RESEARCH

RESEARCH DEPARTMENT OVERVIEW

Research Department Overview
Fall, 2020

Overall Research Department Goals/Priorities

The goal of the research department is to conduct, facilitate and disseminate research that will provide guidance and support to the Council's member districts and other key stakeholders as they work to improve academic achievement and reduce achievement gaps in large urban school districts. The following reports and presentations will be available on our Research Department webpage: http://www.cgcs.org/Research.

Update on Recent Completed Projects/Conferences

17th Annual Curriculum, Research, and Instructional Leaders Meeting

The research and academic teams postponed the 17th Annual Curriculum, Research, and Instructional Leaders Meeting in Atlanta, GA. Given the unprecedented nature of Covid-19, the Council of the Great City Schools decided it was best to postpone the conference until July of 2021 to protect the health of attendees, presenters, and staff. The meeting is traditionally attended by over 175 representatives from Council member districts, staff, and sponsors.

Weekly Covid-19 Research and Assessment Directors Conference Calls

The Council of the Great City Schools began meeting weekly with Research, Evaluation, and Assessment Directors on March 24, 2020 to discuss key decisions and plans given the unprecedented national circumstances associated with Covid-19. CGCS thought it might be useful to provide a forum by which directors could talk in a safe space with colleagues across the country about how they are handling the research and assessment issues that have emerged as districts and states deal with COVID-19. We continue to arrange Zoom meetings (virtual meeting/call) each Tuesday, at 1:00 PM EST to discuss key issues that arise each week. Starting in October 2020, Zoom meetings will be scheduled every other Tuesday, at 1:00 PM EST. Some guiding questions for our conversations are listed below:

- How is your district addressing instruction for students during this period, including English language learners and students with disabilities? Has your research team been asked to evaluate the impact of the closure/distance learning protocols on student learning outcomes?

- What guidance has your state provided about the impact of the school closures on your assessments?
- How are you supporting teachers and administrators with assessing student learning progress during school closures?
- How are you addressing grade-level promotion and graduation?
- How are you planning to assess student learning when school closures end?
- How are you planning to reopen schools?
- What are you asking students, staff, parents, and community members in your district surveys?
- What have you learned from your students, staff, parents, and community surveys?

COVID-10 Research and Assessment Publication

As one of six COVID-19 publications released by the Council of the Great City Schools, a team of research directors from Portland, Indianapolis, Toronto, Charlotte-Mecklenburg, Dallas, Guilford County (Greensboro, NC), Tulsa, Milwaukee, Austin and the Council worked together to write *Ensuring a Data-Driven Approach to Reopening Schools after COVID-19: Recommendations for Research and Assessment*. The full report is provided below.

This document centers research departments as uniquely connecting many other divisions in the central office – curriculum, information technology, student support services, career and technical education, assessment, and facilities – and the data they collect often serve as the glue that holds district operations together. The publication helps define the vital role that research departments play in district operations and planning for reopening schools in the fall. The document specifically addresses how research and assessment can contribute to the reopening of schools in 2020 by:

- Informing Decisions on Programming, Policy, and Budget
- Evaluating District Initiatives
- Identifying Student, Staff, and Community Needs
- Addressing Equity Disparities for Students and Families
- Assessing Impacts on Student Learning Outcomes
- Evaluating Strengths and Weaknesses of Distance Learning Implementation

- Rethinking Policies and Protocols for Calculating Student Enrollment
- Rethinking Transportation and GIS Algorithms
- Temporarily Suspending External Research in Schools

2021 National Assessment of Educational Progress (NAEP)
Trial Urban District Assessment (TUDA)

CGCS has been working closely with the National Assessment Governing Board (Governing Board) to work through any issues and concerns with the administration of the 2021 National Assessment of Educational Progress (NAEP) Trial Urban District Assessment (TUDA). On June 29, 2020, Michael Casserly shared remarks that addressed the sentiments of the TUDA superintendents with the Governing Board. He responded to questions from board members regarding the pending decision to move forward with, or cancel, the 2021 assessment in January. The Governing Board decided at the July board meeting to proceed with the NAEP assessment in January. However, the National Center for Education Statistics, which administers the assessment, has elected to significantly reduce the national sample and will not provide TUDA results for the 2021 assessment. The Council's comments from the meeting are included below.

Trial Urban District Assessment Advisory Task Force to the National Assessment Governing Board

Given the 2017 expansion of the Trial Urban District Assessment (TUDA) program to 27 districts, the Council submitted a technical proposal to the National Assessment Governing Board (Governing Board) to establish a Task Force of local education leaders from TUDA districts. The Council was awarded a contract for a 24-month effort that included the creation, project management, and on-going coordination of the TUDA Task Force. The research team completed the final phase of the requirements for the contract in December 2019. The Council and the Governing Board have now entered into a new contract to continue task force activities for another three years.

The first Task Force provided feedback to the Governing Board, including recommendations on areas of policy, research, and communications related to the TUDA program. The Task Force helped inform, strengthen, and guide the Strategic Vision of the Governing Board and the evolution of the TUDA program. Perhaps the most significant accomplishment of the Task Force was the role of the Council and the group in the development of the new NAEP mathematics framework. Task force members and Council staff formally shared perspectives on the framework, NAEP assessment practices, NAEP national and local communication strategies, and several other topics.

Update on On-Going Projects

Analysis of TUDA Performance and the Influence and Impact of Public Schools on Student Achievement and Urban School Districts

In the spring of 2011, the Council research team published the study *Pieces of the Puzzle: Recent Performance Trends in Urban Districts – A Closer Look at 2009 NAEP Results (An Addendum)*. A portion of that report analyzed the National Assessment of Educational Progress (NAEP) performance of Trial Urban District Assessment (TUDA) performance while adjusting the district performance based on key background variables. The key background variables included race/ethnicity, special education status, English language learner status, free- or reduced-price lunch eligibility, parental education level (grade eight only), and a measure of literacy materials available in the home. The analysis compared the predicted NAEP performance (after controlling for the background variables) to the actual NAEP performance of the districts. The analysis allowed the Council to identify districts that were performing better than expected on the NAEP assessment and beginning to mitigate some of the effects of poverty and other background characteristics of students that typically suppress academic performance.

The lessons learned from that study have prompted the Council research team to replicate the analysis using data from the 2011, 2013, 2015, 2017, and 2019 administrations of NAEP reading and mathematics assessments in grades four and eight. This study not only identifies districts that continue to perform better than expected based on background variables, but when combined with the analysis of the 2009 data, district trends in performance can be examined which provide a very different picture of the changes in district effects over time. For example, Detroit has typically been one of the lowest performing TUDA district, and even when controlling for relevant background variables, Detroit performs lower than expected. However, this analysis revealed that Detroit is one of only a few districts that has made consistent progress on the NAEP assessment each year across multiple grades and subjects (grade eight reading and grade four math). The progress Detroit is making is all but lost in any other analysis of student performance in the district, but indicates that student achievement, though not where it needs to be, is improving. The Council has taken the additional step of applying Census poverty data at the school level to further illuminate the districts that are overcoming the effects of abject poverty and other demographic factors.

Methodology

For this analysis, the research team conducted a hierarchical linear modeling (HLM) analyses to estimate the performance of a district if its demographic profile, in terms of the selected student and school background characteristics, is the same as the average profile

of all students across the country. The analyses put the districts on a more level playing field regarding these characteristics. Based on this HLM analyses (using student and school level data), we computed the expected performance of each district based on their profile in terms of the selected student background characteristics. We subtract the expected performance from the actual performance to calculate the "district effect." We then analyzed the changes in the district effects over the 2009, 2011, 2013, 2015, and 2017 NAEP administrations. We have now added data to the analysis from the 2019 NAEP administration that were released the last week in September 2020.

We revised how we handle what we found were anomalies in the data based on district changes in the identification of Free or Reduced-Price Lunch (FRPL) students. Table 1 illustrates the changes in the identification of FRPL students since 2015. In the Council's analysis, some districts have observed as much as 30 to 40 percentage point changes in the students identified. As a result, the credit that district's receive for educating students in poverty is underestimated and the district effects are subsequently underestimated as well. Consequently, we have incorporated a school-level free or reduced-price lunch rate that is estimated from the NAEP sample or the National Center for Education Statistics' Common Core of Data results.

Based on the NAEP district effect analysis, the Council selected six districts—Boston, Chicago, Dallas, Miami-Dade, San Diego and Washington, DC—that have made substantial progress overcoming the effects of poverty, language, and discrimination on student achievement for site visits. The team conducted site visits in Boston and the District of Columbia Public Schools in Spring 2018. We followed with site visits to Miami-Dade County, Chicago Public Schools, and San Diego Public Schools in the Fall 2018. The last site visit was completed in Dallas in February 2019. The team spoke with a broad cross section of central office and school staff about the factors that led to their success in raising student achievement—particularly with vulnerable student groups. A "counterfactual" district—one that has not demonstrated any growth among these student groups—will also be selected, and the team will visit this district to explore potential differences in practices between districts with varied outcomes.

Using our *Indicators of Success*, we will determine the level of common core implementation in these improving districts in order to investigate whether strong standards implementation work has made a difference in districts' ability to overcome the effects of poverty and language and raise student achievement. We will also explore a broad range of other factors that may have played a role in the achievement outcomes. Based on our findings, we will finalize our NAEP analysis and report by answering the question of how some districts were able to "beat the odds."

A draft report of the initial results of the study has been completed. A final formal report is provided in Achievement Task Force section of this report.

Table 1. TUDA Free or Reduced-Price Lunch Methodology, 2015 – 2019

TUDA	NAEP 2015	NAEP 2017	NAEP 2019
ALB	CEP-ALL	CEP-ALL	CEP-ALL
ATL	**Direct-Only**	CEP-ALL	CEP-ALL
CLA	CEP-ALL	CEP-ALL	CEP-ALL
CLE	CEP-ALL	CEP-ALL	CEP-ALL
DC	CEP-ALL	CEP-ALL	CEP-ALL
AUS	Direct-Plus	Direct-Plus	Direct-Plus
CHA	**CEP-ALL**	**CEP-Direct**	**Direct-Plus**
CHI	Direct-Plus	Direct-Plus	Direct-Plus
DAL	Direct-Plus	Direct-Plus	Direct-Plus
DEN	N/A	Direct-Plus	Direct-Plus
DET	Direct-Plus	Direct-Plus	Direct-Plus
FTW	N/A	Direct-Plus	Direct-Plus
FRE	Direct-Plus	Direct-Plus	Direct-Plus
GUI	N/A	**CEP-Direct**	**Direct-Plus**
HOU	Direct-Plus	Direct-Plus	Direct-Plus
JEF	Direct-Plus	Direct-Plus	Direct-Plus
LOS	Direct-Plus	Direct-Plus	Direct-Plus
MIL	N/A	Direct-Plus	Direct-Plus
NYC	Direct-Plus	Direct-Plus	Direct-Plus
PHI	**CEP-Direct**	**Direct-Plus**	**Direct-Plus**
SAN	Direct-Plus	Direct-Plus	Direct-Plus
BLT	N/A	Direct-Only	Direct-Only
DUV	Direct-Only	Direct-Only	Direct-Only
HIL	Direct-Only	Direct-Only	Direct-Only
MIA	Direct-Only	Direct-Only	Direct-Only
SHE	N/A	**Direct-Only**	**Direct-Only (CEP & Non-CEP schools)**
Boston	**CEP-ALL**	**Direct-Only**	**Direct-Only (CEP & Non-CEP schools)**

Operations and Academic Key Performance Indicators

The board of directors authorized the development of Operations Key Performance Indicators in 2002 and the Academic Key Performance Indicators in the 2014. Several teams of educators from Council member districts crafted a list of desired indicators for operations areas including business services, finance, human resources, and technology and academic areas including general core instruction, special education, and English language learners. The refined set of Academic Key Performance Indicators are designed

to measure the progress among the Council's membership toward improving the academic outcomes for students and include the following:

- Ninth grade algebra completion
- Ninth graders failing one or more core courses
- Ninth graders with a GPA of B or better
- Number of high school students enrolled in advanced placement
- AP exam scores of 3 or higher
- Number of high school students enrolled in AP-equivalent courses
- Four-year high school graduation rate
- Five-year high school graduation rate
- Percent of students with 20 days or more absent from school
- Instructional days per student missed per year due to suspension
- Percent of students identified as needing special education
- Percent of students placed in each general education setting by percent of time

Report. The Council released the request for data for the operations key performance indicators, *Managing for Results in America's Great City Schools 2020*, and the academic key performance indicators, *Academic Key Performance Indicators 2020 Report*, in February. The deadline was originally set for late April 2020, but the research team extended the district deadline for submission to August 7, 2020 to allow districts flexibility given their response to the coronavirus. We did not send reminders of the due dates to districts. Instead, districts who did not submit data in 2020 for one or both of the reports will be encouraged to submit data for 2020 and 2021 next spring. Final reports for the year are provided in Achievement and Management and Governance sections of this report.

Information Technology Update

The Council's research team has developed the first edition of our Academic KPI dashboard. The Council currently collects over 1,000,000 data points and uses those data points to create over 200,000 calculations for our Academic KPIs. We created digital dashboards that visualize more data than previously available in the Academic KPI report. The dashboards allow for longitudinal comparisons for those districts who have submitted survey data across multiple years. Another feature of the dashboard is the ability for districts to compare themselves to peer groups. Peer groups are defined as those districts that have similar student demographics, i.e., district enrollment, FRPL eligibility, ELL status and race/ethnicity. Peer groups allow districts to compare themselves not only to *all* Council districts, but more specifically to Council member districts that share common demographics.

The Council research team is beginning to update the existing data dashboards for more functionality. The research team is unveiling a new Special Education dashboard in

October 2020. Planned updates include the creation of an English learner dashboard to visualize data specifically related to Els. Two examples of the Special Education dashboard are provided below (see Figures 1 and 2).

The Academic KPI dashboards are securely and confidentially available at EdWires.org.

We also relaunched and expanded EdWires.org, a Council-only web application for private online access to files and resources. With the launch is an easy and secure self-registration process: submit your district email address and enter the verification code that is sent to your email from "EdWires by CGCS". Only member district employees with a Council district email domain can log in. Once logged in, CGCS members have immediate access to the Academic KPI dashboards, shared documents from member districts, as well as the confidential KPI ID number for your member district.

In response to the COVID-19 pandemic, the Council created the "COVID-19 Resources" section on our private platform fileshare.edwires.org. This section allows districts to share information amongst themselves relating to the response to COVID-19. Documents include student, staff, and community surveys; parent and community engagement materials; documents on addressing learning loss from curriculum staff; operational considerations for reopening schools; special education documents; and much more. There are private workspaces for sharing sensitive information for district legal teams and for superintendents. The technology team is working on the development of group discussion boards (Edwires Forums) expected to launch in early Spring 2021.

While the listserv is useful for mass communication, the Edwires forum will facilitate smaller discussions. On the forum, members can privately message each other for one-on-one discussions or post to job-alike groups. Members logged in to EdWires.org will be able to enroll in collaboration groups that match their job-alike function and professional interests, as well as task-oriented groups like task forces and working groups. The Council will continue to roll out new and useful improvements to EdWires.org as the memberships' needs evolve.

Figure 1. Sample Special Education Dashboard for Key Performance Indicators, 2017-18

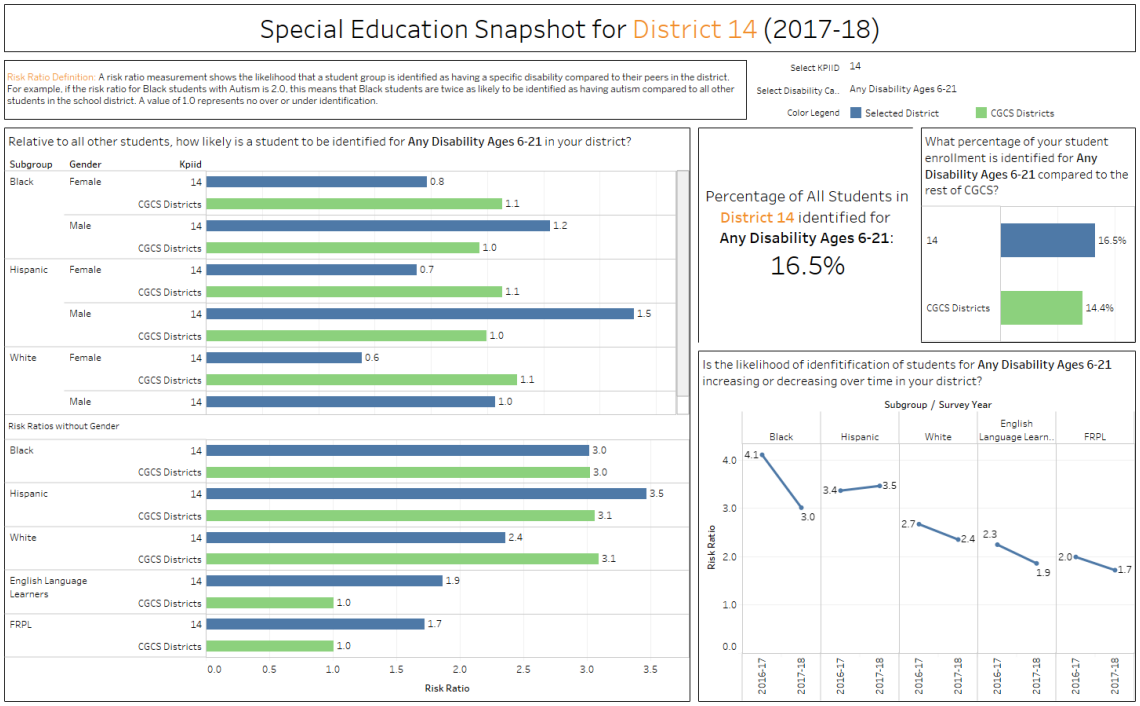
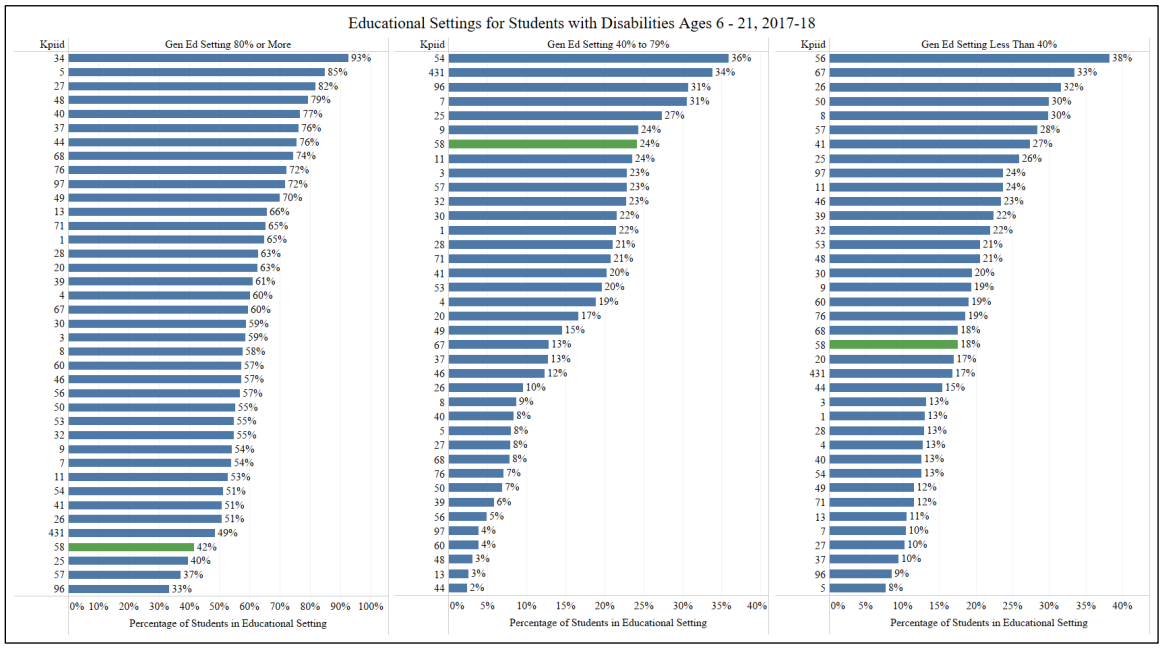


Figure 2. Sample Special Education – Educational Settings - Dashboard for Key Performance Indicators, 2017-18



Update on New Projects

The Annie E. Casey Foundation, National Network of Education Research-Practice Partnerships, Council of the Great City Schools Literature Scan Project

Over the past three months, school districts have faced unprecedented closures and changes to the school year because of the COVID-19 pandemic. These closures have forced school districts to reimagine instructional delivery to students and support for teachers through distance learning and technology. Already existing inequities, such as limited access to technological devices and the internet prior to the pandemic, have likely widened the digital divide between students with lower economic means. Moreover, the current economic environment will result in greater challenges to educational delivery for teachers and schools due to persisting barriers to financial stability.

The 2020-21 school year will present ongoing challenges, some that can be anticipated, and others that will not. The pandemic has forced district leaders to shift plans and thinking daily to maintain student safety and maximize learning opportunities. Over the next year, there will be an increased need for access to the most relevant research by district administrators, principals, and teachers to inform their decision-making and planning. For example, we have already witnessed a heightened demand for evidenced-based practices supporting the implementation of distance learning, including topics related to asynchronous, blended, and full-time distance learning. Given the increased digital divide, it is imperative that these practices include considerations for educating

students in poverty, students with special needs, English language learners, and students of color to help create equitable experiences across the country.

While there is an increasing demand for evidence, districts must still work through their daily challenges, limiting their ability to search for such evidence. Externally prepared literature scans that share evidence-based practices from peer-reviewed research will be critical to supporting evidence-informed decisions all districts will face. Both the Council of the Great City Schools (CGCS) and the National Network of Education Research-Practice Partnerships (NNERPP) are uniquely positioned to fulfill these needs.

Working collaboratively CGCS and NNERPP plan to leverage existing networks of researchers working in research-practice partnerships (RPPs) across the U.S. to produce 2-4 page literature scans on high-need topics identified by district research leaders currently participating in learning communities facilitated by CGCS and NNERPP. The scans will focus on recommendations to support students who are experiencing differential access to technology, and thus, exaggerating existing inequities. Moreover, we expect the scans to help district leaders translate theoretical research into practical, outside-of-the-box applications for traditionally marginalized students that will support the interruption of inequitable opportunities and potential injustices these students face. Because we anticipate these needs to evolve as districts re-open, we propose distributing the production of the scans over a 12-month period to follow the contours of challenges as they arise. Given these considerations, we expect to produce up to 17 literature scans during this 12-month period.

In particular, we plan to focus on two key deliverables:

1. Rapid turnaround literature scans: Using a previously shared literature scan from one of our colleagues at the New York City Department of Education as a key guide, we will produce up to 17 literature scans that will respond to critical evidence needs identified by our district research leader contacts. We anticipate the scans to be of direct utility to time-sensitive decisions that would benefit from evidence.

2. Engagement with district research leaders: Once the scans are completed, we will share them with district research leaders that are members of either CGCS or NNERPP to help support their engagement with the scans. For example, we plan to host learning webinars with these groups so that they may ask questions of each other, ponder the evidence collectively, and identify remaining gaps in their knowledge that may inform future scans.

RAND Corporation and CGCS American School District Panel (ASDP)

The Council has partnered with the Rand Corporation to provide leaders with an opportunity to share their perspectives and contribute to decisions about education policy

and practice. The research team will survey leaders and staff from a representative panel of school districts across the country as well as conduct a complementary set of qualitative studies, following these districts over time to monitor trends.

The surveys will explore a range of district functions, such as curriculum and instruction, professional development supports, and services for students with disabilities. The research will examine district strategy, structure, policy, and practice, and will provide insight into how districts are changing to support school-level problem-solving.

Improving the Lowest Performing Schools under ESSA

The Council research team is conducting a qualitative study of the efforts across 10 Council member districts regarding their programs and practices designed to improve the lowest performing schools in member districts. This data collection effort is designed to catalog the reform efforts in these schools and potentially link any changes in academic outcomes to new reform efforts. The Council research team plans to interview key individuals across districts over a two-year period to identify plans for the lowest performing schools and to ascertain what aspects of the plans resulted in improved practices and student achievement.

Supporting Educators to Align Balanced Literacy Approaches to College and Career Ready Expectations

The Council research team serves as the lead evaluator, in partnership with Student Achievement Partners (SAP) on a Kellogg Foundation grant to improve early literacy achievement in the San Antonio Independent School District (SAISD). The intent of the SAP and CGCS early literacy acceleration work is to significantly improve early reading outcomes for students across the country, particularly those who are of color, living in poverty, and/or English Language Learners. The teams are currently in the first year of the grant, and we have collaborated with SAISD in the development of research questions, goals, and expected outcomes for the project. This evaluation has been suspended due to the coronavirus pandemic.

Online Course on Leading Research Departments in the Great City Schools

The Council research team is currently developing an online course to help research directors meet the evolving demands of research departments in large urban settings. As the roles and responsibilities of research departments change over time, research directors have voiced a need to provide an overview of the various functions of the director of research position. While the functions of a department can vary depending on the organizational structure of a specific school district, there are many common functions that a research department may be asked to fulfill. The aim of this course is to provide a foundation of best practices for leading a research department based on years of working

with big city research departments and feedback from research directors across the Council's membership.

The course, titled "Leading Research Departments in the Great City Schools," will be tailored to directors of research who are either new to the position or veterans who have assumed new responsibilities in their district. The course will focus on five key areas: Leadership, Organizational Structure, and Data Governance; Research and Evaluation; Assessment; Accountability; and Student Information Systems. The Council's research team will create a pilot course and will seek the input of research directors in the coming months to build on the content of the course.

TUDA

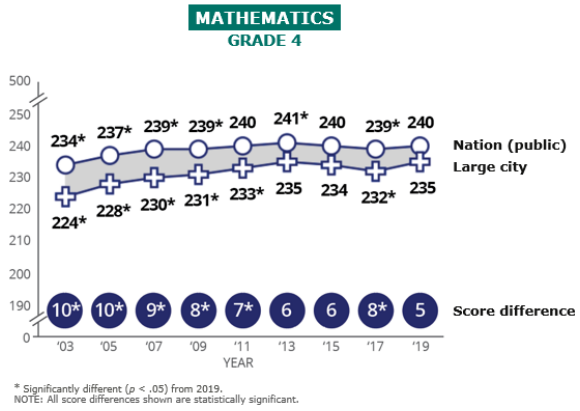
PRE-CORONAVIRUS, BIG GAINS FOR LARGE CITIES AND URBAN DISTRICTS ON NAEP SINCE THE EARLY 2000s

Since the early 2000s, the NAEP program has partnered with some of the largest urban districts in the nation to administer a district-level NAEP assessment, or the Trial Urban District Assessment (TUDA). The most recent results, from 2019, date from before the schooling disruptions caused by the coronavirus pandemic. It is too soon to say the kind of impact the current learning disruptions will have on learning- or how that may look different for students who attend schools in large cities. But, as a baseline, we can review progress they have made since the early 2000's. All results discussed were produced by the online data analysis tool, the NAEP Data Explorer (NDE).

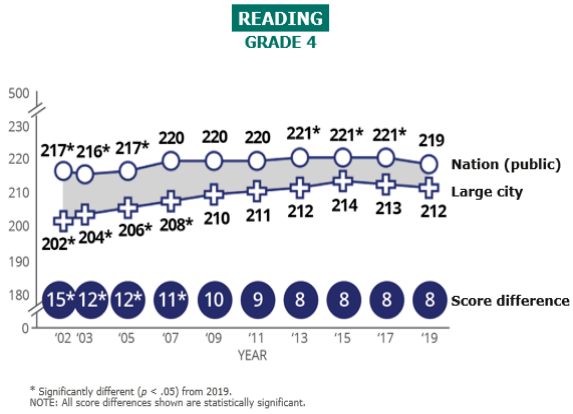
More Gains Overall in Large Cities Than for the Nation

Over the period from the early 2000s to 2019, schools in large cities made significantly greater gains in grades 4 and 8 mathematics and reading than the nation overall. While the nation had 5- to 6-point gains in grades 4 and 8 mathematics, large city schools raised their average scores by 11- to 12-points. In reading, the nation saw 2-point gains in grades 4 and 8 while large cities made gains as large as 10-points. These gains in large cities have reduced score differences with the nation by half or close to half (see figure 1).

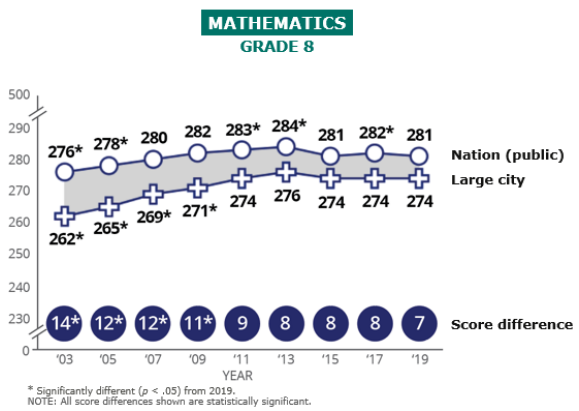
Figure 1



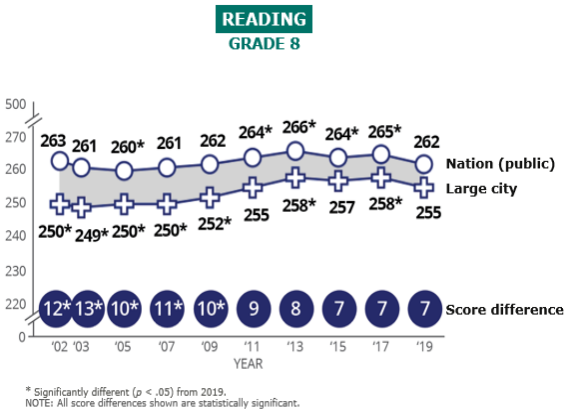
The nation's standard deviation ranged from 28 to 32 between 2003 and 2019. (Or we could use the average = 30.)



