Measuring and Assessing Student Achievement in Urban School Districts

Kristin Blagg and Grace Luetmer

January 2020

Urban school districts—where gaps in academic performance by income and race or ethnicity are often largest—are frequently the epicenter of K–12 education reform efforts and debate. Since 2002, the National Center for Education Statistics (NCES) has tracked the performance of select large urban school districts using the Trial Urban District Assessment (TUDA), a subset of the National Assessment of Educational Performance (NAEP). The TUDA provides a district-level estimate of student performance in fourth- and eighth-grade reading and math. Although TUDA scores provide valuable insight into student achievement within these districts, the interpretation of these scores is complicated by differences between student populations, both across districts and over time. We look at 2017 district TUDA scores and attempt to level the playing field across districts by controlling for student characteristics. We build on work from two earlier Urban Institute publications: "Making the Grade in America's Cities," which looks at 2013 TUDA scores (Blagg 2016), and *Breaking the Curve*, which looks at 2013 NAEP state scores (Chingos 2015).

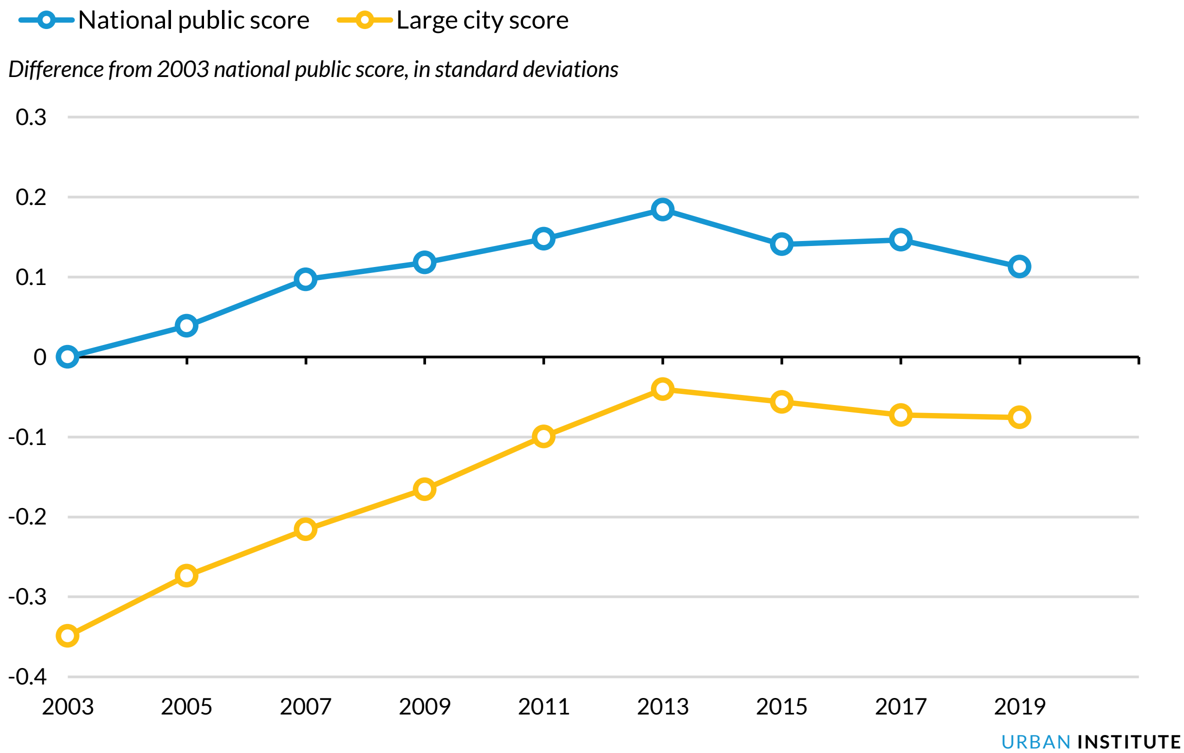
NAEP Scores Have Increased in Large Cities

The NCES defines schools in large cities as those within an urbanized area and a principal city with 250,000 or more residents.[1] This designation is not equivalent to an "inner city" school, but students in large cities are more likely to be students of color and to be from a low-income background. Sixty-eight percent of students in large cities receive free lunch (versus 52 percent in public schools overall), 16 percent of students are English language learners (versus 9 percent), and 80 percent are students of color (versus 51 percent).

FIGURE 1

Main NAEP Score Change, in 2003 Standard Deviations

Averaged across fourth- and eighth-grade mathematics and reading tests



Source: Urban Institute analysis of NAEP data from the Education Data Explorer.
Notes: NAEP = National Assessment of Educational Progress. Beginning in 2017, the assessment was given on digital tablets. In prior years, it was administered on paper.

Wide gaps in student achievement persist in public schools in large cities—the gap between white students and black or Hispanic students is about 20 percent wider than the national average—but student achievement in large cities has generally trended upward. In 2019, the average NAEP score in large cities (across all four tests) was about 0.27 standard deviations higher than in 2003. In contrast, the average NAEP score in public schools nationwide rose only half as much, about 0.11 standard deviations (figure 1).

This long-run increase in student achievement in large cities, even though it has stalled, motivates our examination of academic performance in TUDA districts and of performance in large cities overall. We attempt to tease out how much of this progress has been driven by changes in student demographics (e.g., the share of students from different racial or ethnic backgrounds or students from low-income backgrounds) and how much might be the result of other educational factors (e.g., changes in national, state, or district policy) or other broader trends (e.g., reductions in pollution and improvements in public safety).

What Factors Should We Consider When Comparing Districts?

TUDA districts are invited to participate, in part, because they have high shares of black and Hispanic students and high shares of students eligible for free and reduced-price lunch, but these districts still vary in terms of student needs. Houston has a higher share of English language learners than Atlanta, and Hillsborough County, Florida, in which Tampa is the county seat, reports a smaller share of students receiving free and reduced-price lunch than Cleveland. Comparing raw TUDA scores across these school districts would not account for these student differences.

