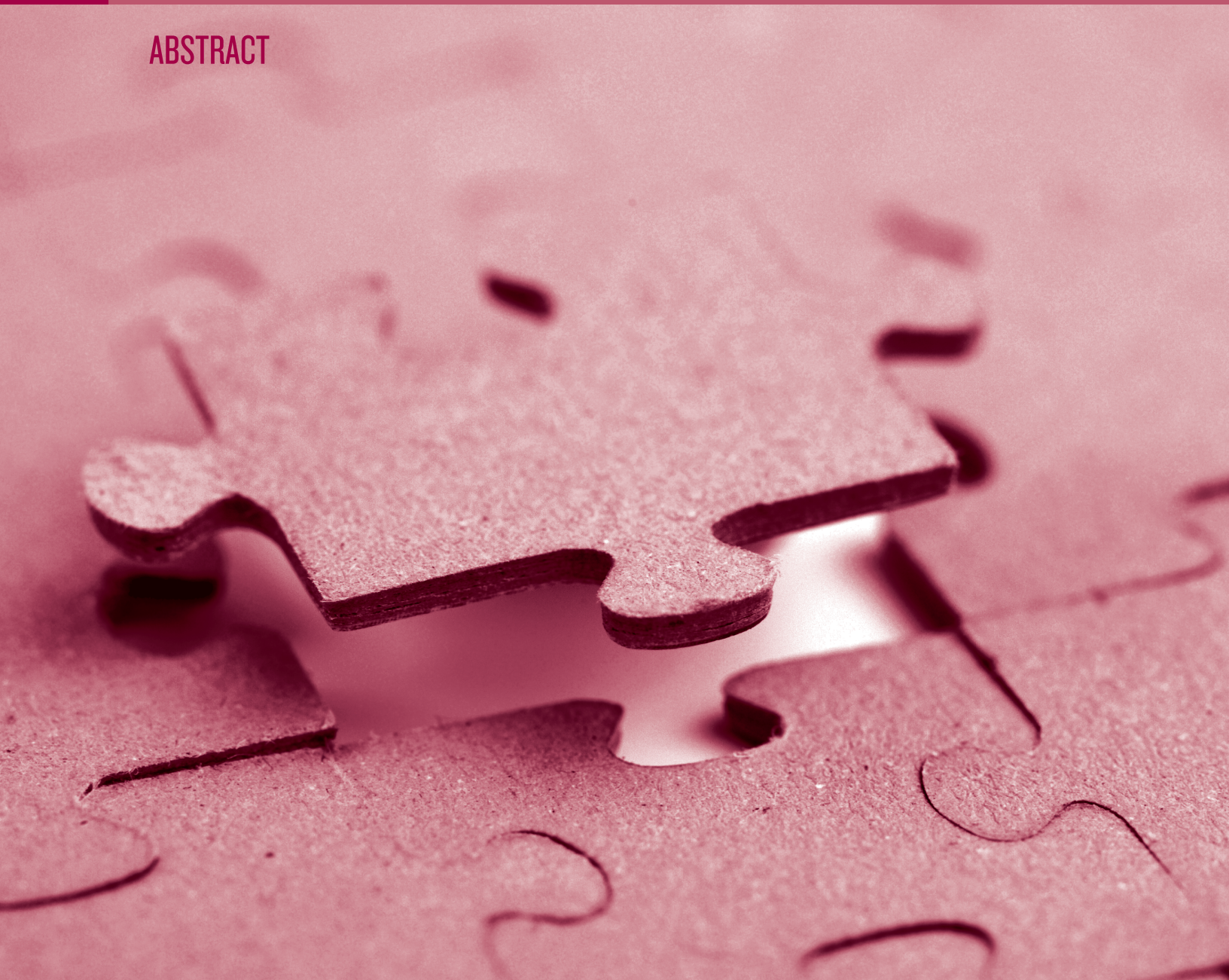


PIECES OF THE PUZZLE

FACTORS IN THE IMPROVEMENT OF URBAN SCHOOL DISTRICTS ON
THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

ABSTRACT



The Council of the Great City Schools

Fall 2011

Abstract

Pieces of the Puzzle

Factors in the Improvement of Urban School Districts on
the National Assessment of Educational Progress

**Council of the Great City Schools and the American Institutes for Research
Fall 2011**

Authors

**Michael Casserly
Ricki Price-Baugh
Amanda Corcoran
Sharon Lewis
Renata Uzzell
Candace Simon**

Council of the Great City Schools

**Jessica Heppen
Steve Leinwand
Terry Salinger
Victor Bandeira de Mello
Enis Dogan
Laura Novotny**

American Institutes for Research

The Council of the Great City Schools thanks The Bill & Melinda Gates Foundation for supporting this project. The findings and conclusions presented herein do not necessarily represent the views of The Foundation.

ACKNOWLEDGMENTS

This report is the product of exceptional teamwork and involved the considerable expertise of both high-quality researchers and experienced practitioners in a mixed-methods analysis of why and how big-city public school systems show progress on the National Assessment of Educational Progress (NAEP). It is the first study of its kind using NAEP, but it will surely not be the last.

I thank Ricki Price-Baugh, the Director of Academic Achievement at the Council of the Great City Schools, for her leadership of the project. Her broad conceptual skills and keen eye for detail were invaluable in the evolution of the study.

The Council's team was also fortunate to have the expertise of Sharon Lewis, the Council's Director of Research, and her team of research managers—Amanda Corcoran, Renata Uzzell, and Candace Simon. Each one played a critical role in analyzing data, reviewing results, and drafting chapters. Thank you.

The team from the American Institutes for Research, led by Jessica Heppen, was terrific in managing and conducting data analysis. Dr. Heppen's expertise was indispensable in keeping the project moving forward and coordinating the endless details a project of this complexity entails. She was joined in the work by Terry Salinger, who had lead responsibility for the reading analysis; Steve Leinwand, who led the work on mathematics; and Laura Novotny, who led the science analysis. Victor Bandeira de Mello and Enis Dogan rounded out the AIR team with their extraordinary technical skills in the analysis of NAEP data.

The ability of the Council and the AIR teams to work together and to test and challenge each other's analyses and conclusions was a unique and critical element of the project's success.

I also thank the research advisory group that provided important guidance to the project as it was getting underway. It consisted of top-flight researchers and practitioners: Peter Afflerbach, professor of education at the University of Maryland; Robin Hall, a principal and an executive director in the Atlanta Public Schools; Karen Hollweg, former director of K-12 science education at the National Research Council; Andrew Porter, dean of the Graduate School of Education at the University of Pennsylvania; Norman Webb, senior research scientist at the Wisconsin Center for Educational Research; and Karen Wixson, professor of education at the University of Michigan.

Finally, I thank Vicki Phillips, director of education at The Bill & Melinda Gates Foundation, for the foundation's generosity in supporting this research. And I thank Jamie McKee, who served as the foundation's program officer and who provided invaluable guidance, advice, and support throughout the project. Thank you.

Michael Casserly
Executive Director
Council of the Great City Schools

Acknowledgments.....00

Chapter 1. Introduction00

Chapter 2. Methodology00

Chapter 3. Analysis of NAEP Results, Trends, and Alignment for Selected Districts00

3a. Reading00

3b. Math00

Chapter 4. Policies, Programs, and Practices of the Selected Districts00

Chapter 5. Recommendations and Conclusions.....00

Bibliography00

Appendix.....00

LIST OF FIGURES

Figure 1. NAEP 4th-grade reading scale score increases in TUDA cities between 2003 and 2009, compared with large-city and national samples00

Figure 2. NAEP 8th-grade reading scale score increases in TUDA cities between 2003 and 2009, compared with large-city and national samples00

Figure 3. NAEP 4th-grade mathematics scale score increases in TUDA cities between 2003 and 2009, compared with large-city and national samples.....00

Figure 4. NAEP 8th-grade mathematics scale score increases in TUDA cities between 2003 and 2009, compared with large-city and national samples.....00

Table 1. Average NAEP reading scale scores of public school students nationwide and large-city public school students in grades 4 and 8, 2003-2009.....	00
Table 2. Average NAEP mathematics scale scores of public school students nationwide and large-city public school students in grades 4 and 8, 2003-2009	00
Table 3. Average NAEP reading scale scores of public school students nationwide and large-city public school students in grades 4 and 8 by student group, 2003-2009.....	00
Table 4. Average NAEP mathematics scale scores of public school students nationwide and large-city public school students in grades 4 and 8 by student group, 2003-2009.....	00
Table 5. TUDA districts showing statistically significant reading gains or losses on NAEP by student group between 2003 and 2009.....	00
Table 6. TUDA districts showing statistically significant mathematics gains or losses on NAEP by student group between 2003 and 2009.....	00
Table 7. District effects by subject and grade after adjusting for student background characteristics, 2009	00
Table 8. Changes in grade 4 NAEP reading subscale scores (significance and effect size measures), by composite, subscale, and district, 2003-2007.....	00
Table 9. Changes in grade 8 NAEP reading subscale scores (significance and effect size measures), by composite, subscale, and district, 2003-2007.....	00
Table 10. Summary statistics on NAEP reading in grade 4	00
Table 11. Summary statistics on NAEP reading in grade 8	00
Table 12. Changes in grade 4 NAEP mathematics subscale scores (significance and effect size measures), by composite, subscale, and district, 2003-2007.....	00
Table 13. Changes in grade 8 NAEP mathematics subscale scores (significance and effect size measures), by composite, subscale, and district, 2003-2007.....	00
Table 14. Summary statistics on NAEP mathematics in grade 4	00
Table 15. Summary statistics on NAEP mathematics in grade 8	00
Table 16. Summary of Key Characteristics of Improving and High Performing Districts versus Districts Not Making Gains on NAEP	00

CHAPTER 1

INTRODUCTION

Overview

America's urban schools are under more pressure to improve than any other institution—public or private—in the nation. Many groups might have folded under the pressure, giving up in the face of mounting criticism. But urban school systems and their leaders are doing the opposite. They are rising to the occasion, innovating with new approaches, learning from each other's successes and failures—there are plenty of both on which to draw—and aggressively pursuing reforms that will boost student academic performance.

There is fresh evidence that the efforts of these urban school systems are beginning to pay off. Reported results from the National Assessment of Educational Progress (NAEP) on the large-city (LC) schools indicate that public schools in the nation's major urban areas made statistically significant gains in both reading and mathematics between 2003 and the most recently reported assessment in 2009 at both grades 4 and 8.

Moreover, an analysis of differences in the rates of improvement of the large cities versus the nation between 2003 and 2009 shows that the gains in reading and mathematics in both fourth and eighth grades were significantly larger in large cities than in the national sample. Large-city schools and the Trial Urban District Assessment (TUDA) districts continue to lag behind national averages for the most part, but these reported NAEP data from 2003 to 2009 indicate that they are making progress and that the progress is over and above what is being seen nationally.¹

This is an abridged, summary report of selected findings from *Pieces of the Puzzle: Factors in the Improvement of Urban School Districts on the National Assessment of Educational Progress*--a comprehensive study prepared by the Council of the Great City Schools in collaboration with the American Institutes for Research (AIR) and with funding from The Bill & Melinda Gates Foundation. The purpose of this report—exploratory as it is—is to present new data on urban school districts that have made significant and consistent gains, have demonstrated high overall performance, or have not produced consistent improvements on NAEP reading and mathematics assessments at grades 4 and 8.

The rationale for looking at these three kinds of districts was to compare and contrast the factors that might be contributing to the achievement of students in each. We have assumed that there was something different to be learned from districts that were improving than from districts showing high performance but not improving or districts with low and stagnant performance.

This report examines factors that might be driving those patterns, how alignment between state or district standards and NAEP, as well as the instructional programs and other features of the districts, might be affecting results, and what may be needed to further improve urban public schooling nationwide. The study also provides a preliminary framework for how future analyses might be conducted as more city school systems participate in TUDA.

