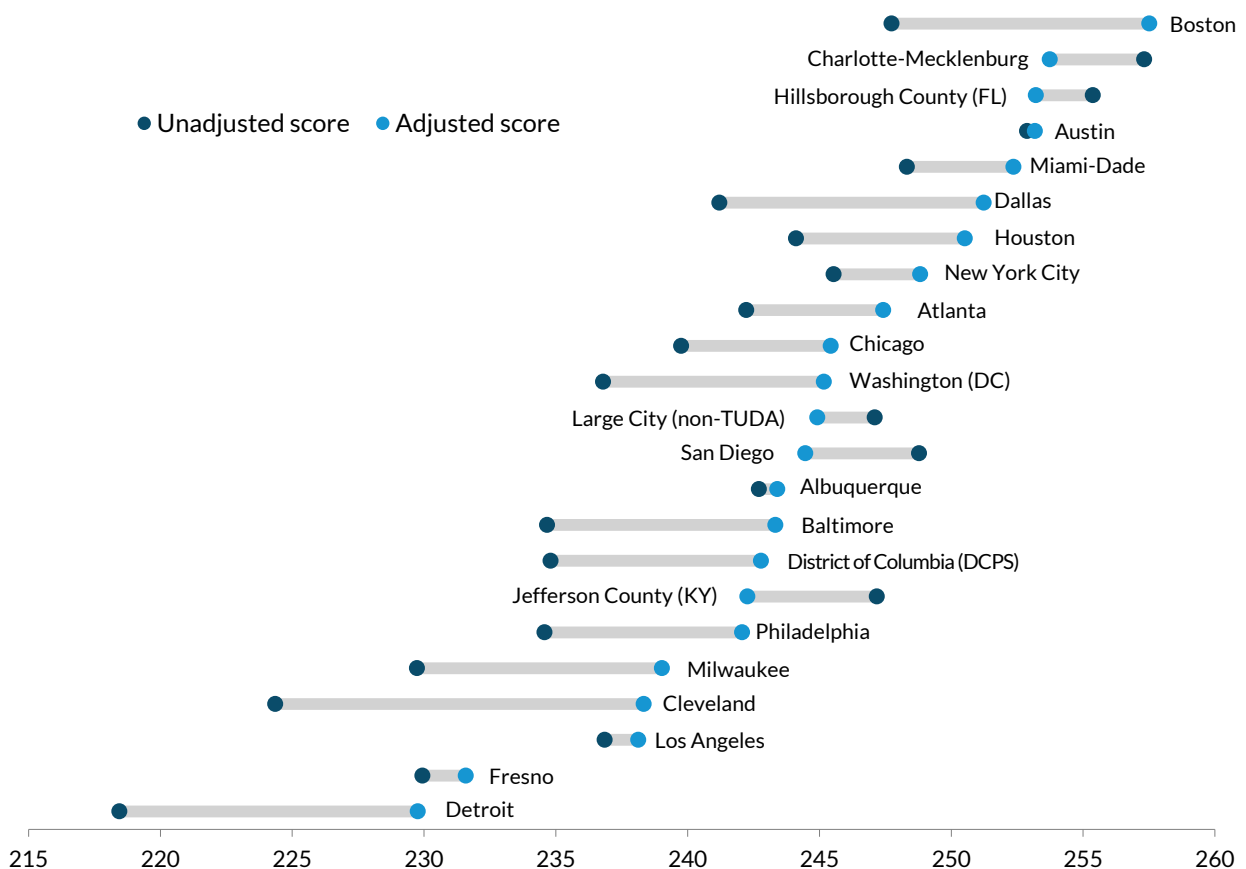
Cities are invited to participate in the TUDA based on selection criteria including district size, percentage of African American or Hispanic students, and percentage of students eligible for free and reduced-price lunch. Because of this selection process, TUDA districts are more likely to serve groups of

students whose demographics pull down their district’s raw scores relative to the average large city, so these cities’ scores tend to be adjusted upward. In a similar way, the scores of students who are part of the large city sample, but not part of a TUDA sample, tend to be adjusted downward.