To level the playing field across TUDA districts, we employ a model that adjusts for demographic differences using restricted-use student-level NAEP data. This model assesses how well students in each district do relative to their demographically similar peers across the country. States are not required to administer the student surveys that provide some of the information needed to perform these adjustments. Although most TUDA districts are located in states that choose to complete the student survey, Colorado opted out of gathering the additional data from students in the 2017 administration of the NAEP. We therefore cannot provide demographically adjusted scores for the Denver TUDA district. For all other TUDA districts, we include factors central to the student's identity and potentially to her performance on NAEP: gender, race or ethnicity, how often English is spoken at home, family structure (e.g., two parents, single parent, or foster care), age at test time, and parents' educational attainment (for eighth-grade students only).

Other student-level factors might be just as critical but potentially subject to decisions and policies initiated by the student's school district or state. These factors include the student's classification as an English language learner, as receiving special education services, and as eligible for free and reduced-price lunch. These classifications could vary by state or even by district.

States are required to administer English language proficiency tests but have flexibility in the test they choose. Six states, including California, New York, and Texas, have developed their own tests, while others have opted into one of three multistate tests.[2] The use of these different tests means that a student could be identified as proficient in English by one state's test but not proficient by another state's test. Researchers have shown that English language learner reclassification standards differ, even between districts in the same state (Cimpian, Thompson, and Makowski 2017). This variability could affect our statistical adjustment. If a student could be reclassified as fluent in one district and not in another, her demographically similar comparison peers would also be different, which could affect our results.

Similarly, there may be variations in how districts and states classify special education students. Districts could have different standards for, or consistency in, special education classification, particularly for designations that involve learning or behavioral difficulties, rather than physical disabilities (Reap and Hanrahan 2017; Singer et al. 1989). Students who are from low-income backgrounds, who are black or Hispanic, or who are underage relative to their peers are overrepresented in special education, and this disproportionality could also vary by state or district (Artiles, Aguirre-Muñoz, and Abedi 1998; Bal, Sullivan, and Harper 2014; Dhuey and Lipscomb 2010; Sullivan 2017). The incentives generated by school funding formulas or state testing programs might further affect disability classification rates (Cullen 2003; Figlio and Getzler 2006).

The classification of students as receiving free and reduced-price lunch—a proxy for low-income status—might be less reliable across states and districts than it once was. Schools and districts with high shares of students from low-income families can use specific provisions under the National School Lunch Program to reduce the burden of collecting forms annually for individual students. Since 1980, high-poverty schools have had the option of certifying students eligible for free meals every two years (provision 1) or making eligibility determinations every four years, provided that all students are served meals regardless of free lunch status (provision 2) (FNS 2002).[3] The passage of the Healthy, Hunger-Free Kids Act of 2010 introduced the Community Eligibility Provision (CEP), which allows schools and districts with high shares of students identified as low income without paper applications (e.g., through participation in the Supplemental Nutrition Assistance Program or Temporary Assistance for Needy Families) to serve all students free lunch (USDA 2015). The program was phased in starting in the 2011–12 school year and has been available nationwide since 2014–15 (USDA 2015).

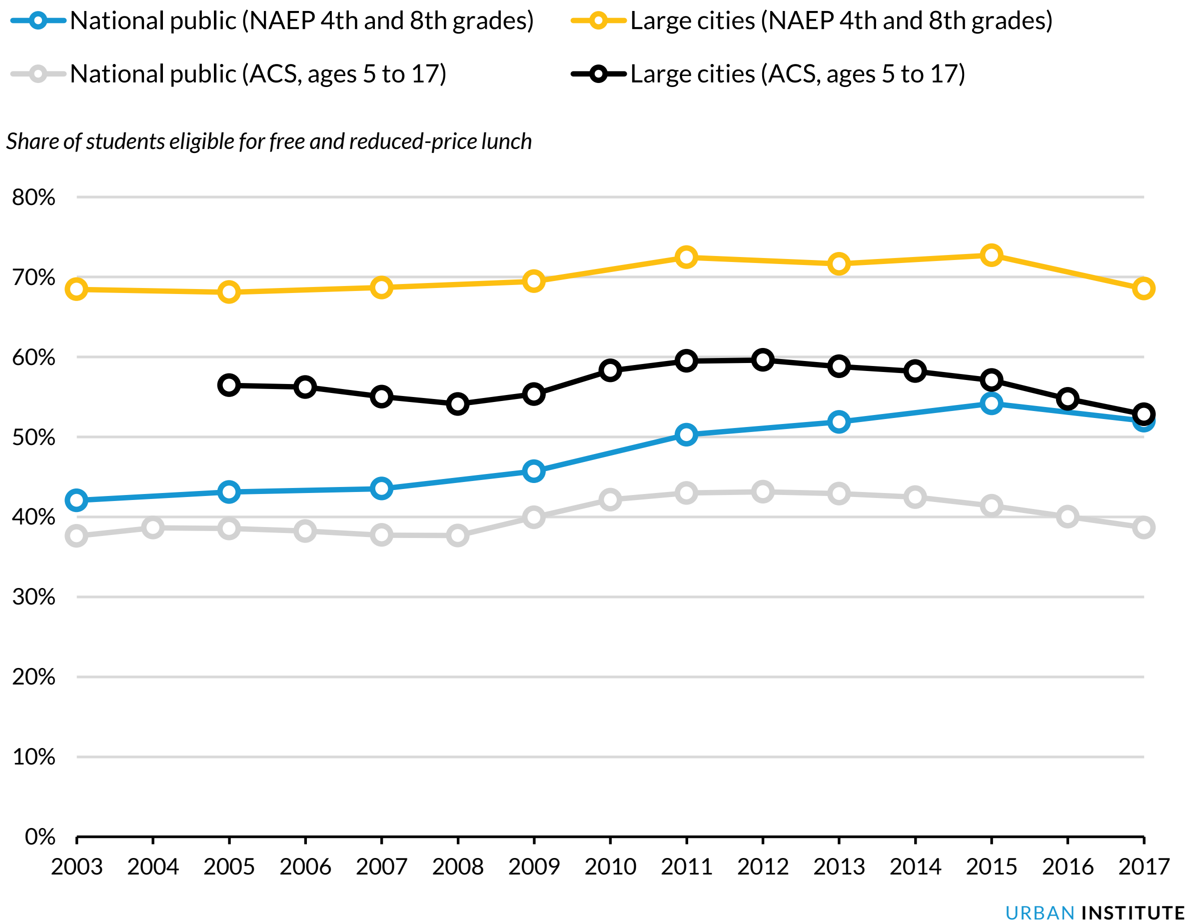
Because of these provisions—particularly the expansion of CEP from 2013 to 2017—NAEP administrators urge caution in interpreting achievement trends among students eligible for free and reduced-price lunch.[4] We observe that the trends in the share of students eligible for free and reduced-price lunch who take NAEP have grown increasingly divorced from the national share of public school students who would be eligible for the program (figure 2). We estimate these trends using the American Community Survey (ACS), building single-year estimates of the share of children ages 5 to 17 who are enrolled in public schools and live in households that earn up to 185 percent of the federal poverty level (i.e., eligible for free and reduced-price school meals).