¹ A chapter detailing demographics and achievement trends in large-city schools and TUDA districts is provided in the full report, *Pieces of the Puzzle: Factors in the Improvement of Urban School Districts on the National Assessment of Educational Progress* (2011) Council of the Great City Schools and AIR, Washington DC

Context

Work on this project began nearly a decade ago, when the Council of the Great City Schools began asking a series of important questions about the improvement of America's major urban school systems.

Were the nation's urban schools, the subject of so much debate and the centerpiece of so many reforms, actually getting better? If so, could we tell which districts were consistently showing significant improvements? What were these improving school districts doing that others were not? And could we apply the lessons learned to urban schools and districts across the country in an attempt to enhance the academic achievement of urban school children across the board?

In 2000, the Council persuaded the National Assessment Governing Board (NAGB) and Congress to oversample big-city school districts during the regular administrations of the National Assessment of Educational Progress (NAEP). The districts that volunteered for the Trial Urban District Assessment (TUDA), as the project came to be known, received district-specific results for the first time in NAEP's history.

The Council of the Great City Schools requested oversampling to demonstrate its commitment and the commitment of its members to high standards and also to procure data (1) to determine whether urban schools were improving academically, (2) to compare urban districts individually and collectively with each other and the nation, and (3) to evaluate the impact of urban reforms in ways that the current 50-state assessment system did not allow.

There is now a critical mass of city school systems participating in NAEP and sufficiently long trend lines on those cities to begin discerning strengths and patterns of student academic growth. This report, the first to use NAEP data for this kind of district-level analysis, also explores the story behind these achievement trends.

One area of investigation involved the alignment between NAEP frameworks and various state and district standards. We asked whether alignments or misalignments affected urban districts' performance on NAEP over time. The project team was interested in determining if a close alignment with state standards hindered or helped their ability to make larger achievement gains as measured by NAEP. This part of the study was intended to inform districts about the possibility that academic progress might be enhanced by better alignment.

A second area of investigation involved the organizational and instructional practices of urban school systems that have shown significant improvements or have consistently outperformed other big-city systems on the NAEP. The project team was interested in studying the conditions under which the gains or the consistently high performance had taken place and seeing how the practices in these school systems might differ in critical ways from those of districts that were not showing substantial progress.

These interconnected areas of inquiry have a common, overarching goal of improving our understanding of the potential of NAEP to inform efforts to improve urban education nationwide, particularly as the new Common Core State Standards are being implemented across the country. This report presents the results from those inquiries.

CHAPTER 2

METHODOLOGY

Research Questions

The principal goal of this research was to answer a series of questions about trends in urban school system academic achievement and to do so using data from NAEP and detailed analysis of local school district practices. The research questions included—

- Are the nation’s large-city schools making significant gains on NAEP and are the gains, if any, greater than those seen nationwide?
- Which of the TUDA districts have been making significant and consistent gains on NAEP in reading and mathematics at the fourth- and eighth-grade levels, both overall and at differing points across the distribution of student achievement scores?
- Which of the TUDA districts outperformed others on the NAEP, controlling for relevant student background characteristics?
- Which of the TUDA districts have made significant and consistent gains on the NAEP in reading and mathematics at the fourth- and eighth-grade levels among student groups defined by race/ethnicity, language, and other factors?
- How have the TUDA districts scored on NAEP subscales in reading and mathematics?² What were their relative strengths and weaknesses across the subscales?
- What was the degree of alignment between the NAEP frameworks in place between 2003 and 2007 in reading, mathematics, and science and the district’s respective state standards? What was the relationship between that alignment and district performance or improvement on the NAEP during those years?
- What instructional conditions and practices were present in districts that made significant and consistent gains on the NAEP? In what ways were their practices different from those of districts showing weaker gains? What are the implications for how urban school districts can improve academically in the future?

Summary of Methodology

Our methodology can be summarized in seven general steps:

First, to answer questions about improvements among large-city schools in the aggregate and how the gains compared with national trends, we examined data from NAEP spanning 2003 to 2007, the latest year available when this project started. The report also examined reported scores from 2003 to 2009.³

² The main report also contains analysis of data on 2005 and 2009 science results

³ The project has also published an addendum to the study with detailed analyses of data from 2007 to 2009.

Second, to answer the detailed questions about NAEP trends in the 11 large-city school systems participating in the Trial Urban District Assessment in 2007, we examined data from 2003, 2005, and 2007 on fourth- and eighth-grade reading and mathematics achievement. All data were analyzed using both reported results and scores that account for differences in exclusion rates, known as “full population estimates.” For some analyses, scores also were adjusted to control for relevant student background characteristics derived from the NAEP background questionnaire.

Third, we selected cities for in-depth analysis based on a multi-step process that involved statistical testing of gains or losses in each time period, from 2003 to 2005, 2005 to 2007, and 2003 to 2007 using both reported results and full population estimates. City school systems were ranked by grade and subject according to the number of times each showed statistically significant improvements across the three time periods. Moreover, trend analyses were conducted at each quintile of the NAEP test-score distribution for each district to determine where students were making significant gains (i.e., Did gains occur across the achievement distribution, or did they occur only at the higher or lower ends of the distribution?).

We used these processes to select one district showing significant and consistent improvements in reading and one in mathematics, as well as one district that lacked such improvement. We also selected another district that outperformed other districts on the 2007 assessment, after controlling for student background characteristics.

In sum, we selected four districts in all—Atlanta, Boston, Charlotte, and Cleveland—for deeper study. While the selection of study districts was based on pre-specified criteria, we conducted additional analyses and determined that the selection of districts did not depend on the kind of analysis we conducted, i.e., reported results vs. full population estimates.⁴ The choice of districts would have been the same using 2009 data.

Fourth, we analyzed NAEP trends by student group for each of the TUDA school systems to ensure that the study districts were not showing gains at the expense of one student group or another. The analysis included trends by race/ethnicity, gender, eligibility for the National School Lunch Program (NSLP-eligible), disability, and language status.

Fifth, to determine whether there were any discernable strengths and weaknesses in reading, mathematics, and science in the four selected districts, we analyzed NAEP data at the subscale and item levels. Because each subscale in NAEP is calibrated separately, subject area by subject area, student performance on different subscales is not directly comparable. Therefore, we computed and compared “effect sizes” for the analysis corresponding to changes in subscale averages or means between 2003 and 2007. We tested which of these changes were statistically significant. We also converted the mean subscale scores to percentiles on the national distribution to allow for additional comparisons of strengths and weaknesses within districts.

Sixth, we examined the alignments in the selected cities between NAEP and the state (and, where applicable, district) standards by looking at NAEP content specifications in reading and mathematics and comparing them to state (and district) standards that were in place in 2007 for grades 4 and 8. Alignment charts were created for each of the four districts that were selected for in-depth analysis. Each chart included actual NAEP specification language and how each respective state and/or district’s content standards matched those specifications in content and at grade level, either completely or partially. Both the NAEP specifications and the content/grade-level matches were then coded for cognitive demand, that is, the difficulty of the tasks represented by the standard statements. Matches and cognitive demand

⁴ See the main report for the methodologies used in the analysis.

codes were determined by two independent “coders” who had been provided specialized training in reliably conducting the comparisons. The results were reviewed by senior content experts. Then, we examined the degree of alignment between the completely matched NAEP specifications and the state/district standards.

Finally, we conducted site visits to the four study districts to determine, retrospectively, the instructional context and practices in place between 2003 and 2007 that could help explain why some of the districts showed more consistent gains or higher performance than others. In so doing, we looked at how the practices of the improving and higher-performing districts differed from the comparison district. On these site visits, the research team conducted extensive interviews of central-office staff (past and present), principals, and teachers; reviewed curriculum and instructional materials; and analyzed additional data.

What Was Not Examined

This research project looked at a considerable number of variables, some of which were quantifiable and some of which were more descriptive and qualitative. This made the study an unusual blend of statistical and case study methodologies. The study was not a controlled experiment, however, from which causality could be determined. In addition, the study was *post hoc* in the sense that it looked backwards and attempted to explain why things appeared to have the effect they did. And, there were areas that we did not examine or quantify that might have a bearing on the ability of some of the districts to make gains on NAEP.

For instance, we were limited in our ability to define, measure, or track teacher quality over the 2003 to 2007 period. In addition, this study did not examine the distribution of teachers across high-need and high-performing schools. The study also did not look at the number of teachers in each district who came from alternative teacher pipelines like Teach for America or the number of teachers that were nationally board certified. Other research suggests that these variables are not likely to explain changes in NAEP results to any significant degree, but we did not examine them to determine their power to affect the outcome of this analysis.

Although the researchers asked questions about pacing guides and other curricular materials during the site visits, this study did not involve classroom visits or other activities that might gauge the extent to which teachers followed pacing guides or introduced state standards in their curriculum.

Finally, our analysis also did not include an examination of the effects of pay-for-performance initiatives in these cities, nor did it explicitly examine such factors as class size, school size, quantifiable measures of parent involvement, school choice, and the use of early-childhood programs, extended-time initiatives, community engagement measures, and other such variables.

CHAPTER 3
**ANALYSIS OF NAEP RESULTS,
TRENDS, AND ALIGNMENT FOR
SELECTED DISTRICTS**

Introduction

This chapter presents our analysis of detailed NAEP achievement and trends between 2003 and 2009. Specifically, we report:

1. Overall changes in reported NAEP reading and mathematics scores between 2003 and 2009 among large cities in the aggregate and changes compared to the nation.
2. Changes in reported NAEP reading and mathematics scores in individual TUDA city school districts between 2003 and 2009, compared to large cities generally and to the nation.
3. Changes in reported NAEP reading and mathematics scores among student groups in the individual TUDA cities between 2003 and 2009, compared to large cities generally and to the nation.
4. Districts that were performing higher or lower than what might be expected statistically based on their student background characteristics.

We then narrowed our focus to NAEP subscale performance trends from 2003 to 2007 in reading and mathematics among the four districts selected for deeper analysis. These results are presented for reading in section 3a and mathematics in section 3b.⁵

For each subject, we then report results of the analysis on the degree of alignment between the state and/or district standards for each of the four selected jurisdictions and the grade 4 and grade 8 NAEP specifications. Specifically, we address two questions:

- What is the degree of content and cognitive demand alignment between the NAEP frameworks and the district's respective state standards?
- What is the relationship between that alignment and district gains or performance on the NAEP?

Overview: Achievement in Large Cities and TUDA Districts

Reading⁶

NAEP data on the large-city (LC) schools indicate that public schools in the nation's major urban areas made statistically significant gains in reading between 2003 and the latest reported testing in 2009 at both grades four and eight. Between 2003 and 2009, reported NAEP scale scores in reading rose in LC from a mean or average of 204 to 210 among fourth graders and increased from 249 to 252 among eighth

⁵ In the main report, an additional section presents science results.

⁶ A new framework for the NAEP reading examination was introduced for the 2009 assessment. The framework presented many changes from the framework that had been in place since 2003, but a bridge study conducted during the 2009 NAEP administration showed that the NAEP trend line for reading could be continued. See http://nces.ed.gov/nationsreportcard/lt/bridge_study.asp for details.

graders. During the same period, reported NAEP scale scores in reading nationwide (a measure that includes students in large cities) moved from 216 to 220 among fourth graders and from 261 to 262 among eighth-graders. (See table 1.)