In a previous Urban Institute report that reported adjusted state NAEP scores, Texas and Florida, which had average overall performance, jumped from the middle of the pack to become the third- and fourth-ranked states when accounting for state demographics (Chingos 2015). In a similar way, some districts “break the curve” and achieve a higher rank amongst TUDA districts when accounting for demographics. Dallas jumps from 13th in average NAEP scores to 6th, and the cities of Washington, DC (inclusive of charters), and Boston both move up five spots, from 16th to 11th and from 6th to 1st, respectively. Also noteworthy, the highest performing districts in this analysis tend to be located in states with high adjusted NAEP scores, such as Massachusetts, Texas, and Florida.

FIGURE 2
TUDA District Performance on the 2013 NAEP, Adjusted for Demographics

Scale score averaged across 4th and 8th grade mathematics and reading tests



Source: Urban Institute analysis of NAEP data

Note: DCPS = District of Columbia Public Schools; NAEP = National Assessment of Educational Progress; TUDA = Trial Urban District Assessment.

Have Districts Improved Performance over Time?

These data show that accounting for the student demographics of TUDA districts can substantially change our assessment of the relative performance of urban school districts. As I've shown, students in large cities overall have posted larger gains in their average performance on NAEP relative to public school students nationally. How much of this change was driven by observable demographic changes, such as an influx of more advantaged students, or a decline in students who speak a language other than English at home? Is the increase in large city performance on NAEP spread equally across all TUDA districts, or have some districts grown more than others? To answer these questions, I perform a demographic adjustment on TUDA cities over time.

Despite the fact that some US cities—such as Atlanta, Minneapolis, Denver, and Washington, DC—have experienced substantial demographic change and gentrification, large cities (and TUDA districts) overall have not seen the same dramatic shifts in resident income and demographics).⁵ Consequently, there are relatively small shifts in the demographics of the student population in most TUDA districts over time. However, given the large differences in the performance of different subgroups of students on the NAEP, even these small demographic shifts could have a measurable effect on overall TUDA scores.

To investigate the gains that large city districts have made, I examine 11 TUDA districts (and public schools in the District of Columbia) that were assessed in both 2005 and 2013. Using a more limited set of control variables that are available for both years (race, age, gender, and amount of English spoken at home), I calculate the increase in scores that might have been expected given changes in student demographics over the eight-year period.⁶ For each TUDA district, I measure the relationship between scores and student-level demographics in 2005, then apply that relationship to the students that were assessed in 2013, essentially predicting their scores based on demographics.⁷