In 2005, our ACS estimate of the share of students identified in the national student population as potentially eligible for free and reduced-price lunch was just 5 percentage points lower than the actual share in the NAEP sample. In 2017, our ACS estimate was 13 percentage points lower than the actual share assessed by NAEP. The share of National School Lunch Program participation in large cities, as assessed by NAEP, has hovered around 70 percent since 2003, though the gap in potential overidentification relative to a more limited ACS large-cities estimate has increased from 12 percentage points to 16 percentage points.

FIGURE 2

Share of NAEP Students Eligible for Free and Reduced-Price Lunch, Relative to ACS Estimates of Students Living in Households That Earn up to 185 Percent of the Federal Poverty Level

Averaged across fourth- and eighth-grade math and reading tests



Source: Urban Institute analysis of NAEP data from the Education Data Explorer and of ACS one-year microdata from IPUMS.
Notes: ACS = American Community Survey; NAEP = National Assessment of Educational Progress. In 2012, the ACS changed its geographical definitions, reducing the number of households that can be identified as within a central or principal city. We use only cities that are consistently identified from 2005 to 2017. This restricted set of cities follows the same trend as the larger cohort of households living in central or principal cities from 2005 to 2012.

Schools and districts might overidentify students as low income, particularly since CEP implementation starting in the in 2011–12. Since the 2015 NAEP administration, states have been instructed to identify students in CEP schools as eligible for free lunch if they would be identified as economically disadvantaged for reporting and accountability purposes under Title I.[5] Thus, one state could identify all students in a CEP school as eligible for free lunch, while another state could rule that only students who are directly certified as participating in a means-tested program such as the Supplemental Nutrition Assistance Program or Temporary Assistance for Needy Families should be deemed eligible for free lunch for NAEP reporting.

The 2017 NAEP data show this flexibility in practice. Some TUDA districts report nearly all students in provision schools as receiving free and reduced-price lunch, while others report a smaller share of their students as eligible. Cleveland and Philadelphia both reported more than 95 percent of students assessed for fourth- and eighth-grade NAEP were enrolled in provision schools. Cleveland reported 99 percent of students eligible for free and reduced-price lunch, while Philadelphia reported 71 percent of students as eligible (appendix table 1).

Building a New Indicator for Free and Reduced-Price Lunch

The misidentification or mismeasurement of any student characteristic across TUDA districts or over time could bias our demographic adjustment. The use of free and reduced-price lunch as a proxy for low-income status is of particular concern because there are few other indicators or proxies for students' household income, particularly in early years of NAEP data. State or district differences in the classification of English language learners and students receiving special education may be mitigated by the inclusion of other variables in our adjustment, such as the amount of English spoken at home and whether the student received accommodations in a regular testing session or in a separate testing session.

Measuring student socioeconomic status, particularly on national assessments such as NAEP, has long been a concern for researchers and policymakers. Although the share of students eligible for free and reduced-price lunch status is correlated with community-level data on socioeconomic status, these indicators are not synonymous (Nicholson et al. 2014). No single measure may be appropriate or adequate. The National Forum on Education Statistics identifies eight potential socioeconomic status measures, which center on such data as eligibility for means-tested programs, reported family income, parents' education and occupation, and community or school district poverty estimates (National Forum on Education Statistics 2015). Considering the development of a new socioeconomic status measure for NAEP, experts convened by the NCES in 2012 recommended developing a composite socioeconomic status measure, incorporating multiple data points to assess a student's family and community background (NCES 2012).

Because we do not have strong proxy indicators for the student's low-income status, and because many students might be misidentified or irregularly identified in the data over time, we build an imputed student-level variable for free and reduced-price lunch eligibility. We substitute this imputed variable in place of the reported free and reduced-price lunch indicator for students enrolled in provision schools.

We use an ordered probit regression model for each year, grade, and test subject to predict the probability that a student is eligible for free, reduced-price, or paid school lunch. Using reported free and reduced-price lunch status for students in nonprovision schools as a base, we model the relationship between free and reduced-price lunch eligibility and the student's other individual, school, and district characteristics within each year, grade, and test.

In all years, we include students' gender, race or ethnicity, English proficiency, language spoken at home, age, and special education status. For eighth-grade students, we include information on parents' educational attainment (not available for fourth-grade students). To determine the resources available in the student's home, we use the student's survey responses to whether she has a home computer and how many books are in her home. In 2013 through 2017, we also include internet access at home. From 2005 to 2011, we include the number of magazines the student has at home.