Table 1. Average NAEP reading scale scores of public school students nationwide and large-city public school students in grades 4 and 8, 2003-2009

Reading	Grade 4					Grade 8				
	2003	2005	2007	2009	Δ	2003	2005	2007	2009	Δ
Overall										
National Public	216	217	220	220*	3***	261	260	261	262*	1***
Large Cities	204	206	208	210**	6***	249	250	250	252**	4***
Gap	12	11	12	10		12	10	11	10	

* Statistically different from large cities ** Statistically different from national public *** Statistically different between 2003 and 2009.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

An analysis of differences in the size of gains of schools in the large cities versus the nation between 2003 and 2009 shows that the increases among the large-city (LC) schools in reading in both fourth and eighth grades were significantly larger than gains in the national sample.⁷ The net difference between the reported scale scores of large-city fourth graders and fourth graders nationwide (which includes large-city fourth graders) narrowed from 12 scale score points in 2003 to 10 scale score points in 2009. At the eighth-grade level, the net difference narrowed from 12 points to 10 points over the same period.

Moreover, the percentage of large-city fourth graders reading at or above basic levels of achievement increased from 47 percent in 2003 to 54 percent in 2009, and those scoring at or above proficient levels increased from 19 percent to 23 percent. The percentage of large-city eighth graders scoring at or above basic levels in reading increased from 58 percent in 2003 to 63 percent in 2009, while those scoring at or above proficient levels increased from 19 percent in 2003 to 21 percent in 2009.⁸

The percentage of fourth graders nationwide reading at or above basic levels of achievement increased from 62 percent in 2003 to 66 percent in 2009, and those scoring at or above proficient levels increased from 30 percent to 32 percent. The percentage of eighth-graders scoring at or above basic levels increased from 72 percent in 2003 to 74 percent in 2009, while those scoring at or above proficient levels remained the same at 30 percent.

In addition, Austin, Boston, and Charlotte outperformed their large-city peers in both fourth and eighth grades in reading in 2009, New York City's fourth graders scored higher than their large-city peers, and Charlotte outperformed their national peers in fourth-grade reading.

Overall, more TUDA districts saw increased reading scores among fourth graders than among eighth graders.⁹ In addition, there were statistically significant reading gains between 2003 and 2007 among large-city fourth graders in the second, third, and fourth quintiles of achievement. In contrast, the nation

⁷ Difference between size or magnitude of gain between 2003 and 2009 in fourth grade equals three scale score points, $p < .05$. Difference between size or magnitude of gain between 2003 and 2009 in eighth grade equals three scale score points, $p < .05$. All comparisons were independent tests for multiple pair-wise comparisons according to the False Discovery Rate procedure. (Differences in scale score gains may be due to rounding.)

⁸ Source: Reading 2009, Trial Urban District Assessment, Results at Grades 4 and 8. National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education (NCES 2010-459), 2010.

⁹ All references to gains or increases in NAEP scores are statistically significant at the $p < .05$ level.

showed a statistically significant improvement across all quintiles.¹⁰ In the eighth grade, the large cities showed no appreciable movement in reading in any quintile, while the nation showed statistically significant declines in the lowest and the two highest quintiles.

Finally, NAEP tests students at the fourth-grade level on their ability to read for literary experience and for information, and at the eighth-grade level on their ability to read for literary experience, for information, and to perform a task. Results tend to be strongly correlated, i.e., students who scored well on one subscale tended to do well on others. While there was considerable variation from city to city in the eighth grade, it appeared that students in the 11 districts were somewhat more likely to do better in reading for literary experience than in reading for information or reading to perform a task.

Mathematics

Public schools in large cities also showed statistically significant gains between 2003 and 2009 in mathematics in both fourth and eighth grades. Over that period, the reported NAEP scale scores of the LC in mathematics increased from 224 to 231 among fourth graders and from 262 to 271 among eighth graders. During the same period, reported NAEP scale scores in mathematics nationwide (which includes students in large cities) increased from 234 to 239 among fourth graders and from 276 to 282 among eighth graders (see table 2). Both sets of gains were statistically significant. (See table 2.)

Table 2. Average NAEP mathematics scale scores of public school students nationwide and large-city public school students in grades 4 and 8, 2003-2009

Mathematics	Grade 4					Grade 8				
	2003	2005	2007	2009	Δ	2003	2005	2007	2009	Δ
Overall										
National Public	234	237	239	239*	5***	276	278	280	282*	6***
Large Cities	224	228	230	231**	7***	262	265	269	271**	9***
Gap	10	9	9	8		14	13	11	11	

* Statistically different from large cities ** Statistically different from national public *** Statistically different between 2003 and 2009.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

An analysis of differences in the size of gains of schools in the large cities versus the nation between 2003 and 2009 shows that the increases in mathematics in both fourth and eighth grades were significantly larger in large cities than in the national sample.¹¹ The net difference between the scale scores of large-city fourth graders and fourth graders nationwide (which included large-city fourth graders) narrowed from 10 scale score points in 2003 to eight scale score points in 2009. At the eighth-grade level, the difference (also statistically significant) narrowed from 14 points to 11 points over the same period.¹²

Moreover, the percentage of large-city fourth graders scoring at or above basic levels of attainment increased from 63 percent in 2003 to 72 percent in 2009, and those at or above proficient levels increased from 20 percent to 29 percent. The percentage of large-city eighth graders scoring at or above basic levels

¹⁰ Distribution of achievement scores across five equally weighted groups. The full quintile analysis is provided in the main report.

¹¹ Difference between size of gain between 2003 and 2009 in fourth grade equals two scale score points, $p < .05$. Difference between size of gain between 2003 and 2009 in eighth grade equals three scale score points, $p < .05$.

¹² Differences between numbers in the text and numbers in the accompanying tables are due to rounding.

increased from 50 percent in 2003 to 60 percent in 2009, while those at or above proficient levels increased from 16 percent in 2003 to 24 percent in 2009.¹³

The percentage of fourth graders nationwide scoring at or above basic levels of attainment in mathematics increased from 76 percent in 2003 to 81 percent in 2009, and those at or above proficient levels increased from 31 percent to 38 percent. The percentage of eighth graders scoring at or above basic levels increased from 67 percent in 2003 to 71 percent in 2009, while those at or above proficient levels increased from 27 percent in 2003 to 33 percent in 2009.

In addition, in 2009, Austin, Boston, Charlotte, Houston, New York City, and San Diego outperformed their large-city peers in mathematics in both fourth and eighth grades. Charlotte students outperformed their national peers in fourth grade, and Austin students outscored their national peers in eighth grade.

In addition, the large cities made more frequent gains in mathematics between 2003 and 2007 (across five quintiles) at both fourth and eighth grade levels than in reading, although there were exceptions. And in contrast to reading, more TUDA districts registered increased mathematics scale scores among eighth graders than among fourth graders.

In fourth grade, large cities showed statistically significant improvements in mean scores at every quintile except quintile 1—the bottom 20 percent. The nation, on the other hand, showed gains in all quintiles. At the eighth-grade level, the large cities posted significant gains in mathematics at every quintile, as did the national sample.

Finally, NAEP mathematics tests assess students in number properties and operations (“number” for short), measurement, geometry, data analysis and probability, and algebra. The analysis of TUDA results indicated considerable variation from city to city, but in general, fourth graders in TUDA districts appeared to score better in geometry, algebra, and number and less well in measurement and data. At the eighth-grade level, TUDA students appeared to do better in geometry and algebra than in number.

City by City Performance Trends among TUDA Districts

We also looked at how individual districts were performing relative to their TUDA peers between 2003 and 2007 and 2003 and 2009. Of the 11 TUDA districts, the Atlanta Public Schools made significant and the most consistent improvements in reading between 2003 and 2007 at both the fourth- and eighth-grade levels, even after adjusting for testing-exclusion rates.^{14, 15} In addition, the Boston Public Schools made significant and the most consistent gains in mathematics between 2003 and 2007 at both fourth- and eighth-grade levels, after adjusting for exclusion rates.

¹³ Source: Math 2009, Trial Urban District Assessment, Results at Grades 4 and 8. National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education (NCES 2010-452), 2009.

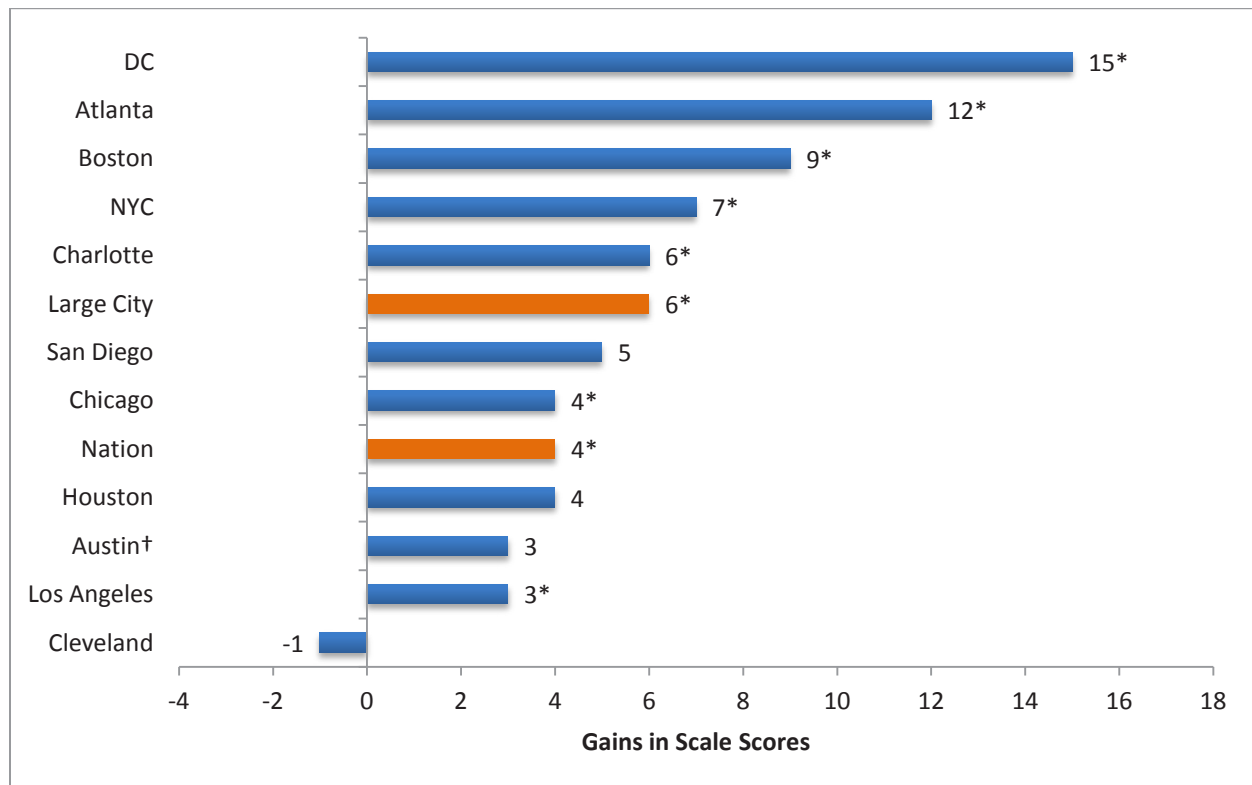
¹⁴ By “most consistent,” the report means that the district had the highest number of statistically significant gains during the periods 2003-2005, 2005-2007, and 2003-2007 using “full population estimates” to adjust for exclusion rates.

¹⁵ A recent state investigation of the Atlanta Public Schools found evidence of cheating on the Georgia state Criterion-Referenced Competency Tests (CRCT), but the investigative report presented no evidence of tampering with the National Assessment of Educational Progress (NAEP) and made no mention of the district’s progress on NAEP. NAEP assessments are administered by an independent contractor (Westat), and Westat field staff members are responsible for the selection of schools and all assessment-day activities, which include test-day delivery of materials, test administration as well as collecting and safeguarding NAEP assessment data to guarantee the accuracy and integrity of results. In addition, an internal investigation by NCES found no evidence that NAEP procedures in Atlanta had been tampered with. For more information on how NAEP is administered, see appendix A in full report.