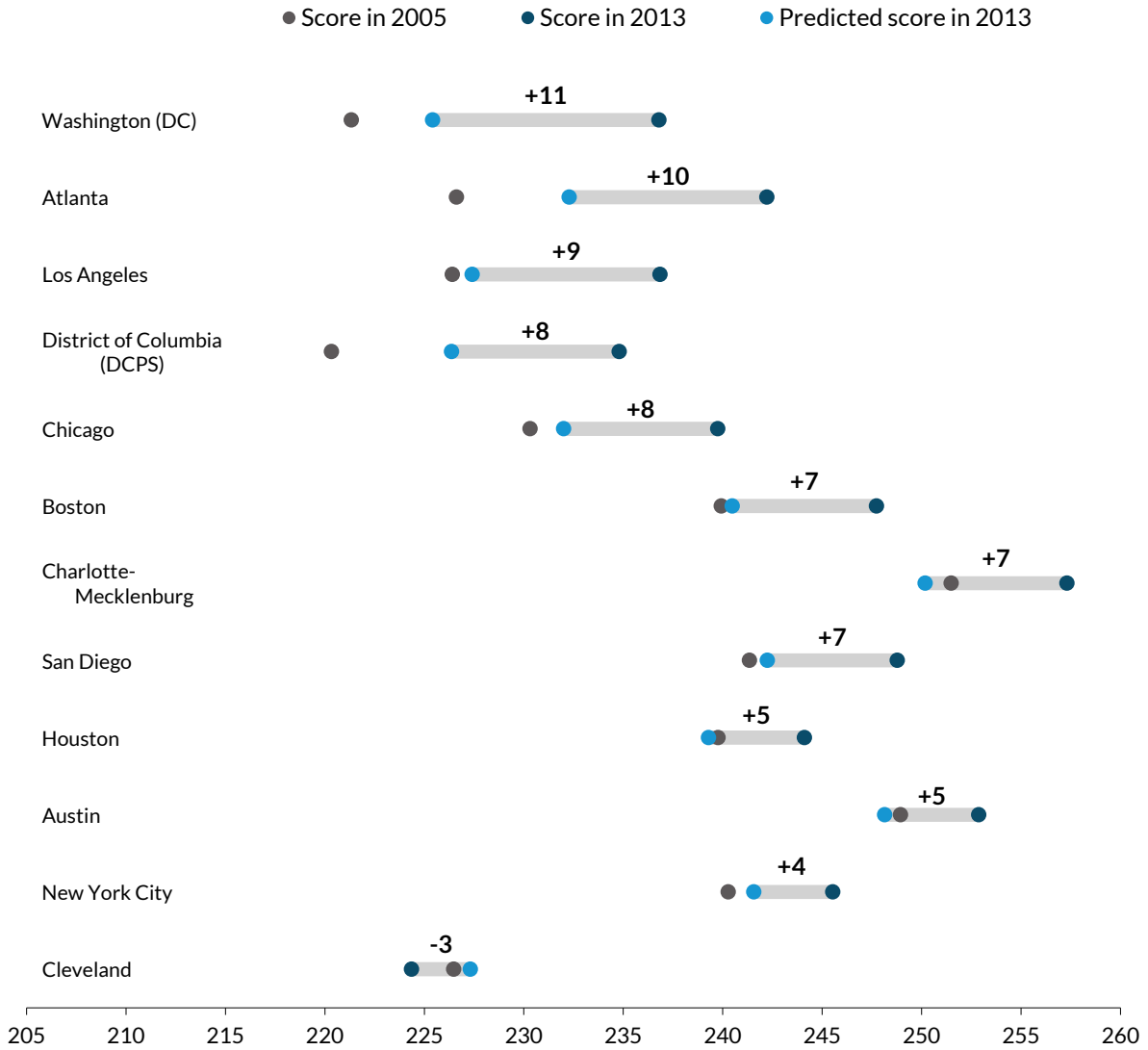
Figure 3 shows the results of this analysis ranked by the size of the achievement gain above the demographic prediction. (The source data for the figure appear later in the brief, in table 2.) The score in 2005 indicates the district's scale score averaged across the four NAEP tests. The other two dots in the graph indicate the predicted 2013 score based on demographic changes and the actual 2013 score. The difference between these two dots is the district's growth above demographic predictions.

Three TUDA districts—Houston, Austin, and Charlotte-Mecklenburg—were predicted to have slight declines in performance relative to 2005, yet nearly all districts (except Cleveland) posted gains over their 2005 performance. Washington, DC, inclusive of the charter sector, posted the highest adjusted growth: an average 11-point scale score gain on top of a predicted gain of 4 points. Atlanta and DC Public Schools were also predicted to have substantial growth based on student demographic changes, and posted gains of roughly 10 and 8 points, respectively, above those predictions. Los Angeles and Chicago had smaller predicted changes (less than 2 points), but still produced sizeable gains (9 and 8 points) over the eight-year period.

FIGURE 3

District Performance on the 2013 NAEP, Predicted versus Actual Score

Scale score averaged across 4th and 8th grade mathematics and reading tests



Source: Urban Institute analysis of NAEP data.

Note: DCPS = District of Columbia Public Schools; NAEP = National Assessment of Educational Progress.

It conceivable that TUDA districts that have lower starting scores may also have larger score changes above demographic predictions. However, similar to the analysis of NAEP score growth in states, there is not an appreciable correlation between the growth of scores above demographic projections and average scale score (Chingos 2015).⁸

Conclusion

Substantial variation in school district performance and growth persists even after accounting for demographic differences in student populations, both across school districts and across time. My examination of TUDA districts over time indicates that a substantial amount of NAEP achievement from 2005 to 2013 is unexplained by demographic changes, particularly in Atlanta, Los Angeles, and Washington, DC. Moreover, I find that districts such as Boston, Charlotte-Mecklenburg, and Hillsborough County post comparatively high TUDA scores when adjusting for student demographics. These results seem to support the finding that school districts could have an effect on student achievement as measured on the NAEP. However, this conclusion comes with several caveats.