We include school-level information in the probit regression—the school's state, proximity to an urbanized area (e.g., city or rural area), the share of students at the school who were reported as eligible for free and reduced-price lunch, and the share of students receiving targeted Title I services. At the district level, we include whether the district is a large city, as defined by NAEP, and the share of children ages 5 to 17 living in poverty in the district, using the US Census Bureau's Small Area Income and Poverty Estimates.

Once we estimate this model for each year, grade, and test subject, we predict each student's probability of eligibility for free, reduced-price, or paid lunch. In our TUDA score adjustment, we use predicted probabilities of lunch status for students in provision schools (e.g., a predicted 70 percent probability of eligibility for free lunch) rather than using these probabilities to randomly assign the student to a single lunch eligibility category. But when we randomly assign a single free, reduced-price, or paid lunch status to each student based on these probabilities, we correctly identify the FRPL status of students in nonprovision schools 60 to 65 percent of the time.

Because of inconsistencies in how states report their 2017 data for provision schools, our imputation does not consistently reduce the share of students receiving free and reduced-price lunch in provision schools (appendix table 1). For example, Philadelphia and Detroit, which have nearly all students in provision schools, are imputed to have more students that would qualify for free and reduced-price lunch than they report. But other TUDA cities see a small drop in the overall share of students imputed to be eligible for free and reduced-price lunch relative to their reported shares.

For TUDA cities that were assessed in both 2005 and 2017, our imputation changes the share of students deemed eligible for free and reduced-price lunch over this period (appendix table 2). For example, Boston reported 79 percent of students eligible in 2005 and 72 percent in 2017. Our imputation adjustment keeps the 2005 share at 79 percent (because no provision schools were assessed) but slightly increases the 2017 share to 74 percent.

Assessing District Performance on the 2017 NAEP

We adjust the 2017 NAEP scores for TUDA districts using multiple models (appendix table 3). Our first model adjusts scores based only on the student's key background characteristics: race or ethnicity, gender, age, amount of English spoken at home, family structure, and parents' educational attainment (for eighth-grade students). Our next two models adjust scores using these variables and either the reported or imputed versions of the free and reduced-price lunch variable. Our final and preferred model incorporates the student's background characteristics, her imputed free and reduced-price lunch status, her status as an English language learner, her eligibility for special education, and the accommodations she received on the NAEP assessment.

When we adjust for these variables, comparing the performance of students who are demographically similar across these characteristics, we reduce the spread of district TUDA scores—the difference between the lowest-scoring and the highest-scoring districts—by about 29 percent (figure 3). But there are still substantial differences in academic performance across TUDA districts. Averaged across the four tests, the difference between the highest adjusted score and the lowest is 27 scale score points, or nearly 1 standard deviation.

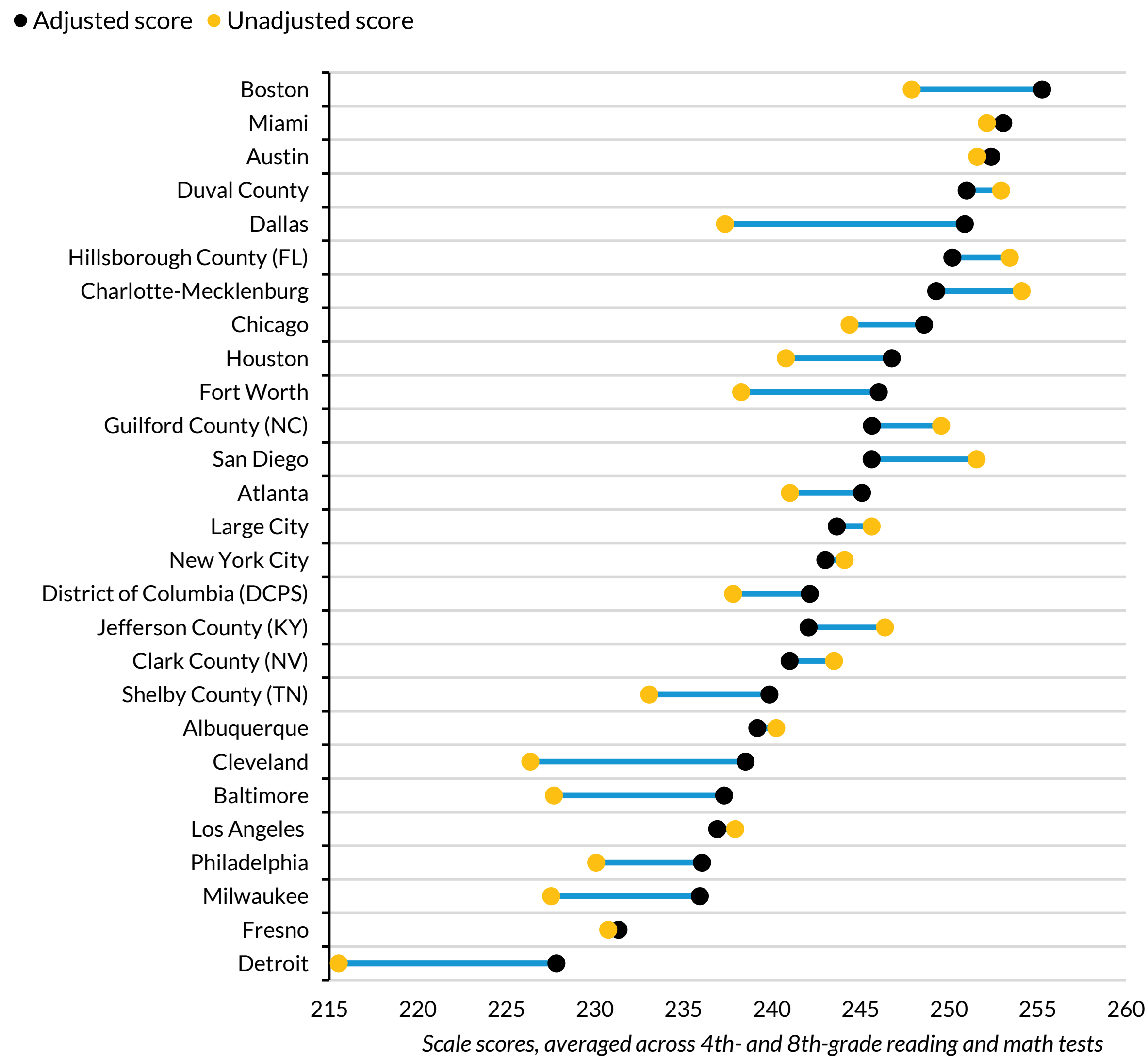
TUDA districts are selected, in part, because they serve higher shares of students of color and low-income students, so TUDA scores tend to be adjusted upward, relative to the performance of students who are not part of the TUDA sample. In previous work, we identified Massachusetts, Florida, and Texas as top performers in adjusted state scores. School districts in these states tend to cluster toward the top of our adjusted TUDA scores as well. Boston, Miami-Dade, and Austin take the top three spots, respectively.