The Charlotte-Mecklenburg Public Schools outperformed all other TUDA districts in reading and mathematics at both grade levels, after controlling for relevant student background characteristics. The district also scored either as high as or higher than the national average and showed student group performance that was higher than peer-group performance nationwide. And finally, the Cleveland Metropolitan School District was the only district among those participating in TUDA in 2007 that failed to make significant gains or that posted significant losses in most subjects and grades between 2003 and 2007, adjusting for exclusion rates. These four districts—Atlanta, Boston, Charlotte, and Cleveland—were chose for in-depth analysis and case study to determine their commonalities and differences.

In addition, the reported NAEP reading scale scores on individual TUDA cities showed significant gains in many cities between 2003 and 2009. Significant reading gains among fourth graders were seen in Atlanta, Boston, Charlotte, Chicago, the District of Columbia (DC), Los Angeles, and New York City (NYC). (See figure 1.)

Figure 1 NAEP 4th-grade reading scale score increases in TUDA cities between 2003 and 2009, compared with large-city and national samples



† Austin did not participate in TUDA in 2003, so figure shows change from 2005 to 2009.

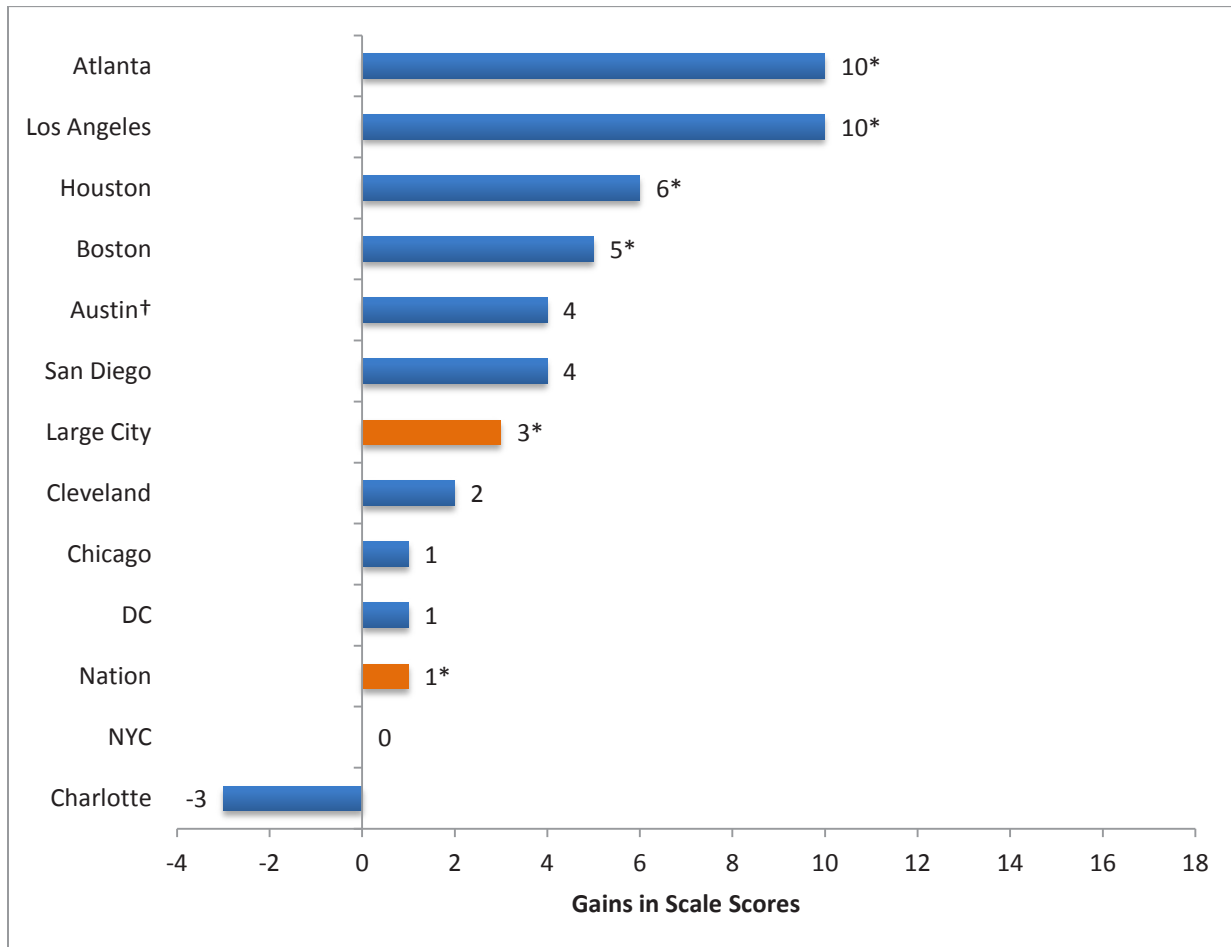
Note: Beginning in 2009, the results for charter schools were not included in a district's TUDA results if they were not included in a district's Adequate Yearly Progress (AYP) data. The results affect only DC.

* Significant difference ($p < .05$) between 2003 and 2009.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

And significant gains between 2003 and 2009 in reported reading scale scores among eighth graders were seen in Atlanta, Boston, Houston, and Los Angeles. (See figure 2.)

Figure 2 NAEP 8th-grade reading scale score increases in TUDA cities between 2003 and 2009, compared with large-city and national samples



† Austin did not participate in TUDA in 2003, so figure shows change from 2005 to 2009.

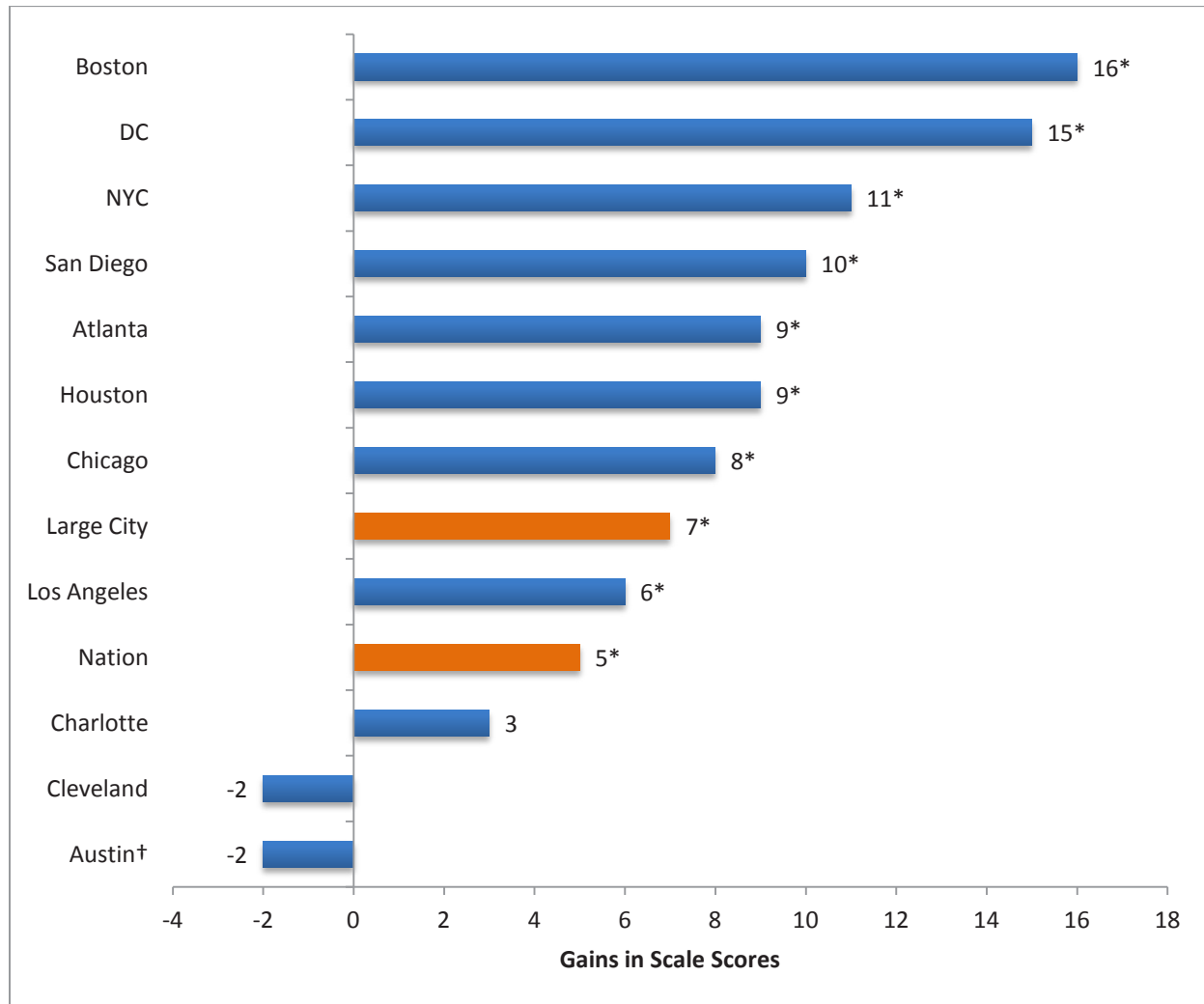
Note: Beginning in 2009, the results for charter schools were not included in a district's TUDA results if they were not included in a district's Adequate Yearly Progress (AYP) data. The results affect only DC.

* Significant difference ($p < .05$) between 2003 and 2009.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

In addition, the reported NAEP mathematics data on individual TUDA cities showed significant gains in many cities. Significant mathematics gains among fourth graders between 2003 and 2009 were seen in Boston, the District of Columbia (DC), New York City (NYC), San Diego, Atlanta, Houston, Chicago, and Los Angeles. (See figure 3.)

Figure 3 NAEP 4th-grade mathematics scale score increases in TUDA cities between 2003 and 2009, compared with large-city and national samples



† Austin did not participate in TUDA in 2003, so figure shows change from 2005 to 2009.

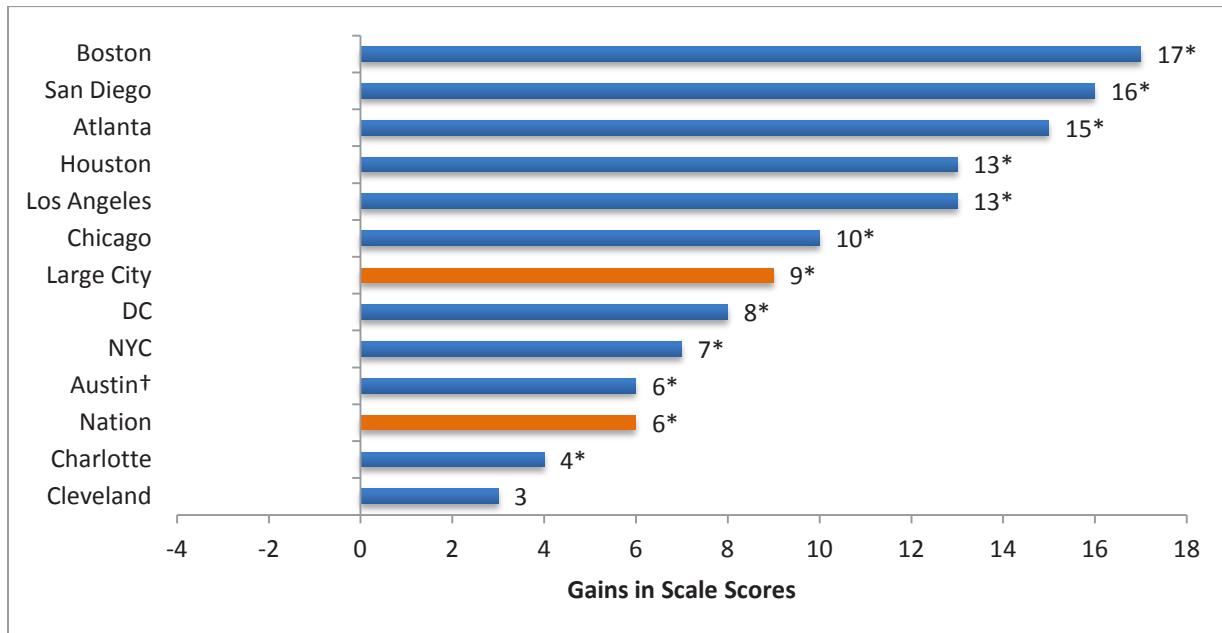
Note: Beginning in 2009, the results for charter schools were not included in a district's TUDA results if they were not included in a district's Adequate Yearly Progress (AYP) data. The results affect only DC.