The nation's standard deviation ranged from 35 to 39 between 2002 and 2019. (Or we could use the average = 37.)



The nation's standard deviation ranged from 36 to 40 between 2003 and 2019. (Or we could use the average = 37.)



The nation's standard deviation ranged from 34 to 38 between 2002 and 2019. (Or we could use the average = 35.)

Most TUDAs Made Larger Gains Than the Nation

Across mathematics and reading at grades 4 and 8, most districts with TUDA data reaching back to the early 2000s[1] had score gains that were significantly larger than the nation's gains (see figure 2). These districts are:[2]

- Atlanta made gains of 16 points in mathematics grade 4, 24 points in mathematics grade 8, 18 points in reading 4, and 19 points in reading 8;
- District of Columbia (DCPS) made gains of 30 points in mathematics grade 4, 26 points in mathematics grade 8, 24 points in reading 4, and 11 points in reading 8;
- Chicago made gains of 18 points in mathematics grade 4, 21 points in mathematics grade 8, and 15 points in reading 4;
- Los Angeles made gains of 16 points in mathematics grade 8, 14 points in reading 4, and 11 points in reading 8;
- Boston made gains of 14 points in mathematics grade 4 and 17 points in mathematics grade 8;
- San Diego made gains of 14 points in mathematics grade 4 and 18 points in mathematics grade 8;
- Charlotte made gains of 9 points in mathematics grade 8; and
- Houston made gains of 10 points in mathematics grade 8.

[1] The first TUDA *mathematics* assessment was administered in 2003, and ten districts participated in the assessment in 2003: Atlanta, Boston, Charlotte, Chicago, Cleveland, DCPS, Houston, Los Angeles, New York City, and San Diego. The first TUDA *reading* assessment was administered in 2002, and five districts participated in the assessment in 2002: Atlanta, Chicago, DCPS, Houston, and Los Angeles.

[2] The ordering of the TUDAs is based on the number of score gains significantly larger than the national public's across the four subjects/grades (in descending order). For TUDAs with the same number of score gains across the subjects/grades, alphabetic order is used.

Figure 2

Mathematics 4 2003-2019 Score Change		Mathematics 8 2003-2019 Score Change		Reading 4 2002-2019 Score Change		Reading 8 2002-2019 Score Change	
Jurisdiction	**Score Change**	**Jurisdiction**	**Score Change**	**Jurisdiction**	**Score Change**	**Jurisdiction**	**Score Change**
DCPS	↑30*	**DCPS**	↑26*	**DCPS**	↑24*	**Atlanta**	↑19*
Chicago	↑18*	**Atlanta**	↑24*	**Atlanta**	↑18*	**Los Angeles**	↑11*
Atlanta	↑16*	**Chicago**	↑21*	**Chicago**	↑15*	**DCPS**	↑11*
San Diego	↑14*	**San Diego**	↑18*	**Los Angeles**	↑14*	**Large City**	↑4*
Boston	↑14*	**Boston**	↑17*	**Large City**	↑10*	**Chicago**	↔4
Large City	↑11*	**Los Angeles**	↑16*	**New York City**	↔6	**Houston**	↔1
Houston	↑8	**Large City**	↑12*	National Public	↑3	National Public	↔-1
Los Angeles	↑8	**Houston**	↑10*	**Houston**	↔-2		
National Public	↑6	**Charlotte**	↑9*				
New York City	↑4	**New York City**	↑7				
Charlotte	↑4	National Public	↑5				
Cleveland	↔3	**Cleveland**	↔#*				

↑ 2019 score is significantly higher ($p < .05$) than 2003/2002 score.
↔ 2019 score is not significantly different from 2003/2002 score.
↓ 2019 score is significantly lower ($p < .05$) than 2003/2002 score.

Rounds to zero.
* Average score change for the district or large city is significantly different ($p < .05$) than the average score change for National Public.

Summary

While students across the nation continued to outscore students in large cities in NAEP reading and mathematics as of 2019, students in large cities overall and several TUDA districts have made exceptional progress since the early 2000s in closing the achievement gap with the nation. Gains made by students in large cities have been greater than the gains seen throughout the nation as a whole. Most of the TUDA districts that were a part of the first administrations in the early 2000's scored significantly higher than the nation over this same time period, with Atlanta and DCPS having scoring gains that were significantly higher than the nation in all four subject-grade combinations.

While we anticipate school disruptions related to the pandemic may influence the next rounds of NAEP results for urban schools, we cannot speculate how that may affect the long-term improvement we've observed in students in large city districts. As more data become available, NCES will continue to analyze and report on how NAEP data can be a tool in understanding how our 4th and 8th graders around the country are learning.

NAGB

Remarks to the National Assessment Governing Board
By
Michael Casserly, Executive Director
Council of the Great City Schools

May 29, 2020

- Thank you. I am Michael Casserly, Executive Director of the Council of the Great City Schools. It is nice to be with you again.

- I will be brief since I know there may be questions or comments.

- It has been 20 years since I came before this group at the old Washington Hotel in the fall of 2000 to propose that big city schools be allowed to be oversampled in order to yield district-level scores.

- After we proposed what became known as the Trial Urban District Assessment of NAEP (TUDA), we went to Capitol Hill to change the law and to garner the appropriations to fund the initiative. We also rounded up the first cohort of volunteers to participate.

- We did all this for three reasons--

  --We wanted to know if we as the nation's largest urban school systems were making progress
  --We wanted to compare ourselves to similar districts across state lines, which other assessment systems did not allow
  --And, we wanted a way to figure out what was working in our districts and what was not.

- Over the years, we have been able to grow the program to serve 27 big-city school districts—with others waiting in line.

- The data from TUDA has been enormously beneficial to us. I know that the press uses the results mostly for ranking purposes, but we use it to guide our reforms and gauge our strategies.

- We have also used our results to help determine why and how some of our big city school districts make more progress than others.

- We appreciate the challenge facing this board about whether and how to conduct the 2021 NAEP assessment. We face multiple challenges on this front as well.

- In preparation for this meeting, we were asked to poll our 27 TUDA-participating districts to determine their thoughts about moving forward with next Spring's testing.

- We asked our superintendents about their preferences among five options—

- One, we asked, would you like for your district to participate in TUDA in 2021 in the same way as in 2019—that is, a full test administration that would yield a districtwide estimate in reading and math in fourth and eighth grades and all disaggregated results that you typically see when you take the exam?

- Two, would you like your district to participate in TUDA in 2021 but with a reduced sample that would yield a districtwide estimate of reading and math in fourth and eighth grades but would not yield disaggregated results for poor students, ELLs, students with disabilities, and others?

- Three, Would you like your district to participate solely in the regular state sample in 2021, which would yield a national estimate of reading and math in fourth and eighth grades for the large city schools—but would not yield a district-specific estimate or any disaggregated results for your district?

- Four, would you like to postpone all NAEP testing for states and TUDAs until 2022?

- And five, do you have another idea?

- We received responses from all 27 TUDA districts.

- The tally of responses indicated that 21 of the 27 superintendents—78 percent—preferred option #4—postpone the 2021 test until 2022.

- Six of the 27—or 22 percent—preferred option #1—move forward in 2021 as usual.

- No one preferred options two or three—that is, reduced samples that yielded no disaggregated results. And no one had a better idea.

- In general, the TUDA superintendents saw little value in taking NAEP if it did not yield the disaggregated data that they value.

- The preferences were based on several concerns.

- First, we are not sure when or if our schools will be open, given the vagaries of the virus. In some cases, we may not have control over whether we are open or closed, since governors and mayors sometimes make those decisions for us.

- Two, most of our city school systems will be offering a mix of in-person, remote, and hybrid instruction. The same students in grades four and eight will not necessarily be in their schools and classrooms on any given day. No place is likely to look the same from day-to-day. To test the entire sample, NAEP teams might have to be in a school for two to three days in a row—and still they would not be able to test students who choose to remain home.

- Three, surveys that our districts are doing suggest that 25 to 35 percent of parents do not plan to send their children back to face-to-face instruction anytime soon. We do not think that the demographics of students choosing to stay home will necessarily match those participating in face-to-face or hybrid instruction, which will affect the nature and quality of the NCES sample.

- Fourth, some of our districts are having to move students to other schools with lower enrollments or were closed in order to create more social distancing. Which students are where may not match what NCES is expecting to sample. And these configurations may change over the school year. Furthermore, test administrators may find that some sampled schools don't have rooms large enough to accommodate test takers with social distancing or that those rooms are being used to handle students who are being spread out through the building.

- Fifth, most of our school districts will be restricting outside visitors, vendors, contractors, community members and others to reduce virus exposure to students. These restrictions are likely to be different from school-to-school and district-to-district and may involve health-screening procedures that are unlike federal guidelines that contractors may be using. In addition, the NCES vendors will have to clean every device any time it changes hands from one student to another.

- Sixth, NAEP test administrators often move from school to school in the same district, presenting the possibility that they would spread the virus into multiple schools—presenting both health and liability risks for NAGB.

- Seven, if anyone in a school becomes sick, then that school is likely to be shut down at least temporarily with little to no notice to NAEP test administrators.

- Eight, adding voluntary exams would likely extend the testing window from March into April or May, when state summative exams—if they are given—are being administered. This would create a level of burden on the schools—on top of everything else—that superintendents think is intolerable.

- Nine, NAGB and NCES could not spend enough on their testing vendors to mitigate these and other problems. You could throw an extraordinary amount of money at these problems and still not have a sample that yielded usable data.

- And finally, moving forward with a national assessment in January and February, possibly during the peak months of the traditional flu season and possible spikes in coronavirus outbreaks, could make the Governing Board and NCES appear tone deaf to the needs of students, teachers, and schools across the country. The message that you risk sending is that you value testing over the health and well-being of students and staff.

- In short, we think there are significant risks to the reliability and validity of the sample because student counts will not be stable this upcoming school year; there are risks to the quality of the data it yields; and there are risks to the reputation of NAGB and NAEP if the predictable comes to pass. Somewhere between a quarter and a third of the sample may not be available, and it is highly questionable about whether parents who have kept their children home will send them back into school for a test that does not count.

- Even the six of the 27 TUDA districts that wanted NAEP to proceed indicated that these were major risks.

- We would very much like the data that NAEP could yield about the extent of learning loss from this pandemic, but even in the best of circumstances, we would not be getting results at least until late fall 2021—too late to do much about them, even if they were valid and complete.

- For these reasons, we respectfully request a postponement in the 2021 NAEP testing like has already been done with the international assessments.

Thank you.

NAEP DISTRICT EFFECTS STUDY

MIRRORS OR WINDOWS:

How Well Do Large City Public Schools Overcome the Effects of Poverty and Other Barriers?

About the Council of the Great City Schools

The Council of the Great City Schools is the only national organization exclusively representing the needs of urban public-school districts. Composed of 76 large city school districts, its mission is to promote the cause of urban schools and to advocate for inner-city students through legislation, research, instructional support, leadership, management, technical assistance, and media relations. The organization also provides a network for school districts sharing common problems to exchange information and to collectively address new challenges as they emerge to deliver the best education for urban youth.

Mirrors or Windows: How Well Do Large City Public Schools Overcome the Effects of Poverty and Other Barriers?

2020

Table of Contents

Mirrors or Windows: How Well Do Large City Public Schools Overcome the Effects of Poverty and Other Barriers?

Introduction

One of the most consistent and long-standing relationships in social science research is the one between poverty and student academic performance. In nearly every case, the evidence demonstrates that student achievement declines as poverty rises. At least as far back as the Coleman report (1966), research has suggested that poor students do not do as well in school as students whose parents are better off financially and educationally. More recently, a study by Reardon (2016) showed similar results and concluded that the gap between high- and low-income students may have widened between the 1980s and the early 2000s.

At the same time, education has been depicted by countless politicians, philosophers, scientists, and advocates as the ticket out of poverty. Education is thought to be society's main engine for smoothing out its inequities. In fact, Horace Mann once stated, "Education then, beyond all other devices of human origin, is the great equalizer of the conditions of men, the balance-wheel of the social machinery." To be sure, schooling aspires to level the playing field for rich and poor alike. Immigrant and native born. Commoner and blue-blood. But is it?

It is not likely that these two themes are true at the same time. Either schools help overcome the effects of poverty and other barriers or they reflect those inequities. Either schools serve to perpetuate society's inequities, or they serve to overcome them. Either schools work to level the playing field or they keep opportunity at bay. As noted Chicago journalist Sydney Harris once asserted, "The whole purpose of education is to turn mirrors into windows."

Our question in this report is a straightforward one: Are urban public schools, which have the largest numbers and concentrations of poor students in the nation, mirrors or windows?

Do urban public schools overcome the effects of poverty and other barriers or do they simply reflect them? Do urban public schools do a better job at overcoming the effects of poverty on achievement than public schools generally? Do some urban public-school districts do a better job at overcoming these effects than other urban school districts? Who are they? Are urban school districts getting any better at overcoming these effects over time or are they producing the same results they have always produced? What is the difference between urban school districts that appear to be 'beating the odds' and those that are not progressing? What are these more effective urban school districts doing that other urban school districts are not doing? Finally, are there similarities among urban school systems that have not shown as much progress, and what are the lessons we might learn from them?

These are questions that are infrequently asked in the research or answered in a way that gives urban schools better guidance about what they need to be doing differently. Instead, most research is backward leaning in the sense that it helps explain why things in the past looked like they did. This study will lean forward, and it will attempt to show where to look for clues using differences in how school districts perform over time.

To conduct this analysis, the Council of the Great City Schools used data from the National Assessment of Educational Progress (NAEP) and looked at the effects of not only poverty but also language status, parental education, disability, literacy materials in the home, and race to answer many of the questions above. We predict statistically what results might be likely based on these variables, and we compare those predictions against actual results over six separate administrations of NAEP between 2009 and 2019.

In other words, we created a 'district effect' or 'value-added' measure using NAEP data to determine whether urban school districts are producing enough "educational torque" to overcome poverty and other long-standing effects to any degree and to ascertain how they are doing it. We also look at districts that are not making as much progress and discuss their commonalities. In these ways, we attempt to discern whether

public education, urban public education, in particular, is a force for upward social mobility or whether it simply reflects and perpetuates the inequities that society creates.

Demographics of Large City and Not Large City Schools

Members of the Council of the Great City Schools educate disproportionately large numbers of the nation's students facing barriers to their educational success. The 76 cities whose school districts are members of the Council are home to about 17.4 percent of the U. S. population (56,863,400 of 326,474,013 est.). Their school districts enrolled some 7.8 million students in 2016-17 or about 15 percent of the nation's public elementary and secondary school enrollment.

This report primarily looks at the educational performance of Large City schools using data from the National Assessment of Education Progress (NAEP)[1]. In general, the Council's membership comprises the bulk of the Large City variable in NAEP, a variable that we use extensively in this report. Reading and math performance on NAEP are controlled statistically for relevant background variables summarized earlier, i.e., race/ethnicity groups, national school lunch program (FRPL), Census poverty, students with disabilities (IEP), English language learners (ELL), literacy materials in the home, and parent education level for students in grade eight. Relevant background variables are defined in more detail in subsequent sections, but generally they were selected because previous research indicated that they consistently predict student outcomes.

Our analysis looks at two distinct, mutually exclusive, and not-overlapping types of schools We compare the results of NAEP test takers[2] in Large City schools with the results of test-takers not in Large City Schools. Students not in Large City Schools includes test takers in private schools and U. S. territories who take the NAEP assessment. Both categories include charter schools identified within the jurisdiction, but NAEP data on charter schools are not coded in a way that would allow one to determine which charters are governed by regular public-school districts and which ones are chartered and operated independently. Consequently, in this analysis, Large City and Not Large City schools include district-authorized charters, charters authorized by others, and independent charters.

We start the analysis by looking at the student demographic characteristics of Large City schools and Not Large City schools. One should keep in mind that the demographics of school types in the fourth grade are slightly different from demographics in the eighth grade. Exhibits 1 through 5 summarize critical demographic characteristics of the two types of schools reported in the NAEP data for Large City and Not Large City schools.

Data in Exhibit 1 shows that Large City schools had an aggregate enrollment in 2017 that was 22 percent African American, 46 percent Hispanic, and 20 percent white. The percent of African American students in Large City schools declined from 27 percent to 22 percent over the period, while Hispanic students increased from 43 percent to 46 percent. percent.)

Exhibit 1. Percentages of NAEP fourth grade math test takers by race and type of school, 2009 to 2017.

	% Black					% Hispanic					% White				
	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017
Large City Schools	27%	25%	24%	22%	22%	43%	45%	44%	47%	46%	21%	20%	22%	20%	20%
Not Large City Schools	14%	14%	14%	14%	13%	19%	20%	22%	22%	23%	61%	59%	58%	56%	54%

Source: NAEP Data Explorer (NDE) based on NAEP reported demographics for mathematics.

[1] Definition of Large City

[2] The analysis uses test-takers in math in both fourth and eighth grades rather than test takers in English language arts, because the numbers of ELA test-takers is likely to be more skewed by testing exclusions related to English proficiency or disability status.

By contrast, African American students made up about 13 percent of the fourth-grade enrollments of schools that were not in Large Cities in 2017. Hispanic students made up approximately 23 percent and white students made up about 54 percent. Between 2009 and 2017, schools outside the Large Cities became slightly more Hispanic and somewhat less white.

Exhibit 2. Percentages of NAEP fourth grade math test takers by FRPL-status, Language-status, and IEP status and type of school, 2009 to 2017.

	% FRPL					% ELLs					% IEPs				
	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017
Large City Schools	71%	74%	73%	74%	70%	20%	22%	20%	21%	21%	11%	11%	12%	13%	13%
Not Large City Schools	43%	48%	50%	51%	47%	8%	9%	9%	10%	9%	12%	12%	13%	14%	13%

Source: NAEP Data Explorer (NDE) based on NAEP reported demographics for mathematics.

The NAEP data also show that the percent of fourth-grade students in Large City schools who were free and reduced-price lunch eligible in 2017 was 70 percent, about the same level as in 2009. (Exhibit 2.) The percent of these students who were not in large cities was 47 percent in 2017, an uptick from 43 percent in 2009. In other words, the enrollment of FRPL students in 2017 was about 50 percent higher in Large Cities than in Not Large Cities.

In addition, NAEP data on fourth grade English Language Learners (ELLs) show that these students made up 21 percent of the enrollment in Large City schools in 2017, about the same as in 2009. The enrollment in Not Large Cities was about nine percent ELLs in 2009, approximately the same as in 2009.

Finally, NAEP data in 2017 showed fourth grade students with Individualized Education Plans (IEPs) comprised some 13 percent of the Large City school sample, the same as the Not Large City sample. Both school types showed slight increases in their proportions of students with IEPs over the study period, 2009 to 2017.

Eighth grade NAEP data showed similar patterns to those in the fourth grade. African American students made up approximately 21 percent of students in Large Cities and 12 percent of students in Not Large Cities. Both settings showed drops in the percent of African American students. In addition, Hispanic students made up approximately 45 percent of the enrollments in Large Cities, compared to 21 percent in Not Large Cities. The percent of Hispanic students in both settings increased between 2009 and 2017.

Finally, white students made up about 21 percent of the enrollments of Large City schools in 2017, compared with 58 percent in Not Large Cities. The proportion of white students in Not Large Cities declined appreciably between 2009 and 2017, while the percent in Large Cities remained about the same.

Exhibit 3. Percentages of NAEP eighth grade math test takers by race and type of school, 2009 to 2017.

	% Black					% Hispanic					% White				
	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017
Large City Schools	26%	25%	25%	25%	21%	42%	44%	43%	44%	45%	22%	21%	21%	21%	21%
Not Large City Schools	14%	14%	13%	13%	12%	17%	19%	20%	21%	21%	63%	60%	59%	58%	58%

Source: NAEP Data Explorer (NDE) based on NAEP reported demographics for mathematics.

At the eighth-grade level, the data also indicated that the portion of students who were FRPL-eligible was slightly lower than that at the fourth-grade level. (Exhibit 4.) About 65 percent of eighth graders in Large

Cities were FRPL eligible in 2017, as were 42 percent in Not Large Cities. In other words, eighth grade students in Large Cities were about 55 percent more likely to be poor than students in Not Large Cities.

Exhibit 4. Percentages of NAEP Eighth grade math test takers by FRPL-status, Language-status, and IEP status and type of school, 2009 to 2015.

	% FRPL					% ELLs					% IEPs				
	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017
Large City Schools	66%	69%	69%	71%	65%	12%	11%	11%	12%	12%	11%	11%	12%	13%	13%
Not Large City Schools	39%	44%	46%	48%	42%	5%	5%	4%	5%	5%	10%	10%	12%	12%	12%

Source: NAEP Data Explorer (NDE) based on NAEP reported demographics for mathematics.

In addition, the eighth-grade data indicate that the percentages of ELL students in Large City schools remained at the same level (12 percent) between 2009 and 2017. (Exhibit 4.) Only about five percent of students in Not Large Cities were ELLs in 2017, a level that was unchanged from 2009.

The percentage of eighth-grade students with IEPs in Large City schools in 2017 was 13 percent, the same level as among fourth graders, a level that showed some increase over 2009. (Exhibit 4.) The enrollments in Not Large Cities among eighth graders with IEPs was about 12 percent, an uptick from 2009.

Finally, we examined data on the education levels of parents of students in Large City and Not Large Cities. (Exhibit 5) The data on this NAEP background variable were available only on eighth graders, not fourth graders. The results of the analysis showed that the percent of Large City school parents who did not finish high school was about 10 percent in 2017, compared to approximately six percent among Not Large City parents. At the other end of the education scale, some 42 percent of Large City School parents were college graduates in 2017, compared with 57 percent among parents in Not Large Cities. In both settings, there were declines in the percentages of school parents who did not finish college and increases in the percentages of school parents who were college graduates.

Exhibit 5. Percentages of NAEP Eighth Grade Math Test-Takers Whose Parents Had Differing Levels of Educational Attainment, 2009 to 2017.[3]

	Did Not Finish High School					Graduated High School					Graduated College				
	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017	2009	2011	2013	2015	2017
Large City Schools	13%	12%	11%	12%	10%	17%	17%	17%	17%	17%	35%	37%	38%	38%	42%
Not Large City Schools	7%	7%	7%	7%	6%	17%	17%	16%	16%	14%	47%	49%	50%	50%	57%

Source: NAEP Data Explorer (NDE) based on NAEP reported demographics for mathematics.

In sum, the NAEP data indicate that the demographics of Large City schools and Not Large City schools were substantially different from one another. Large City schools tended to be more predominantly African American and Hispanic than Not Large City Schools. In addition, Large City schools were more likely to have higher enrollments of poor students and ELLs. Finally, Large City Schools tended to have larger percentages of parents who did not finish high school and lower percentages of parents who had not graduated from college than Not Large Cities. The percentages of students with IEPs were similar in both settings, although there may be differences in the types and severity of disabilities between the two types of schools.

[3] The variable is defined as "at least one parent."

Methodology

In 2010, the Council of the Great City Schools along with the American Institutes of Research analyzed the results of the National Assessment of Education Progress (NAEP) in a way that had not been done previously (Dogan, et al., 2011). The two prominent research questions of that study were:

1. How did urban districts participating in the Trial Urban District Assessment (TUDA) in 2009 compare to other districts when one controlled for relevant background variables?
2. How did urban districts participating in the Trial Urban District Assessment (TUDA) in 2009 perform, compared to their statistically expected performance based on relevant background variables?

To answer these questions, the study compared the performance of each district against other districts after adjusting for specified student background characteristics, i.e., race/ethnicity, special education status, English language learner status, eligibility for free- or reduced-price lunch under the National School Lunch Program, the highest level of education attained by either parent, and information on the availability of written materials and computers in a student's home. The analysis employed a methodology used elsewhere in the literature (e.g., Braun, Jenkins, and Grigg, 2006). A regression analysis was conducted to estimate the "expected" performance of an urban district against a national sample of other public-school students controlling for variations in these demographic characteristics.

Next, each district's actual performance was compared to the expected performance for that district. The difference between the two (actual vs. expected) was called a "district effect." Positive effects indicated that the district was performing better than expected statistically and negative effects indicated that the district was performing below what was expected statistically.

A similar methodology using NAEP restricted-use data from 2009, 2011, 2013, 2015, and 2017 was used in this report. Comparable student background variables were used to calculate "adjusted" NAEP scale scores in TUDA districts using HLM analysis and make comparisons between actual and statistically expected scores. This study compared the performance of each district against other districts after adjusting for specified student background characteristics, i.e., race/ethnicity, special education status, English language learner status, the highest level of education attained by either parent, and information on the availability of written materials and computers in a student's home. However, to control for poverty, differences in school-level free or reduced-price lunch rates and the percentage of families in the school's zip code were included the HLM analysis.

In 2010, Congress, as part of the Healthy, Hunger-Free Kids Act, authorized the Community Eligibility Provision (CEP) to allow schools and local educational agencies (LEAs) in low-income areas to provide free breakfast and lunch to all students. The CEP program was available to a small group of states in July 2014 and nationwide in 2014 (School Year 2014-15). Table 1 shows that TUDA districts began to apply different methodology for identifying and reporting free or reduced-price lunch eligibility for students in 2015. As a result, the research team noted that different decisions regarding school lunch eligibility for students inhibited the comparability of calculated expected scores for districts across years. Further, the team found that using the traditional NAEP free or reduced-price lunch indicator in 2013, 2015, and 2017 significantly influenced the direction and magnitude of school district adjusted scores. Consequently, the research team applied two school-level variables to the analysis and removed the traditional student-level free or reduced-price lunch variable from the analysis. The variables included:

Level 1 – Student Level Variables

- **Race/ethnicity**

  In the NAEP files, student race/ethnicity information is obtained from school records and classified according to six categories: *White, Black, Hispanic, Asian/Pacific Islander, American Indian/Alaska*

Native, or *unclassifiable*. When school-reported information was missing, student-reported data from the Student Background Questionnaire were used to establish student race/ethnicity. Using restricted NAEP data sets, we categorized as *unclassifiable* students whose race-ethnicity based on school-records was *unclassifiable* or *missing* and (1) who self-reported their race as *multicultural* but not *Hispanic* or (2) who did not self-report race information.

- **Special education status**

  Student has an Individualized Educational Program (IEP), for reasons other than being gifted or talented; or is a student with a Section 504 Plan.

- **English language learner status**

  Student is currently classified as an English language learner and is receiving services.

- **Parental Education**

  Highest level of education attained by either parent: *did not complete high school*, *graduated high school, had some education after high school*, or *graduated college*. This indicator is only available for grade 8 students.

- **Literacy Materials**

  The presence of reading materials in the home is associated with both socioeconomic status and student achievement. The measure reported in 2009 was based on questions in both grade 4 and grade 8 in the *Student Background Questionnaires,* which asked about the availability of computers, newspapers, magazines, and more than 25 books in the home. Between 2009 and 2015, the *Student Background Questionnaire* changed and a different combination of items was used to calculate a summary score of how many materials were present. In 2011, the items included the availability of computers, magazines, and more than 25 books in the home (newspapers were dropped as a survey item). In 2013, 2015, and 2017 the items included the availability of computers in the home, the availability of the internet, and more than 25 books in the home (magazines were dropped as a survey item). A summary score was created to indicate how many of these types of literacy materials were present in the home.[4]

Level 2 – School Level Variables

- **School free or reduced-price lunch eligibility rates**

  To level the influence of changing free or reduced-price lunch rates across districts, the research team chose to employ a school level, rather than a student level, school lunch indicator. Researchers did so by comparing the percentage of free or reduced-price lunch students reported in the National Center for Education Statistics Common Core of Data (CCD) files in the NAEP years prior to the CEP program and the NAEP reported free or reduced-price lunch percentages. When the values were within five percentage points of each other, researchers used the NAEP results for schools as the school level factor. However, for large discrepancies in the data (values well above or well below the 2012-13 school year), the CCD school lunch rate was used for the analysis.

- **School Zip Code Poverty Estimates – Percentage of Family Incomes Less Than $15,000 per year**

  As discussed later in this document, abject poverty has been shown to impair student academic outcomes. To further control for the influence of abject poverty across school districts, the research

[4] This summary score has been used for reporting NAEP background variables for several years and has been shown to be associated with students' achievement scores. (See for example, NAEP 1996 Mathematics Cross-State Data Compendium.)

team incorporated the percentage of families making less than $15,000 per year in a school's physical zip code as a school level poverty factor. The zip code data was taken from the U. S. Census Bureau American Community Survey estimates for each of the NAEP assessment years.