The cities that receive the largest upward demographic adjustment include Dallas (with an average 14-point increase in scale scores), Detroit (12 points), and Cleveland (12 points). These districts serve higher shares of students from demographic backgrounds who tend to score less well on NAEP. But these larger adjustments do not always mean that a district is performing well compared with the other demographically adjusted districts. For example, Dallas moves closer to the top with this adjustment, but Detroit still lingers at the low end of the scale even after a statistical adjustment.

FIGURE 3

TUDA District Performance on the 2017 NAEP, Adjusted for Demographics

Scale score averaged across fourth- and eighth-grade mathematics and reading tests



URBAN INSTITUTE

Source: Urban Institute analysis of NAEP data.

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

Evaluating Changes in District Performance from 2005 to 2017

Demographically adjusted 2017 TUDA scores can identify districts that are serving their students well relative to the performance of students with similar backgrounds and needs. But these scores might mask another important measure of school district improvement: growth in academic achievement over time.

We have shown that NAEP scores for students in large cities have increased faster than the national average. This improvement could be attributable to improvements in instruction or curriculum or other district changes, but it could be because of changing student demographics. If a district enrolls a higher share of students from more advantaged backgrounds over time (e.g., more students whose parents have a college degree or fewer students who are English language learners), TUDA score increases could be driven by these demographic changes, rather than by changes in policy or other factors.

To untangle this relationship, we look at the 11 TUDA districts that were assessed in both 2005 and 2017. We estimate the relationship between a student's TUDA score and her demographic characteristics within the TUDA district in 2005 and predict the district's score in 2017 based on changes in these characteristics. For example, if a district sees an increase in the share of students that, based on their demographic characteristics, tend to do better on the NAEP assessment in 2005, we would predict that its 2017 score would be higher. If the district's actual scores are higher than this prediction, the district's students scored better, on average, than would be predicted by their demographically similar peers from 2005.

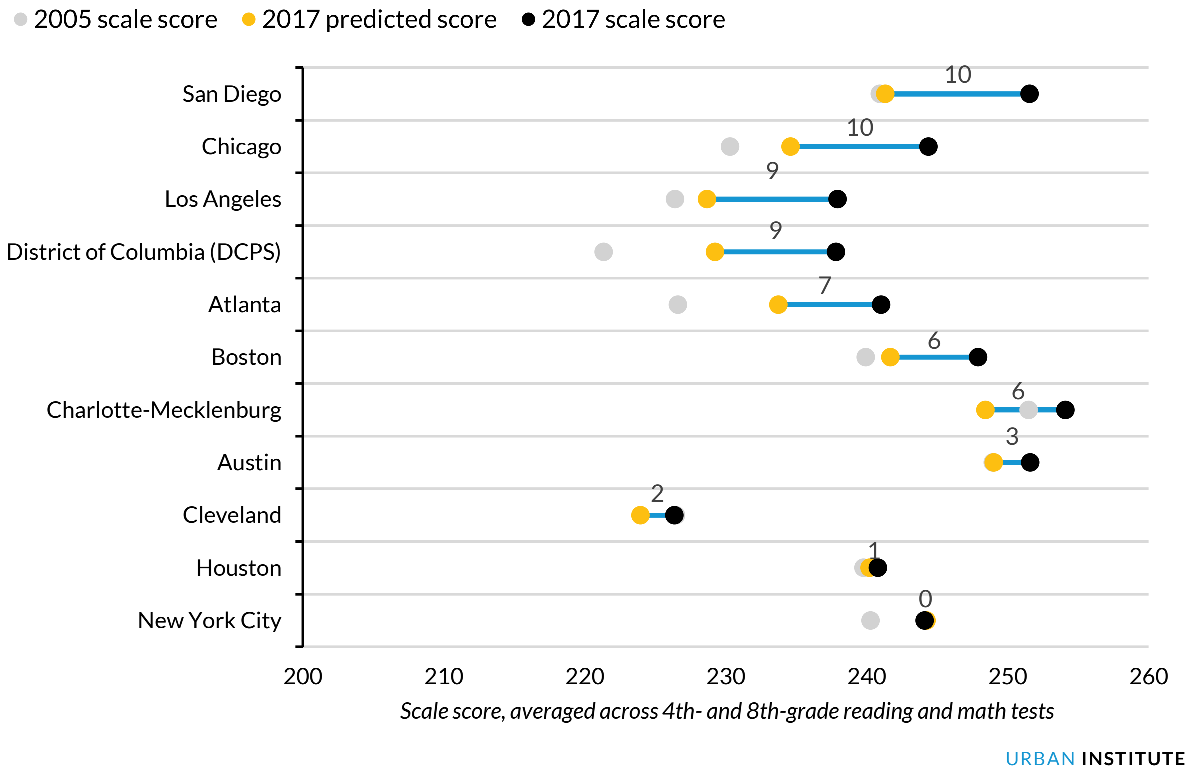
For this analysis, we use a more limited set of control variables that are available in both years of data: race or ethnicity, gender, age, amount of English spoken at home, parents' educational attainment (eighth-grade only), and our imputed free and reduced-price lunch variable. Because we are assessing the performance of students relative to their 2005 district peers only, our estimation could be particularly sensitive to policy changes around classifying students as English language learners or as eligible for special education. For example, if a district tightened its criteria for reclassifying English language learners over this period—resulting in higher shares of students classified as English language learners over time even though the 2005 and 2017 cohorts have similar levels of fluency—we would incorrectly predict a lower score in 2017. Similarly, if a district is working to reduce the overidentification of students with special needs, we would incorrectly predict a higher score. To avoid attributing change in scores to these state or district policy choices, we omit these variables from our analysis of score changes over time.