* Significant difference ($p < .05$) between 2003 and 2009.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

Significant mathematics gains among eighth graders between 2003 and 2009 were seen in every TUDA city except Cleveland. (See figure 4.)

Figure 4 NAEP 8th-grade mathematics scale score increases in TUDA cities between 2003 and 2009, compared with large-city and national samples



† Austin did not participate in TUDA in 2003, so figure shows change from 2005 to 2009.

Note: Beginning in 2009, the results for charter schools were not included in a district's TUDA results if they were not included in a district's Adequate Yearly Progress (AYP) data. The results affect only DC.

* Significant difference ($p < .05$) between 2003 and 2009.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

Student Groups

Next, we analyzed performance trends by selected student groups, and found that over the 2003-2009 period, large-city districts generally improved the reading and math scores of key student groups.

Table 3. Average NAEP reading scale scores of public school students nationwide and large-city public school students in grades 4 and 8 by student group, 2003-2009

Reading	Grade 4					Grade 8				
	2003	2005	2007	2009	Δ	2003	2005	2007	2009	Δ
African American										
National Public	197	199	203	204*	7***	244	242	244	245*	1***
Large Cities	193	196	199	201**	8***	241	240	240	243**	2***
White										
National Public	227	228	230	229	2***	270	269	270	271	1***
Large Cities	226	228	231	233	7***	268	270	271	272	4***
Hispanic										
National Public	199	201	204	204*	5***	244	245	246	248*	4
Large Cities	197	198	199	202**	5***	241	243	243	245**	4
Asian/Pacific Islander										
National Public	225	227	231	234*	9***	268	270	269	273*	5***
Large Cities	223	223	228	228**	5	260	266	263	268**	8***

NSLP-eligible										
National Public	201	203	205	206*	5***	246	247	247	249	3***
Large Cities	196	198	200	202**	6***	241	243	242	244	3***

* Statistically different from large cities ** Statistically different from national public *** Statistically different between 2003 and 2009.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

Table 4. Average NAEP mathematics scale scores of public school students nationwide and large-city public school students in grades 4 and 8 by student group, 2003-2009

Mathematics	Grade 4					Grade 8				
	District	2003	2005	2007	2009	Δ	2003	2005	2007	2009
African American										
National Public	216	220	222	222*	6***	252	254	259	260*	8***
Large Cities	212	217	219	219**	7***	247	250	254	256**	9***
White										
National Public	243	246	248	248*	5***	287	288	290	292	5***
Large Cities	243	247	249	250**	7***	285	288	292	294	9***
Hispanic										
National Public	221	225	227	227	6***	258	261	264	266	8***
Large Cities	219	223	224	226	7***	256	258	261	264	8***
Asian/Pacific Islander										
National Public	246	251	254	255	9***	289	294	296	300	11***
Large Cities	246	247	251	233	7	281	289	291	299	18***
NSLP-eligible										
National Public	222	225	227	228*	6***	258	261	265	266*	8***
Large Cities	217	221	223	225**	8***	252	256	260	262**	10***

* Statistically different from large cities ** Statistically different from national public *** Statistically different between 2003 and 2009.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

Most notably, the scale scores of African American students, white students, and NSLP-eligible students in large cities and nationwide rose significantly in both reading and mathematics at both the fourth- and eighth-grade levels. (See tables 3 and 4.) Reported NAEP math scale scores of Hispanic students also increased among both fourth and eighth graders. Yet while reading scale scores rose significantly among Hispanic fourth grade students, the gain in scale scores among Hispanic eighth graders in reading was not significant either in large cities or nationwide. And while large cities and the nation improved both the reading and math scores of Asian/Pacific Islander students in the eighth grade, at the fourth-grade level the change in scale scores among large-city Asian/Pacific Islander students was not found to be significant in either reading or mathematics.

In addition to these overall trends, the reported TUDA data showed that numerous districts made statistically significant progress in reading and mathematics with critical student groups, including African American, Hispanic, Asian American, NSLP-eligible students, limited English proficient students, and students with disabilities. (See tables 5 and 6 below.) In fact, between 2003 and 2009, a majority of districts improved both the reading and math scale scores of their African American students and NSLP-eligible students.

Table 5. TUDA districts showing statistically significant reading gains or losses on NAEP by student group between 2003 and 2009

City/Grade	Black		Hispanic		Asian		White		NSLP		LEP		SPED	
	4	8	4	8	4	8	4	8	4	8	4	8	4	8
Atlanta	↑	↑	—	—	—	—			↑	↑	—	—		
Austin†	↑				—	—								↑
Boston	↑		↑						↑	↑				↑
Charlotte									↑					
Chicago					—	—			↑					
Cleveland					—	—								
D.C.	↑		↑		—	—			↑		↑			
Houston	↑			↑	—	—		↑	↑	↑	↑			↓
Los Angeles				↑						↑	↓		↓	↑
New York City	↑				↑				↑					↑
San Diego												↓	↓	
National Public	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑			↑	↑
Large City	↑	↑	↑	↑		↑	↑	↑	↑	↑				↑

↑ Significant positive ↓ Significant negative - Reporting standard not met (too few students) † Data from 2005 to 2009

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

Table 6. TUDA districts showing statistically significant mathematics gains or losses on NAEP by student group between 2003 and 2009

Mathematics	Black		Hispanic		Asian		White		NSLP		LEP		SPED	
	4	8	4	8	4	8	4	8	4	8	4	8	4	8
Atlanta	↑	↑	—	—	—	—			↑	↑	—	—		↑
Austin†		↑		↑	—	—		↑		↑		↑		
Boston	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑		↑	↑
Charlotte		↑					↑			↑				
Chicago	↑	↑	↑	↑	—	↑		↑	↑	↑		↑		↑
Cleveland					—	—					—	—		
D.C.	↑	↑	↑	↑	—	—	↑		↑	↑	↑	—	↑	
Houston		↑	↑	↑	—	—		↑	↑	↑	↑	↑	↓	↓
Los Angeles		↑	↑	↑		↑			↑	↑				↑
New York City	↑	↑	↑		↑	↑	↑		↑	↑	↑		↑	↑
San Diego		↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑		↑
National Public	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑		↑	↑
Large City	↑	↑	↑	↑		↑	↑	↑	↑	↑	↑		↑	↑

↑ Significant positive ↓ Significant negative - Reporting standard not met (too few students) † Data from 2005 to 2009

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007, and 2009.

These areas where individual city school districts are making significant achievement gains, particularly with key student groups, are important to highlight because they show the capacity of urban districts to overcome historic barriers and meet critical educational challenges.

District Effects

Finally, we examined which districts were performing higher or lower than what might be expected statistically based on their student background characteristics.¹⁶ Positive effects indicate the district was performing higher among the 11 TUDA participants than expected statistically in 2009; negative effects indicate that the district was performing lower than expected relative to the other districts.¹⁷

In other words, the result is a “district effect” that cannot be explained by differences in student background characteristics, but still might include more than the district itself.¹⁸ In general—

- In grade four reading, the results indicated that district effects were positive and significant in Austin, Boston, Charlotte, Houston, and New York City; and were negative and significant in Chicago, Cleveland, the District of Columbia, and Los Angeles. Results were not different from what was predicted in Atlanta and San Diego.
- In grade eight reading, the results indicated that district effects were positive and significant in Austin, Boston, Charlotte, and Houston; and were negative and significant in the District of Columbia and Los Angeles. Results were not different from what was predicted in Atlanta, Chicago, Cleveland, New York City, and San Diego.
- In grade four mathematics, the results indicated that district effects were positive and significant in Austin, Boston, Charlotte, Houston, and New York City; and were negative and significant in Chicago, Cleveland, the District of Columbia, and Los Angeles. Results were the same as predicted in Atlanta and San Diego.
- In grade eight mathematics, the results were positive and significant in Austin, Boston, Charlotte, Houston, and New York City; and they were negative and significant in Cleveland, the District of Columbia, and Los Angeles. Results were the same as predicted in Atlanta, Chicago, and San Diego.

This component of the analysis did not measure change or improvement over time nor did it account for a district’s starting point in 2003. For example, Atlanta and Cleveland had similar scores in 2003, but Atlanta moved significantly to predicted levels performance by 2009, while Cleveland continued to show performance below predicted levels (see table 7).

¹⁶ A full description of the methodology employed in the statistical analysis of district effects is available in the full report. Results from 2007 are presented in the full report; results from 2009 are presented in the addendum, tables D-1, D-2, D-3, and D-4.

¹⁷ District effect is the difference between district mean and statistically expected district mean.

¹⁸ The student background variables used in this analysis explained between 35 and 40 percent of the variance from the mean performance depending on subject and grade tested.

Table 7. District effects by subject and grade after adjusting for student background characteristics, 2009*

	Reading Grade 4	Reading Grade 8	Mathematics Grade 4	Mathematics Grade 8
Atlanta	0.9	2.8	-1.4	-1.1
Austin	6.5*	6.1*	8.3*	14.4*
Boston	8.6*	6.6*	8.2*	12.1*
Charlotte	6.2*	2.5*	7.3*	8.4*
Chicago	-4.6*	1.5	-5.7*	0.0
Cleveland	-12.4*	-2.1	-10.5*	-2.6*
District of Columbia	-5.9*	-7.3*	-6.0*	-8.4*
Houston	4.7*	2.2*	9.3*	11.1*
Los Angeles	-6.3*	-1.9*	-6.2*	-6.1*
New York City	7.2*	-0.4	6.9*	2.7*
San Diego	0.2	-1.2	0.9	-2.0

* District effect is significantly different from zero.

Changes in Subscale Performance from 2003 to 2007

As we indicated previously, Atlanta, Boston, Charlotte, and Cleveland were selected for deeper study. Atlanta was selected for its significant and consistent gains in reading achievement, Boston was chosen for gains in math, and Charlotte was picked for high performance in reading and math. Cleveland was chosen because of its weak gains in both subjects. This deeper analysis begins with an examination of changes in composite and subscale reading performance between 2003 and 2007 in the four districts and compares them to subscale results for the large cities (LC) and the national public school sample. Table 8 shows the results for fourth-grade reading and table 9 shows results for the eighth grade. (Note that reading to perform a task is not assessed at grade 4.) The changes are shown in terms of effect size and statistical significance to indicate the direction and magnitude of change in performance on composite reading and its subscales during the 2003–2007 study period.

Table 8. Changes in grade 4 NAEP reading subscale scores (significance and effect size measures), by composite, subscale, and district, 2003-2007

	Atlanta	Boston	Charlotte	Cleveland	LC	National Public
Composite Reading	↑ 0.28	↔ 0.12	↔ 0.09	↔ 0.09	↑ 0.10	↑ 0.09
Literary	↔ 0.24	↔ 0.08	↔ 0.03	↔ 0.05	↑ 0.07	↑ 0.05
Information	↑ 0.30	↔ 0.17	↑ 0.15	↔ 0.12	↑ 0.13	↑ 0.12

Key: LC=Large Cities. ↑ Significant positive ↔ Not significant ↓ Significant negative

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 and 2007 Reading Assessments.