Although I have adjusted for a broad range of observable student characteristics, I cannot rule out the possibility that there are unobservable differences in student populations that affect achievement on the NAEP. For example, due to limitations of NAEP data, I cannot directly control for students' household income, family attitudes toward academic achievement, or student mobility between schools or districts. In addition, I cannot control for broader city-level factors, such as pollution, crime rates, or other environmental factors that could have an effect on academic performance.

It is also important to emphasize that, to the extent that district policy changes are responsible for NAEP score changes, this type of analysis cannot identify which policy changes were most important. Districts often simultaneously adopt several education-related policies, such as changes in the availability of early childhood education, human capital policies, and local funding formulas. State policy, such as accountability systems, may also contribute to the observed variation between districts in different states.

The TUDA program is valuable for researchers because it provides the opportunity to understand and evaluate student achievement within and between large city school districts. Making causal claims using TUDA data is almost always unwarranted, but my analysis suggests that urban- and district-level policy potentially have an important role to play in student achievement outcomes. This analysis highlights districts that warrant further study to understand why their students perform better or worse than their demographic peers in cities around the country.

Data Notes and Tables

This report draws on restricted-use, student-level NAEP data on the 2005 and 2013 administrations of fourth- and eighth-grade reading and math tests. These tests are given every two years to a nationally representative sample of US students

For the analysis of the 2013 data, I use a rich set of student-level control variables that are drawn from administrative records and a student survey. The variables used and the coding of them is as follows:

- SEX: gender (male or female)

- DRACEM: race and ethnicity as reported by the student (white, black, Hispanic, Asian American or Pacific Islander, American Indian or Alaska Native, or multiple)
- SLUNCH: eligibility for the federal free and reduced-price lunch program (not eligible, eligible for reduced-price lunch, eligible for free lunch, or other or missing)
- LEP: student classified as an English language learner (yes or no)
- IEP: student classified as having a disability (yes or no)
- Age on February 1 of testing year, using date of birth estimated as 15th day of birth month [BMONTH] in birth year [BYEAR], with ages more than two years from the mean weighted national age recorded to the mean
- ACCOMCD: whether the student received an accommodation (no accommodation, accommodation in regular testing session, or accommodation in separate testing session)
- B018201: how often a language other than English is spoken at home (never, once in a while, about half of the time, or all or most of the time)
- B0268A1 through B0268F1: family structure, measured as which parents child lives with (mother and father, mother only, father only, mother and other parent or guardian, father and other parent or guardian, foster parent or parents, or other or missing).

For the analysis of the change in average scale scores between 2005 and 2013, I used a more limited set of variables (race, age, gender, and amount of English spoken at home), coded in the same way as above.