Our analysis indicates that some TUDA districts have made large gains that cannot be fully explained by changes in the student population (figure 4 and appendix table 4). For example, in 2005, Chicago had an average scale score, across the four tests, of 230.3 (figure 4, gray). Based on student demographic changes over 12 years, we predict that Chicago's score would increase by 4.3 scale points to 234.6 (yellow). But Chicago's score increased by 14.1 scale score points (black), or 9.8 points above what was predicted. This finding echoes data from 2009 to 2014, which found that Chicago students' test scores improved faster than those for the average US student (reardon and Hinze-Pifer 2017). We also observe large gains above what would have been predicted for traditional public schools in San Diego, Los Angeles, DC Public Schools, and Atlanta.

FIGURE 4

District Performance on the 2017 NAEP, Predicted versus Actual Score

Scale score averaged across fourth- and eighth-grade mathematics and reading tests



Source: Urban Institute analysis of NAEP data.

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

Conclusion

Differences in observable demographic characteristics do not fully explain the variation in TUDA test scores across districts and over time. Even when accounting for student background characteristics, there are still substantial differences in district performance on NAEP.

Although it is tempting to attribute these changes to school district policies and practices, we cannot exclude various unobservable factors that might contribute to this finding. These unobservable factors could include differences in the generosity of the local social safety net, differences in exposure to crime, differences in the occurrence of natural disasters, or other environmental variables that affect academic outcomes. Thus, this analysis can only broadly point to districts where student achievement is higher or has risen over time and cannot point to specific policy levers as the cause.

Further, although we have controlled for many observable characteristics, unobservable student-level characteristics (e.g., a student's personal motivation or her family's perspective on academic achievement) could contribute to these differences.

The TUDA assessment program provides a valuable yardstick for policymakers who want to understand how their school district performs relative to the academic achievement of other US students. Our demographic adjustments make this comparison easier by attempting to level the playing field in terms of the needs of the students each district serves. This analysis points to districts that have relatively higher student achievement, both compared with other districts and over time, prompting further study of whether we can attribute these changes to certain school district policies or practices.

Appendix

TABLE A.1

District FRPL Share, as Reported on the 2017 NAEP

Share averaged across fourth- and eighth-grade mathematics and reading tests

	Share of students in provision schools	Eligible for FRPL, in nonprovision school	Eligible for FRPL, in provision school	Overall eligible for FRPL (reported)	Overall eligible for FRPL (imputed)
Albuquerque	56%	43%	100%	70%	62%
Atlanta	70%	38%	100%	76%	68%
Austin	0%	57%	-	55%	55%
Baltimore	96%	77%	70%	67%	77%
Boston	98%	75%	71%	72%	74%
Charlotte-Mecklenburg	53%	25%	55%	41%	52%
Chicago	86%	86%	80%	82%	78%
Clark County (NV)	23%	64%	100%	71%	68%
Cleveland	96%	100%	100%	99%	88%
Dallas	91%	96%	93%	92%	86%
Denver	2%	68%	97%	67%	67%
Detroit	86%	79%	77%	76%	84%
Duval County	64%	34%	61%	49%	58%
Fresno	98%	89%	86%	86%	84%
Fort Worth	56%	63%	86%	79%	78%
Guilford County (NC)	57%	38%	58%	47%	56%
Hillsborough County (FL)	2%	55%	75%	55%	55%
Houston	49%	64%	86%	76%	76%
Jefferson County (KY)	87%	43%	69%	64%	63%
Los Angeles	56%	59%	66%	62%	71%
Large City	34%	58%	87%	66%	63%
Miami	0%	72%	-	72%	72%
Milwaukee	98%	69%	89%	82%	82%
New York City	50%	64%	74%	70%	63%
Philadelphia	98%	67%	73%	71%	82%
San Diego	40%	44%	97%	64%	60%
Shelby County (TN)	91%	74%	58%	58%	78%
District of Columbia (DCPS)	59%	32%	100%	73%	62%

Source: Urban Institute analysis of NAEP data.

Note: DCPS = DC Public Schools; FRPL = free and reduced-price lunch; NAEP = National Assessment of Educational Progress; TUDA = Trial Urban District Assessment.