Table 9. Changes in grade 8 NAEP reading subscale scores (significance and effect size measures), by composite, subscale, and district, 2003-2007

	Atlanta	Boston	Charlotte	Cleveland	LC	National Public
Composite Reading	↑ 0.16	↔ 0.04	↔ -0.07	↑ 0.19	↔ 0.03	↔ -0.01
Literary	↔ 0.12	↔ -0.05	↔ -0.06	↔ 0.15	↔ 0.01	↔ 0.00
Information	↔ 0.17	↔ 0.09	↔ -0.01	↔ 0.21	↔ 0.05	↔ 0.00
Perform a Task	↑ 0.19	↔ 0.10	↓ -0.16	↔ 0.14	↔ 0.04	↓ -0.04

Key: LC=Large Cities. ↑ Significant positive ↔ Not significant ↓ Significant negative

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 and 2007 Reading Assessments.

We see that fourth graders in Atlanta made statistically significant gains on their composite reading score between 2003 and 2007, the only district among the four to show a gain on this measure.

In fact, Atlanta's composite score effect size was approximately three times larger than that of both the large-city (LC) and the national public sample. During the study period, Atlanta also showed significant

gains on the subscale for reading for information. In Charlotte, there was significant gain on one subscale only, reading for information, but it was only half the effect size seen in Atlanta. Subscale scores in Boston and Cleveland did not change significantly on either of the two subscales or on the composite measure.

In grade 8 reading, Atlanta again made significant gains on the composite reading measure and made significant gains on reading to perform a task.

Atlanta's composite effect size was some five times greater than that of the LC and sixteen times greater than the national public sample. Boston did not show any significant improvement on any of the three subscales. Charlotte showed a significant loss in the subscale of reading to perform a task. Subscale scores in Cleveland did not change significantly on any of the subscales, although it posted a significant gain on the eighth-grade composite measure.¹⁹

Summary of Analysis of Reading Standards Alignment and NAEP Results

Our analysis showed that content and cognitive-demand alignment was not high between NAEP reading specifications in grades 4 and 8 and state and district standards in Atlanta, Boston, Charlotte, and Cleveland.

In grades 4 and 8, the complete and partial content match²⁰ of district/state standards to NAEP ranged from 37 percent (Massachusetts in grade 8) to 80 percent (Charlotte in grade 4), with most hovering around 50 percent. However, the complete matches in grade 4 and 8 never exceeded 67 percent (Charlotte in grade 4) with most matches being below 40 percent.

Generally, the greatest degree of complete and partial alignment was in reading for literary experience in grade 4. In grade 8, the degree of complete and partial alignment appeared similar in reading for literary experience and in reading for information, although there was a greater range of matches with reading for information. The analysis indicated that making “reader/text connections” was the least aligned aspect across all reading subscales in both grades.

In addition, the level of cognitive demand on completely matched standards was higher in grade 8 than in grade 4 in the selected jurisdictions.

Finally, there was little obvious connection between the content and cognitive matches with NAEP reading and overall gains or reported scale scores during the study period. (See tables 10 and 11.)

¹⁹ Cleveland did not show significant reading gains in eighth grade when analyzed with full population estimates.

²⁰ Content match refers to the percentage of district/state standards that aligned to NAEP specifications either completely or partially.

Table 10. Summary statistics on NAEP reading in grade 4

Study District	2003-07 Effect Size Change and Significance	2007 Unadjusted Composite Percentile	Percentage Complete Content Match with NAEP	Weighted Cognitive Demand Mean for Complete Content Matches
Atlanta	0.28↑	33	39%	2.1
Boston	0.12↔	36	39%	2.1
Charlotte	0.09↔	50	67%	2.4
Cleveland	0.09↔	25	39%	2.0
LC	0.10↑	--	--	--
National Public	0.09↑	50	--	1.9

Key: LC=Large Cities, ↑ Significant positive, ↔ Not significant, ↓ Significant negative

In fourth grade, Atlanta was the only one of the selected districts to see a significant increase in reading, yet it had the same percentage of complete content matches with NAEP as did Boston and Cleveland (39 percent), two districts that saw no significant increase in NAEP reading scores. The three districts also appeared to have similar cognitive demand levels. It is interesting, however, that the district with the highest overall percentile in fourth-grade reading, Charlotte, was also the district with the highest percentage of complete content matches and the highest weighted cognitive demand mean.

In eighth grade, Atlanta and Cleveland saw significant increases in reading scores (although Cleveland did not see increases using the full population estimates); however, the degree of content matches in Atlanta appeared similar to Boston, which saw no significant reading score increase. Cleveland had content matches that appeared similar to Charlotte, which saw no reading increases. Again, Charlotte had the highest overall percentile score in eighth-grade reading on NAEP, and its state appeared to have the highest content match with NAEP and the highest weighted cognitive mean.

Table 11. Summary statistics on NAEP reading in grade 8

Study District	2003-07 Effect Size Change and Significance	2007 Unadjusted Composite Percentile	Percentage Complete Content Match with NAEP	Weighted Cognitive Demand Mean for Complete Content Matches
Atlanta	0.16↑	29	40%	2.4
Boston	0.04↔	38	35%	2.5
Charlotte	0.07↔	45	59%	2.8
Cleveland	0.19↑	30	56%	2.3
LC	0.03↔	--	--	--
National Public	-0.01↔	50	--	1.9

Key: LC=Large Cities, ↑ Significant positive, ↔ Not significant, ↓ Significant negative

Changes in Subscale Performance from 2003 to 2007

The mathematics analysis begins with an examination of changes in subscale performance between 2003 and 2007 in the four selected districts described earlier and compares them to subscale results for the large cities (LC) and the national public school samples. Table 12 shows the results for fourth-grade mathematics and Table 13 for eighth grade. The changes are shown in terms of statistical significance and effect size to indicate the direction and magnitude of change in performance by subscale during the 2003–2007 study period.

Table 12. Changes in grade 4 NAEP mathematics subscale scores (significance and effect size measures), by composite, subscale, and district, 2003-2007

	Atlanta	Boston	Charlotte	Cleveland	LC	National Public
Composite Math	↑ 0.27	↑ 0.52	↔ 0.08	↔ 0.03	↑ 0.20	↑ 0.18
Number	↑ 0.23	↑ 0.52	↔ 0.04	↔ 0.04	↑ 0.19	↑ 0.17
Measurement	↔ 0.18	↑ 0.46	↔ -0.03	↔ 0.06	↑ 0.16	↑ 0.15
Geometry	↑ 0.41	↑ 0.52	↑ 0.35	↔ -0.04	↑ 0.21	↑ 0.19
Data	↑ 0.30	↑ 0.40	↔ 0.05	↔ 0.04	↑ 0.20	↑ 0.23
Algebra	↑ 0.30	↑ 0.38	↔ 0.09	↔ -0.03	↑ 0.18	↑ 0.14

↑ Significant positive ↔ Not significant ↓ Significant negative

Note: NAEP subscales are not all reported on the same metric; hence, gains on subscales are not comparable. Therefore, the numeric values of the changes in subscales are not represented in this table.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 and 2007 Mathematics Assessments.

Table 13. Changes in grade 8 NAEP mathematics subscale scores (significance and effect size measures), by composite, subscale, and district, 2003-2007

	Atlanta	Boston	Charlotte	Cleveland	LC	National Public
Composite Math	↑ 0.34	↑ 0.38	↑ 0.10	↔ 0.13	↑ 0.18	↑ 0.11
Number	↑ 0.22	↑ 0.29	↔ 0.06	↔ -0.09	↑ 0.08	↑ 0.06
Measurement	↑ 0.50	↑ 0.33	↔ 0.11	↔ 0.03	↑ 0.16	↑ 0.06
Geometry	↔ 0.31	↑ 0.34	↔ 0.07	↔ 0.12	↑ 0.18	↑ 0.10
Data	↑ 0.30	↑ 0.35	↔ 0.11	↔ 0.11	↑ 0.18	↑ 0.11
Algebra	↑ 0.29	↑ 0.43	↔ 0.09	↑ 0.34	↑ 0.23	↑ 0.16

↑ Significant positive ↔ Not significant. ↓ Significant negative

Note: NAEP subscales are not all reported on the same metric; hence, gains on subscales are not comparable. Therefore, the numeric values of the changes in subscales are not represented in this table.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 and 2007 Mathematics Assessments.

We see that fourth graders in Atlanta made statistically significant gains in math composite scores and in four of the five subscales (all except measurement). Boston improved on the composite measure and in all five subscales in grade 4 with effect sizes that were two to three times larger than those of both the large cities (LC) and the national sample. Charlotte saw a significant gain only in geometry and did not see any change in the composite measure. The composite and subscale scores in Cleveland did not change significantly between 2003 and 2007 in any of the five areas.

In grade 8 math, three of the four jurisdictions made statistically significant gains on the composite measure. Boston improved on the composite measure and in all content areas, and Atlanta improved on the composite measure and in four of five areas (all except geometry). Cleveland showed a significant gain only in algebra, but not in the composite score. Mean scores in Charlotte did not change significantly in any of the five content areas between 2003 and 2007, but it did show a significant gain on the composite measure. The effect sizes in Boston were two to three times larger than the LC or the national public sample. At both grade levels in Atlanta and Boston, effect sizes on the composite measure and the individual subscales were generally greater than those of either the LCs or the national public schools.

Summary of Analysis of Math Standards Alignment and NAEP Results

Our analysis of alignment in both content and cognitive demand showed consistent results. (See tables 14 and 15.) Overall, the content match appeared similarly low in grade 4 and grade 8, although there was greater variability in grade 8.

Although the complete and partial matches on the NAEP standards never fell below 50 percent in mathematics, only at grade 8 in Cleveland did the content match exceed 80 percent. However, analyses of the complete matches provided a different picture. At grade 4, complete matches were at or below 50 percent in the four cities, and at grade 8 none exceeded 56 percent.

Finally, there is little obvious connection between the content and cognitive matches with NAEP mathematics and overall gains or reported scale scores during the study period.

In fourth grade, Atlanta and Boston were the only selected districts to see significant increases in mathematics, yet both districts had lower complete content matches than Charlotte and Cleveland, which saw no significant increases in NAEP math scores. Moreover, the cognitive demand means of all four districts appeared to be similar.

As in reading, Charlotte had the highest percentile measure in mathematics and what appeared to be the highest overall level of complete content matches.

Table 14. Summary statistics on NAEP mathematics in grade 4

Study District	2003-07 Effect Size Change and Significance	2007 Unadjusted Composite Percentile	Percentage Complete Content Match with NAEP	Weighted Cognitive Demand Mean for Complete Content Matches
Atlanta	0.27↑	28	38%	2.0
Boston	0.52↑	39	38%	2.0
Charlotte	0.08↔	54	46%	2.0
Cleveland	0.03↔	20	40%	1.9
LC	0.20↑	--	--	--
National Sample	0.18↑	50	--	1.8

Key: LC=Large Cities, ↑ Significant positive, ↔ Not significant, ↓ Significant negative

In eighth grade, Atlanta, Boston, and Charlotte saw significant increases in mathematics scores, but the districts had complete content matches that ranged from 24 percent in Charlotte to 45 percent in Boston. In addition, Cleveland, which showed no gain in math, had the highest level of complete content matches. All four districts appeared to have similar weighted cognitive demand codes.

Again, Charlotte had the highest percentile in math but had content matches that appeared lower than the other three districts and also had cognitive demand means that were similar to the other districts.