Exhibit 6. Free or Reduced-Price Lunch Methodology Used by TUDA Districts, 2015, 2017, 2019

TUDA District	NAEP 2015	NAEP 2017	NAEP 2019
Albuquerque	CEP-ALL	CEP-ALL	CEP-ALL
Atlanta	Direct-Only	CEP-ALL	CEP-ALL
Clark County	CEP-ALL	CEP-ALL	CEP-ALL
Cleveland	CEP-ALL	CEP-ALL	CEP-ALL
DC	CEP-ALL	CEP-ALL	CEP-ALL
Austin	Direct-Plus	Direct-Plus	Direct-Plus
Charlotte	CEP-ALL	CEP-Direct	Direct-Plus
Chicago	Direct-Plus	Direct-Plus	Direct-Plus
Dallas	Direct-Plus	Direct-Plus	Direct-Plus
Denver	N/A	Direct-Plus	Direct-Plus
Detroit	Direct-Plus	Direct-Plus	Direct-Plus
Fort Worth	N/A	Direct-Plus	Direct-Plus
Fresno	Direct-Plus	Direct-Plus	Direct-Plus
Guilford County	N/A	CEP-Direct	Direct-Plus
Houston	Direct-Plus	Direct-Plus	Direct-Plus
Jefferson County	Direct-Plus	Direct-Plus	Direct-Plus
Los Angeles	Direct-Plus	Direct-Plus	Direct-Plus
Milwaukee	N/A	Direct-Plus	Direct-Plus
New York City	Direct-Plus	Direct-Plus	Direct-Plus
Philadelphia	CEP-Direct	Direct-Plus	Direct-Plus
San Diego	Direct-Plus	Direct-Plus	Direct-Plus
Baltimore	N/A	Direct-Only	Direct-Only
Duval County	Direct-Only	Direct-Only	Direct-Only
Hillsborough County	Direct-Only	Direct-Only	Direct-Only
Miami-Dade County	Direct-Only	Direct-Only	Direct-Only
Shelby County	N/A	Direct-Only	Direct-Only (CEP & Non-CEP schools)
Boston	CEP-ALL	Direct-Only	Direct-Only (CEP & Non-CEP schools)

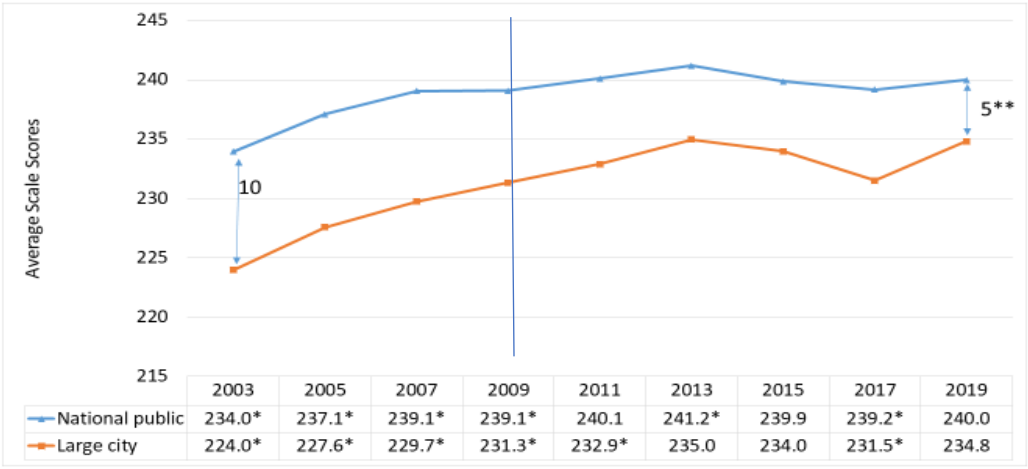
Source: National Center for Education Statistics, National Assessment of Educational Progress, 2019.

The reader should note that information on race/ethnicity, school lunch, and ELL and disability status come from the school and are available for all students. However, data on background characteristics for students who did not participate in NAEP are not available, i.e., excluded students or students who are not tested do not complete the *Background Questionnaire*. Therefore, data on *reading materials in the home* and *parent education* are only available for the tested populations. Consequently, the calculation of adjusted scores controlling for background characteristics was conducted on the reported sample only.

Analysis of National Assessment of Educational Progress Data

The Council of the Great City Schools initiated the Trial Urban District Assessment of NAEP in the fall of 2000. The initiative was piloted in 2002 in math. And in 2003, Large City schools participated in both reading and math assessments. The voluntary effort involves the over-sampling of students in each participating district to obtain a district-level estimate of reading and math performance in grades four and eight. Over the years, results on the assessment show that Large City schools have not only improved their performance but that they have improved faster than the nation at large, narrowing the gaps between the nation's urban schools and the national average. The remainder of our analysis begins with results from the 2009 testing. (See graphs below.)

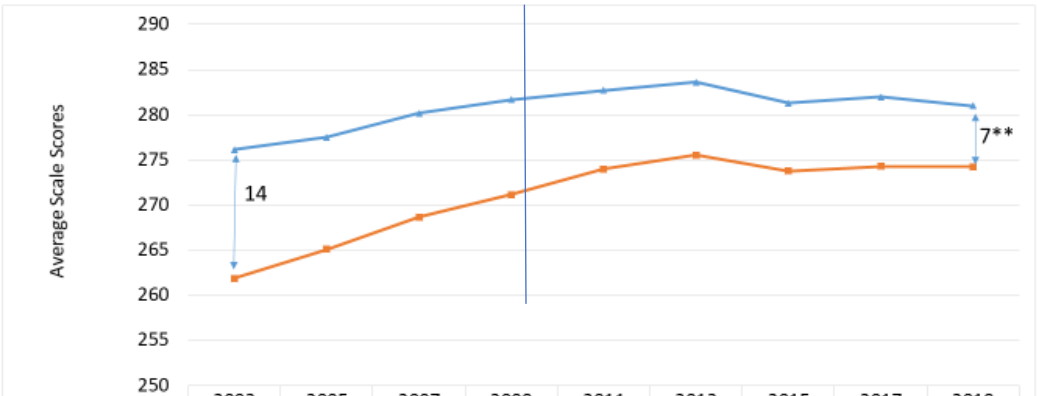
Average Scale Scores on NAEP Mathematics in Grade 4 for Public School Students, 2003-2019



	2003	2005	2007	2009	2011	2013	2015	2017	2019
National public	234.0*	237.1*	239.1*	239.1*	240.1	241.2*	239.9	239.2*	240.0
Large city	224.0*	227.6*	229.7*	231.3*	232.9*	235.0	234.0	231.5*	234.8

*Value is significantly different ($p < .05$) from the value for the same jurisdiction and student group in 2019.
**Gap is significantly different ($p < .05$) from the gap in 2003.

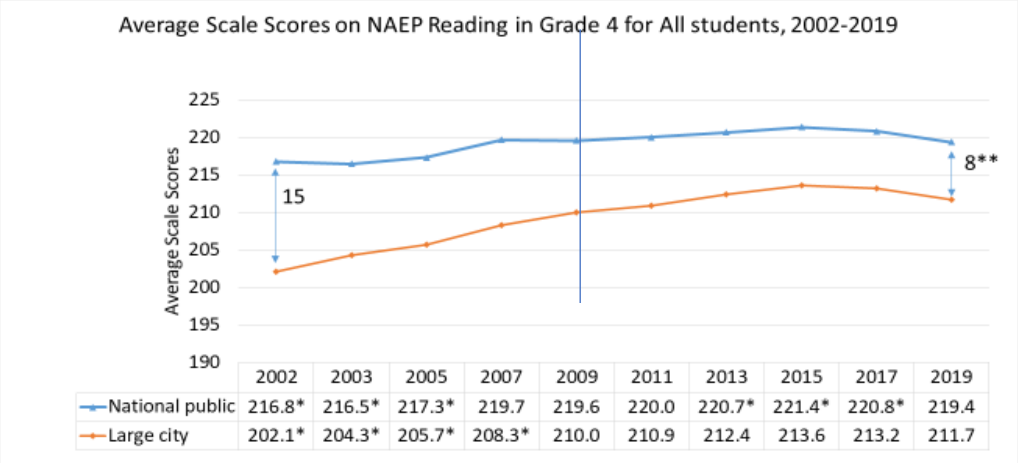
Average Scale Scores on NAEP Mathematics in Grade 8 for Public School Students, 2003-2019



	2003	2005	2007	2009	2011	2013	2015	2017	2019
National public	276.1*	277.5*	280.2	281.7	282.7*	283.6*	281.3	282.0*	281.0
Large city	261.9*	265.1*	268.7*	271.2*	274.0	275.5	273.8	274.3	274.2

*Value is significantly different ($p < .05$) from the value for the same jurisdiction and student group in 2019.
**Gap is significantly different ($p < .05$) from the gap in 2003.

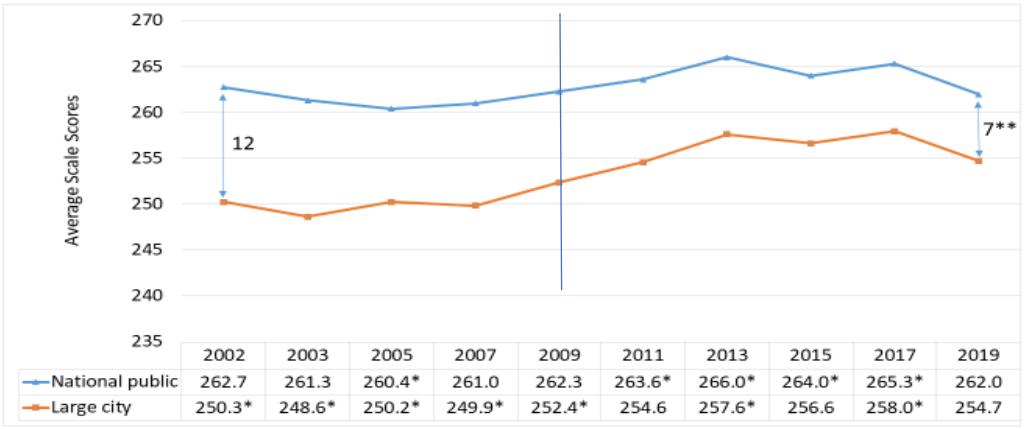
Average Scale Scores on NAEP Reading in Grade 4 for Public School Students, 2002-2019



	2002	2003	2005	2007	2009	2011	2013	2015	2017	2019
National public	216.8*	216.5*	217.3*	219.7	219.6	220.0	220.7*	221.4*	220.8*	219.4
Large city	202.1*	204.3*	205.7*	208.3*	210.0	210.9	212.4	213.6	213.2	211.7

*Value is significantly different ($p < .05$) from the value for the same jurisdiction and student group in 2019.
**Gap is significantly different ($p < .05$) from the gap in 2002.

Average Scale Scores on NAEP Reading in Grade 8 for Public School Students, 2002-2019



	2002	2003	2005	2007	2009	2011	2013	2015	2017	2019
National public	262.7	261.3	260.4*	261.0	262.3	263.6*	266.0*	264.0*	265.3*	262.0
Large city	250.3*	248.6*	250.2*	249.9*	252.4*	254.6	257.6*	256.6	258.0*	254.7

*Value is significantly different ($p < .05$) from the value for the same jurisdiction and student group in 2019.
**Gap is significantly different ($p < .05$) from the gap in 2002.

The analysis for this new study compared actual NAEP performance levels for Large City school districts and individual TUDA districts in 2009, 2011, 2013, 2015, and 2017 to predicted NAEP reading and mathematics performance (after controlling for the background variables outlined above) in grades four and eight. Comparisons were also made to Not Large City schools. The analysis allowed the Council to identify districts that were performing better than expected on the NAEP assessment and those who were performing under expectation. In other words, we could estimate over time whether Large City schools and others were getting better at mitigating the effects of poverty and other variables that typically suppress academic performance.

Exhibit 7 shows the actual performance for individual TUDA districts, Large City, and Not Large City schools, so the reader can see unadjusted results. Note that Albuquerque, Dallas, and Hillsborough County began participating in TUDA in 2011 and trends are reported on them for only four assessment cycles.

Duval County began participating in 2015 and the Milwaukee public schools did not participate in 2015. Clark County, Denver, Fort Worth, Guilford County, and Shelby County began in 2017 and only have calculations in one year.

Exhibit 7. Actual Scale Scores of TUDA Districts, Large City Schools, and Not Large City Schools, 2009 to 2017.

	2009				2011				2013				2015				2017			
	Math		Reading		Math		Reading		Math		Reading		Math		Reading		Math		Reading	
Jurisdiction	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8
Albuquerque					235	275	209	254	207	256	235	274	231	271	207	251	230	270	207	255
Atlanta	225	259	209	250	228	266	212	253	214	255	233	267	228	266	212	252	231	265	214	254
Austin	240	287	220	261	245	287	224	261	221	261	245	285	246	284	220	261	243	283	217	263
Baltimore	222	257	202	245	226	261	200	246	204	252	223	260	215	255	199	243	215	255	197	243
Boston	236	279	215	257	237	282	217	255	214	257	237	283	236	281	219	258	233	280	217	261
Charlotte	245	283	225	259	247	285	224	265	226	266	247	289	248	286	226	263	244	287	225	260
Chicago	222	264	202	249	224	270	203	253	206	253	231	269	232	275	213	257	232	276	211	259
Clark County																	230	272	213	258
Cleveland	213	256	194	242	216	256	193	240	190	239	216	253	219	254	197	240	214	257	196	237
Dallas					233	274	204	248	205	251	234	275	238	271	204	250	234	268	201	246
Denver																	229	272	214	258
Detroit	200	238	187	232	203	246	191	237	190	239	204	240	205	244	186	237	200	246	182	235
D.C. (DCPS)	220	251	203	240	222	255	201	237	206	245	229	260	232	258	214	245	231	262	213	246
Duval County													243	275	225	264	248	275	226	263
Fort Worth																	230	269	206	248
Fresno	219	258	197	240	218	256	194	238	196	245	220	260	218	257	199	242	221	255	203	244
Guilford County																	240	276	222	260
Hillsborough County					243	282	231	264	228	267	243	284	244	276	230	261	245	277	227	265
Houston	236	277	211	252	237	279	213	252	208	252	236	280	239	276	210	252	235	273	205	249
Jefferson County	233	271	219	259	235	274	223	260	221	261	234	273	236	272	222	261	233	271	221	261
Los Angeles	222	258	197	244	223	261	201	246	205	250	228	264	224	263	204	251	223	267	207	254
Miami	236	273	221	261	236	272	221	260	223	259	237	274	242	274	226	265	245	274	229	261
Milwaukee	220	251	196	241	220	254	195	238	199	242	221	257					216	254	195	245
New York City	237	273	217	252	234	272	216	254	216	256	236	274	231	275	214	258	229	275	214	258
Philadelphia	222	265	195	247	225	265	199	247	200	249	223	266	217	267	201	248	214	260	197	248
San Diego	236	280	213	254	239	278	215	256	218	260	241	277	233	280	216	262	237	283	222	264
Shelby County																	225	257	203	248
Large City Schools*	231	271	210	252	233	274	211	255	235	276	212	258	234	274	214	257	232	274	213	258
Not Large City Schools‡	241	284	221	264	242	284	222	265	242	285	222	268	241	283	223	265	241	284	222	267

* Includes district-authorized charters

‡ Includes charters authorized by others and independent charters

The raw data show that Large City schools generally scored below Not Large City schools by between nine and ten scale score points in 2017—depending on grade and subject. Individual TUDA school districts also showed extensive variation.

However, comparing these results without statistically controlling for background variables is only one way to look at these data. For instance, comparing Detroit and Charlotte-Mecklenburg on raw scores clearly indicates that one scores higher than another, but they have vastly different demographics and quite different challenges. To sort out these distinctions and how they might mask how districts perform and whether they improve, we asked a series of research questions—

- Are Large City schools performing at the same level as, above, or below statistical expectations in reading and math on NAEP in fourth and eighth grades after adjusting for differences in demographic characteristics? In other words, do urban public schools overcome—to any degree—the effects of poverty and other barriers or do they simply reflect those characteristics?
- Are Large City schools getting better at overcoming these effects over time (2009, 2011, 2013, 2015, and 2017)? Which school districts appear to be overcoming these effects the most?
- Do Large City schools do a better job at overcoming the effects of poverty and other variables on achievement than schools outside the cities?
- Do some urban public-school districts do a better job at overcoming these effects than other urban school districts? Who are they?
- Are there any fundamental differences between urban school districts that overcome these effects compared with ones who do not?
- What are the urban school districts that seem to be overcoming these affects doing that other urban school districts are not doing? Are there common features of urban school districts that are not showing progress yet?

To answer these questions, this study compared the performance of each district or type of school against other districts and school types after adjusting for student background and school characteristics. A HLM analysis estimated the performance of a district or type of school had its demographic profile been the same as the average profile of all districts or jurisdictions in the nation using the NAEP restricted data set for each of the study years. The methodology to estimate the adjusted mean scores is based on a two-level, students and schools, HLM model. In the mixed effects model:

Let y_{ijv} be plausible value[5] v of student j in district (or school type) i, and

X_{ijk} be the demographic characteristic k of student j in district (or school type) i.

Assume the mean plausible value for student j in district i, $y_{ij\bullet}$, can be expressed as a function of an overall mean achievement μ , a differential effect α_i associate with district (or school type) i, and differential effects β_k associate with characteristic k of student j in district or school type i:

$$y_{ij\bullet} = \mu + \alpha_i + \sum \beta_k X_{ijk} + e_{ij} \,, \qquad [1]$$

where μ is the overall mean,

[5] Plausible values are imputed values that resemble individual test scores and have approximately the same distribution as the latent trait being measured. Plausible values were developed as a computational approximation to obtain consistent estimates of population characteristics in assessment situations where individuals are administered too few items to allow precise estimates of their ability. Plausible values represent random draws from an empirically derived distribution of proficiency values that are conditional on the observed values of the assessment items and the background variables. The random draws from the distribution represent values from the distribution of scale scores for all adults in the population with similar characteristics and identical response patterns. These random draws or imputations are representative of the score distribution in the population of people who share the background characteristics of the individual with whom the plausible value is associated in the data.

α_i is the district (or school type) *i* effect, and

β_k is the effect of demographic characteristic *k* of student *j* in district (or school type) *i*.

Letting the subscript • indicate average, then the average scale score in district (or school type) *i* is expressed as

$$y_{i\bullet\bullet} = \mu + \alpha_i + \sum \beta_k \ X_{i\bullet k} + e'_i, \qquad [2]$$

Subtracting [2] from [1] we can estimate the analysis [3]

$$z_{ij} = y_{ij\bullet} - y_{i\bullet\bullet} = \sum \beta_k [X_{ijk} - X_{i\bullet k}] + e''_{ij} \qquad [3]$$

and obtain estimates of β_k directly, without any contamination from α_i because α_i has been subtracted out before the analysis. With the estimates $\hat{\beta}_k$, we compute the average effect of the demographic characteristics of student *j* in district (or school type) *i*.

$$\hat{y}_{ij\bullet} = \sum \hat{\beta}_k [X_{ijk} - X_{\bullet\bullet k}] \qquad [4]$$

where $X_{\bullet\bullet k}$ is the overall mean of $X_{\bullet\bullet k}$.

The adjusted score, y'_{ijv} is estimated by subtracting $\hat{y}_{ij\bullet}$ from each y_{ijv}:

$$y'_{ijv} = y_{ijv} - \hat{y}_{ij\bullet} \qquad [5]$$

The adjusted score, $y'_{i\bullet\bullet}$ is the critical statistic for the analysis. It is an estimator for $\mu + \alpha_i$, and we can estimate its standard error by the usual NAEP procedures. Note that $\mu + \alpha_i$ is the overall mean plus the effect of district (or school type) *i*. It is what the mean of district (or school type) *i* would be if the mean of all demographics in district (or school type) *i* were the same as the overall mean.

The hierarchical model used in the current study calculates this statistic by applying two Level 2 random factors and four mixed Level 1 factors. In the HLM model, rather than treating each student as varying from the overall mean plausible value, we estimate the mean of all student means for each school, noted below as $\gamma_{00.}$ The full HLM model is represented by:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{1j} + \ldots + \beta_{2j}X_{2j} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}(SCHOOLLN_j) + \gamma_{02}(LESS15K_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(SCHOOLLN_j) + \gamma_{12}(LESS15K_j) + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}(SCHOOLLN_j) + \gamma_{22}(LESS15K_j) + u_{2j}$$

Where SCHOOLLN is the school free or reduced-price lunch rate and LESS15K is the percentage of families in the school zip code with a household income less than $15,000 per year.

Next, the expected performance of each district and school type—based on the selected student background characteristics—was computed. Each district's actual performance was then compared to the expected performance for that district or comparison group. The difference between the two was called a "district effect" or group effect. Significant positive effects indicated that a district or group was performing better than expected statistically, and significant negative effects indicated that the district or group was performing worse than expected statistically. The actual model for the analysis is:

Level-1 Model

$$MRPCM1_{ij} = \beta_{0j} + \beta_{1j}*(LITERACY_{ij}) + \beta_{2j}*(IEPNOT_{ij}) + \beta_{3j}*(LEPNOT_{ij}) + \beta_{4j}*(SRACEAA_{ij}) + \beta_{5j}*(SRACEH_{ij}) + \beta_{6j}*(SRACEASP_{ij}) + \beta_{7j}*(SRACEAI_{ij}) + \beta_{8j}*(SRACEMR_{ij}) + e_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(SCHOOLLN_j) + \gamma_{02}*(LESS15K_j) + u_{0j}$$

Next, the expected performance of each district and school type—based on the selected student background characteristics—was computed. Each district's actual performance was then compared to the expected performance for that district or comparison group. The difference between the two was called a "district effect" or group effect. Significant positive effects indicated that a district or group was performing better than expected statistically, and significant negative effects indicated that the district or group was performing below what was expected statistically.

Variance Accounted for by the HLM Analysis

Exhibit 8 estimates the variance, or the R-squared value, explained by the background variables for each of the HLMs calculated on the *national* sample. The variances in the national sample ranged from a low of 0.2966 to a high of 0.3838. A recent presentation by Ward, Broer, and Jewsbury (2017) estimated explained variance at about 0.306 when using similar background variables. Their R-squared values were consistent with the values reported in this study.

Exhibit 8. Percent of variance (R^2) Explained by Relevant Background Variables for the Total NAEP Sample of Students (Public and Non-public) by Subject and Grade, 2009 to 2017

R^2 Values for All Students in NAEP Sample by Grade and Subject				
	Math		Reading	
Year	Grade 4	Grade 8	Grade 4	Grade 8
2009	0.2966	0.3530	0.3031	0.3471
2011	0.3198	0.3607	0.3390	0.3498
2013	0.3457	0.3733	0.3802	0.3712
2015	0.3367	0.3838	0.3777	0.3671
2017	0.3391	0.3754	0.3557	0.3552
Δ	+0.0425	+0.0224	+0.0526	+0.0081

‡ Includes district-authorized charters, charters authorized by others, and independent charters

In addition to the significance of these variables in explaining overall NAEP results, the analysis suggests that the power of these variables in predicting results has increased over time. In each subject-grade combination, the R-squared value increases somewhat between 2009 and 2017.

Limitations of this and similar analysis

Several limitations in the current study—and other similar studies—should be mentioned. First, both the adjusted and expected performance numbers are estimates based on variables that research indicates affect student achievement. Some of these variables are beyond the control of educators and policy-makers even though they affect performance. Still, the purpose of this study was to determine the extent to which Large City schools were overcoming their effects.

Second, there may be other variables related to achievement that were not controlled for in this analysis. Some of these variables are not measured in NAEP, and possibly some are not measurable at all. A district effect is the product of our best estimate of whether a district or school type was performing differently than expected given its student profile on a limited number of variables measured in NAEP. We did not look at other background variables like spending levels in part because previous studies have not shown them to be as powerful in predicting performance as the ones we did choose. Still, there is room for additional analyses on such variables.

Third, comparing school types at any grade level ignores the fact that students may enter the formal educational process at very different achievement levels. Consequently, attempts to control for differences using various student characteristics or attempts to match students based on background variables will not always account for other differences that affect student achievement. For example, parents who enroll their children in Large City or Not Large City schools may have very different parenting practices. Research (e.g., Wilder, 2014; Jeynes, 2012; Hill & Tyson, 2009; Patall, Cooper, & Robinson, 2008; Senechal & Young, 2008; Jeynes, 2007; Erion, 2006; Jeynes, 2005; Jeynes, 2003; Fan & Chen, 2001) indicates that differences in parental involvement and expectations have a significant impact on student achievement, yet many studies, including this one, do not adequately account for these differences except to the extent that we look at parental education levels and literacy materials in the home.

Fourth, this study was not able to parse the differences between charter schools that were authorized by school districts, those that were authorized by other entities, and those that were entirely independent. NAEP does not code charter schools in a way that would allow analyses of this type.

Fifth, this analysis does not control for differences in such in-school variables as teacher experience or school size. Other studies have shown that these variables show little impact on difference between school types (see, e.g., Braun, Jenkins, & Grigg, 2006), although these variables may have effects in other types of analyses.

Finally, differences in concentrations of poverty are likely to affect comparisons as well. (See, for example, Orfield & Lee, 2005 for a discussion of concentrated poverty). This study attempts to explain some of this effect by looking at income levels within jurisdictions with Census data, but additional analyses are needed.

Results of Analysis

This section answers study questions posed in the previous section. First, we look at "district effects" using the 2017 restricted-use NAEP data set. Second, we look at trends city-by-city and across cities using NAEP restricted-use data from 2009, 2011, 2013, 2015, and 2017. Third, we more carefully examine the poverty levels in cities whose school districts show district effects above and below what might be expected statistically. Fourth, we compare the performance of large city school districts to others.

(a) Actual vs. Expected (Adjusted) Mean NAEP Performance

Exhibits nine through 12 show the actual mean scale scores of districts and school types in 2017, the expected mean of the same groups after adjusting for relevant background variables, and the overall "district effect" of individual cities and various school types. Comparable tables for 2009, 2011, 2013, and 2015 are available in Appendix A. Again, the district effect is the difference between the actual performance and the adjusted performance. A positive effect suggests that the entity is scoring higher than one would expect statistically given its demographic characteristics; a negative effect suggests that the entity is scoring lower than one would expect statistically given its demographic characteristics. Zero is the point at which an entity scores exactly what one would expect statistically—suggesting that the entity is more likely to reflect its demographic characteristics than overcome them.

In grade four reading (Exhibit 9), many individual TUDA districts—the focus of this study—nominally out-scored their expected performance in 2017 after adjusting for relevant background variables. Individual

city effects ranged from a high of +18.51 in Denver to a low of -13.20 in Detroit. Overall, 17 of 27 cities (Atlanta, Austin, Boston, Charlotte-Mecklenburg, Chicago, Dallas, Denver, the District of Columbia, Duval County, Fort Worth, Guilford County, Hillsborough County, Houston, Jefferson County, Miami-Dade County, New York City, and San Diego) on which there were NAEP data on grade four reading in 2017 had positive district effects; and 10 of 27 had nominally negative district effects.

Large City Schools had an aggregate positive effect of +2.25, compared to an aggregate effect of Not Large Cities of +0.90.

In grade 8 reading (Exhibit 10), Large City schools had a district effect of +1.32 and individual cities varied. Individual cities ranged from +11.90 in Boston to a low of -8.11 in Fresno. Overall, 10 of 26 cities (Atlanta, Austin, Boston, Chicago, Dallas, Duval County, Hillsborough County, Miami-Dade County, New York City, and San Diego) on which there were NAEP data in 2017 had positive district effects; and 16 of 26 had negative district effects. Not Large City schools had an aggregate district effect of +0.44 in 2017.

Exhibit 11 shows that Large City schools had an effect of +1.57 in fourth grade math and individual cities showed considerable variation. For instance, cities ranged from a high of +13.06 in Dallas to a low of -13.41 in Detroit. Overall, 16 of 27 cities (Atlanta, Austin, Boston, Charlotte-Mecklenburg, Chicago, Dallas, Denver, the District of Columbia, Duval County, Fort Worth, Guilford County, Hillsborough County, Houston, Miami-Dade County, and San Diego) posted positive effects; and 11 had negative effects. Not Large City Schools had an aggregate effect in 2017 of +1.83. (See subsequent discussion of adjustments to fourth and eighth grade math scores based on effects of college and career-ready standards.)

Exhibit 12 shows that Large City schools overall had a positive effect, +3.61, in eighth grade mathematics, while individual cities varied from a high of +17.31 in Boston to a low of -10.60 in Fresno. Some 15 of 26 cities (Atlanta, Austin, Boston, Charlotte-Mecklenburg, Chicago, Cleveland, Dallas, Duval County, Fort Worth, Guilford County, Hillsborough County, Houston, Miami-Dade County, New York City, and San Diego) on which there were NAEP data in 2017 had positive district effects; and 11 of 26 had negative effects. The aggregate Not Large City effect in 2017 was +2.19.