TABLE 1

Unadjusted and Adjusted 2013 NAEP Scores, by Grade and Subject

TUDA district	2013 RAW SCORES				2013 ADJUSTED SCORES			
	4th Grade		8th Grade		4th Grade		8th Grade	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading
Albuquerque	234.5	206.6	273.8	256.0	235.2	207.5	274.7	256.2
Atlanta Public Schools	233.1	214.3	266.8	254.7	237.2	215.3	276.6	260.6
Austin Independent School District	245.0	220.8	284.6	261.2	245.3	222.0	284.8	260.6
Baltimore City Public Schools	222.9	204.3	259.8	251.8	234.1	208.0	274.3	256.9
Boston School District	236.9	214.4	283.1	256.5	244.5	225.3	293.9	266.3
Charlotte-Mecklenburg Schools	247.4	226.4	289.0	266.4	244.1	221.1	286.8	263.0
Chicago Public Schools	230.5	206.2	268.9	253.5	235.2	211.0	276.4	259.1
Cleveland Metropolitan School District	216.3	189.7	252.7	238.8	228.1	201.7	270.7	253.0
District of Columbia (DCPS)	228.6	205.7	260.3	244.6	235.4	210.6	271.8	253.3
Dallas	234.2	204.7	274.6	251.3	245.0	216.9	284.1	258.9
Detroit Public Schools	204.3	189.9	240.1	239.4	215.0	200.1	254.6	249.4
Fresno Unified School District	219.7	195.9	259.7	244.6	222.0	198.4	260.8	245.1
Hillsborough County (FL)	242.8	227.8	283.7	267.1	240.6	225.5	282.2	264.5
Houston Independent School District	235.9	207.9	280.5	252.2	242.5	215.6	286.9	257.0
Jefferson County Public Schools (KY)	233.7	220.9	273.5	260.6	230.8	213.5	269.8	255.0
Large City (non-TUDA)	236.3	212.9	279.0	260.2	235.9	211.0	276.1	256.7
Los Angeles Unified School District	228.5	204.8	264.3	249.8	230.1	207.5	265.2	249.7
Miami-Dade County Public Schools	237.4	223.1	273.8	259.0	241.2	227.9	278.2	262.1
Milwaukee Public Schools	221.4	198.7	257.3	241.5	228.3	206.3	270.0	251.5
New York City Public Schools	235.8	216.3	273.6	256.4	237.4	219.3	277.8	260.8
School District of Philadelphia	223.4	199.9	266.5	248.5	229.8	206.1	276.4	256.0
San Diego Unified School District	240.9	217.8	276.9	259.6	237.6	215.1	270.5	254.7
Washington (DC)	228.6	205.6	265.3	247.7	236.1	211.7	277.2	255.7

Source: Urban Institute analysis of NAEP data.

Note: DCPS = District of Columbia Public Schools; NAEP = National Assessment of Educational Progress; TUDA = Trial Urban District Assessment.

TABLE 2

2013 Predicted NAEP Scores and 2005 and 2013 Actual Scores

	2005 SCALE SCORES				2013 PREDICTED SCORES				2013 SCALE SCORES			
	4th Grade		8th Grade		4th Grade		8th Grade		4th Grade		8th Grade	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Math	Reading
Atlanta Public Schools	220.8	201.0	245.1	239.6	226.0	207.9	251.0	244.3	233.1	214.3	266.8	254.7
Austin Independent School District	242.0	216.6	280.8	256.5	242.7	216.3	279.5	254.1	245.0	220.8	284.6	261.2
Boston School District	229.4	207.3	269.9	253.2	230.5	208.8	270.1	252.6	236.9	214.4	283.1	256.5
Charlotte-Mecklenburg Schools	244.5	221.5	280.6	259.5	242.4	220.7	279.7	258.0	247.4	226.4	289.0	266.4
Chicago Public Schools	215.5	198.4	258.1	249.2	217.8	201.6	258.8	249.8	230.5	206.2	268.9	253.5
Cleveland Metropolitan School District	220.0	197.1	249.1	239.7	220.1	197.6	249.9	241.6	216.3	189.7	252.7	238.8
District of Columbia (DCPS)	210.2	190.5	243.8	236.9	216.0	197.1	249.7	242.7	228.6	205.7	260.3	244.6
Houston Independent School District	233.2	210.6	267.3	248.1	234.2	208.9	266.8	247.3	235.9	207.9	280.5	252.2
Los Angeles Unified School District	220.2	195.5	250.4	239.5	221.0	196.9	251.9	239.8	228.5	204.8	264.3	249.8
New York City Public Schools	230.6	212.9	266.5	251.1	233.1	214.8	267.4	251.0	235.8	216.3	273.6	256.4
San Diego Unified School District	232.4	207.6	270.3	253.4	233.9	208.8	271.2	255.0	240.9	217.8	276.9	259.6
Washington (DC)	211.1	190.8	245.2	238.2	215.0	194.8	249.5	242.4	228.6	205.6	265.3	247.7

Source: Urban Institute analysis of NAEP data.

Note: DCPS = District of Columbia Public Schools; NAEP = National Assessment of Educational Progress.