Table 15. Summary statistics on NAEP mathematics in grade 8

Study District	2003-07 Effect Size Change and Significance	2007 Unadjusted Composite Percentile	Percentage Complete Content Match with NAEP	Weighted Cognitive Demand Mean for Complete Content Matches
Atlanta	0.34↑	25	32%	2.1
Boston	0.38↑	44	45%	2.1
Charlotte	0.10↑	51	24%	2.0
Cleveland	0.13↔	25	56%	2.1
LC	0.18↑	--	--	--
National Sample	0.11↑	50	--	2.0

Key: LC=Large Cities, ↑ Significant positive, ↔ Not significant, ↓ Significant negative

CHAPTER 4
POLICIES, PROGRAMS,
AND PRACTICES OF THE
SELECTED DISTRICTS

Introduction

The four TUDA districts that were selected for case studies based on their performance on NAEP were different from each other in many ways, but the three districts that showed either large gains in performance or higher scores than other districts—Atlanta, Boston and Charlotte—shared many similarities in terms of their political context, instructional focus, and reform agenda. The three districts also differed from the one district—Cleveland—we examined for its weak trends on NAEP.

This chapter compares and contrasts the policies, programs, and practices of these four districts during the 2003 to 2007 period and summarizes the observations and interpretations that the study teams of urban education and content experts made during their site visits to each of the districts.²¹ (See table 16 at the end of the chapter for a summary of key characteristics of district reforms.) Detailed case studies of Atlanta, Boston, and Charlotte-Mecklenburg are provided in the full report.

Atlanta

Atlanta showed significant and consistent gains in reading throughout the study period.²² The findings of the study team’s site visit suggested that the district benefited from a literacy initiative launched in 2000. The initiative was well-defined, sustained over a long period of time, built around a series of comprehensive school reform demonstration models (CSR), and bolstered by a system of regionally based School Reform Teams (SRTs) deployed to provide services directly to schools and assist them in meeting performance targets. Atlanta’s schools had some latitude to choose their own reading programs, and the district leveraged this school-by-school latitude to build ownership for reforms at the building level. At the same time, the district, which closed approximately 20 mostly low-performing schools during the study period, laid out clear, research-based strategies and “best practices” for how literacy would be taught throughout the school system, creating a common vocabulary for reading instruction and providing extensive site-based and cross-functional support through literacy coaches and professional

²¹ Site visit findings on Cleveland were augmented and checked against a study that the Council of the Great City Schools conducted of the instructional practices of the district in 2005, *Foundations for Success in the Cleveland Municipal School District, Report of the Strategic Support Team of the Council of the Great City Schools*, Fall 2005. In addition, the site visit findings on Charlotte-Mecklenburg were augmented and checked against the case study that the Council conducted with MDRC as part of the report, *Foundations for Success: Case Studies of How Urban School Systems Improve Student Achievement*, September 2002.

²² A recent state investigation of the Atlanta Public Schools found evidence of cheating on the Georgia state Criterion-Referenced Competency Tests (CRCT), but the investigative report presented no evidence of tampering with the National Assessment of Educational Progress (NAEP) and made no mention of the district’s progress on NAEP. NAEP assessments are administered by an independent contractor (Westat), and Westat field staff members are responsible for the selection of schools and all assessment-day activities, which include test-day delivery of materials, test administration as well as collecting and safeguarding NAEP assessment data to guarantee the accuracy and integrity of results. In addition, an internal investigation by NCES found no evidence that NAEP procedures in Atlanta had been tampered with. For more information on how NAEP is administered, see appendix A in full report.

development. Atlanta also began to emphasize writing and the development of literacy skills across the curriculum from the early years of its literacy initiative (around 2003).

Mathematics reforms, on the other hand, lagged behind literacy reforms in Atlanta by several years, only starting in earnest around 2006. Not surprisingly, the district showed uneven growth in math achievement between 2003 and 2007, although its math improvements were notable when compared with other TUDA districts. Some of this gain in mathematics may have been due in part to the school system's progress in reading and its efforts to infuse reading across the curriculum.

Boston

As noted earlier in this report, Boston was selected for study because it showed significant and consistent gains in mathematics. The Boston site visit revealed a strong instructional focus on math in the school district during the study period.

Interestingly, Boston began much of its current reforms in 1996 in the area of literacy rather than mathematics, but this 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. In fact, the district's literacy work was not even placed organizationally inside the curriculum unit for much of the study period. 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 mathematics, according to interviewees who were asked directly about why math gains outstripped reading progress.

Boston's math leadership team was able to learn from the difficulties faced by the literacy initiative and began implementing a common, challenging, concept-rich core mathematics program (*Investigations* at the elementary level and *Connected Math* in the middle grades) in 2000. Boston pursued a multi-staged, centrally defined, and well-managed roll-out over several years and provided strong, sustained support and oversight for implementation of its math reforms despite a lack of immediate improvements systemwide. Success came despite the fact that, according to Council staff members who have tracked efforts in many urban school systems, these programs have proven difficult to implement in other cities.

Charlotte-Mecklenburg

While Charlotte did not demonstrate the same gains as Atlanta or Boston in NAEP reading and mathematics over the study period, the district maintained consistently high performance at or above national averages from 2003 to 2007. Charlotte was selected for study because, after controlling for student background characteristics such as poverty and race/ethnicity, it out-performed all other TUDA districts in reading and mathematics in 2007.

In the early 1990s, Charlotte was among the first school districts in the nation to develop and implement standards of learning, and it built a strong accountability system for meeting these standards, including implementing "balanced scorecards" in the mid and late 1990s as a data tool to track and manage school- and department-specific goals that were aligned to systemwide priorities.

Charlotte also replaced its site-based management approach in the late 1990s with a more centrally defined system, employing a standardized, managed-instructional approach to improve student

achievement across the board. The central office was particularly focused on providing on-site support and oversight for its lowest-performing schools, mandating the implementation of prescriptive reading (*Open Court*) and math (*Saxon Math*) programs and offering incentives for teachers and staff to move to struggling sites in an effort to ensure the highest quality of education was provided to students. At the same time, the district implemented programs intended to address the differing needs of students along the continuum of achievement.

Cleveland

In contrast with the other districts, Cleveland was chosen because of its consistently flat achievement on NAEP assessments in both reading and mathematics during the study period, with the exception of eighth-grade reading. In Cleveland, a number of factors seemed to limit the district's ability to advance student achievement on NAEP, even though the district and its leadership team worked hard to turn the district around between 1998, when the district was taken over by the state and put under mayoral control, and late 2006, when a new superintendent assumed responsibility. The chief executive officer during much of the study period labored to clean up a school system that had been plagued for years by dysfunctional school board governance, weak management, ineffective instruction, financial and operational problems, and other systemic issues.

Much of this CEO-led work was instrumental in helping the district pass a construction bond, enhance community engagement, reduce operating debt, and raise state test scores in the elementary grades. But the efforts were not strong enough to move student performance on NAEP.

Until 2005, there was no functional curriculum in place to guide instruction. The school district's instructional program remained poorly defined, and the system had little ability to build the capacity of its schools and teachers to deliver quality instruction. The district also lacked a system for holding its staff and schools accountable for student progress in ways that other study districts were implementing at the time. In the judgment of the site-visit team, the outcome was a weak sense of ownership for results and little capacity to advance achievement on a rigorous assessment like NAEP.

In addition, the district suffered unusually large budget cuts during the study period that resulted in the layoff of hundreds of teachers and the “bumping” of many others. During the study period, the district was also moving toward smaller learning communities and K-8 schools, with what many individuals in the district at the time described as “too much speed and too little expertise, professional development or support.” Amidst these cuts and changes, principals did not have the authority to hire their own teachers, and little professional development to teachers and principals accompanied the transitions.

While each of the districts included in this report faced considerable instructional, financial, and political challenges during the study period, these forces seemed to derail the educational reform initiatives in Cleveland, weakening the district's instructional efforts and undercutting its ability to produce better outcomes on NAEP.

Cross-cutting themes

Despite their differences, there were a number of traits and themes common among the improving or high-performing districts—and clear contrasts with the experiences and practices documented in Cleveland. These themes fell under six broad categories:

- *Leadership and Reform Vision.* Atlanta, Boston, and Charlotte each benefited from strong leadership from their school boards, superintendents, and curriculum directors. These leaders were

able to unify the district behind a vision for instructional reform and then sustain that vision for an extended period.

- *Goal setting and Accountability.* The higher-achieving and most consistently improving districts systematically set clear, systemwide goals for student achievement, monitored progress toward those instructional goals, and held staff members accountable for results, creating a culture of shared responsibility for student achievement.
- *Curriculum and Instruction.* The three improving or high-performing districts also created coherent, well-articulated programs of instruction that defined a uniform approach to teaching and learning throughout the district.
- *Professional Development and Teaching Quality.* Atlanta, Boston, and Charlotte each supported their programs of instruction with well-defined professional development or coaching to set direction, build capacity, and enhance teacher and staff skills in priority areas.
- *Support for Implementation and Monitoring of Progress.* Each of the three districts designed specific strategies and structures for ensuring that reforms were supported and implemented districtwide and for deploying staff to support instructional programming at the school and classroom levels.
- *Use of Data and Assessments.* Finally, each of the three districts had regular assessments of student achievement and used these assessment data and other measures to gauge student learning, modify practice, and target resources and support.

Leadership and Reform Vision

Atlanta, Boston, and Charlotte all benefited from the sustained leadership of unified, reform-minded school boards and strong superintendents who had a clear focus on instruction. In each city, the superintendent and school board worked collaboratively over a sustained period to pursue change and improvement in student academic achievement. Consequently, each of these leadership teams was able to focus the organization and the community away from battles over politics and school governance and onto the business of instruction, developing and communicating a shared vision for instructional reform and clear, measurable objectives for districtwide growth. And all three districts went to great lengths to ensure that the right people were in the right place at the right time to drive these reforms.

In Atlanta, for example, districtwide reform was championed by a strong superintendent who came to the city in 1999 steeped in the reform experiences of other major urban school districts. She made teaching and learning her focus from the beginning and brought a clear vision for districtwide improvement, strong leadership and instructional skills, communications expertise, and high expectations for student achievement and adult performance. She worked over several years to build consensus for reform on the elected school board and to break the district's past negative culture. The board's leadership was further enhanced by the city's business community, which worked alongside the superintendent to build a school board that could work with the administration on academic improvement. This coalescence of forces attracted substantial investments and grants from national philanthropic organizations like the GE Foundation, the Panasonic Foundation, and The Bill & Melinda Gates Foundation, which helped seed and support the reforms.

Boston, meanwhile, benefited from the consensus and support of a strong, mayor-appointed school board led by a board president who had strong working relations with the former and current superintendents.

The board used its mandate for improvement to spearhead a comprehensive school improvement plan in 1996 that focused on strengthening student achievement and advancing standards-based instructional practice. In fact, much of the plan remains intact, though with substantial enhancements in reading, under the leadership of the current superintendent.

In Charlotte, a relatively stable school board worked with the superintendent to ensure support for an aggressive instructional reform agenda even when the board was not always unified on other issues. In the early 1990s, Charlotte was one of the nation's early leaders and innovators in the standards movement, and the district benefited subsequently from a series of strong superintendents who focused on instructional issues even as the district was settling one of the nation's longest running court-ordered school desegregation cases.

In addition to the school board and superintendent, another essential element in the reform agendas of the three districts was the strategic hiring and placement of instructional leaders in key leadership roles. In fact, by most accounts, Charlotte's approach to reform was guided by the core belief that people more than programs made the difference. District leadership systematically selected central-office instructional staff they felt were committed to student achievement and had a record of success.

Atlanta also developed what the site-visit team found to be an extremely strong and deep cadre of central-office staff members—including the deputy superintendent for instruction, director of reading, and director of mathematics—as well as principals with considerable expertise in instructional programming. These staff leaders formed the core of the instructional team that the superintendent used to implement and drive reforms.

Similarly, Boston hired a former principal to lead curriculum and instruction, a math leader with national experience and considerable expertise, and other experts skilled at building partnerships and overseeing the strategic rollout of a new, concept-rich math program, paying particular attention to the management of change in the implementation process.