Exhibit 9. Grade Four Reading Actual Performance, Expected Performance, and District Effects in 2017

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	206.83	212.06	-5.23
Atlanta	213.96	210.27	3.70
Austin	216.74	211.03	5.71
Baltimore	197.37	203.54	-6.17
Boston	217.15	203.58	13.57
Charlotte	224.89	218.87	6.01
Chicago	211.26	208.11	3.15
Clark County	213.38	214.25	-0.87
Cleveland	196.41	197.30	-0.88
Dallas	201.10	200.67	0.43
Denver	213.93	195.42	18.51
Detroit	181.52	194.73	-13.20
District of Columbia (DCPS)	213.00	210.45	2.55
Duval County	225.62	218.37	7.25
Fort Worth	205.91	202.27	3.65
Fresno	202.96	205.75	-2.79
Guilford County	222.03	216.84	5.19
Hillsborough County	227.23	217.98	9.25
Houston	205.31	204.44	0.87
Jefferson County	220.88	217.76	3.12
Los Angeles	207.50	210.53	-3.03
Miami	228.92	214.58	14.34
Milwaukee	195.23	203.25	-8.02
New York City	214.38	211.05	3.33
Philadelphia	197.33	205.52	-8.19
San Diego	221.69	213.96	7.73
Shelby County	203.14	203.29	-0.15
Large City Schools*	213.24	210.99	2.25
Not Large City Schools□	222.42	221.52	0.90

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit 10. Grade Eight Reading Actual Performance, Expected Performance, and District Effects in 2017

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	255.17	257.43	-2.26
Atlanta	254.29	252.14	2.14
Austin	262.97	259.27	3.70
Baltimore	242.73	247.70	-4.97
Boston	261.87	249.97	11.90
Charlotte	260.64	262.56	-1.92
Chicago	258.93	254.32	4.61
Clark County	258.54	258.77	-0.23
Cleveland	237.76	241.88	-4.12
Dallas	246.47	239.75	6.72
Denver	--	--	--
Detroit	235.85	239.75	-3.90
District of Columbia (DCPS)	246.73	252.29	-5.56
Duval County	263.29	261.78	1.51
Fort Worth	248.59	248.67	-0.08
Fresno	244.60	252.71	-8.11
Guilford County	259.89	261.31	-1.43
Hillsborough County	265.16	261.70	3.47
Houston	249.60	252.02	-2.42
Jefferson County	260.94	263.03	-2.09
Los Angeles	254.78	256.98	-2.20
Miami	261.26	257.26	4.00
Milwaukee	245.04	247.41	-2.38
New York City	259.24	256.04	3.19
Philadelphia	249.37	250.64	-1.27
San Diego	265.43	263.18	2.25
Shelby County	247.92	248.83	-0.91
Large City Schools*	257.97	252.14	1.32
Not Large City Schools‡	266.83	259.27	0.44

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit 11. Grade Four Mathematics Actual Performance, Expected Performance, and District Effects in 2017

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	229.90	232.79	-2.89
Atlanta	231.14	226.23	4.91
Austin	243.32	231.99	11.34
Baltimore	215.36	220.08	-4.71
Boston	233.33	224.48	8.85
Charlotte	243.87	235.78	8.09
Chicago	231.81	227.63	4.18
Clark County	230.13	233.50	-3.36
Cleveland	214.37	215.22	-0.85
Dallas	233.77	220.71	13.06
Denver	228.76	217.84	10.92
Detroit	199.89	213.30	-13.41
District of Columbia (DCPS)	230.80	227.77	3.02
Duval County	247.50	234.78	12.72
Fort Worth	230.47	223.58	6.89
Fresno	221.42	224.97	-3.55
Guilford County	240.03	233.97	6.06
Hillsborough County	244.64	235.65	8.99
Houston	235.25	225.60	9.65
Jefferson County	233.31	234.71	-1.40
Los Angeles	223.14	230.30	-7.15
Miami	244.99	232.70	12.29
Milwaukee	215.88	221.72	-5.85
New York City	229.22	230.16	-0.94
Philadelphia	214.33	223.04	-8.71
San Diego	237.51	234.82	2.69
Shelby County	224.71	220.92	3.79
Large City Schools*	231.52	229.95	1.57
Not Large City Schools‡	240.81	238.98	1.83

* Includes district-authorized charters

‡ Includes charters authorized by others and independent charters

Exhibit 12. Grade Eight Mathematics Actual Performance, Expected Performance, and District Effects in 2017.

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	269.84	270.44	-0.59
Atlanta	265.15	263.11	2.04
Austin	283.34	274.04	9.30
Baltimore	255.84	258.49	-2.64
Boston	280.38	263.07	17.31
Charlotte	287.78	277.15	10.63
Chicago	275.88	266.27	9.61
Clark County	272.82	273.94	-1.12
Cleveland	257.62	251.83	5.79
Dallas	268.25	254.60	13.65
Denver	--	--	--
Detroit	245.58	249.98	-4.40
District of Columbia (DCPS)	263.39	264.67	-1.28
Duval County	275.62	274.45	1.18
Fort Worth	268.47	262.36	6.11
Fresno	254.95	265.55	-10.60
Guilford County	277.01	274.42	2.59
Hillsborough County	277.35	275.58	1.77
Houston	273.49	265.32	8.16
Jefferson County	270.95	276.40	-5.45
Los Angeles	266.99	269.93	-2.93
Miami	274.03	269.83	4.20
Milwaukee	254.40	259.58	-5.18
New York City	275.35	270.24	5.11
Philadelphia	260.78	263.21	-2.42
San Diego	283.50	279.90	3.60
Shelby County	256.98	259.99	-3.01
Large City Schools*	274.30	270.69	3.61
Not Large City Schools‡	283.52	281.33	2.19

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

(b) Trends in Overcoming Poverty and Other Variables

Exhibits 13 through 16 show the district effects for all TUDA districts across all five assessment periods (2009, 2011, 2013, 2015, and 2017) in grades four and eight, reading and math. These data are meant to answer the question about whether Large City schools were getting better at overcoming the effects of poverty, language, and other demographic variables or not.

In grade four reading, most cities had district effects that were above expectations, and several improved those effects between 2009 and 2017. In 2017, there were 17 cities that showed overall positive effects and 10 had negative effects. Of the districts with positive effects in 2017, five had improved since 2009--Boston, Chicago, the District of Columbia, Miami-Dade County, and San Diego. One other district—Duval County—had gains between 2015 and 2017, the only two years they were tested. Two districts with negative effects in 2017 improved between 2009 and 2017—Cleveland and Fresno. And two districts--Chicago and the District of Columbia—moved from having a negative district effect in 2009 to having a positive one in 2017. (Exhibit 13)

In grade eight reading (Exhibit 14), 10 cities had positive effects in 2017. Of these cities, five showed larger effects in 2017 than in 2009—Atlanta, Chicago, Dallas (since 2011), New York City, and San Diego. Duval County showed gains in their effects over the two years on which there are data. There were 16 districts with negative district effects in 2017. Of these districts, two showed a larger effect in 2017 than in 2009, even though they remained in negative territory—Detroit and Milwaukee. And only one district—San Diego--moved from below the zero line in 2009 to above it in 2017. The remaining districts showed slippage.

In grade four mathematics (Exhibit 15), 16 of the 27 TUDA districts performed better than expected in 2017. Five of these districts—Atlanta, Chicago, the District of Columbia, Duval County (over two testing cycles), and Miami—showed gains in 2017 over and above their effects in 2009. Some 11 other districts had negative district effects in 2017. Two of which, Cleveland and Detroit, showed gains over and above 2017—even though they remained in negative territory throughout the period. Two districts—Chicago and the District of Columbia—went from below the line to above the line between 2009 and 2017.

Finally, in grade eight mathematics (Exhibit 16), 15 of 26 TUDA districts performed better than expected in 2017. Of those, nine—Atlanta, Boston, Charlotte, Chicago, Cleveland, Dallas, Duval County (over two testing cycles), Miami-Dade County, and San Diego—had larger effects in 2009 than in 2017. In addition, 11 cities showed a negative district effect in 2017. Five of these districts (Baltimore, the District of Columbia, Detroit, Los Angeles, and Milwaukee) showed somewhat higher district effects in 2017 than in 2009. The remaining districts slipped in their district effects. No city moved from a negative district effect in 2009 to a positive one in 2017.

Overall, there were several notable trends. Chicago, for instance, showed improvements in all four assessments (i.e., reading, math, fourth grade, and eighth grade) from 2009 to 2017. Boston posted increased district effects on three assessments, as did the Cleveland, the District of Columbia, Miami-Dade County, and San Diego. And several districts showed gains across two assessment areas. In addition, several districts went from a negative district effect in 2009 to a positive one in 2017 in at least one assessed area—Chicago, the District of Columbia, and San Diego.

(c) *Influence of Abject and Concentrated Poverty*

An initial review of results after adjusting for relevant background variables indicated that they may not adequately control for poverty. The question emerged about whether the Free & Reduced-Price Lunch-eligibility measure used by NAEP sufficiently differentiated poverty levels or took adequate account of deep or abject poverty. The National School Lunch Act in 1946 created the modern school lunch program though the U. S. Department of Agriculture, and about 7.1 million children were participating in it by the end of its first year, 1946-47. By 1970, 22 million children were participating, and by 1980 the figure was

nearly 27 million. In 2012, more than 31.6 million children were participating in the National School Lunch Program.

The program provides free meals to eligible children in households with income at or below 130 percent of the federal poverty guidelines, and reduced-price meals to eligible children in households with income above 130 percent and at or below 185 percent of poverty. Unfortunately, as the number of participating students rose and the income categories remained the same, the lunch-eligibility data became less and less able to differentiate the very poor from the poor and near-poor.

The distinction between levels of poverty becomes important as we look at which districts are most able to overcome the effects of poverty and other barriers—and conversely, which ones have a more difficult challenge. Exhibit 17 shows the difference in abject poverty across districts. Later in this analysis, one will see that despite progress, districts like Baltimore, Cleveland, Detroit, Milwaukee, Philadelphia, and others with high levels of abject poverty have a more difficult time rising above statistical expectations.

Using free and reduced priced lunch as a proxy for poverty has been an acceptable and frequently used measure in many research studies, but it has flaws. In fact, the measure has become increasingly challenging because of the new Community Eligibility Provision (CEP). The CEP is a meal service option for schools and school districts in low-income areas. A key provision of *The Healthy, Hunger Free Kids Act* (HHFKA, Public Law 111-296, December 13, 2010), CEP allows the highest poverty schools to serve breakfast and lunch at no cost to all enrolled students without the burden of collecting household applications. Instead, schools that adopt CEP are reimbursed using a formula (1.6 times direct certification) based on the percentage of students participating in other means-tested programs, such as the Supplemental Nutrition Assistance Program (SNAP) and Temporary Assistance for Needy Families (TANF).

As a result, a school that may have 85 percent of its students eligible for free and reduced priced lunch will serve 100 percent of students. Obviously, the change has been important for ensuring that students have adequate nutrition, but the new provision has been problematic for researchers trying to measure poverty or use it in their analyses. The changes, for instance, have affected the ability to maintain trend lines in poverty levels and obtain accurate counts of students in poverty. Researchers have tried to use a combination of direct certification, census poverty data using geocodes, imputed variables, and prior information to determine a best metric, but the attempts have not always been fully successful.

Finally, poverty thresholds in the federal free and reduced-price lunch data do not vary by geography or economic cost living factors, although other adjustments can be made. They also do not count for students who are at or below the 100 percent poverty threshold. And poverty rates are compounded in cities where the costs of living vary (e.g., New York City vs. Des Moines).

The table below (Exhibit 19) shows income levels for TUDA districts according to bands of income below $50,000 annually, using Census income data for 2015. For the purposes of this analysis, abject poverty is annual income below $10,000. We also use that measure in combination with annual income below $50,000. Unfortunately, the Census data cannot be juxtaposed against all the NAEP variables used in this study.

Exhibit 13. Percentage of Households by Income Level in TUDA Districts, 2015

	Less than $10,000	$10,000 to $14,999	$15,000 to $24,999	$25,000 to $34,999	$35,000 to $49,999	Total Percent of Families
Detroit City School District	21.7	10.2	16.9	12.7	13.6	75.1
Cleveland Municipal School District	20.5	10.6	17.1	12.5	13.5	74.2
Fresno Unified School District	11.5	9.4	16.0	13.4	14.5	64.8
Milwaukee School District	12.2	8.7	15.1	12.9	14.5	63.4
Philadelphia City School District	14.2	7.9	13.0	11.6	13.6	60.3
Fort Worth Independent School District	9.9	7.1	13.3	12.2	14.0	56.5
Baltimore City Public Schools	13.1	7.5	11.6	11.1	13.0	56.3
Dallas Independent School District	9.6	6.5	13.1	12.2	14.9	56.3
Miami-Dade County School District	10.6	6.8	13.3	11.1	14.1	55.9
Guilford County Schools	8.1	5.8	12.3	12.2	15.0	53.4
Shelby County School District	9.7	6.2	12.7	11.1	13.2	52.9
Houston Independent School District	9.1	6.4	12.8	10.8	13.3	52.4
Duval County School District	8.7	5.6	10.9	11.6	15.1	51.9
Albuquerque Public Schools	9.1	5.8	12.3	11.2	13.4	51.8
Atlanta City School District	12.8	6.3	11.1	9.4	12.0	51.6
Jefferson County School District	8.5	6.0	11.3	10.8	14.6	51.2
Chicago Public School District 299	11.1	5.9	11.6	10.0	12.4	51.0
Los Angeles Unified School District	7.9	6.9	12.0	10.5	12.8	50.1
Hillsborough County School District	7.7	5.4	11.3	10.6	14.3	49.3
Clark County School District	6.7	4.6	10.4	11.4	15.2	48.3
New York City	10.4	6.1	10.5	8.9	11.4	47.3

	Less than $10,000	$10,000 to $14,999	$15,000 to $24,999	$25,000 to $34,999	$35,000 to $49,999	Total Percent of Families
Denver County School District 1	8.4	5.2	9.6	10.1	13.4	46.7
Boston School District	12.0	7.3	9.3	7.2	10.2	46.0
Austin Independent School District	7.9	4.5	9.3	9.6	13.6	44.9
Charlotte-Mecklenburg Schools	6.4	4.4	9.4	10.3	13.7	44.2
San Diego City Unified School District	6.3	4.9	9.0	8.5	12.2	40.9
District of Columbia Public Schools (DCPS)	10.2	4.2	7.4	6.7	9.6	38.1

What is clear from the data is that TUDA districts with NAEP scores in reading and math below expectations in 2015 in all four subject-grade combinations (reading, math, grade 4, and grade 8) also had unusually high poverty rates. See Exhibit 20. This suggests that districts with particularly low-income levels and high concentrations of such poverty are much less likely to produce a positive district effect in reading and math performance. (More on this in the case study section of this report.)

Exhibit 14. TUDA Districts with Negative District Effects in Four Areas and Their Abject Poverty Levels, 2015

	District Effect in Grade 4 Reading	District Effect in Grade 8 Reading	District Effect in Grade 4 Math	District Effect in Grade 8 Math	Percent of Families below $10,000	Percent of Families below $50,000
Detroit	-18.20	-6.43	-19.70	-13.90	21.7	75.1
Cleveland	-5.93	-2.39	-4.15	-0.97	20.5	74.2
Fresno	-10.16	-11.71	-12.84	-14.10	11.5	64.8
Milwaukee*	-7.91	-6.48	-7.44	-5.75	12.2	63.4
Philadelphia	-12.62	-5.80	-13.85	-2.43	14.2	60.3
Baltimore	-9.01	-3.31	-10.73	-3.16	13.1	56.3
Los Angeles	-6.13	-3.60	-7.41	-7.04	7.9	50.1

*District Effects data for 2013

By and large, this effect appears to apply to districts with populations with incomes below $10,000 annually of at least 10 percent and incomes below $50,000 of at least 50 percent. All districts in Exhibit 18, except Los Angeles, have these characteristics. At the same time, there are districts with both demographic conditions that have at least one or more positive district effects. In fact, Dallas, Miami-Dade County, and Chicago have four positive district effects--reading and math in both fourth and eighth grades. And Atlanta has two. Interestingly, Chicago has gone from below the zero line to above it in two areas between 2009 and 2017—fourth grade reading and fourth grade math.

Figure 15. Trends in District Effects in Grade Four Reading by City, 2009 to 2017

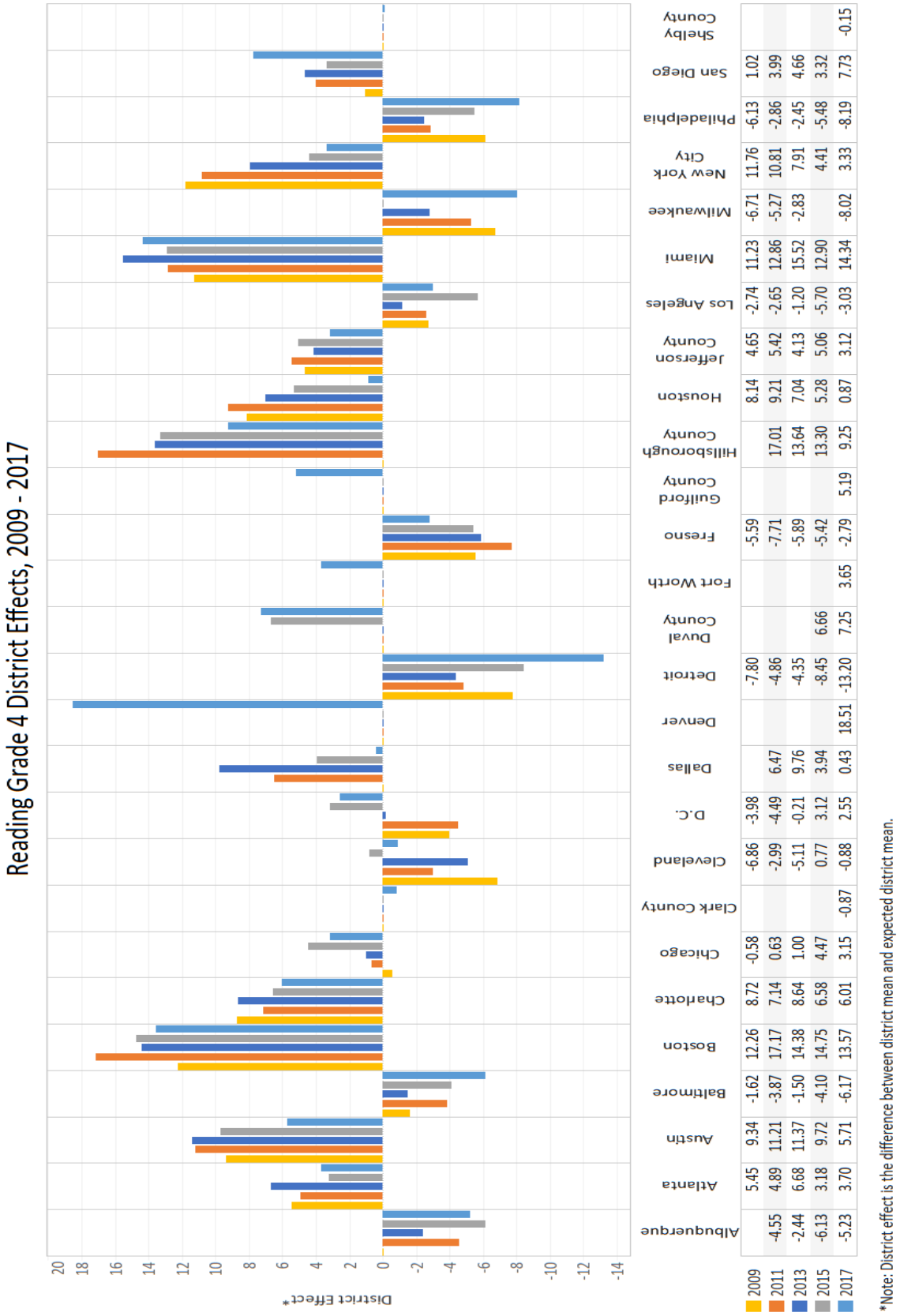
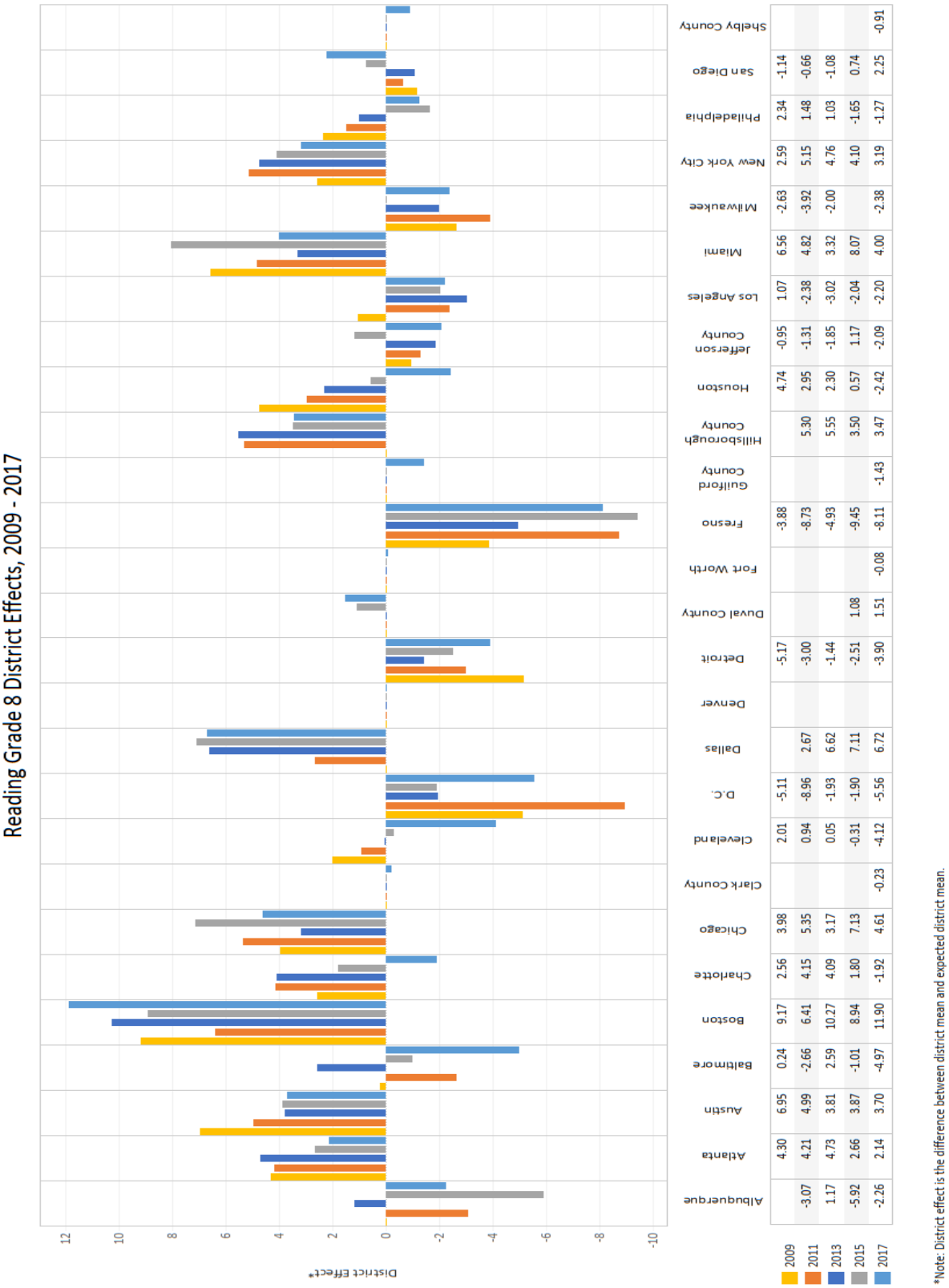


Figure 16. Trends in District Effects in Grade Eight Reading by City, 2009 to 2017



	2009	2011	2013	2015	2017
Albuquerque		-3.07	1.17	-5.92	-2.26
Atlanta	4.30	4.21	4.73	2.66	2.14
Austin	6.95	4.99	3.81	3.87	3.70
Baltimore	0.24	-2.66	2.59	-1.01	-4.97
Boston	9.17	6.41	10.27	8.94	11.90
Charlotte	2.56	4.15	4.09	1.80	-1.92
Chicago	3.98	5.35	3.17	7.13	4.61
Clark County					-0.23
Cleveland	2.01	0.94	0.05	-0.31	-4.12
D.C.	-5.11	-8.96	-1.93	-1.90	-5.56
Dallas		2.67	6.62	7.11	6.72
Denver					
Detroit	-5.17	-3.00	-1.44	-2.51	-3.90
Duval County				1.08	1.51
Fort Worth					-0.08
Fresno	-3.88	-8.73	-4.93	-9.45	-8.11
Guilford County					-1.43
Hillsborough County		5.30	5.55	3.50	3.47
Houston	4.74	2.95	2.30	0.57	-2.42
Jefferson County	-0.95	-1.31	-1.85	1.17	-2.09
Los Angeles	1.07	-2.38	-3.02	-2.04	-2.20
Miami	6.56	4.82	3.32	8.07	4.00
Milwaukee	-2.63	-3.92	-2.00		-2.38
New York City	2.59	5.15	4.76	4.10	3.19
Philadelphia	2.34	1.48	1.03	-1.65	-1.27
San Diego	-1.14	-0.66	-1.08	0.74	2.25
Shelby County					-0.91

Figure 17. Trends in District Effects in Grade Four Mathematics by City, 2009 to 2017

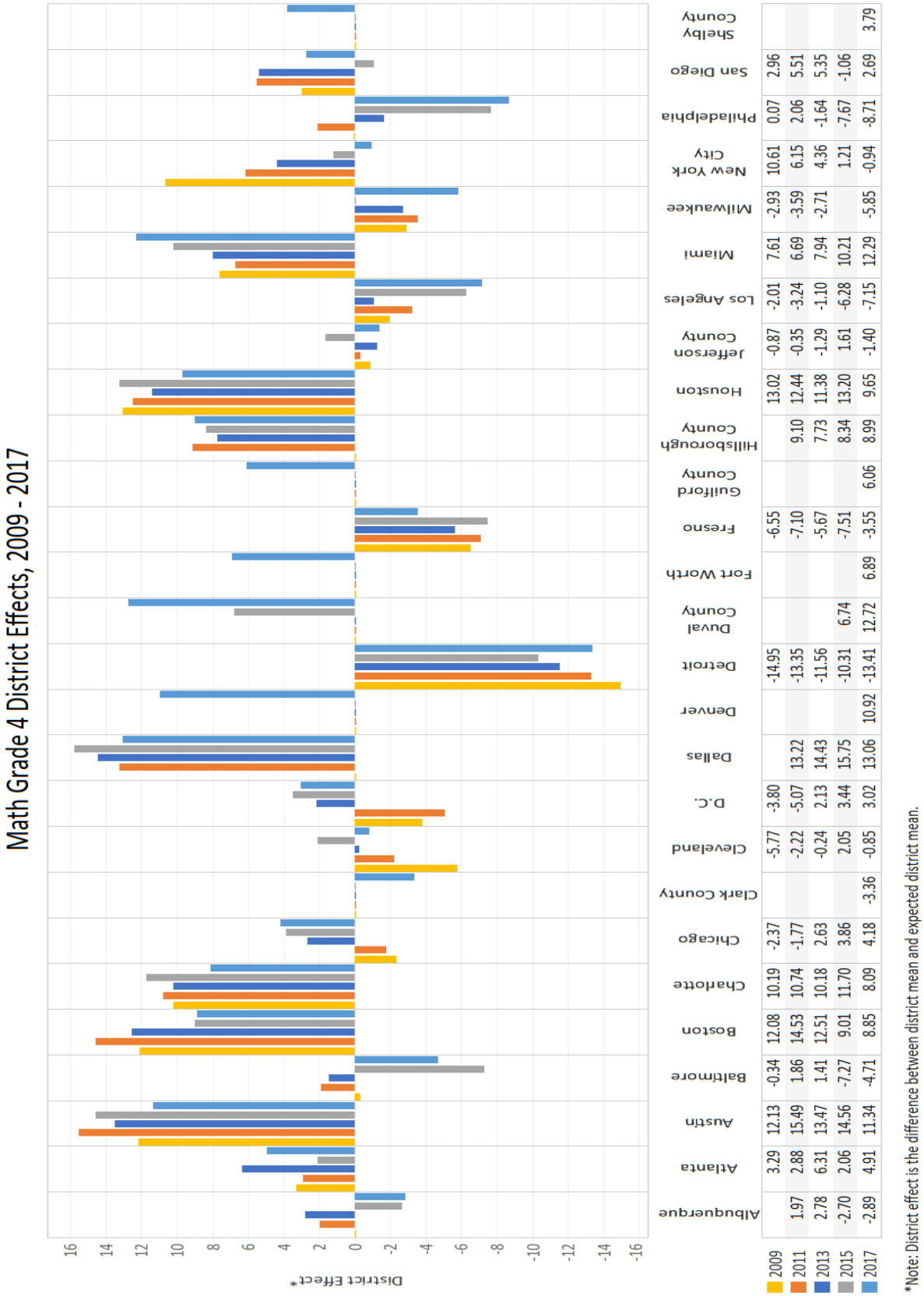
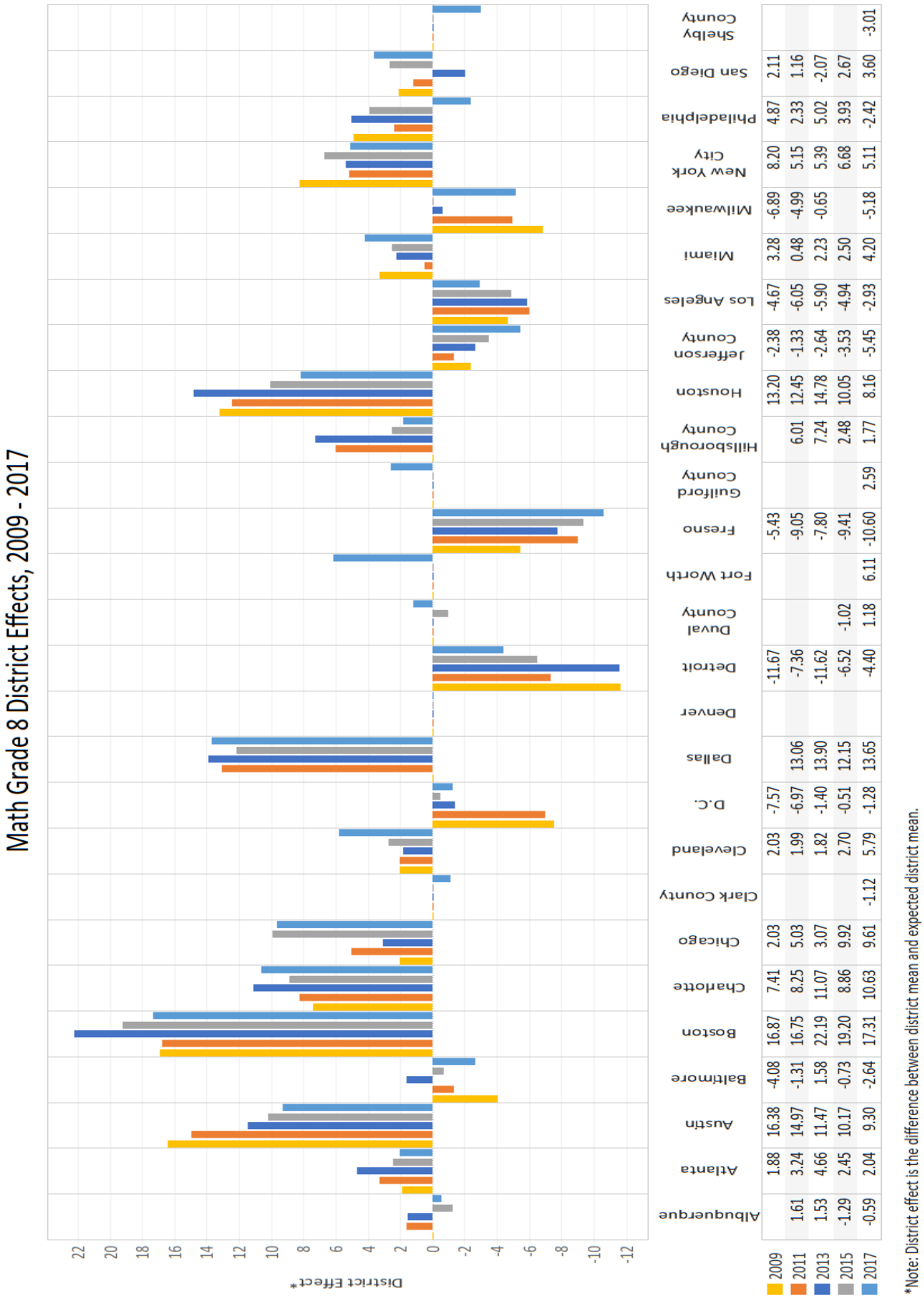


Figure 18. Trends in District Effects in Grade Eight Mathematics by City, 2009 to 2017



(d) Raw Scale Scores vs. District Effects

The following exhibits compares how the districts ranked with each other when looking at NAEP raw scores versus the "effects" that each district produced. The rankings were sometimes dramatically different. The tables also show the numbers of districts that produced an effect that was larger than the national average.