TABLE A.2

Change in District FRPL Share from 2005 to 2017 NAEP

Share averaged across fourth- and eighth-grade mathematics and reading tests

	2005			2017		
	Share of students in provision schools	Eligible for FRPL (reported)	Eligible for FRPL (imputed)	Share of students in provision schools	Eligible for FRPL (reported)	Eligible for FRPL (imputed)
Atlanta	0%	76%	76%	70%	76%	68%
Austin	0%	55%	55%	0%	55%	55%
Boston	5%	79%	79%	98%	72%	74%
Charlotte-Mecklenburg	0%	46%	46%	53%	41%	52%
Chicago	0%	83%	83%	86%	82%	78%
Cleveland	82%	100%	78%	96%	99%	88%
Houston	0%	73%	73%	49%	76%	76%
Los Angeles	23%	81%	77%	56%	62%	71%
Large City	9%	59%	57%	34%	66%	63%
New York City	32%	85%	70%	50%	70%	63%
San Diego	23%	59%	56%	40%	64%	60%
District of Columbia (DCPS)	3%	73%	71%	59%	73%	62%

Source: Urban Institute analysis of NAEP data.

Note: DCPS = DC Public Schools; FRPL = free and reduced-price lunch; NAEP = National Assessment of Educational Progress.

TABLE A.3

Raw and Adjusted TUDA Scores for 2017

Scores for all four tests, adjusted for different student characteristics

	2017 Raw Scores				2017 Adjusted Scores (Background Characteristics)				2017 Adjusted Scores (Background + Reported FRPL)				2017 Adjusted Scores (Background + Imputed FRPL)				2017 Adjusted Scores (Background + Imputed FRPL + ELL and Special Education)			
	G4		G8		G4		G8		G4		G8		G4		G8		G4		G8	
	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R
Albuquerque	230	207	270	255	228	207	269	253	229	209	270	254	228	208	269	253	227	206	270	253
Atlanta	231	214	265	254	233	216	275	260	234	217	277	262	233	216	275	261	233	215	274	260
Austin	243	217	283	263	240	215	281	260	239	214	280	259	239	214	280	259	244	220	283	263
Baltimore	215	197	255	243	221	203	270	253	221	202	269	252	223	204	270	253	223	203	271	253
Boston	233	217	280	261	234	219	283	264	235	220	283	264	235	220	284	265	239	225	288	269
Charlotte-Mecklenburg	244	225	287	260	241	222	285	258	238	218	283	256	240	221	285	257	239	219	283	256
Chicago	232	211	276	259	232	213	281	262	234	215	283	263	234	215	283	263	235	216	282	262
Clark County (NV)	230	213	272	258	226	210	270	256	228	213	271	257	228	212	271	257	226	210	271	257
Cleveland	214	196	257	237	219	202	270	245	224	207	274	249	222	205	273	248	223	206	275	250
Dallas	234	201	268	246	237	207	278	254	240	210	281	256	240	210	280	256	243	211	286	263
Detroit	200	182	246	235	207	188	260	245	207	189	259	245	209	191	261	247	209	191	262	249
Duval County	248	226	275	263	245	223	276	262	243	221	275	261	245	223	276	262	245	223	275	261
Fresno	221	203	255	244	219	203	256	245	223	207	258	247	223	207	259	248	221	204	256	245
Fort Worth	230	206	269	248	232	210	278	254	234	211	278	255	234	211	279	255	236	215	278	255
Guilford County (NC)	240	222	276	260	236	218	273	255	234	216	272	254	236	218	274	256	235	217	274	256
Hillsborough County (FL)	245	227	277	265	240	224	275	261	240	223	274	260	240	223	274	260	240	223	275	262
Houston	235	205	273	249	237	209	281	254	238	210	281	254	238	211	281	255	240	213	280	254
Jefferson County (KY)	233	221	271	261	228	215	267	256	230	217	268	257	230	217	269	258	229	216	268	256
Los Angeles	223	207	267	254	221	207	266	253	221	207	266	252	222	209	267	254	222	208	266	253
Large City	232	214	277	260	229	211	275	257	230	212	276	258	229	212	275	258	230	212	275	258
Miami	245	229	274	261	247	232	277	262	246	232	278	263	246	232	279	263	244	229	277	262
Milwaukee	216	195	254	245	218	198	262	251	221	201	264	252	220	201	264	252	221	202	266	254
New York City	229	214	275	258	224	210	270	254	226	212	271	255	225	211	270	254	228	214	273	257
Philadelphia	214	197	260	248	216	200	267	253	217	201	267	253	219	203	269	255	218	202	269	255
San Diego	237	222	283	264	230	215	273	257	231	217	274	258	231	217	273	257	232	219	273	258
Shelby County (TN)	225	203	257	248	229	207	267	254	227	205	265	253	230	208	269	256	229	206	268	256
District of Columbia (DCPS)	231	213	262	246	233	215	269	250	234	216	270	251	232	214	268	249	233	214	271	251

Source: Urban Institute analysis of NAEP data.

Note: DCPS = DC Public Schools; ELL = English language learners; FRPL = free and reduced-price lunch; G4 = fourth grade; G8 = eighth grade; M = math; R = reading; TUDA = Trial Urban District Assessment.

TABLE A.4

Raw and Adjusted TUDA Scores for 2005 and 2017

Scores for all four tests, predicted and actual for 2017 based on 2005 performance

	2005 Scale Scores				2017 Predicted Scores				2017 Scale Scores			
	G4		G8		G4		G8		G4		G8	
	M	R	M	R	M	R	M	R	M	R	M	R
Atlanta	221	201	245	240	227	209	253	245	231	214	265	254
Austin	242	217	281	257	243	216	281	256	243	217	283	263
Boston	229	207	270	253	231	209	272	254	233	217	280	261
Charlotte-Mecklenburg	244	221	281	259	240	218	278	257	244	225	287	260
Chicago	216	198	258	249	220	204	262	253	232	211	276	259
Cleveland	220	197	249	240	216	191	249	239	214	196	257	237
Houston	233	211	267	248	235	209	268	249	235	205	273	249
Los Angeles	220	196	250	239	222	197	254	241	223	207	267	254
New York City	231	213	267	251	229	206	268	252	232	214	277	260
San Diego	232	208	270	253	234	216	272	255	229	214	275	258
District of Columbia (DCPS)	211	191	245	238	232	205	273	255	237	222	283	264

Source: Urban Institute analysis of NAEP data.