Notes

1. Although the large city subgroup designation was available in starting in 2003, I have opted to start the graph from 2005 to ensure comparability with the subsequent sections of the report, which use the 2005 TUDA results.
2. "Nation's Report Card Frequently Asked Questions," National Center for Educational Statistics, accessed June 15, 2016, <http://www.nationsreportcard.gov/faq.aspx>.
3. Kevin Drum, "Test Scores in New York City Are Nothing to Write Home About," *Mother Jones*, August 5, 2013, <http://www.motherjones.com/kevin-drum/2013/08/test-scores-new-york>; Patrick O'Donnel, "Small Gains for the Cleveland Schools Stand Out, as NAEP Scores Fall for Ohio and the Nation," *The Plain Dealer* (blog),

October 28, 2015,

http://www.cleveland.com/metro/index.ssf/2015/10/small_gains_for_the_cleveland_schools_stand_out_as_naep_scores_fall_for_ohio_and_the_nation.html; Susan Aud Pendergrass, “Los Angeles Charter Schools Outperform District-Run Schools in 2015 NAEP TUDA Results,” National Alliance for Public Charter Schools blog, March 10, 2016, <http://blog.publiccharters.org/los-angeles-charter-schools-outperform-district-run-schools-in-2015-naep-tuda-results>.

4. These control variables are the same as the variables used to calculate adjusted state scores in an earlier Urban Institute report (Chingos 2015). However, that adjustment also included several home-based variables (internet access, number of books in home, having one’s own room, having a dishwasher or clothes dryer in the home). I opted to exclude those variables because they are likely too sensitive to individual city infrastructure to be useful for cross-district comparisons (for example, having a clothes dryer in New York City may indicate a different socioeconomic status than having a clothes dryer in Jefferson County, Kentucky).
5. Nathaniel Baum-Snow and Daniel Hartley, “Demographic Changes in and Near US Downtowns,” *Economic Trends* (Federal Reserve Bank of Cleveland), June 5, 2015, <https://www.clevelandfed.org/newsroom-and-events/publications/economic-trends/2015-economic-trends/et-20150605-demographic-changes-in-and-near-us-downtowns.aspx>; Mike Maciag, “Gentrification in America Report,” *Governing*, February 2015, <http://www.governing.com/gov-data/census/gentrification-in-cities-governing-report.html>.
6. Although information on students’ special education status, limited English proficiency status, and free and reduced-price lunch status was available in both years, I have opted not to include those controls because their measurement over time is subject to district- and state-level policy changes (for example, school lunch eligibility was expanded during this period because of direct certification and community eligibility provisions). As a check on the results, I included a control for parents’ education level, available for eighth-grade students, as a proxy for socioeconomic status. This alternate specification did not appreciably change the results; on average, the scores changed by less than two-tenths of a scale score point, and the adjusted scores were highly correlated ($r > 0.99$).
7. Two of the 11 districts—San Diego and District of Columbia (DCPS)—assessed charter school students in 2005 but not in 2013, because of an NCES change enacted in 2009. To maintain comparability with the rest of the districts in figure 2, I have used the only the traditional public school student sample (i.e., excluding charter school students) in our 2005 predictions. This does not qualitatively alter the findings.
8. Because I have a sample size of just 11 TUDA districts (and the District of Columbia) that were assessed in both 2005 and 2013, I cannot rule out the possibility that additional years of TUDA data may illuminate a relationship between magnitude of scores and score change above demographic predictions. However, I also do not observe a consistent correlation between growth above prediction and average scale score when examining the four individual NAEP tests (fourth and eighth grade reading and math) for these districts (correlations between change above demographic prediction and the district’s 2005 scale scores range from -0.57 to 0.19 across the four tests).

References

- 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.
- Sable, Jennifer, Chris Plotts, and Lindsey Mitchell. 2010. *Characteristics of the 100 Largest Public Elementary and Secondary School Districts in the United States: 2008–09*. Report 2011-301. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

About the Author

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. Blagg holds a BA in government from Harvard University, an MEd from Hunter College, and an MPP from Georgetown University.

Acknowledgments

This brief was funded by the Urban Institute. The views expressed are those of the author 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 author thanks Matthew Chingos for his guidance and helpful comments in producing this brief. The author also thanks Erik Rodriguez for his helpful feedback.



2100 M Street NW
Washington, DC 20037
www.urban.org

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.

Copyright © June 2016. Urban Institute. Permission is granted for reproduction of this file, with attribution to the Urban Institute.