Ranking of TUDA Districts on 4th Grade Math Scale Scores and District Effects, 2017

Raw Scale Score	District Effects
Duval County	Dallas
Charlotte-Mecklenburg	Duval County
Austin	Miami-Dade County
Miami-Dade County	Austin
Hillsborough County	Denver
Guilford County	Houston
National Public	Hillsborough County
San Diego	Boston
Houston	Charlotte-Mecklenburg
Dallas	Fort Worth
Boston	Guilford County
Jefferson County	Atlanta
Chicago	Chicago
Atlanta	District of Columbia
District of Columbia	Shelby County
Albuquerque	San Diego
Clark County	National Public
Fort Worth	Cleveland
Denver	New York City
New York City	Jefferson County
Shelby County	Albuquerque
Los Angeles	Clark County
Fresno	Fresno
Milwaukee	Baltimore
Baltimore	Milwaukee
Cleveland	Los Angeles
Philadelphia	Philadelphia
Detroit	Detroit

Ranking of TUDA Districts on 8th Grade Math Scale Scores and District Effects, 2017

Raw Scale Score	District Effects
Charlotte	Boston
Austin	Dallas
National Public	Charlotte
San Diego	Chicago
Boston	Austin
Hillsborough County	Houston
Chicago	Fort Worth
Guilford County	Cleveland
New York City	New York City
Duval County	Miami-Dade County
Miami-Dade County	San Diego
Houston	Guilford County
Denver	National Public
Clark County	Atlanta
Jefferson County	Hillsborough County
Albuquerque	Duval County
Fort Worth	Albuquerque
Dallas	Clark County
Los Angeles	DC
Atlanta	Philadelphia
DC	Baltimore
Philadelphia	Shelby County
Shelby County	Los Angeles
Cleveland	Detroit
Baltimore	Milwaukee
Fresno	Jefferson County
Milwaukee	Fresno
Detroit	Atlanta

Ranking of TUDA Districts on 4th Grade Reading Scale Scores and District Effects, 2017

Raw Scale Score	District Effects
Miami-Dade County	Denver
Hillsborough County	Miami-Dade County
Duval County	Boston
Charlotte	Hillsborough County
San Diego	San Diego
Guilford County	Duval County
Jefferson County	Charlotte
National Public	Austin
Austin	Atlanta
Boston	Guilford County
Denver	Fort Worth
New York City	New York City
Atlanta	Chicago
Clark County	Jefferson County
DC	DC
Chicago	National Public
Los Angeles	Houston
Albuquerque	Dallas
Fort Worth	Shelby County
Houston	Clark County
Fresno	Cleveland
Shelby County	Fresno
Dallas	Los Angeles
Baltimore	Albuquerque
Philadelphia	Baltimore
Cleveland	Milwaukee
Milwaukee	Philadelphia
Detroit	Detroit

Ranking of TUDA Districts on 8th Grade Reading Scale Scores and District Effects, 2017

Raw Scale Score	District Effects
National Public	Boston
Hillsborough County	Dallas
San Diego	Chicago
Duval County	Miami
Austin	Austin
Miami-Dade County	Hillsborough
Jefferson County	New York City
Boston	San Diego
Charlotte	Atlanta
Guilford County	Duval County
Chicago	National Public
Denver	Fort Worth
Clark County	Clark County
New York City	Shelby County
Albuquerque	Philadelphia
Atlanta	Guilford County
Los Angeles	Charlotte
Houston	Jefferson County
Philadelphia	Los Angeles
Fort Worth	Albuquerque
Shelby County	Milwaukee
DC	Houston
Dallas	Detroit
Milwaukee	Cleveland
Fresno	Baltimore
Baltimore	DC
Cleveland	Fresno
Detroit	

(e) Effects of College and Career-Ready Standards

One of the abiding questions that some observers have asked involves the effects of college and career-ready standards on NAEP results. This question has emerged because of the apparent slow-down in NAEP gains over the last several years, particularly in mathematics. To answer the question, the National Center for Educational Statistics conducted an analysis of differences in NAEP math content and the content of state assessments that were generally aligned with the standards.[6] The main research question was, "How would 2013, 2015, 2017, and 2019" mathematics grade 4 and grade 8 TUDA mean scores change if NAEP subscales were weighted according to the content focus of selected state assessments." Only TUDA districts in selected states were analyzed.

Results of the analysis showed that the reweighting of NAEP mathematics scale scores changed the means in grades 4 and 8 for the nine TUDA districts analyzed. (See Exhibits 17-18.)

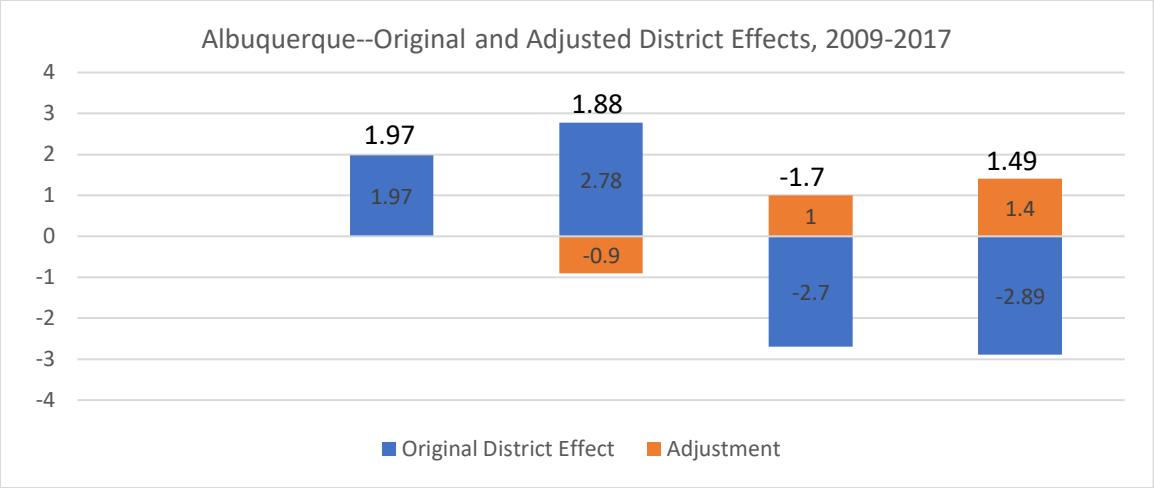
Exhibit 19. Reported and Reweighted TUDA Means for Grade 4 Mathematics by Year

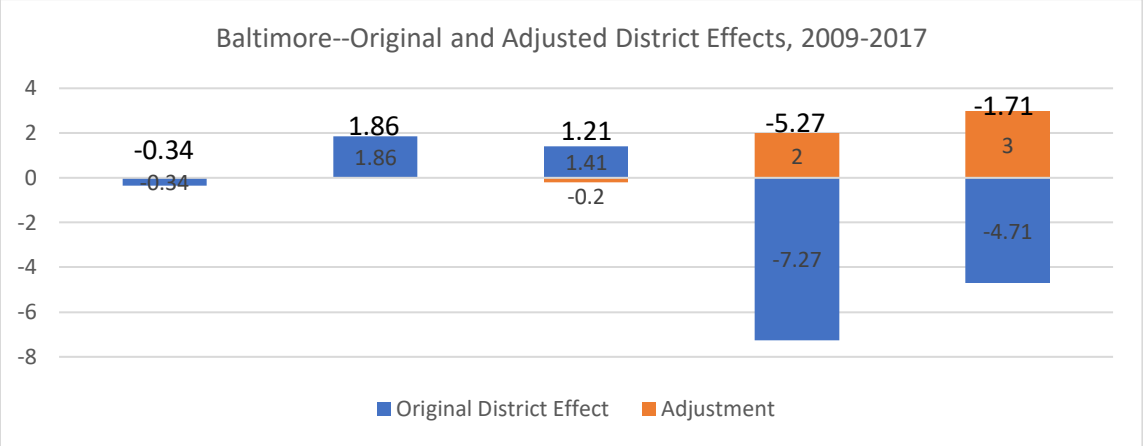
	Reported Scale Score				Reweighted Scale Score			
District	2013	2015	2017	2019	2013	2015	2017	2019
Albuquerque	234.5	230.6	229.8	229.8	233.6	231.6	231.2	231.2
Baltimore	222.9	215.0	215.3	216.5	222.7	217.0	218.3	218.5
Boston	236.9	235.5	233.3	233.8	237.4	236.6	234.5	234.8
Chicago	230.5	231.9	231.8	232.5	229.3	233.0	233.9	234.7
Clark County	NA	NA	230.2	234.5	NA	NA	231.8	236.9
DC	228.6	232.2	230.8	235.3	229.1	234.7	232.9	238.7
Fresno	219.7	217.7	221.4	224.0	222.1	220.5	226.0	227.1
LA	228.5	224.2	223.1	223.6	231.3	226.5	226.4	225.9
San Diego	240.9	232.8	237.4	240.2	242.8	235.2	241.0	244.2
Median Diff.					0.49	2.18	2.08	2.30
Mean Diff.					0.73	1.90	2.54	2.42

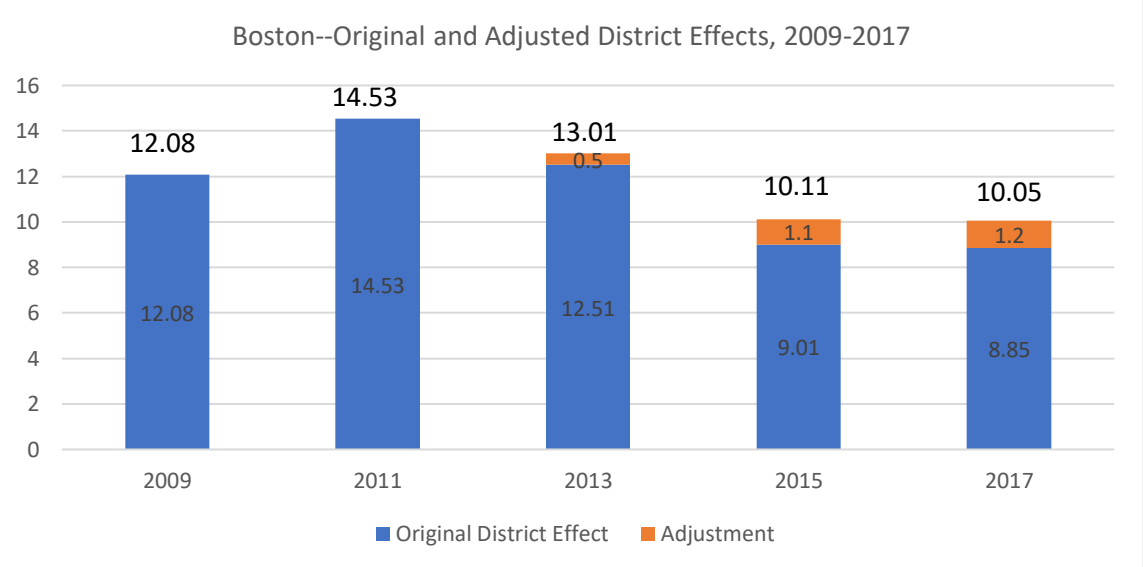
Exhibit 20. Reported and Reweighted TUDA Means for Grade 8 Mathematics by Year

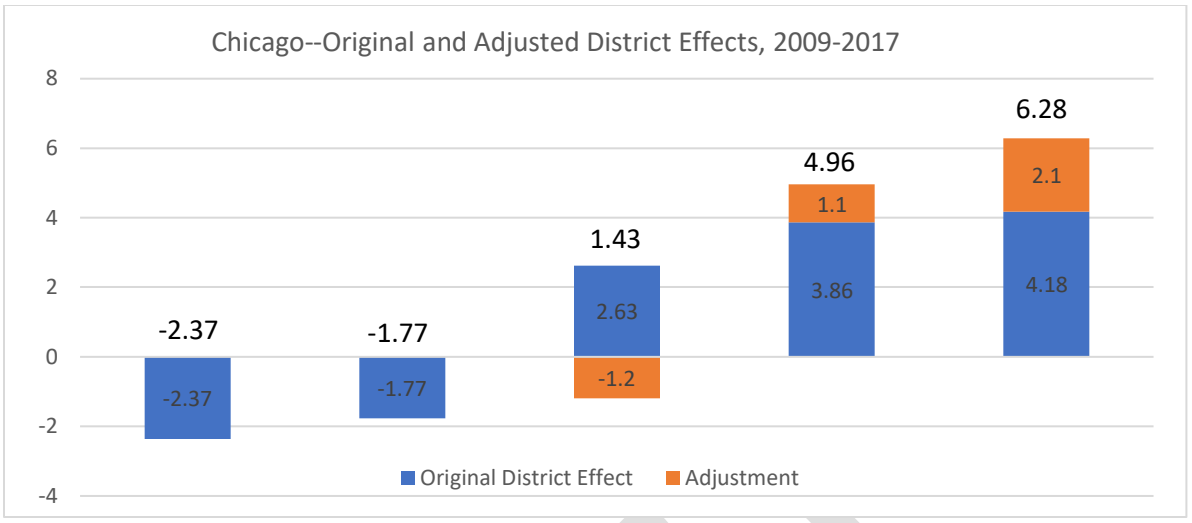
	Reported Scale Score				Reweighted Scale Score			
District	2013	2015	2017	2019	2013	2015	2017	2019
Albuquerque	273.8	270.7	269.6	266.8	274.2	271.0	270.4	267.8
Baltimore	259.8	255.2	255.5	254.1	260.0	255.2	255.9	255.0
Boston	283.1	281.1	279.7	278.8	283.4	281.9	280.6	279.7
Chicago	268.9	274.9	275.6	275.3	269.3	275.7	276.7	276.2
Clark County	NA	NA	272.2	271.6	NA	NA	273.8	273.5
DC	260.3	258.4	262.0	268.6	260.2	259.0	262.9	269.9
Fresno	259.7	256.9	254.6	253.5	261.9	257.6	256.3	254.9
LA	264.3	263.5	266.8	260.7	266.6	265.0	269.4	262.8
San Diego	276.9	280.4	282.8	282.6	278.7	281.9	284.2	284.7
Median Diff.					0.41	0.76	1.02	1.28
Mean Diff.					0.94	0.78	1.27	1.39

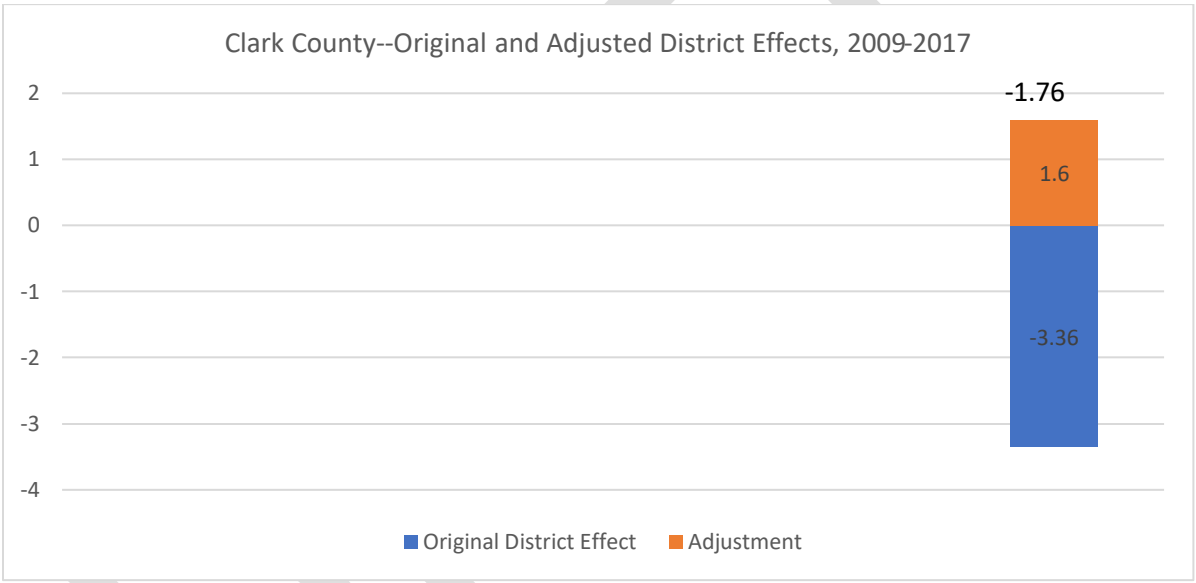
[6] *Appendix: Analysis of Recent NAEP TUDA Mathematics Results Based on Alignment to State Assessment Content*, National Center for Educational Statistics, 2019

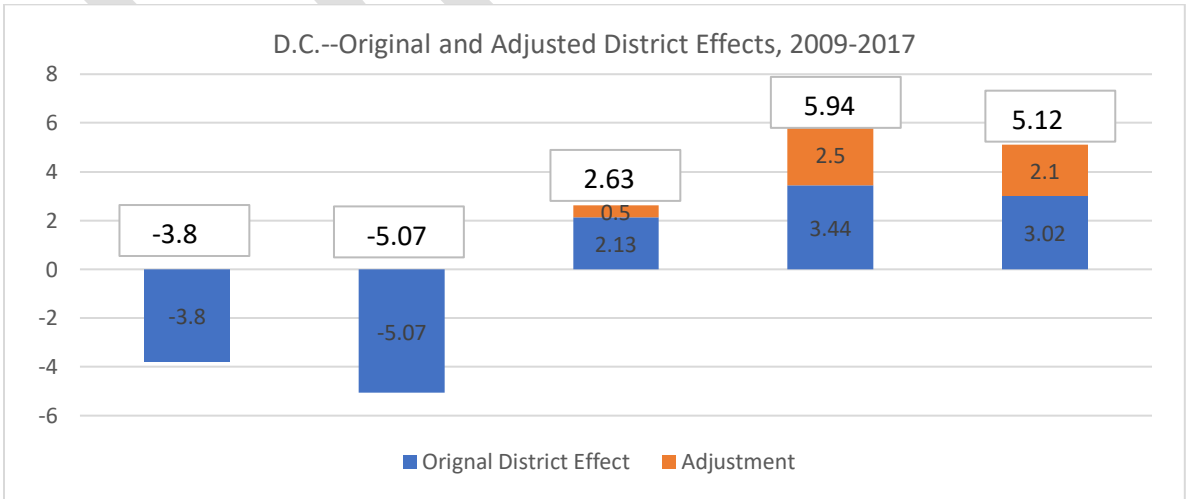

Albuquerque--Original and Adjusted District Effects, 2009-2017
4
3
2
1
0
-1
-2
-3
-4
1.97
1.88
-1.7
1.49
1.97
2.78
-0.9
1
-2.7
1.4
-2.89
Original District Effect
Adjustment

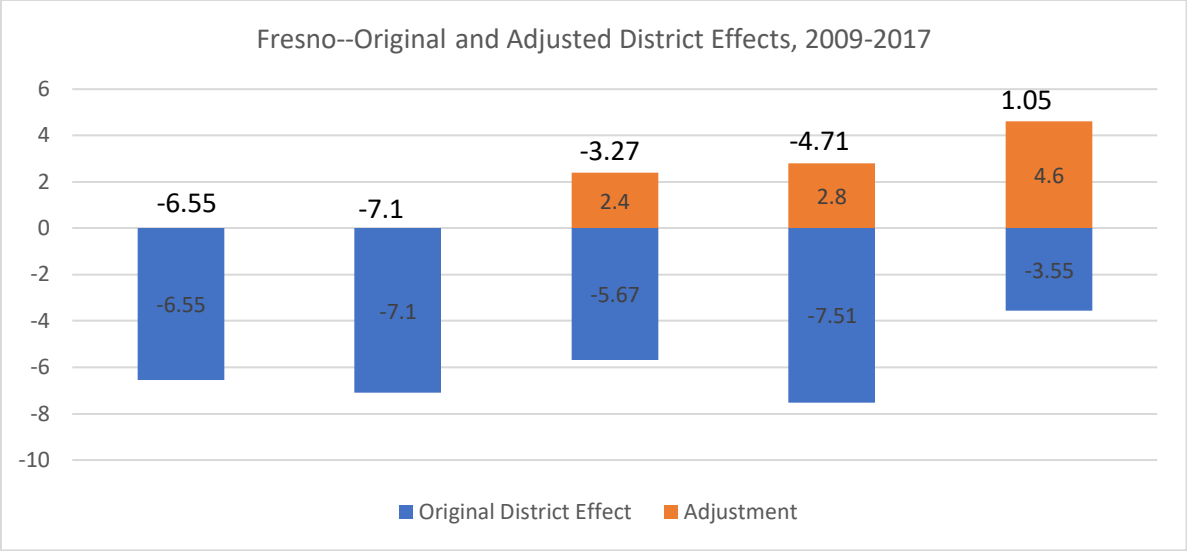

Baltimore--Original and Adjusted District Effects, 2009-2017
4
2
0
-2
-4
-6
-8
-0.34
1.86
1.21
-5.27
-1.71
-0.34
1.86
1.41
-0.2
2
-7.27
3
-4.71
Original District Effect
Adjustment

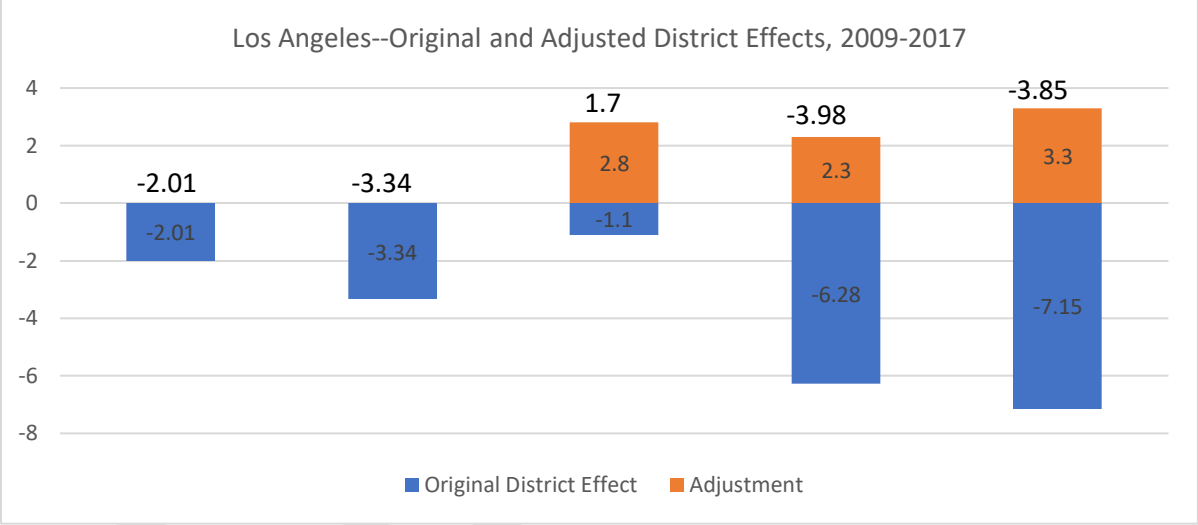

Boston--Original and Adjusted District Effects, 2009-2017
16
14
12
10
8
6
4
2
0
12.08
14.53
13.01
10.11
10.05
12.08
14.53
0.5
12.51
1.1
9.01
1.2
8.85
2009
2011
2013
2015
2017
Original District Effect
Adjustment

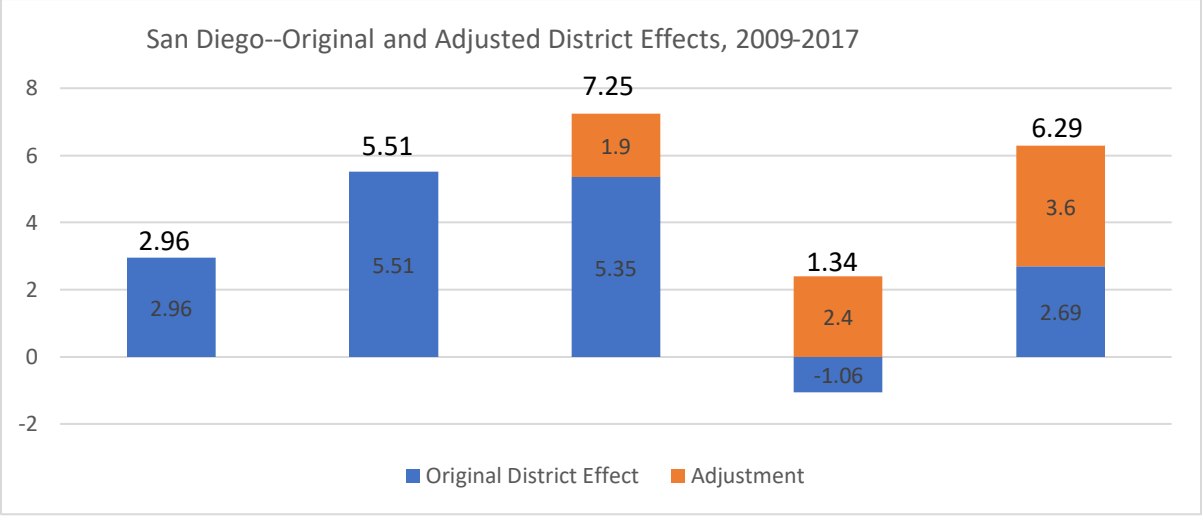

Chicago--Original and Adjusted District Effects, 2009-2017
8
6
4
2
0
-2
-4
-2.37
-1.77
1.43
4.96
6.28
-2.37
-1.77
2.63
-1.2
3.86
1.1
4.18
2.1
Original District Effect
Adjustment


Clark County--Original and Adjusted District Effects, 2009-2017
2
1
0
-1
-2
-3
-4
-1.76
-3.36
1.6
Original District Effect
Adjustment


D.C.--Original and Adjusted District Effects, 2009-2017
8
6
4
2
0
-2
-4
-6
-3.8
-5.07
2.63
5.94
5.12
-3.8
-5.07
2.13
0.5
3.44
2.5
3.02
2.1
Orignal District Effect
Adjustment


Fresno--Original and Adjusted District Effects, 2009-2017
6
4
2
0
-2
-4
-6
-8
-10
-6.55
-7.1
-3.27
-4.71
1.05
-6.55
-7.1
2.4
-5.67
2.8
-7.51
4.6
-3.55
Original District Effect
Adjustment


Los Angeles--Original and Adjusted District Effects, 2009-2017
4
2
0
-2
-4
-6
-8
-2.01
-3.34
1.7
-3.98
-3.85
-2.01
-3.34
2.8
-1.1
2.3
-6.28
3.3
-7.15
Original District Effect
Adjustment


San Diego--Original and Adjusted District Effects, 2009-2017
8
6
4
2
0
-2
2.96
5.51
7.25
1.34
6.29
2.96
5.51
1.9
5.35
2.4
-1.06
3.6
2.69
Original District Effect
Adjustment

(f) Comparing Large City and Not Large City School Trends

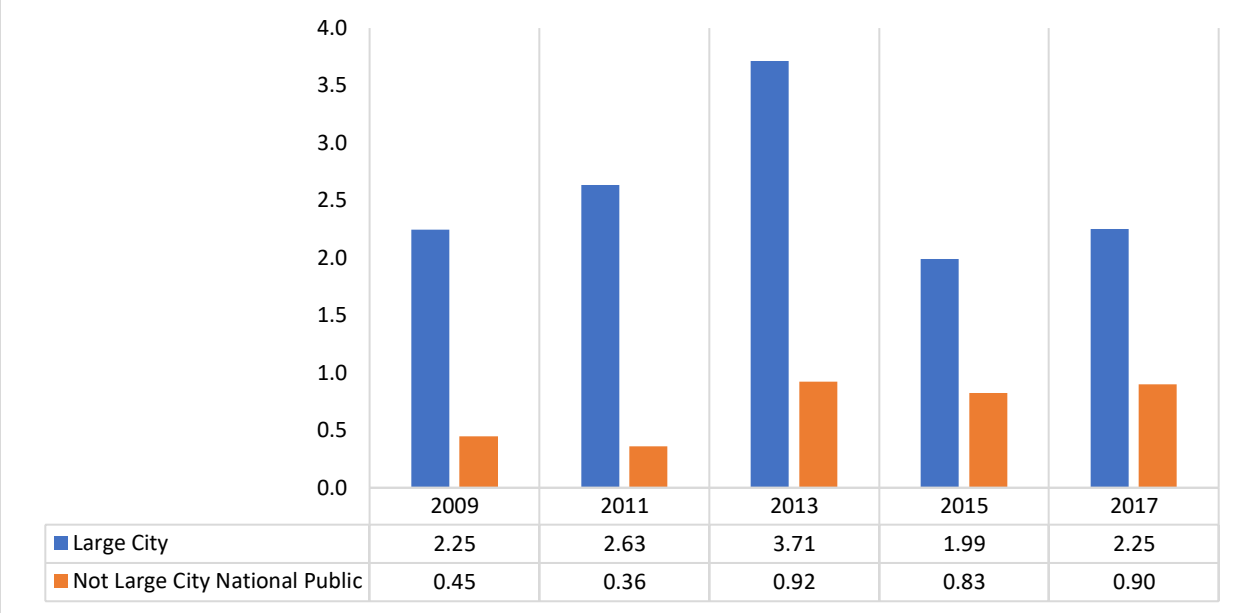
This section examines how large city school districts participating in TUDA performed compared to Not Large City Schools. Results of the data analysis are shown in Exhibits 20 through 23. The results show several things. One, in 2017, the district effect was larger in Large City schools in three out of four areas—fourth grade reading, eighth grade math, and eighth grade reading. Only in fourth grade math did Not Large City Schools produce a larger district effect. There will be additional discussion of this in the next section.