Note: DCPS = DC Public Schools; G4 = fourth grade; G8 = eighth grade; M = math; R = reading; TUDA = Trial Urban District Assessment.

Notes

[1] “The NAEP Glossary of Terms,” US Department of Education, Institute of Education Sciences, National Center for Education Statistics, accessed July 19, 2019, https://nces.ed.gov/nationsreportcard/glossary.aspx.

[2] “English Language Proficiency (ELP) Assessments,” New America, accessed July 19, 2019, https://www.newamerica.org/education-policy/topics/english-learners/dual-language-learners/dll-assessment/english-language-proficiency-assessments/non-consortia-states/.

[3] See also “National School Lunch Program: Provisions 1, 2, and 3,” US Department of Agriculture, Food and Nutrition Service, last updated May 6, 2014, https://www.fns.usda.gov/school-meals/provisions-1-2-and-3.

[4] “Interpreting NAEP Mathematics Results,” US Department of Education, Institute of Education Sciences, National Center for Education Statistics, accessed July 19, 2019, https://nces.ed.gov/nationsreportcard/mathematics/interpret_results.aspx.

[5] From correspondence with staff at the National Center for Education Statistics Assessments Division, National Assessment Branch.

References

Artiles, Alfredo J., Zenaida Aguirre-Muñoz, and Jamal Abedi. 1998. “Predicting Placement in Learning Disabilities Programs: Do Predictors Vary by Ethnic Group?” *Exceptional Children* 64 (4): 543–59.

Bal, Aydin, Amanda L. Sullivan, and John Harper. 2014. “A Situated Analysis of Special Education Disproportionality for Systemic Transformation in an Urban School District.” *Remedial and Special Education* 35 (1): 3–14.

Blagg, Kristin. 2016. “Making the Grade in America’s Cities: Assessing Student Achievement in Urban Districts.” Washington, DC: Urban Institute.

Chingos, Matt. 2015. *Breaking the Curve: Promises and Pitfalls of Using NAEP Data to Assess the State Role in Student Achievement*. Washington, DC: Urban Institute.

Cimpian, Joseph R., Karen D. Thompson, and Martha B. Makowski. 2017. “Evaluating English Learner Reclassification Policy Effects across Districts.” *American Educational Research Journal* 54 (1_suppl): 255S–278S.

Cullen, Julie Berry. 2003. “The Impact of Fiscal Incentives on Student Disability Rates.” *Journal of Public Economics* 87 (7–8): 1557–89.

Dhuey, Elizabeth, and Stephen Lipscomb. 2010. “Disabled or Young? Relative Age and Special Education Diagnoses in Schools.” *Economics of Education Review* 29 (5): 857–72.

Figlio, David N., and Lawrence S. Getzler. 2006. “Accountability, Ability, and Disability: Gaming the System?” In *Improving School Accountability*, vol. 14, edited by Timothy J. Gronberg and Dennis W. Jansen, 35–49. Bingley, United Kingdom: Emerald Group Publishing Limited.

FNS (Food and Nutrition Service). 2002. *Provision 2 Guidance: National School Lunch and School Breakfast Programs.* Alexandria, VA: US Department of Agriculture, FNS.

National Forum on Education Statistics. 2015. *Forum Guide to Alternative Measures of Socioeconomic Status in Education Data Systems*. Washington, DC: US Department of Education, National Center for Education Statistics.

NCES (National Center for Education Statistics). 2012. *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation*. Washington, DC: US Department of Education, NCES.

Nicholson, Lisa M., Sandy J. Slater, Jamie F. Chriqui, and Frank Chaloupka. 2014. “Validating Adolescent Socioeconomic Status: Comparing School Free or Reduced-Price Lunch with Community Measures.” *Spatial Demography* 2 (1): 55–65.

Reap, Thomas, and Jo Ann Hanrahan. 2017. "Crossing the Line: Exploring Equity in Special Education across the United States." Frontline Research and Learning Institute.

reardon, sean f., and Rebecca Hinze-Pifer. 2017. "Test Score Growth among Public School Students in Chicago, 2009–2014." Stanford, CA: Stanford Center for Education Policy Analysis.

Singer, Judith D., Judith S. Palfrey, John A. Butler, and Deborah Klein Walker. 1989. "Variation in Special Education Classification across School Districts: How Does Where You Live Affect What You Are Labeled?" *American Educational Research Journal* 26 (2): 261–81.

Sullivan, Amanda L. 2017. "Wading through Quicksand: Making Sense of Minority Disproportionality in Identification of Emotional Disturbance." *Behavioral Disorders* 43 (1): 244–53.

USDA (US Department of Agriculture). 2015. "The Community Eligibility Provision (CEP): What Does It Mean for Your School or Local Education Agency?" Washington, DC: USDA.

About the Authors

Kristin Blagg is a research associate in the Center on Education Data and Policy at the Urban Institute. Her research focuses on K–12 and postsecondary education. Blagg has conducted studies on student transportation and school choice, student loans, and the role of information in higher education. In addition to her work at Urban, she is pursuing a PhD in public policy and public administration at the George Washington University. Blagg holds a BA in government from Harvard University, an MSEd from Hunter College, and an MPP from Georgetown University.

Grace Luetmer is a research analyst in the Center on Education Data and Policy. She graduated from Boston College with a BA in economics and from Carnegie Mellon University with an MS in public policy and management.

Acknowledgments

This brief was funded by the Bill & Melinda Gates Foundation. 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 urban.org/fundingprinciples.

We thank Joshua Goodman and participants at the Harvard Kennedy School's Education Cities conference for comments on an earlier draft. We also thank Matthew Chingos for his helpful feedback and David Hinson for editorial assistance.

URBAN INSTITUTE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org