Two, Large City schools showed uniform improvement in its district effects between 2009 and 2017 but more uneven trends between 2013 and 2017. Between 2009 and 2017, Large City schools did not show any gain or loss in its district effects in fourth grade reading. In fourth grade math, the district effects with Large City schools declined from +3.16 in 2009 to +1.57 in 2017. At the eighth-grade level in reading, the district effects among Large City schools improved from +0.52 in 2009 to +1.32 in 2017. And in eighth-grade math, Large City schools improved their district effects from +2.52 in 2009 to +3.61 in 2017. In other words, Large City schools have generally improved their ability to overcome the effects of the background variables measured in at least two out of four areas and held steady in one. It is also notable that Large City schools showed higher district effects than Not Large Cities in every grade, subject, and year except for fourth grade math in 2017.

While Large City schools almost universally showed larger district effects than Not Large Cities, the Not Large City schools also showed gains. In fourth grade reading, Not Large City schools improved their district effects from +0.45 in 2009 to +0.90 in 2017. In fourth grade math, Not Large Cities showed gains in their district effects from +1.21 in 2009 to +1.83 in 2017. In eighth grade reading, Not Large Cities improved their district effects from -1.00 in 2009 to +0.44 in 2017. And in eighth grade math, Not Large Cities showed gains in their district effects from +1.07 in 2009 to +2.19 in 2017.

In other words, Not Large City schools were more likely to reflect the demographic variables measured than did Large Cities but by 2017 both Large City schools and Not Large City schools in the aggregate were showing results that were at least somewhat better than statistically expected by 2017. This is a promising development for schools in both settings. Of note, however, is the sizable additional district effect that Large City schools have over Not Large City schools, except in fourth grade math. In fourth grade reading, the Large City schools have a district effect in 2017 that is 2.5 times greater than Not Large City schools; 3.0 times greater in eighth grade reading; and 1.65 times greater in eighth grade math.

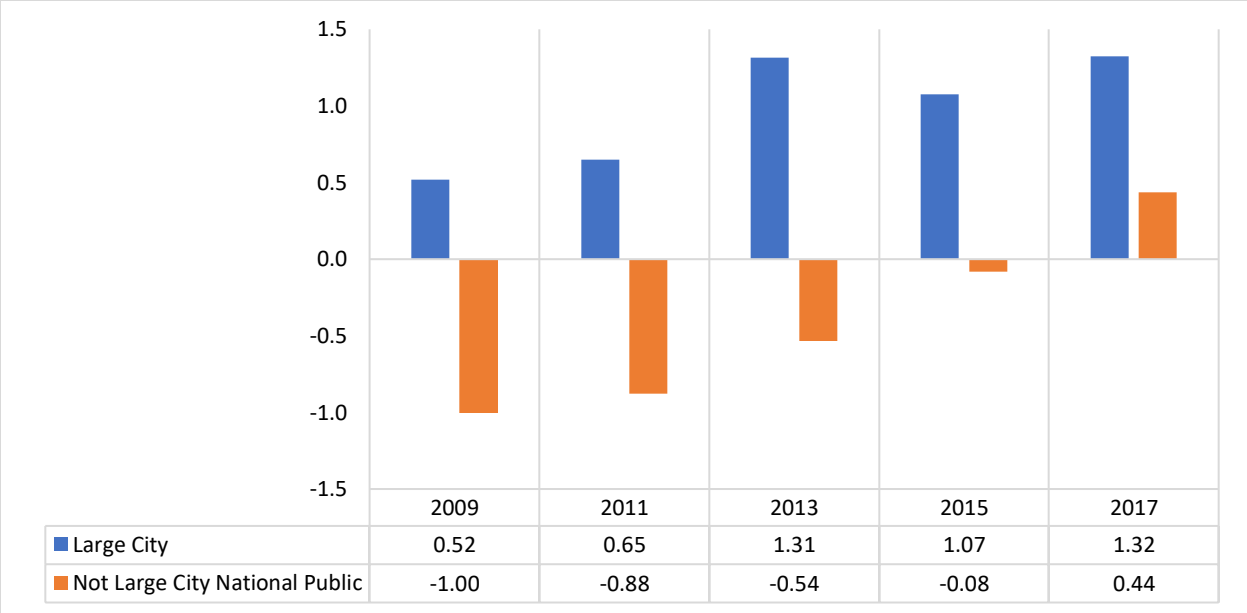
Exhibit 21. Trends in District Effects in Grade Four Reading on NAEP by School Type, 2009 to 2017



	2009	2011	2013	2015	2017
Large City	2.25	2.63	3.71	1.99	2.25
Not Large City National Public	0.45	0.36	0.92	0.83	0.90

*District effect is significantly different from zero.
‡ Includes district-authorized charters, charters authorized by others, and independent charters

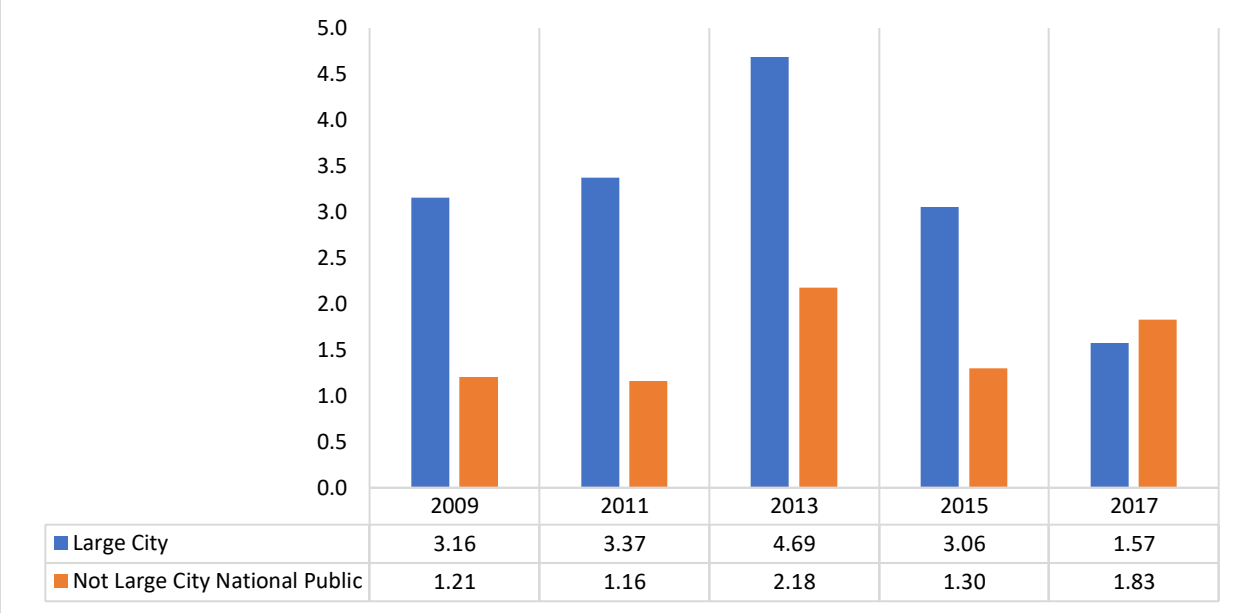
Exhibit 22. Trends in District Effects in Grade Eight Reading on NAEP by School Type, 2009 to 2017



	2009	2011	2013	2015	2017
Large City	0.52	0.65	1.31	1.07	1.32
Not Large City National Public	-1.00	-0.88	-0.54	-0.08	0.44

*District effect is significantly different from zero.
‡ Includes district-authorized charters, charters authorized by others, and independent charters

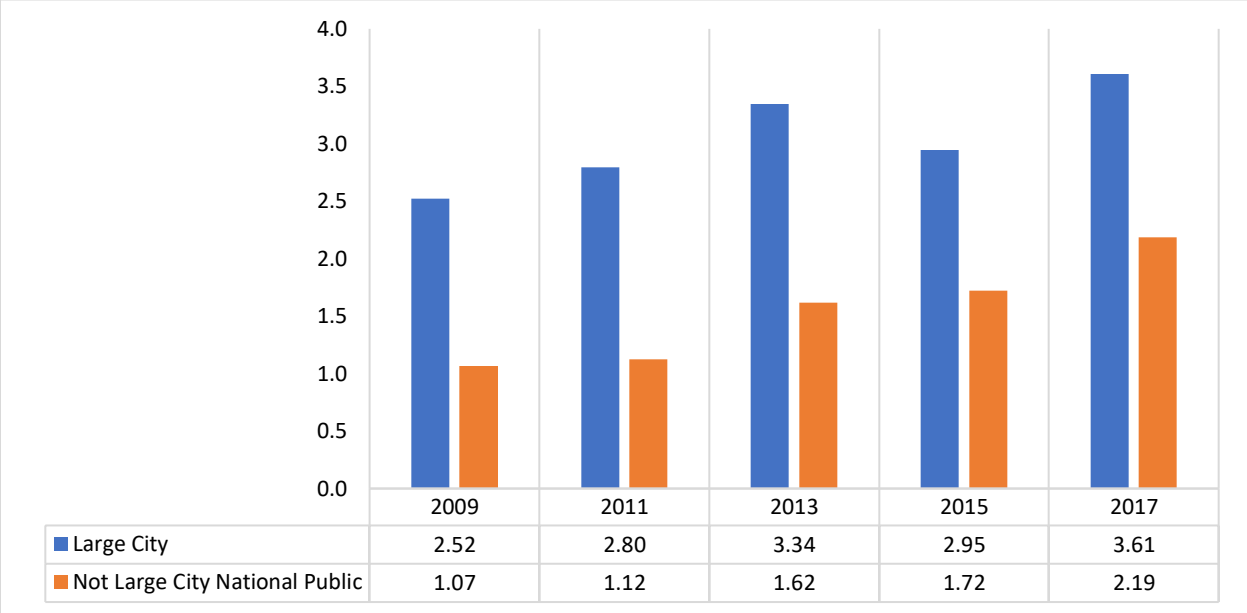
Exhibit 23. Trends in District Effects in Grade Four Math on NAEP by School Type, 2009 to 2017



	2009	2011	2013	2015	2017
Large City	3.16	3.37	4.69	3.06	1.57
Not Large City National Public	1.21	1.16	2.18	1.30	1.83

*District effect is significantly different from zero.
‡ Includes district-authorized charters, charters authorized by others, and independent charters

Exhibit 24. Trends in District Effects in Grade Eight Math on NAEP by School Type, 2009 to 2017



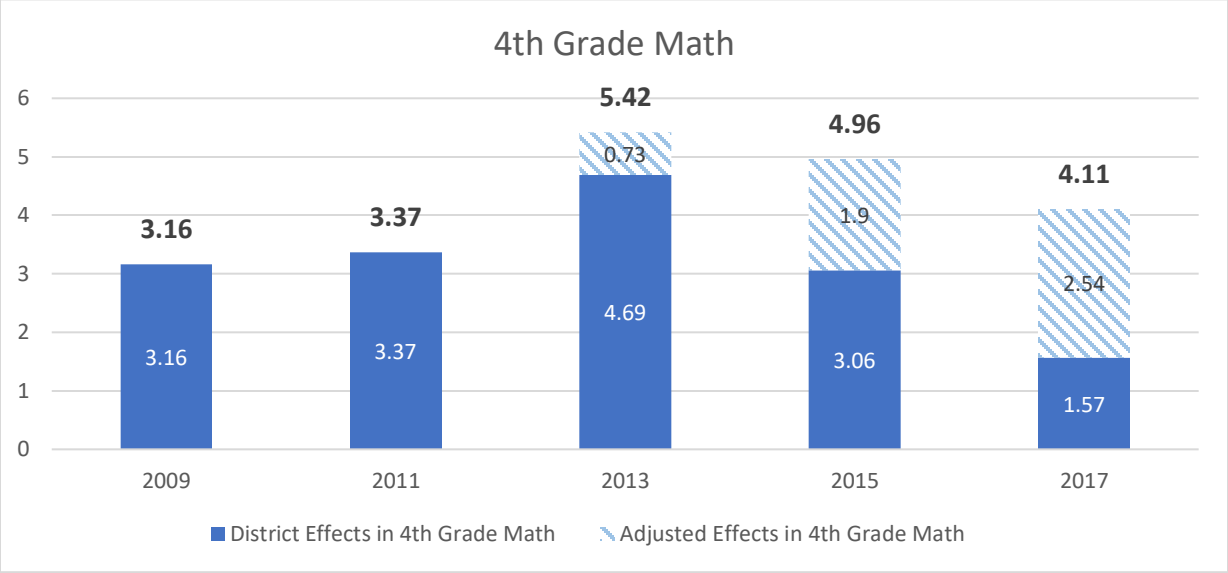
	2009	2011	2013	2015	2017
Large City	2.52	2.80	3.34	2.95	3.61
Not Large City National Public	1.07	1.12	1.62	1.72	2.19

*District effect is significantly different from zero.
‡ Includes district-authorized charters, charters authorized by others, and independent charters

(g) Combined District Effects and Median Adjustments

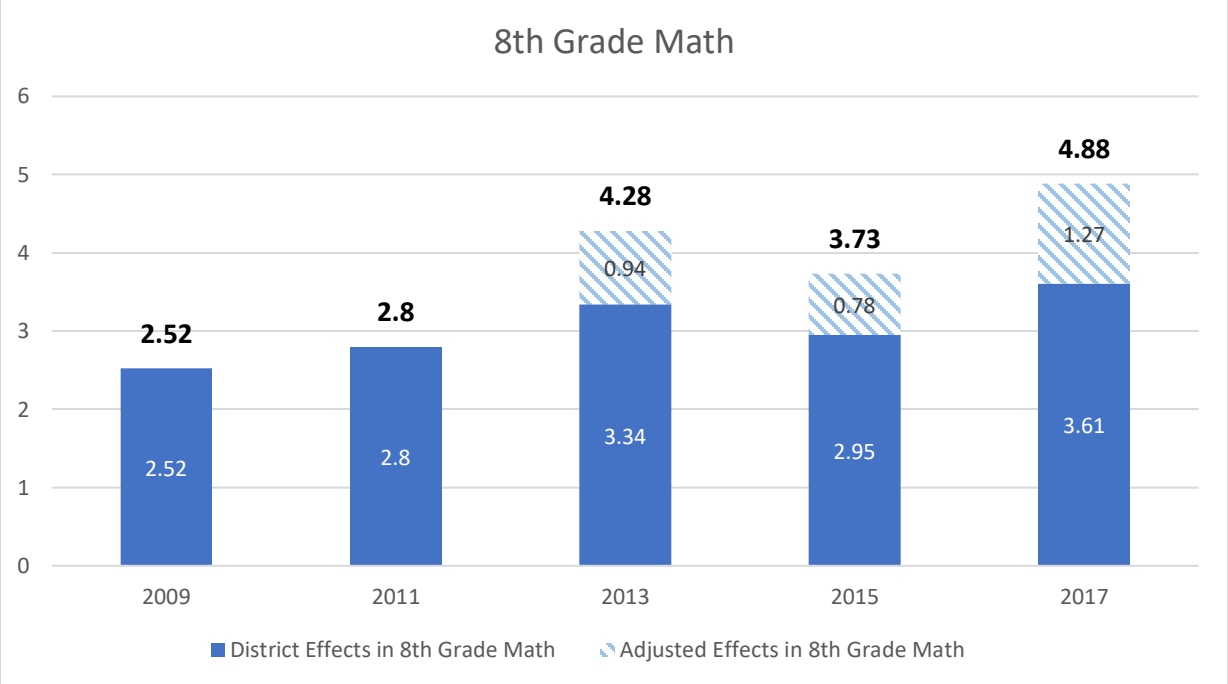
The combination of the aggregate district effects and the median adjustments to the effects allows one to see a clearer possible trend line in the performance of Large City schools. Large city schools saw their overall effects on fourth grade math improve steadily from 2009 through 2015 before dipping in 2017 and then rising again in 2019.

Exhibit 25. Combined District Effects in 4th Grade Math and Adjustments to Scale Scores



At the eighth-grade level, the possible trend line in math steadily improved between 2009 and 2017 once one took into account the adjustments to the scale scores.

Exhibit 26. Combined District Effects in 8th Grade Math and Adjustments to Scale Scores



Case Studies: How Districts Improved

Our next step was to go beyond identifying districts making outsized academic progress on NAEP to the *how.* How were some of these districts overcoming barriers and improving student achievement, and how can we apply these lessons more broadly? Are there approaches or strategies these districts are using that could inform the work of other major urban school systems?

To answer these questions, the Council embarked on a qualitative research effort to better understand the practices that might have driven the higher levels of performance and student growth observed in our statistical analysis. Between May 2018 and February 2019, the project team visited six districts: Boston Public Schools, Chicago Public Schools, the Dallas Independent School District, the District of Columbia Public Schools, Miami-Dade County Public Schools, and the San Diego Unified School District. Each of the districts were chosen for slightly different reasons, but all of them demonstrated results that were above expectations or results that showed substantial improvement between 2009 and 2017.

- Boston demonstrated consistent results in fourth and eighth grade reading and math that were well above statistical expectations in all areas. The district showed some fluctuation in scores between 2009 and 2017, but every year was significantly above expectations in both grades and subject areas.

- Chicago showed reading and math results in fourth and eighth grades that were above expectations in 2017. Moreover, Chicago was the only district that showed gains in district effects in all four grade/subject combinations. It was also one of the few districts that showed gains between 2009 and 2017 that went from below expectations to above. Performance at the eighth-grade level was consistently above expectations over the study period.

- Dallas showed reading and math results that were above expectations at the eighth-grade level. Notably, Dallas had unusually high rates of abject poverty compared to other city school systems that met or exceeded expectations.

- The District of Columbia had gains like those seen in Chicago. The district went from below expectations to above expectations between 2009 and 2017 in fourth grade reading and math. Results at the eighth-grade level were below expectations in both reading and math, but the district showed progress over the study period.

- Miami-Dade County also showed results that were above expectations in fourth and eighth grade reading and math in all years. The district demonstrated substantial gains in both subjects and grades over the study period.

- San Diego was one of the districts that showed gains from below expectations to above expectations in a grade/subject combination between 2009 and 2017. It also showed substantial gains in three grade/subject combinations. In 2017, San Diego was above expectations in all grades and subjects.

In addition, the team conducted multiple visits to a 'counterfactual' district. The Council selected this district to study based on its chronically low achievement and stalled progress. During the review, the Council team noted several clear contrasts between this district and the other six districts that helped put an even finer point on the patterns and practices we were observing in other sites. These contrasts were both striking and potentially informative for other districts seeking to address instructional challenges and make systemic improvements in teaching and learning. In addition, this report discusses commonalities across a number of districts whose results were below statistical expectations

After selecting these school districts, the Council's academic and research staff conducted site visits to each city. During each visit, the project team interviewed the superintendent, chief academic officer, director of research and assessment, director of professional development, and head of district turnaround efforts, as well as focus groups of curriculum staff and content area experts, coaches or other school support staff, principal supervisors, principals, and teachers. We reviewed organizational charts, strategic plans, professional development plans, and sample curriculum documents. In a handful of districts, we also visited schools and debriefed school and district leaders following our walk-throughs. Finally, the Council team analyzed an extensive array of data on each district, in addition to the data shown in the previous chapters, to better understand the nature and extent of district performance and improvement.

While the six case study districts had very different contexts and histories of reform, there were several common features and practices that appeared to be connected to the progress seen in student performance on NAEP across these cities. These shared factors included—

Strong and stable leadership focused on instruction

The relative stability of leadership was cited as a key factor in the progress made by several of the site-visit districts. At a time of increasing leadership turnover in districts throughout the country, the relatively long tenures of superintendents in districts such as Miami, where Alberto Carvalho has been superintendent since 2008, and San Diego, where Cindy Marten has been superintendent since 2013, has enabled these districts to pursue a consistent and sustained reform agenda over the years.

In Dallas, Superintendent Michael Hinojosa's first term spanned six years, from 2005 to 2011. Coming on the heels of a string of relatively short-lived leaders, this period was referred to by staff as a time of "instructional healing" in which the district was able to refocus its attention on teaching and learning and find the momentum necessary to drive instructional reform. When Hinojosa then returned to Dallas as superintendent in 2015, his historical knowledge of the district enabled him to quickly regain this momentum and continue moving the work forward. Staff in the district now commonly refer to his first and second terms as "Hinojosa 1.0" and "Hinojosa 2.0."

We also observed that the impact of strong, longstanding leaders can affect a district for years. In Boston, staff still cite the impact of Tom Payzant's 11 years as superintendent, and the culture of accountability that was built during that time.

Moreover, many of the districts benefitted from the stability of their curriculum and instruction leaders. The tenures of Janice Jackson, chief academic officer and then CEO of the Chicago Public Schools; Brian Pick, chief academic officer in the D.C. Public Schools; Marie Izquierdo, chief academic officer of the Miami-Dade County Public Schools; Ivonne Durant, chief academic officer in Dallas; and Linda Davenport, math director of the Boston Public Schools serve as examples. The longevity of their instructional leadership teams has allowed these districts to maintain a consistent instructional approach and to build on this approach over time even when there were transitions in the superintendents of those districts.

It is important to note, however, that it is not simply the *stability* of leadership that has yielded academic improvements in these cities, because one can find TUDA districts in our analysis where superintendent tenures were relatively long (i.e., over three years) and student achievement did not improve. Leaders in districts that did improve, on the other hand, brought strength, primacy, and focus to their instructional programming for a sustained period and allocated the time and resources necessary to improve it.

In fact, districts like DCPS, Chicago, and Boston showed us that progress can be maintained and even accelerated despite leadership churn if a district sustains its focus on instruction and retains its broad

instructional strategy.[7] In DC, which had five chancellors over some 12 years, there was both consistency and intentionality in the sequencing of reforms. Starting with Michelle Rhee in 2007, the focus of the district's reform efforts initially was on human capital, accountability, and building an effective teacher corps. This helped to create an overall environment where there was a perceived "brain gain"—talented people coming into the district because they saw an opportunity to turn around a once-failing system. Rhee's deputy and then successor Kaya Henderson expanded on this teacher-centered reform agenda. The district had reached a point where it had weeded out many of its weakest teachers, so the next step was to further enhance the capacity of the remaining teachers by equipping them with the necessary curricular resources, guidance, and training. Over the Henderson years this focus expanded with school-based structures, new materials, and the content expertise necessary to help teachers effectively implement the district's curricular resources. In other words, while the work evolved, each subsequent leader approached the district's past efforts and successes as an important foundation for their work, all the while remaining focused on what was needed to further improve instruction. Chicago offers another similar story of a district that has sustained and advanced its reforms across multiple superintendents.

This idea of strong leadership being defined by a focus on instruction prompted another big-picture observation. While in some districts the board of education was a full partner with the administration in improving district instruction, effectively supporting and monitoring district efforts to boost student achievement; in other places, boards appeared to add little value. Where they were partners in the work, the board and the superintendent were largely on the same page about the district's instructional vision and theory of action, and the board provided effective oversight and accountability for meeting the system's academic goals. In other cases, school boards were too focused on their own internal divisions and agendas to accelerate (or even impact) the administration's work to boost student outcomes. In these instances, the boards can take credit for hiring effective CEO's, but can take little credit for the academic gains that those superintendents and their staff attained.

Finally, in each of the districts we visited, strong, instruction-focused leadership was nurtured not only at the central office, but throughout the organization with the empowerment and support of principals and principal supervisors. In fact, several of the case-study districts reported that their instructional visions and theories of action were built, in part, around *school leaders as the levers of change*. As conduits between the district and schools, principal supervisors—in particular—were increasingly seen as critical to ensuring the success of this approach.

For example, when asked about factors driving district progress in Chicago, staff throughout the organization cited the fact that there was "genuine principal leadership" in the district. But the district took a more strategic approach than just deploying strong school leaders and hoping for district transformation. Principals were empowered to make decisions that were right for their communities—a situation that has been in place in Chicago since the late 1980s—but the district ensured *via* its new network structure and Network Chiefs that principals were sufficiently supported, coached, and held accountable for results. In other words, Chicago used its network structure and principal supervisors to realign its organizational structure around the instructional focus it wanted to achieve.

Area Superintendents in San Diego also described a strong, hands-on relationship with principals, meeting with them regularly throughout the year to review school-wide progress and help determine goals. In our interviews with the district leadership team, they told us that they believed it was the support and oversight structure of the school system that allowed for their site-based approach to work (when it doesn't necessarily work in other districts). "We don't need top down assessment to know if we are making progress because we have such a strong connection to schools through the Area Superintendents," they explained.

[7] This same lesson was learned some years ago in the Charlotte-Mecklenburg public schools, which had several superintendents but who all sustained the same overall academic theories of action.

Importantly, in order to ensure that principal supervisors are equipped to effectively advance school leadership and capacity in this way, their roles have been explicitly and intentionally redefined around instruction. Where in past years, principal supervisors oversaw a host of administrative and operational issues, these districts (and others across the country) have taken a number of steps (including narrowing spans of control, rewriting supervisor job descriptions, reallocating operational responsibilities to other staff or offices, and providing professional development in coaching) that fundamentally refocused their work with schools and principals around bolstering instructional effectiveness. In addition to Chicago and San Diego, Dallas, Miami, and the District of Columbia all did this to one extent or another.

High standards and common instructional guidance and support

It also appeared from our site visits that academic standards played a role in the improvement of some of the districts we examined. For instance, leadership of the Chicago and the District of Columbia public schools used the onset of college- and career-readiness standards to rethink and refocus their entire academic program. This was also at least partially the case with the Miami-Dade County schools.

The data suggest that there was also a distinctive "state effect" in places like Massachusetts, Florida, Texas, and North Carolina. Boston was a clear beneficiary of the state's historically high standards in addition to its own local efforts. This also appears to be the case in Miami, Hillsborough County, and Duval County. On the other hand, Dallas and other Texas cities did not adopt the academic standards that other states were putting into place, but they did make it clear what they wanted taught across their systems in ways that helped boost their overall academic performance.

This practice of better articulating what districts expected from their instructional programs was at the heart of their standards-based or curriculum reforms. Each of the districts we visited clearly communicated their instructional expectations at each grade level, including what high quality instruction and student work should look like. This was true regardless of whether they formally adopted the new standards or used a common district curriculum; all of them clarified their instructional expectations. For example, while San Diego does not have a traditional district curriculum, they do require each school to have a "Guaranteed Viable Curriculum"[8] that meets the district's requirements. They also lay out for schools the 'critical concepts' they expect to be covered at each grade level, and work with schools to develop units of study to ensure that this common understanding is employed in every classroom.

In another case, Miami-Dade County Public Schools provided teachers with detailed, standards-aligned pacing guides embedded with links to relevant instructional materials and resources. "*What* our children are going to learn is non-negotiable," explained an instructional leader in the district. But while the content was determined by the district, the "how" was left up to the classroom teacher, with more detail provided for those teachers who needed it. The district also provided a curated set of options in terms of instructional materials. This not only helped ensure the use of high quality, vetted materials, it also allowed the district to better support schools in using these materials. As one district staff member pointed out, "We can't support at scale if there is a cornucopia of materials."

Similarly, to drive instructional coherence and consistency in Dallas the central office releases instructional units every six weeks called Six Weeks at a Glance (SWAG). In addition to clearly laying out instructional expectations across core subjects over a six-week period, they are released six weeks *in advance* to allow teachers plenty of lead time to prepare. These units are accompanied by training sessions to provide teachers with a chance to dive into an upcoming unit, experience a modeled strategy, collaborate, and plan (although this training is on a voluntary basis). Teachers also can explore the SWAG and work through each unit in

[8] This concept was popularized by Robert Marzano in his book, "What Works in Schools" and refers to the pacing of how a curriculum is applied so that students have the opportunity to learn it. (Curriculum+opportunity-to-learn+time=A Guaranteed Viable Curriculum.)

their professional learning communities and have access to on-site coaching support and an online bank of videos of teachers using the lessons in classrooms.

Moreover, the district carefully monitors implementation through school and classroom visits, during which they look at whether a teacher is following the scope and sequence, what texts they have selected, and what strategies they are using with students. Since all district curriculum guidance and resources are online, lead staff members also have access to analytics that can tell them who is using the materials, what they are using, and which resources are used the most. Moreover, they field a user survey with every unit they publish and use the results and feedback they receive to further refine their guidance and support.

In DCPS, this unifying vision for instructional quality is referred to as "instructional oneness." The district provides principals with a clear picture—and even exemplars—of what high quality instruction should look like in the classroom. Teachers report getting more guidance than ever before. The teachers the Council team interviewed explained that in the past there had been a revolving door of textbooks and initiatives, with very little support or direction from the central office. Now, with the advent of IMPACT (the accountability system), LEAP (the district's teacher leadership development initiative), and resources such as an instructional video bank, they feel they understand the district's expectations and how to meet them.

In fact, DC is in the process of moving even more toward a centralized or normalized definition of its expectations for curriculum and instruction. In addition to a district curriculum, there are now required units of study and exemplars in each content area. As one instructional leader explained, while there was a shared district curriculum before, it looked drastically different from classroom to classroom and school to school. The district is therefore addressing this unevenness by ramping up the amount and content-specificity of its support for teachers.

Chicago is also moving toward a universal district curriculum, although schools will be able to opt out and use their own if they can show that it meets standards and is producing results. Like some of the other districts, the district provides schools with a curated set of instructional materials to choose from, and the guidance they need in selecting appropriate grade-level materials. The district has also created a "Knowledge Center"—an online clearinghouse with thousands of resources created by framework specialists. Unlike other online databases we've encountered, the district vets the materials that are posted to the Knowledge Center, ensuring that they are high quality and aligned to district standards.

This centralization of instructional expectations, resources, and guidance was described in more than one district as "autonomy with guardrails," and appeared to be based on the general acknowledgement that while pure site-based autonomy may work for some high performing districts with high capacity and experienced principals, it doesn't work for all districts and schools—and it doesn't always work everywhere or every time that systemic academic improvement is needed. This means that there needs to be greater definition, specificity, and support, as well as a norming of standards and instructional practice across all schools in a district in order to ensure higher quality and greater equity across a very mobile student body. At the same time, many districts grant increased autonomy to principals based on performance. Dallas, for example, defines their instructional approach as "managed instruction with earned empowerment." Chicago's approach is similar.

Moreover, although it is referred to here as "centralization," this standardization of instructional expectations is often described by central office staff as the district becoming more service oriented, and it has by and large led to greater support for schools in these districts. In Chicago, for example, staff report that "Supporting schools is our charge. Strategic planning revolves around the question, 'How is our work going to impact students/teachers?'" Another district leader pointed out that "the district's focus on what goes on in the classroom shouldn't be underestimated. Staffing, assignment, structure—ultimately what matters is what goes on in the classroom." The bottom line, in other words, was that empowerment without support, resources, and clear communication of district expectations won't drive growth on its own.

In fact, in San Diego, this service orientation led the superintendent to dismantle the two-sided structure of the system—operations vs. academics—in favor of a design that put principals at the center of the work. The message this structure was designed to convey was that everyone's chief responsibility is to support schools, principals, and teachers.

Teacher/leader quality

The strength of teachers and principals was another defining feature across the six districts, and the result of intentional district human capital strategies on the part of district leaders to boost the capacity of schools to make instructional improvements. In Boston, for example, high teacher pay likely contributes to both the high quality of teachers and low teacher turnover. In addition, the policy of mutual consent hiring (phased in around 2010) allowed school leaders more choice in selecting teachers, and it is credited with creating better matches between teachers and schools. In DCPS, as discussed previously, the first phase of the district's recent reform efforts was largely a human capital strategy, whereby weak teachers were removed and effective or potentially effective teachers were identified using the district's new evaluation system, IMPACT. The district subsequently transitioned into leadership development, although they acknowledge that this is an area they wish they had addressed earlier in the reform process.

The Chicago Public Schools, on the other hand, made the pivot toward a leadership development focus about eight years ago, putting them ahead of the curve. One of the most important changes they made was to introduce an additional layer of screening in addition to state certification to determine suitable principal candidates, who are then selected by parents and communities. This screening process has evolved over time, but it has remained a rigorous undertaking that requires candidates to present a portfolio of work, complete a written exam, and participate in a set of interviews where they are asked to respond to various scenarios and leadership challenges. According to district staff, this process has successfully raised the quality of the candidate pool, and it has enabled the district to imbed district-defined expectations, competencies, and beliefs about what makes a strong school leader into the selection process.

Similarly, in its human capital work Miami-Dade County first focused on strengthening its principal ranks and finding school leaders that reflected the district's priorities. The district also placed a special focus on the staffing and leadership of fragile schools. In the early phases of their reform work, the district identified effective teachers using a value-added measure charting progress over three to five years, and then recruited these teachers to work at struggling schools. They also moved other teachers out of these high-needs sites, at times using involuntary transfers.

Dallas's pay-for-performance model—the Teacher Excellence Initiative—also focuses on identifying the most effective teachers and paying them significantly more to work in high-need schools—specifically, the district's Accelerating Campus Excellence (ACE) schools. Moreover, the district mounted a systemwide effort to identify and deploy bilingual teachers as it built out its dual language model across the district.

In addition to these strategies aimed at recruiting, retaining, and effectively deploying high quality teachers and principals, many of the districts we visited focused on the development of teachers and future leaders. DCPS, for example, partners with outside organizations such as Relay Graduate School to support teacher candidate residencies in district schools, while Chicago established the Chicago Leadership Collaborative (CLC), a partnership between the district and leading principal development programs to create a pipeline of highly qualified leaders to meet the district's needs. Other districts, such as San Diego, also offer mentors to new principals, as well as providing teachers and vice principals with opportunities for growth and leadership roles at the school level.

In fact, Chicago's early focus on growing the leadership capacity of classroom, school, and network leaders has endowed them with a deep leadership "bench"— as evidenced by the fact that the district's current CEO, CAO, and many other chief positions have been filled internally with instructional staff who have

risen through the ranks and now bring a wealth of expertise and experience at multiple organizational levels to their roles as district leaders.

Professional Development and Other Capacity Building Measures

In addition to centralized curricular guidance and human capital strategies, the six study districts employed a variety of other strategies aimed at school-based capacity building. This can be seen, for example, in the reorientation of the role of principal supervisors (as discussed earlier in this chapter), as well as the widespread use of teacher leaders, school-based instructional leadership teams, building and network-level instructional coaches, and professional learning communities (PLCs) in most of the districts we visited.

School-based support structures such as instructional leadership teams and PLCs exist in many districts around the country. However, it is the level of intentionality and focus that really set the study districts apart. In Chicago, teachers described a transition during which they began getting clearer signals from the central office that school-level instructional leadership team meetings mattered, and schools became more accountable for selection, capacity building, and support of their teacher leaders. Chicago also employed PLCs and professional learning summits modeled after their common core implementation strategy of providing training and then employing teacher leaders to bring that training back to their buildings, providing site-based professional development tied to both school-level strategic plans and district strategic goals.

Miami-Dade County, meanwhile, hosts annual Synergy Summer Institutes, a week-long professional development course attended by teams of school staff. The institute is designed to provide these school leadership teams with the opportunity to study data together, reflect on current practices, identify the essential practices that should be sustained or enhanced during the upcoming school year, and take part in strategic planning to ensure continuous improvement at their school sites.

San Diego and Dallas had the most well-articulated PLCs we saw, which are closely monitored and supported by the district. In fact, in San Diego PLCs appear to have affected the whole culture of the school system and were cited by district and school staff alike as perhaps the most important factor driving the district's progress. As in Chicago, the evolution of PLCs was the result of intentional guidance and messaging from the central office. One principal, for example, described for the Council team the evolution of PLCs at her site from conversations about evaluation to sessions that are now devoted to collaborative problem-solving, providing her with an invaluable opportunity to work and learn alongside her teachers. According to district and school leaders, this structure helped the district drill down on Tier 1 instruction and its effectiveness.

Of course, just having PLCs in place is not enough to achieve instructional growth. Without clear guidance on what the district's expectations are for the time spent in PLCs and training on how to effectively lead collaborative, content-driven work sessions, PLCs in other systems often amount to glorified staff meetings rather than meaningful opportunities to improve teachers' instructional practice and build capacity at the school level.

Another unique and even somewhat counterintuitive strategy that serves to build school capacity in San Diego was the district's requirement that schools develop their own formative assessments. In past years when there was a district-mandated interim assessment, staff found that teachers would give it but not necessarily use the data. So, while this process took up a lot of schools' time to develop, they acknowledged that the process builds not only expertise, but ownership of formative assessment data where it was needed most. Of course, there were numerous guardrails in place. Area superintendents, for example, met with principals quarterly to review school-wide progress and help determine goals, and teachers received support in developing formative assessments through school-based instructional leadership team meetings, PLCs, and meetings with school and area leadership. The downside was that the district did not have the benefit

of aggregate assessment results over the course of the school year, but leadership concluded that its regular school and classroom monitoring gave them the information they needed.

In DCPS, LEAP (LEarning together to Advance our Practice) is another prime example of a district strategy for building school-based capacity. Through a weekly cycle of professional development in small, site-based, content-specific professional learning communities (LEAP Teams) led by content experts (LEAP Leaders), the district is aiming to develop on-the-ground expertise in teaching the DCPS Common Core-aligned curriculum.

In Dallas, meanwhile, principals and teachers cited the tremendous value of school-based support staff and structures such as Campus Instructional Coaches and Campus Instructional Leadership Teams (CILT) made up of principals, assistant principals, and core teachers. Yet while coaches and school-based instructional leadership teams were certainly not unique to this district, it is the level of support and structure that sets this district apart. The CILT teams in Dallas receive intensive, content-specific training with the academic department six times throughout the year to ensure that they are prepared to lead the learning at their respective campuses, while a corps of Instructional Lead Coaches serve as the "coaches of coaches," providing ongoing professional development and support for the campus-based coaches to ensure that the support that they, in turn, provide to teachers is consistent and aligned to the district's vision and standards for high-quality instruction.

Ultimately, the success of these capacity-building efforts was grounded in a common vision for instructional excellence, a clear set of expectations of what students should know and at what level of depth, and implementation that created ownership and buy-in among principals and teachers.

Acting at scale

Another similarity we observed across the case study districts was a shared belief that systemwide results could only come from systemwide change. Rollouts of reform initiatives, curricular materials, and programming (including implementation of college- and career-readiness standards) were therefore undertaken at scale in many – if not all— of these districts.

In Miami-Dade County, for instance, Superintendent Alberto Carvalho explained that he doesn't believe in pilots. His strategy for districtwide reform instead involved spending a lot of time planning, but then acting at scale in order to remove all vestiges of past practice. "If you want improvement at scale, act at scale (with deep planning)," he told the Council team. "The only way to overcome the gravitational pull of the status quo is to execute forcefully."

Of course, acting at scale took on many different dimensions across districts. In Miami-Dade County they phased in instructional reforms and new academic standards by grade level, but at scale across all schools. In Chicago, the rollout of the district's new literacy program was executed across the board, while in the area of mathematics they adopted a grade six through eight "bridge."

Importantly, the Council team concluded after visiting each of these districts that it was not only the scale of the work that ultimately determined their success, but the level of coherence and support for these rollouts that made the biggest impact. In this way instructional reform initiatives or new curricula adopted districtwide benefit from the shared focus and effort of staff throughout the organization working together toward common goals and expectations. This unifying instructional vision was critical in places like DCPS as they rolled out districtwide initiatives from the Cornerstone Units to LEAP. Similarly, in Boston the rollout of a new concept-rich core math program in 2000 was undergirded by a unifying instructional philosophy and sustained support, professional development, and oversight for implementation over several years. As noted in the 2011 Council report *Pieces of the Puzzl*e, the district's reading reforms did not benefit from the unanimity of approach observable in the district's (later) work in math.

> "The district's literacy program, which was built around a Reading and Writing Workshop (RWW) model during the study period, appeared to be less well- defined and less focused than the district's math reforms. In addition, the study team noted from interviews with teachers and district leaders that philosophical differences at the central office level over approaches to literacy instruction contributed to a lack of coherence in reading instruction districtwide…For example, while the district used its Reading First grants to adopt a common reading program for 34 of its schools—Harcourt's Trophies— most Boston schools had their choice of reading programs, and some opted out of using any specific published series. These differences led to a greater unevenness in reading program implementation than in math, according to interviewees who were asked directly about why math gains outstripped reading progress."[9]

Accountability and Collaboration

In a point related to teacher and leader quality, the rollout of accountability systems was cited as a key lever for change across the six study districts. As mentioned previously, the IMPACT system in DCPS was the centerpiece of the district's human capital strategy for building a stronger teacher corps. In addition to helping identify effective and ineffective teachers, this practice of holding everyone—including principals, assistant principals, instructional coaches, etc.—accountable for student growth reportedly helped to focus everyone on the primary goal of supporting instruction and to building an overall culture of responsibility.

In Dallas, which was a pioneer in the use of value-added data, growth in the use of classroom and school effectiveness indices played an important role in driving shared accountability for student results. Like IMPACT, these measures were controversial at first as they provided a quantitative measure of teacher effectiveness based on student achievement data. However, over time they became more accepted since they compared students in each classroom to other similar kids in the district. The classroom and school effectiveness indices are now used in the district's evaluation instruments for teachers and principals, as well as in the district's pay-for-performance initiative (TEI, or the Teacher Excellence Initiative).

Similarly, the school accountability system in Chicago was often the first factor cited by school leaders and staff in the district's progress. Interviewees reported that the evaluation tools for both teachers and principals took a deep look at what was happening in classrooms and measure success in terms of student growth. These evaluation tools in turn helped to norm the work of teachers and to create high standards and clear expectations for instruction across schools. In fact, everyone in the district is evaluated in some measure on student growth, and this has helped build a sense of urgency and shared responsibility for student progress.

Interestingly, this culture of accountability that has been built across districts has come hand in hand with increased collaboration. Leaders and staff in several of the sites the Council team visited discussed an intentional shift from competition to teamwork—a shift that could be seen in everything from how principal supervisors worked together with the curriculum department and other central office departments to the practice of connecting principals and teachers across schools. In Chicago, for example, staff reported that collegiality in general across the organization has improved drastically—despite several teacher strikes. They have seen the vertical and horizontal exchange of information increase dramatically, and a shift toward more inclusive, cross-functional strategic planning. Staff at both the central office and school levels report that "everyone is accessible—everyone returns calls."

This service orientation has in turn nurtured an environment of sharing lessons learned and resources across schools. The network chiefs (Chicago's principal supervisors) see it as part of their job to create opportunities for collaboration and to promote cross-pollination between schools and networks. The

[9] *Pieces of the Puzzle: Factors in the Improvement of Urban School Districts on the National Assessment of Educational Progress*. Council of the Great City School, 2011

Council team heard the same thing in Washington, DC, where instructional superintendents see the systemic sharing of lessons learned and effective practices as a key part of their role, describing themselves as "facilitators of the learning principals do with one another."

In fact, the Council team observed that the role of principal supervisors—discussed earlier in this chapter—was a key mechanism by which many districts helped further accountability, communication, and collaboration districtwide. Despite differences in organizational structure from district to district, principal supervisors served as a conduit between the central office and schools, allowing districts to communicate district standards, instructional expectations, and priorities while helping to identify which school sites required additional support and what opportunities existed for greater collaboration and sharing of effective practices.

In all, accountability in these districts is being redefined in these districts away from the more mechanistic, administrative accountability that one saw under the No Child Left Behind Act towards one that was oriented around a shared culture of responsibility for improving student outcomes.

Challenges as Opportunities

One interesting characteristic that we observed across many of the districts was the resilience and resourcefulness each district demonstrated in the face of change, challenge, or adversity. In Miami-Dade County, for example, the economic crisis of a decade ago is credited by district leaders as having "opened the door" to a wave of instructional and operational reforms, including greater centralization of curricular guidance and resources as a way to save on costs and support schools in the most effective and efficient manner.

This ability to respond constructively to new circumstances could perhaps be seen most clearly in the districts' responses to the adoption of new, rigorous academic standards in states across the country. Districts such as Boston, Chicago, Miami, and Washington D.C., for example, were among the earliest adopters of the Common Core State Standards or similar state-specific college- and career-readiness standards. San Diego even petitioned for a waiver from the California Standards Test (CST) so they could phase in the common-core-aligned SBAC (Smarter Balanced Assessment Consortium) ahead of other districts in the state.

Instructional leaders and staff at each site talked about seizing the opportunity provided by the standards to advance instructional coherence across the system. While some of these districts were already well underway in their instructional improvement efforts, the introduction of the common core or other college- and career-readiness standards helped these districts connect the work of supporting higher-quality instruction to assessment and evaluation. Interviewees also cited the value of the shared work and learning that came as staff throughout the organization unpacked and implemented the instructional shifts that the standards prescribed. In fact, the process of adopting districtwide standards was commonly described as having helped "even out" the support provided to teachers and principals across networks, as everyone worked to get onto the "same page" in terms of both common core content and pedagogy.

In each school district we visited, the successful implementation of college- and career-readiness standards was dependent on communication and close collaboration between the school management structure, the curriculum staff, and leaders at the central office. These districts worked cross-functionally to support implementation through multi-pronged strategies involving professional development, curriculum guidance and materials, instructional reviews, data reporting, and teacher and principal evaluation. Of course, standards alignment hasn't always led to student gains in other districts, and in one district leader's opinion this is because there is often not enough investment of time, effort, and resources in the implementation process. Progress, in other words, is not a function of declared alignment to rigorous standards, but of alignment in practice, which requires sustained monitoring and support to ensure that instructional changes made at the systems level reach all classrooms.

Support for struggling schools and students

Finally, some districts may have seen gains in part as a result of an explicit emphasis on support for struggling students, English learners, and students with disabilities. In Chicago, the district's implementation of multi-tiered systems of support (MTSS) and its efforts to support the examination of student-level data and the use of these data to inform strategies were likely factors in their progress on NAEP. In Miami, principals reported becoming more deliberate in their approach to reaching struggling students, as well as the increased use of disaggregated data and the development of strategies, interventions, and support based on understanding how different students learn.

The San Diego Unified School District has developed a particularly robust focus on individual students and the examination of student work. This is the result of a districtwide effort undertaken some years ago to study the experiences of struggling students at their schools and to identify what it revealed in terms of instructional and support needs. A significant part of the time teachers and administrators spend conducting school and classroom walk-throughs and in professional learning communities is now spent discussing individual students, looking at student work, and using these data to design lesson plans around the specific needs of the lowest performing students in each classroom for every lesson.

As compared to this somewhat common focus on struggling students, the Council team found that districts varied much more in their approach to struggling schools and school turnaround efforts. DCPS, for example, did not articulate a clear school turnaround strategy, instead focusing its efforts on programming and instruction systemwide—along with an effort targeted on African American male students.

In contrast, Dallas, the district in this study with the highest concentration of students in extreme poverty, has a particularly strong focus on resource allocation based on equity. The district uses an "intensity of poverty" index based on census block data to identify schools with particularly high needs, looking not only at poverty but generational poverty. A common sentiment echoed in conversations with staff throughout the organization was that "schools that need more should get more—in time, treasure, talent," and this could be seen in the district's emphasis on ensuring that struggling schools serving high numbers of poor students, African American students, and English learners received increased levels of campus-based support, additional resources, and effective teachers and principals.

A primary example of this resource allocation strategy in Dallas was the district's Accelerating Campus Excellence (ACE) initiative. The ACE initiative targeted the district's most historically failing schools—i.e., those with five years or more of not meeting state accountability requirements—and provided them with intensive additional resources that included strategic staffing (paying the most effective teachers to work at these schools *via* the district's pay-for-performance model TEI); prescriptive, data-driven instructional practices; increased monitoring and feedback; schoolwide systems for Social Emotional Learning; extended learning time; and investments in school and classroom upgrades.

In addition to this school-based strategy, Dallas also has a robust effort to improve the academic performance of its African American students, particularly its male students. The effort encompasses a combination of early childhood participation, staff diversification, strategic partnerships, single-gender schools, an African American studies program, mentoring, and enhanced instruction, along with other initiatives. The Dallas superintendent is held explicitly accountable on his annual evaluation for progress with these students. The district also has a parallel effort focused on Mexican American students.

In San Diego, meanwhile, the district identifies its highest needs schools as "focus schools." Oversight for these schools is distributed evenly—each area superintendent has six focus schools. And although district staff report that focus schools have the same level of autonomy as other sites, they also report spending more time at these schools, conducting more classroom walkthroughs, and working intensely with them in developing and sustaining their 'Guaranteed Viable Curriculum' and ensuring that the district's 'critical concepts' are covered at each grade level.

Miami also cites its focus on "fragile" schools—and the alignment of resources to meet student needs at these sites—as one of the main pillars of its district improvement strategy. In addition to deploying the most effective teachers and leaders to these schools, the district directs greater support and resources to these sites.

Moreover, Miami employs the unique strategy of pairing its support for struggling schools with its school choice initiative. Roughly 72 percent of Miami-Dade County students are now involved in a choice program of some sort, and students have over 1,000 choice options. Their approach, as described to the Council team, is to support struggling schools by increasing student engagement using niche programming. In other words, these schools and programs are designed specifically to appeal to parents, students, and communities, and district staff refer to this strategy as "demand-driven reform and innovation."

Like Miami-Dade County, Dallas also uses choice schools and programs to meet the needs of struggling schools, as well as to incentivize parents to remain in the district. There are currently waiting lists at each of the district's 25 P-TECH (Pathways to Technology Early College High School) and ECHS (Early College/Collegiate High Schools) campuses, and the district offers a range of other choice options, including over 50 two-way dual language schools and over 30 magnet school programs.

Finally, Miami also focuses efforts on its African American male students in a way that is like Dallas and the District of Columbia.

Community Investment and Engagement

Another notable feature of many of the school districts we visited was the active engagement and investment of community organizations, educational groups, foundations, businesses, and local colleges and universities—particularly in Boston, Chicago, and Miami.

Boston Public Schools, in particular, benefits from having a high concentration of educational institutions located in the city. School and district staff alike cited investments made in after- school and summer enrichment opportunities for area students as an important factor in students' progress and sustained achievement. One district leader estimated that some 80 percent of Boston students have benefitted from some sort of outside investment. This high concentration of colleges and universities also means a plethora of training programs and residencies for teacher candidates.

In Chicago there were similar investments in after-school activities and programs for kids. In addition, the school district's relationship with the UChicago Consortium on School Research ensures that district staff and leadership have access to a wealth of data on Chicago schools, and was cited as a key factor in helping the district sustain its commitment to its new accountability system, which was initially met with both internal and external resistance.

Miami also has an impressive array of community partners that the system relies on to provide support. The district has arranged hundreds of organizations and companies to provide summer intern opportunities for students, including offerings ranging from the American Dental Center to the Miami Arts & Academics Youth Summer Camp. The Miami-Dade County Public Schools also has a vast array of other community partners like the First National Bank of South Miami, American Airlines, and the Mexican American Council to provide support services.

While these partnerships and investments were critical sources of support and resources for city schoolchildren, what was equally important is that these districts were intentional about the investments made in – and on behalf of – their schools. Programs were vetted to ensure that they were consistent with district objectives and approaches, and staff dedicated time and focus to coordinating and connecting these investments so that schools weren't overwhelmed with redundant programming or mixed messaging on district instructional priorities.

A Counterfactual District

Although the counterfactual school district that the Council examined does not participate in TUDA, the Council's hands-on work with the district took place over roughly the same period as the team was conducting other site visits for this study. It should be noted that the purpose of the Council's multiple visits to this district—to provide technical assistance to a district that was struggling—was different from the research-driven visits made to Miami, Chicago, Boston, Washington D.C., San Diego, and Dallas. Since this time, the school district has hired a new superintendent and implemented several of the Council's recommendations to improve its programming and support for schools. Nonetheless, during the period of the study the team noted a number of clear contrasts—most notably in the areas of capacity building, instructional focus, and accountability—between this district, which has seen low and largely stagnant student achievement in recent years, and those that had seen growth. There are also likely parallels between this district and some of the urban school districts that were shown in our statistical analysis to fall below expectations.

Instructional Focus

Unlike the clear instructional vision and strategic, sequenced reforms we observed in the other districts, the counterfactual district appeared to lack a coherent strategy or working theory of action for improving student achievement districtwide or for moving failing schools out of that status. Although the district had a document called, "Theory of Action for Change, 2014," the Council team saw little evidence that it substantially drove the work of the district, and during the initial visit staff members that the team interviewed could not describe what the district's strategy was for improving academic performance.

Perhaps as a result of this imprecise instructional vision the district lacked the focus the other districts demonstrated on developing strong Tier 1 programming. Instead, the district was focused disproportionately on interventions with its lowest 25 percent of students. These interventions were ill-defined and differentially applied from school to school and from area to area within the district, and they were not evaluated for effectiveness. This strategy appeared to be done to garner extra state accountability points, but in doing so the district was missing an important segment of students—those between the lowest 25 percent and proficiency—and so even as an intervention strategy it was failing to move schools out of "failing" status. Moreover, it was undermining support for effective Tier 1 instruction to boost student achievement.

The district did employ learning walks, as we saw in other districts, but these appeared to be focused more on observing student engagement, classroom climate, and procedures than on the content and rigor of instruction. This contributed to the district's inability to monitor and improve the quality of instruction. In addition, the results of the walk-throughs did not appear to be used beyond the school to inform broader patterns of systemic needs or to improve districtwide strategies. In other words, the Council team saw no evidence that walk-through data were aggregated across schools, feeder patterns, or regions to inform broader systemwide improvements in curriculum, interventions, or professional development. The lack of district coherence was further evidenced by the fact that district network leaders each had a different set of strategies and plans for improving student achievement based only on their individual areas of expertise or experience. There also did not appear to be any districtwide resources or exemplars to guide instructional administrators and teachers about the level of rigor and student work expected in specific grade levels and content areas.

Capacity building

Perhaps the most conspicuous difference between the counterfactual district and the other districts we visited was in the area of capacity building. Whereas other districts invested time, energy, and focus on human capital strategies aimed at building up the quality of teachers and leaders, the counterfactual district made a number of decisions that ended up diluting the quality of their people, creating inconsistencies in

the district's instructional expectations, and limiting their capacity as a school system to support schools. For example, some years ago the school district's leadership decided to dismantle the school system's curriculum department in favor of outsourcing key instructional functions, like the development of curriculum materials, guidance, and some local testing activities. This not only left them beholden to outside vendors and responsible for an annual subscription fee for access to their own instructional materials, it deprived staff of the critical learning and capacity-building process of developing curriculum and providing instructional support and guidance to their own schools.

The district was working to re-establish its curriculum office when the Council team arrived, but the impact of this past decision was still evident. In our work with districts over the years we have observed that the strength of district staff and instructional leadership is critical to a school system's ability to adapt to challenges and move the system forward academically. So, while none of the districts we visited were immune to controversy or leadership turnover, this district was less equipped than other districts to weather the various upheavals it was facing.

Moreover, despite this history and severe funding shortages, the counterfactual district continued to rely heavily on outside vendors to provide materials and support services. While all of the districts we visited worked with outside vendors in some capacity, leadership and staff in the other case study districts explicitly cited a move away from "buying stuff to fix our problems," focusing their efforts and scarce funding instead on building internal capacity and investing in people. In the counterfactual district, however, the team ultimately concluded that the district's unusually high rates of teacher and staff turnover were likely due to the general lack of support for teachers, which is typically the reason why teachers leave. Moreover, while most of the other districts were intentional in their efforts to recruit and hire high quality teachers and leaders, this district lacked any sort of a teacher or leader pipeline program and the human resources department had delegated its primary function— identifying and hiring qualified teachers—to principals.

Accountability

A third main area of contrast between the counterfactual district and the other six study districts was in the area of accountability. Staff in each of the other districts spoke at length about a cultural shift toward shared accountability—a shift often founded on quantitative measures of student growth that held staff *throughout the organization* responsible for student progress. At the time of the Council's visit to this district, however, the district lacked any mechanism for holding personnel responsible for improving student academic outcomes. The personnel evaluation instrument that the district used was the Educator and Administrator Professional Growth System, which was the instrument endorsed by the State Department of Education as the framework for teacher and administrator evaluations. Principals, for example, were evaluated on five domains and 19 total elements. Each of these domains and elements included examples of evidence that could be used to demonstrate where principals were on a four-point evaluation scale, but none of the examples included actual student outcomes. The district's teacher evaluation systems also did not include concrete measures of student outcomes or progress. The district's procedure for evaluating central office administrative staff also graded performance across a series of domains and elements—none of which involved measures of districtwide student outcomes or their improvement.

This lack of accountability also marked the counterfactual district's relationship with its partners and vendors. On the other hand, staff in Miami-Dade County, for example, look at return-on-investment for all supplemental materials purchased and implemented by the district. Moreover, a set of district-developed 'Essential Questions' are sent to all vendors, who are required to show usage data and data on how they have met the promises and objectives they set out to accomplish. If they do not meet these criteria, the district doesn't renew the annual contract. Vendor accountability for results in the counterfactual district, conversely, was nearly non-existent.

Historical and Racial Context

In addition to issues of abject poverty, discussed earlier in this report, issues related to race, the historical legacy of discrimination, and urban investments may also inform the student performance levels analyzed and presented here. To be clear, the counterfactual district is not unique in this respect, but indicative of a history shared by cities across the country. This pattern is also found in Baltimore, Detroit, Milwaukee, and Philadelphia. In these and many other American "legacy" cities, African American communities were subject to sustained legal isolation, oppression, and a lack of investment that left many of these communities, which made up large segments of their respective cities, without the social and economic capital they needed to support educational progress.

The segregation and "redlining" of many African American communities in these cities over considerable time made it increasingly difficult for individuals of color to buy homes, borrow against the value of their homes, or start businesses or improve their properties. The result in many places was that owner occupancy was reduced, property values were lowered, housing quality slipped, and racial segregation increased. Many of these communities also saw the exit of grocery stores, gas stations, movie theaters, and banks that further isolated the communities and lowered the quality of life. In sum, the lack of investments in these cities left their communities without the wherewithal to compete with other better endowed locations.

This context clearly took a toll on schools in these communities and cities. The reduction in property values alone reduced the financial investment in schools; increased jobless rates meant that families were unable to provide the educational tools that many other families would have taken for granted; and violence that may have become endemic in some places made for learning climates that were suboptimal.

The names of the neighborhoods affected from city-to-city differed, but the effects were the same. Whether it was the Fairfield neighborhood in Baltimore; Forest Park in Detroit; Triangle North in Milwaukee; or Strawberry Mansion in Philadelphia, the systematic deprivation of resources and investment in these and other neighborhoods left schools and other institutions that residents rely on unable to serve and support them.

In this context, the inability of districts to make academic gains, demonstrated by district effects in 2017 that were substantially below what might be expected statistically, is hardly surprising. The track record of other major city school systems who share some of this same history suggests that more time is needed for the improvement process to take root. City school districts like Cleveland and Memphis, which have demographic characteristics similar to these four other cities but have been pursuing their current reform efforts for a longer period of time, have shown improvements over the years. Atlanta and the District of Columbia, moreover, have sustained their reform initiatives for even longer and both show substantial gains on the National Assessment of Educational Progress. This suggests that gains are possible once the right leadership and supports are in place. But it takes longer than a couple of school years to address the effects of centuries of accumulated oppression and disenfranchisement.

Time alone, of course, will not be sufficient, but in combination with the right set of improvement strategies like those outlined earlier in this report, sustained effort may be what these districts need in order to show gains.

Discussion and Conclusions

The ability of the nation's large urban school districts to overcome poverty, discrimination, language barriers, and other challenges is critical in the struggle to guarantee all students access to educational and social opportunity. It is therefore important, as urban educators, to examine the extent to which urban public schools are "beating the odds"—prevailing over these inequities to raise student achievement, rather than simply reflecting or perpetuating the opportunity gaps that exist across the country.

It is clear from our analysis that large city school systems are, in fact, doing a better job outpacing projected achievement and growth—i.e., adding value to the education of its students—compared to public schools *writ large*. Some big city school systems are more successful in this than others, but urban public schools in general are producing results that are greater than statistical expectations.

To be sure, not every urban school district that is beating the odds has followed the same path. We have observed different theories of action, varying approaches, and seemingly contrasting programming. These districts also present us with a wide array of different political, historical, and organizational contexts. Boston, for example, has benefitted from being in a high-performing state with consistently high standards. It has also seen mostly stable leadership over the years, retained their teachers longer than many districts and built their instructional capacity, worked to turn around some of it lowest-performing schools, and created and sustained a high-quality math program across the entire system.

Dallas has also benefitted from relatively stable leadership and clear academic goals and has built an accountability system based on those goals. In addition, Dallas has centrally defined its curriculum and instructional expectations, boosted professional development around those expectations, created performance incentives, built professional learning communities, focused on schools with cross-generational poverty, adroitly used its dual-language programming as both a parental incentive and as an instructional improvement strategy, and relied on exacting data to inform progress.

Miami-Dade County used many of the same strategies that one sees in Dallas. The district has enjoyed unusually long and successful leadership. It centrally defined its curriculum and instructional expectations and employed an "earned autonomy" theory of action. Like Dallas and Chicago, it acted at scale to get improvements at scale. Miami-Dade County also created a "value-added" system to identify its best teachers and incentivize them to teach in the most difficult schools. Moreover, the district expanded its Advanced Placement offerings, melded its choice offerings with its school turnaround efforts, developed strong lesson plans, boosted early-childhood programming, and used a very strong data system to boost performance.

In San Diego, the strategy looked substantially different. The district's leadership had not been stable until recently. It created a "leading from the middle" theory of action with no district chief academic officer or centralized curriculum, but it does have very well-articulated instructional expectations and one of the best-developed professional learning community systems we have ever seen. Those PLC's are long-standing and critical to the district's ability to boost staff capacity and set expectations for instructional quality.

In Chicago, the district used the onset of college- and career-readiness standards as a galvanizing event to rethink the quality of its instructional program and worked relentlessly on their district, regional, and school-based leadership to build a more coherent academic program. It also used professional learning communities, a longer school day, and a more centralized instructional program designed around the standards. Like other large urban districts like Boston, Dallas, and Miami-Dade County, Chicago implemented its reforms at scale in a staged manner that avoided isolated pilot programs. In addition, the system relied on good data systems and partnerships with external research groups to inform what was working and what wasn't, and it bolstered its overall accountability systems.

Finally, the District of Columbia used a mayoral takeover in the same way that Miami-Dade County used their extraordinary budget cuts and Chicago used new college- and career-ready standards as galvanizing events. The district used the work of two chancellors back-to-back to create leadership stability and pursued reforms in two differing stages: one devoted to human capital and the second defined around instructional excellence. It also created a more centralized instructional program and a clear set of high-quality academic expectations around which it built its professional development and instructional coaching. Unlike other districts that saw major gains, D.C. did not emphasize work in its lowest-performing schools or have a robust data system by which to inform progress at a tactical level. On the other hand, the district enhanced the quality of its instructional program and its teaching force in ways that many others did not.

The counterfactual district that Council staff examined had few if any of the strategies or reforms that the more successful districts had. Its leadership was unstable and weak; its organizational structure was incoherent; it had no system of accountability; its instructional program was poorly defined and did not clarify for teachers what was expected; and it had no way to enhance the capacity of its people to do the work. At the same time, the counterfactual district shared many of the same historic dynamics that other major city school systems struggling to get traction under their reforms show.

The findings from this report suggest several conclusions. One, any analysis of NAEP--or other student achievement results--that does not take into consideration the effects of poverty, race, ELL status, disability status, literacy materials in the home, and family education levels is likely to produce incomplete results and an only partial understanding of student attainment. The background variables used in this analysis explain around forty percent of the differences in student achievement scores on NAEP and provide substantial context to the results. Other variables, like the historic context of the cities, are not as measurable but surely as important.

Two, the data suggest that efforts to account for the effects of poverty using free or reduced-price lunch may fall short of capturing the full impact of abject and concentrated poverty on student outcomes. In addition, the free and reduced-price lunch data reported by various outlets is becoming substantially unstable and unusable. Researchers should be very careful in using those data without first questioning their stability over time. Moreover, it is clear from this analysis that districts with large percentages of students living in households with annual incomes below $10,000 and $50,000 face a more difficult set of challenges than other urban school systems in producing a "value-added" effect that is higher than statistical expectations.

Three, several TUDA districts demonstrated consistently that they were overcoming the influence of identified student background characteristics on achievement. Atlanta, Austin, Boston, Charlotte, Chicago, Dallas, Denver, Hillsborough County, and Miami-Dade County were among the districts that consistently out-performed expected levels.

Four, the data are clear that Large City schools—in the aggregate—are producing results on NAEP that exceed statistical expectations. Moreover, the data are clear that Large City schools are now producing results that generally exceed the ability of Not Large City schools to overcome the effects of the measured background characteristics.

Five, the data suggest that Large City Schools in half of the subjects/grades tested have gotten better at overcoming the effects of the background variables over time. In one subject/grade, there was no movement, and in one subject/grade combination there was slippage—fourth grade math.

Six, we wanted to put the changes in urban school performance in context, because we were unclear about whether the results urban schools were producing were better or worse than anyone one else. Does this mean that urban public schools have higher results than the average public school across the nation? No. The typical public school across the nation has higher NAEP scores than do the Large City schools. But the results do suggest that Large City schools do a better job of overcoming the effects of poverty, language,

discrimination, disability, and differences in family education than the average school does. Put another way, urban public schools appear to produce greater instructional torque than does the typical school.

We should be clear that none of the improving districts we have described in this report have reached the promised land. Much of their reforms are a work in progress. And while there were some key similarities among the districts we studied, there was no single, shared strategy or formula that can be definitively tied to their gains. If there was a "secret sauce," it was that these districts used varying theories of action, strategies, and programs to do one fundamental thing: improve the quality of instruction in their classrooms. This central endeavor was often aided by stable leadership, clear curricular expectations, aligned organizational structures, defined and shared accountability systems, and capacity-building mechanisms. But each of these components were employed in the service of improving instruction—something we do not always see in other districts.

This central finding is like findings in two previous iterations of studies conducted by the Council on why and how some urban school systems improve faster than others. This new study asks a more complicated set of questions than do those earlier studies, but the results are remarkably consistent. Large City schools have not overcome the barriers before them entirely, otherwise results would be even higher, but the data in this study suggest that Large City schools are doing a better job of overcoming the effects of poverty and potentially moving student out of that status and into the middle class than most schools.. We hope that this study not only provides a useful and timely exploration of what and how districts are beating the odds, but points the way to how more progress can be made in overcoming—to some degree—the inequities experienced by too many of our nation's urban schoolchildren.

Over the last decade, large city school districts have narrowed the achievement gap with the nation at large, but what is new here is that urban public schools are doing a better job of overcoming the effects poverty, English language proficiency, and other factors that often limit student outcomes. To be sure, there is a great deal of work to be done, but urban public schools are doing a better job of opening the windows of opportunity rather than simply mirroring the inequities that students so often face.

Appendices

Appendix A

District/Jurisdiction Actual Scaled Score, Expected Scaled Score, and "District Effects" for 2009, 2011, 2013, 2015

Exhibit A-1. Grade Four Reading Actual Performance, Expected Performance, and District Effects, 2009

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Atlanta	209.16	203.71	5.45
Austin	220.35	211.01	9.34
Baltimore	201.99	203.61	-1.62
Boston	215.02	202.76	12.26
Charlotte	224.51	215.80	8.72
Chicago	202.19	202.78	-0.58
Cleveland	193.75	200.61	-6.86
Detroit	187.27	195.07	-7.80
District of Columbia (DCPS)	203.46	207.44	-3.98
Fresno	197.28	202.86	-5.59
Houston	211.39	203.25	8.14
Jefferson County	219.43	214.78	4.65
Los Angeles	197.41	200.15	-2.74
Miami	221.16	209.93	11.23
Milwaukee	196.02	202.73	-6.71
New York City	216.81	205.05	11.76
Philadelphia	195.18	201.31	-6.13
San Diego	212.83	211.81	1.02
Large City Schools*	210.04	207.79	2.25
Not Large City Schools□	221.43	220.98	0.45

* Includes district-authorized charters

‡ Includes charters authorized by others and independent charters

Exhibit A-2. Grade Eight Reading Actual Performance, Expected Performance, and District Effects, 2009

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Atlanta	249.95	245.65	4.30
Austin	261.63	254.68	6.95
Baltimore	245.90	245.67	0.24
Boston	257.78	248.61	9.17
Charlotte	259.92	257.36	2.56
Chicago	249.50	245.52	3.98
Cleveland	242.60	240.59	2.01
Detroit	233.32	238.49	-5.17
District of Columbia (DCPS)	241.26	246.37	-5.11
Fresno	240.11	243.99	-3.88
Houston	252.21	247.47	4.74
Jefferson County	258.56	259.51	-0.95
Los Angeles	244.39	243.32	1.07
Miami	260.94	254.38	6.56
Milwaukee	241.70	244.33	-2.63
New York City	253.15	250.57	2.59
Philadelphia	247.39	245.05	2.34
San Diego	254.89	256.04	-1.14
Large City Schools*	252.36	251.85	0.52
Not Large City Schools□	264.09	265.09	-1.00

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-3. Grade Four Mathematics Actual Performance, Expected Performance, and District Effects, 2009

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Atlanta	225.35	222.05	3.29
Austin	240.46	228.33	12.13
Baltimore	222.21	222.55	-0.34
Boston	236.23	224.15	12.08
Charlotte	244.94	234.75	10.19
Chicago	221.88	224.25	-2.37
Cleveland	213.48	219.25	-5.77
Detroit	199.76	214.71	-14.95
District of Columbia (DCPS)	219.99	223.79	-3.80
Fresno	218.93	225.49	-6.55
Houston	235.79	222.77	13.02
Jefferson County	232.83	233.70	-0.87
Los Angeles	221.90	223.90	-2.01
Miami	236.34	228.73	7.61
Milwaukee	219.93	222.86	-2.93
New York City	237.47	226.86	10.61
Philadelphia	221.57	221.50	0.07
San Diego	236.30	233.33	2.96
Large City Schools*	231.32	228.16	3.16
Not Large City Schools□	240.61	239.40	1.21

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-4. Grade Eight Mathematics Actual Performance, Expected Performance, and District Effects, 2009

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Atlanta	259.52	257.63	1.88
Austin	287.55	271.17	16.38
Baltimore	257.64	261.72	-4.08
Boston	280.45	263.58	16.87
Charlotte	282.77	275.36	7.41
Chicago	263.88	261.85	2.03
Cleveland	256.00	253.98	2.03
Detroit	238.95	250.62	-11.67
District of Columbia (DCPS)	251.87	259.44	-7.57
Fresno	258.76	264.19	-5.43
Houston	276.89	263.70	13.20
Jefferson County	271.28	273.66	-2.38
Los Angeles	258.73	263.40	-4.67
Miami	273.05	269.77	3.28
Milwaukee	251.80	258.69	-6.89
New York City	274.73	266.54	8.20
Philadelphia	264.80	259.93	4.87
San Diego	280.38	278.27	2.11
Large City Schools*	271.17	268.65	2.52
Not Large City Schools□	283.58	282.51	1.07

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-5. Grade Four Reading Actual Performance, Expected Performance, and District Effect, 2011

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	208.92	213.47	-4.55
Atlanta	211.62	206.73	4.89
Austin	223.63	212.41	11.21
Baltimore	200.50	204.36	-3.87
Boston	217.00	199.83	17.17
Charlotte	224.19	217.05	7.14
Chicago	203.27	202.64	0.63
Cleveland	192.54	195.53	-2.99
Dallas	203.66	197.19	6.47
Detroit	191.00	195.86	-4.86
District of Columbia (DCPS)	201.02	205.51	-4.49
Fresno	194.27	201.98	-7.71
Hillsborough County	230.83	213.83	17.01
Houston	213.04	203.84	9.21
Jefferson County	222.79	217.37	5.42
Los Angeles	200.60	203.25	-2.65
Miami	221.01	208.14	12.86
Milwaukee	195.49	200.76	-5.27
New York City	216.39	205.58	10.81
Philadelphia	198.75	201.61	-2.86
San Diego	215.41	211.42	3.99
Large City Schools*	210.90	208.27	2.63
Not Large City Schools□	221.83	221.47	0.36

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-6. Grade Eight Reading Actual Performance, Expected Performance, and District Effects, 2011

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	254.33	257.40	-3.07
Atlanta	252.66	248.45	4.21
Austin	261.95	256.96	4.99
Baltimore	246.61	249.27	-2.66
Boston	255.14	248.73	6.41
Charlotte	264.90	260.75	4.15
Chicago	253.19	247.84	5.35
Cleveland	240.51	239.57	0.94
Dallas	247.65	244.99	2.67
Detroit	237.03	240.03	-3.00
District of Columbia (DCPS)	237.99	246.95	-8.96
Fresno	238.32	247.06	-8.73
Hillsborough County	264.51	259.21	5.30
Houston	252.81	249.86	2.95
Jefferson County	259.94	261.25	-1.31
Los Angeles	246.59	248.97	-2.38
Miami	260.06	255.24	4.82
Milwaukee	239.04	242.96	-3.92
New York City	255.09	249.94	5.15
Philadelphia	247.43	245.95	1.48
San Diego	256.76	257.42	-0.66
Large City Schools*	254.58	253.93	0.65
Not Large City Schools□	265.32	266.19	-0.88

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-7. Grade Four Mathematics Actual Performance, Expected Performance, and District Effects, 2011

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	235.47	233.50	1.97
Atlanta	228.14	225.26	2.88
Austin	245.39	229.90	15.49
Baltimore	225.59	223.73	1.86
Boston	237.24	222.71	14.53
Charlotte	246.86	236.13	10.74
Chicago	223.76	225.53	-1.77
Cleveland	215.82	218.04	-2.22
Dallas	232.83	219.61	13.22
Detroit	203.17	216.52	-13.35
District of Columbia (DCPS)	221.82	226.90	-5.07
Fresno	217.74	224.84	-7.10
Hillsborough County	243.33	234.22	9.10
Houston	237.04	224.60	12.44
Jefferson County	235.24	235.59	-0.35
Los Angeles	223.26	226.50	-3.24
Miami	235.51	228.82	6.69
Milwaukee	219.55	223.14	-3.59
New York City	234.46	228.31	6.15
Philadelphia	225.31	223.25	2.06
San Diego	238.94	233.43	5.51
Large City Schools*	232.89	229.52	3.37
Not Large City Schools□	241.55	240.39	1.16

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-8. Grade Eight Mathematics Actual Performance, Expected Performance, and District Effects, 2011

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	275.11	273.50	1.61
Atlanta	265.99	262.75	3.24
Austin	287.38	272.41	14.97
Baltimore	261.54	262.86	-1.31
Boston	282.14	265.39	16.75
Charlotte	285.46	277.20	8.25
Chicago	270.50	265.47	5.03
Cleveland	256.10	254.10	1.99
Dallas	274.29	261.23	13.06
Detroit	246.46	253.82	-7.36
District of Columbia (DCPS)	256.21	263.19	-6.97
Fresno	256.62	265.67	-9.05
Hillsborough County	282.26	276.25	6.01
Houston	279.54	267.08	12.45
Jefferson County	274.46	275.79	-1.33
Los Angeles	261.04	267.09	-6.05
Miami	271.86	271.37	0.48
Milwaukee	254.40	259.40	-4.99
New York City	272.67	267.52	5.15
Philadelphia	265.28	262.95	2.33
San Diego	278.73	277.57	1.16
Large City Schools*	273.97	271.17	2.80
Not Large City Schools‡	284.39	283.27	1.12

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-9. Grade Four Reading Actual Performance, Expected Performance, and District Effects, 2013

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	206.55	209.00	-2.44
Atlanta	214.28	207.60	6.68
Austin	220.81	209.44	11.37
Baltimore	204.26	205.76	-1.50
Boston	214.40	200.02	14.38
Charlotte	226.44	217.80	8.64
Chicago	206.15	205.15	1.00
Cleveland	189.66	194.78	-5.11
Dallas	204.65	194.89	9.76
Detroit	189.71	194.06	-4.35
District of Columbia (DCPS)	205.73	205.94	-0.21
Fresno	195.85	201.75	-5.89
Hillsborough County	227.86	214.22	13.64
Houston	207.83	200.79	7.04
Jefferson County	220.94	216.81	4.13
Los Angeles	204.85	206.04	-1.20
Miami	223.11	207.59	15.52
Milwaukee	198.71	201.54	-2.83
New York City	216.27	208.36	7.91
Philadelphia	199.93	202.38	-2.45
San Diego	217.77	213.11	4.66
Large City Schools*	212.43	208.72	3.71
Not Large City Schools□	222.39	221.47	0.92

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-10. Grade Eight Reading Actual Performance, Expected Performance, and District Effects, 2013

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	256.42	255.25	1.17
Atlanta	254.87	250.14	4.73
Austin	261.72	257.91	3.81
Baltimore	252.52	249.93	2.59
Boston	257.22	246.95	10.27
Charlotte	266.99	262.90	4.09
Chicago	253.75	250.58	3.17
Cleveland	239.25	239.20	0.05
Dallas	251.67	245.05	6.62
Detroit	239.61	241.05	-1.44
District of Columbia (DCPS)	245.55	247.48	-1.93
Fresno	245.40	250.33	-4.93
Hillsborough County	267.50	261.95	5.55
Houston	252.50	250.20	2.30
Jefferson County	260.69	262.54	-1.85
Los Angeles	250.18	253.20	-3.02
Miami	259.16	255.84	3.32
Milwaukee	242.74	244.74	-2.00
New York City	256.78	252.01	4.76
Philadelphia	248.72	247.69	1.03
San Diego	259.97	261.05	-1.08
Large City Schools*	257.63	256.31	1.31
Not Large City Schools‡	267.54	268.08	-0.54

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-11. Grade Four Mathematics Actual Performance, Expected Performance, and District Effects, 2013

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	234.53	231.76	2.78
Atlanta	233.10	226.79	6.31
Austin	244.97	231.49	13.47
Baltimore	222.87	221.46	1.41
Boston	236.87	224.36	12.51
Charlotte	247.35	237.18	10.18
Chicago	230.50	227.87	2.63
Cleveland	216.27	216.51	-0.24
Dallas	234.22	219.79	14.43
Detroit	204.25	215.82	-11.56
District of Columbia (DCPS)	228.61	226.48	2.13
Fresno	219.69	225.36	-5.67
Hillsborough County	242.80	235.07	7.73
Houston	235.90	224.52	11.38
Jefferson County	233.70	234.99	-1.29
Los Angeles	228.46	229.56	-1.10
Miami	237.40	229.46	7.94
Milwaukee	221.45	224.16	-2.71
New York City	235.84	231.48	4.36
Philadelphia	223.38	225.03	-1.64
San Diego	240.88	235.53	5.35
Large City Schools*	234.96	230.27	4.69
Not Large City Schools‡	242.49	240.31	2.18

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-12. Grade Eight Mathematics Actual Performance, Expected Performance, and District Effects, 2013

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	273.89	272.36	1.53
Atlanta	267.19	262.53	4.66
Austin	285.00	273.53	11.47
Baltimore	260.72	259.13	1.58
Boston	283.76	261.57	22.19
Charlotte	289.43	278.36	11.07
Chicago	269.29	266.22	3.07
Cleveland	253.26	251.44	1.82
Dallas	274.84	260.95	13.90
Detroit	240.00	251.63	-11.62
District of Columbia (DCPS)	260.76	262.16	-1.40
Fresno	260.05	267.85	-7.80
Hillsborough County	284.07	276.82	7.24
Houston	280.70	265.92	14.78
Jefferson County	273.57	276.21	-2.64
Los Angeles	264.90	270.80	-5.90
Miami	273.98	271.75	2.23
Milwaukee	257.62	258.27	-0.65
New York City	274.11	268.72	5.39
Philadelphia	267.03	262.02	5.02
San Diego	277.54	279.61	-2.07
Large City Schools*	275.52	272.18	3.34
Not Large City Schools□	285.08	283.46	1.62

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-13. Grade Four Reading Actual Performance, Expected Performance, and District Effects in 2015

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	206.88	213.01	-6.13
Atlanta	212.18	208.99	3.18
Austin	220.02	210.29	9.72
Baltimore	199.07	203.17	-4.10
Boston	219.46	204.72	14.75
Charlotte	225.58	219.00	6.58
Chicago	213.09	208.62	4.47
Cleveland	196.81	196.04	0.77
Dallas	213.91	200.10	3.94
Detroit	204.03	194.90	-8.45
District of Columbia (DCPS)	186.45	210.79	3.12
Duval County	225.27	218.61	6.66
Fresno	198.95	204.36	-5.42
Hillsborough County	229.65	216.35	13.30
Houston	209.55	204.27	5.28
Jefferson County	221.95	216.89	5.06
Los Angeles	204.43	210.13	-5.70
Miami	226.41	213.51	12.90
Milwaukee	--	--	--
New York City	214.01	209.60	4.41
Philadelphia	200.53	206.02	-5.48
San Diego	215.91	212.59	3.32
Large City Schools*	213.63	211.64	1.99
Not Large City Schools□	223.04	222.21	0.83

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-14. Grade Eight Reading Actual Performance, Expected Performance, and District Effects in 2015

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	251.66	257.58	-5.92
Atlanta	252.87	250.21	2.66
Austin	262.14	258.27	3.87
Baltimore	244.27	245.28	-1.01
Boston	258.71	249.78	8.94
Charlotte	263.86	262.06	1.80
Chicago	257.15	250.02	7.13
Cleveland	240.79	241.09	-0.31
Dallas	245.83	243.19	7.11
Detroit	250.30	240.30	-2.51
District of Columbia (DCPS)	237.79	247.73	-1.90
Duval County	264.39	263.31	1.08
Fresno	242.51	251.96	-9.45
Hillsborough County	261.54	258.05	3.50
Houston	252.02	251.45	0.57
Jefferson County	261.83	260.65	1.17
Los Angeles	251.28	253.32	-2.04
Miami	265.22	257.15	8.07
Milwaukee	--	--	--
New York City	258.61	254.51	4.10
Philadelphia	248.65	250.31	-1.65
San Diego	262.29	261.54	0.74
Large City Schools*	256.65	255.57	1.07
Not Large City Schools‡	265.47	265.55	-0.08

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-15. Grade Four Mathematics Actual Performance, Expected Performance, and District Effects in 2015

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	230.58	233.28	-2.70
Atlanta	228.09	226.03	2.06
Austin	246.14	231.58	14.56
Baltimore	214.91	222.18	-7.27
Boston	235.53	226.52	9.01
Charlotte	247.82	236.12	11.70
Chicago	231.92	228.06	3.86
Cleveland	219.15	217.10	2.05
Dallas	237.93	222.18	15.75
Detroit	204.64	214.95	-10.31
District of Columbia (DCPS)	232.24	228.80	3.44
Duval County	242.80	236.06	6.74
Fresno	217.68	225.19	-7.51
Hillsborough County	243.61	235.27	8.34
Houston	238.71	225.51	13.20
Jefferson County	235.75	234.15	1.61
Los Angeles	224.19	230.47	-6.28
Miami	242.10	231.89	10.21
Milwaukee	--	--	--
New York City	231.05	229.84	1.21
Philadelphia	217.45	225.12	-7.67
San Diego	232.76	233.82	-1.06
Large City Schools*	233.99	230.93	3.06
Not Large City Schools‡	241.14	239.84	1.30

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters

Exhibit A-16. Grade Eight Mathematics Actual Performance, Expected Performance, and District Effects in 2015

TUDA/ Jurisdiction	Actual Mean	Expected Mean	District Effect
Albuquerque	270.85	272.14	-1.29
Atlanta	266.59	264.14	2.45
Austin	284.34	274.18	10.17
Baltimore	256.07	256.81	-0.73
Boston	282.46	263.27	19.20
Charlotte	286.57	277.71	8.86
Chicago	275.32	265.40	9.92
Cleveland	254.62	251.92	2.70
Dallas	271.20	259.05	12.15
Detroit	244.69	251.21	-6.52
District of Columbia (DCPS)	260.84	261.35	-0.51
Duval County	274.90	275.92	-1.02
Fresno	257.32	266.73	-9.41
Hillsborough County	276.04	273.57	2.48
Houston	276.63	266.58	10.05
Jefferson County	271.92	275.45	-3.53
Los Angeles	264.01	268.95	-4.94
Miami	274.74	272.24	2.50
Milwaukee	--	--	--
New York City	276.67	269.99	6.68
Philadelphia	267.50	263.57	3.93
San Diego	281.26	278.59	2.67
Large City Schools*	273.78	270.83	2.95
Not Large City Schools□	282.80	281.08	1.72

* Includes district-authorized charters
‡ Includes charters authorized by others and independent charters