By most accounts from interviewees in each city—Atlanta, Boston, and Charlotte—these instructional leadership teams had excellent technical and programmatic skills and were open to and eager for change and innovation, and staff members at all levels who were passionate about the reforms.

Also important in Atlanta, Boston, and Charlotte was sustaining a commitment to the district's vision for reform and its implementation throughout the jurisdiction. Despite initial pushback from teachers who disliked the systematic approach of the reading program in Atlanta, the district pressed forward with the implementation of its literacy reforms and gained and sustained teacher support over a number of years.

In Boston, the district's math reforms also met with considerable initial resistance and a lack of immediate results districtwide over the first several years. But the school board and superintendent resisted efforts to change course and abandon the new math program. Instead, the district redoubled its rollout efforts, engaging and communicating with schools and the community around the strategic plan and building broad-based understanding and ownership in the direction and success of the city's public schools.

Charlotte also experienced initial resistance to its reforms but stayed the course until results were evident. The district was able to do this even as it saw turnover among some of its leadership and staff.

Interestingly, Cleveland—like the three other study districts—had a long-serving, reform-minded superintendent during the study period. The city also had a mayor-appointed school board, but that board did not have the same decision-making authority that Boston's mayor-appointed body had. The

superintendent vetted her decisions through the school board, but the board did not have the power to reverse her decisions.

Many in Cleveland saw the superintendent as a visionary leader. She improved the district's standing on state indicators, started to break down some of the organizational silos that had characterized the district for many years, improved student attendance and graduation rates, initiated a literacy program, and made other substantial instructional reforms that the district had never seen before. But, ultimately, the district lacked a well-defined and coherent theory of action or a strong underlying program of instruction to guide its reforms.

Instead, the district let principals shape their schools' instructional efforts with little guidance, oversight, or technical assistance from the central office. The consistency of instructional reforms may have been further undermined by staff that was not as strong as those the research teams observed in the other three districts. In addition, the district saw numerous changes in central-office instructional staff members during the study period, and this turnover was accompanied by ever-changing tactical agendas and programs that added to the inconsistency in program implementation.

Overall, this lack of coherence at the program level led to an instructional effort that, while an improvement over the past, remained incapable of boosting academic performance on anything other than state tests. The district, in fact, did show substantial gains on the Ohio Proficiency Test (OPT) in reading, math, and science until it was phased out in 2005. Once it was replaced with the more rigorous Ohio Achievement Test (OAT), Cleveland showed only modest gains in mathematics and little progress in reading in grades 3 through 8 during the remainder of the study period.

Goal Setting and Accountability

The ability of the school districts to set clear academic goals and hold school and district staff accountable for instructional improvement was a common element of reforms in Atlanta, Boston, and Charlotte. Each district articulated systemwide targets for improvement, as well as school-specific goals, promoting collaboration among staff at all levels to reach these goals. These achievement goals and standards of performance were generally clear, measurable, and communicated throughout the organization. In addition, the transparency of these goals helped create widespread buy-in for new programs and a culture of ownership for student achievement.

Atlanta had perhaps the most explicit goal-setting and accountability system of the districts we studied, enacting a two-tiered goal structure aimed not only at reducing the number of students in the lowest-performing categories or increasing the numbers reaching proficiency on the state test, but at driving improvements across the achievement spectrum for all students. This two-tiered system may be related to this study's findings that Atlanta's students made gains in all quintiles on NAEP reading between 2003 and 2007.

The Atlanta superintendent and all district senior staff—including executive directors of the regional School Reform Teams—were placed on performance contracts tied to the attainment of districtwide academic targets on state tests. Each school, in turn, had specific achievement targets calculated by the district and based on a formula tied to districtwide goals for improvement. These measures were integrated into the performance evaluations of teachers, administrators, and principals, with bonuses provided for meeting or exceeding goals.

Goal setting in Boston also became more explicit and more school-based as the district's data system improved and annual target-setting under No Child Left Behind (NCLB) was put into place. But the district's accountability system during this period was defined around a mutual ownership of results that

emerged among the leadership staff over time as the system improved its capacity. Except, in part, for the superintendent's evaluation, personnel evaluations in Boston were not tied to student scores per se, but the review and analysis of student performance data reportedly led to candid conversations with staff and principals about where improvements were needed. In addition, the district was using a state index that gave credit for movement across multiple performance levels—as in Atlanta—a practice that may have contributed to Boston's math gains among all subgroups and across all quintiles.

Charlotte also had a strong goal-setting and staff-accountability system that fell somewhere between Atlanta's and Boston's in its explicitness. For example, Charlotte had concrete academic achievement targets as well as equity goals that each school was required to meet and a balanced scorecard system that was used to monitor progress, but the district's accountability system did not carry explicit punitive consequences. Charlotte's culture of high standards and collaboration helped instill a strong sense of shared responsibility for student achievement. At the district level, senior staff met with the superintendent on a regular basis, and these conversations revolved around student data and how instruction could be modified for better results.

In comparing accountability systems, it is important to keep in mind that Atlanta started its reforms with student achievement levels much lower than did Boston and Charlotte. It is not unusual for very low-performing urban school districts to begin their reforms by putting into effect more explicit targets and accountability systems than districts that are farther ahead or that have been implementing their reforms for longer periods. This more explicit initial strategy by lower-performing districts is often pursued as a way to build capacity and model excellence in ways that the district may not have seen before.

Yet, although the accountability systems in these three districts—Atlanta, Boston, and Charlotte—differed somewhat in their explicitness, there was a strong sense of ownership for results and shared responsibility for student progress that was not present in Cleveland. In fact, a recurring theme in interviews with staff members in Atlanta, Boston, and Charlotte was that all knew they were making progress, but they were often their own toughest critics about the work left to do.

In contrast with the other three districts, Cleveland had an approach to goal setting and accountability that did not go much beyond meeting NCLB safe-harbor targets, according to district-level staff members interviewed by the research team. School-based staff that the site-visit team interviewed also indicated there was little support or monitoring of progress at school sites by the central office, which had very few instructional staff members. Principals were evaluated only minimally on student academic gains, and school staff and teacher evaluations were not linked to student achievement during the study period.

There was also no mechanism to hold central-office staff responsible for districtwide gains in Cleveland. Rapid turnover of leadership and staff during the study period may also have weakened confidence in and ownership of reforms, and staff members throughout the organization evidenced little personal responsibility for improvement. In fact, a focus group of teachers expressed the opinion that the district, its policies, and personnel often reflected very low expectations for student achievement.

Curriculum and Instruction

Although the three improving or high-performing study districts did not necessarily employ uniform academic programs or materials at each school, each had district-defined teaching and learning objectives that laid out what students were expected to know and be able to do at various grade levels.

In Atlanta, for example, the district's reform efforts began by rethinking what was going on in classrooms and then redesigning administrative and structural supports in a process the district termed "Flipping the Script." Schools were given the latitude to choose among a list of district-approved literacy programs and

Comprehensive School Reform Demonstration (CSRD) models, as long as the schools consistently met their site-specific growth targets. While other districts have a hard time supporting multiple reading and math programs from school to school, Atlanta was able to support a range of programs by focusing on districtwide learning objectives and a uniform instructional philosophy and by building an organizational structure that provided ongoing and intensive technical assistance directly to schools around each program the schools selected.

Along the way, the district developed a clear, systemwide curriculum articulating what students were to be taught—something that did not exist prior to 2000—and implemented a full-day kindergarten program. Long-serving staff members interviewed by the research team credited the district’s gains less to any one instructional reform model than to an overall instructional program that was coherent, disciplined, standards-based, and sustained over time.

Charlotte also designed and successfully enacted a comprehensive literacy plan for the teaching of reading and writing during the study period, adopting a core curriculum based mainly on the North Carolina Standard Course of Study and the *Open Court* reading program. This program was supplemented with a strong writing initiative, an important addition that staff and community members interviewed by the site visit team widely credited with improving student literacy and achievement across the curriculum. The district was also among the first in the nation to mandate a 90-minute reading block, and it employed basal texts and supplemental and enrichment materials designed to meet the full range of students’ literacy needs.

Boston, on the other hand, adopted a districtwide curriculum in 2000 as the foundation of its math program—a decision that proved crucial to ensuring consistency and coherence in math instruction throughout the district. This curriculum, anchored by *TERC Investigations* at the elementary school level and *Connected Mathematics* in middle schools, emphasized moving students beyond memorizing math procedures and algorithms to developing a deeper conceptual understanding of the material, a focus that may have contributed to district gains on the NAEP mathematics assessment, according to the district’s math director.

Boston also bolstered the new math programs with supplemental materials, including additional instruction in math language, 10-minute math sessions devoted to specific topic areas of need, “math facts” handouts, and homework packets. In addition, the central office set a districtwide, designated time for math instruction—70 minutes, which consisted of 60 minutes for core instruction and 10 additional minutes devoted to reviewing routine math facts and procedures. And every school was charged with having a math plan. During this time, the district was also implementing a full-day kindergarten program and a series of pre-k centers with state funds and mayoral support that incorporated a pre-k math program designed by the authors of *Investigations* and accompanied by professional development in mathematics for teachers.

Importantly, all three districts—Atlanta, Boston and Charlotte—worked to ensure close alignment between their instructional programs and state standards and frameworks, creating comprehensive curriculum and framework documents to unpack and clarify state standards and working closely with publishers to identify and address gaps in programs and materials. None of the three districts, however, explicitly used the NAEP frameworks beyond comparing their progress with other TUDA districts.

A coherent, fully articulated program of instruction was not developed by Cleveland during the study period, although the district put into place the Cleveland Literacy System and adopted the Harcourt *Trophies* reading basal in selected grades. In fact, the district did not have a published curriculum in place when the new superintendent took office as school district CEO in late 2006. In the absence of a defined curriculum or unifying set of learning standards, the district and its teachers leaned heavily on state

standards and textbook adoptions as the main arbiters of what students would learn. There was some use of textbook materials and lesson plans built around the standards, but not everyone used them, and the new reading series was adopted initially only for grades k-3 owing to a lack of resources for use in other grades.

In addition, it was clear to the site-visit team that Cleveland had not taken the appropriate steps to identify and address the gaps between these instructional materials in both reading and mathematics and the state standards, which we saw in the previous chapter, were better aligned to the NAEP frameworks than other districts and states studied. As a result, schools used a wide range of materials to implement the standards, which in turn appeared to result in poor cohesion of instructional programs overall and inconsistent use of standards of teaching and learning throughout the district. In addition, the district did not provide ongoing support in the use of adopted materials, according to interviewees. And the district did not appear to have a well-defined intervention strategy for children when they fell behind.

It was interesting that, at the middle school level, Cleveland used the same math program that Boston had so thoughtfully rolled out, but restricted its use to schools that were covered by a National Science Foundation grant without integrating it into the broader districtwide math program. The program was used to train about 240 teachers in some 24 schools and emphasized the building of algebra skills among middle-school teachers, an activity that may be related to the improvement in the district's eighth-grade algebra strand.

Professional Development and Teaching Quality

Professional development and teaching quality also played an important role in ensuring the effective implementation of cohesive instructional programs in the three districts. Although approaches and programs differed from site to site, the site-visit team found that each district was proactive and thoughtful in providing professional development and in putting support structures into place to build staff capacity to deliver quality instruction. The districts were clear about defining quality instruction and expecting teachers and administrators to deliver it, using consistent professional development, “professional learning community” strategies, or coaches to support new curricula and programming.

Atlanta, for instance, started its professional development reforms around implementation of the CSRD models and then enlisted the Consortium on Reading Excellence (CORE) in 2000 to help define and drive high-quality, research-based literacy programming and practices systemwide. The district, which allowed principals to hire their own teachers, provided site-based and nearly universal professional development in literacy instruction through CORE to all district staff and teachers, thereby creating a common theoretical framework, vocabulary, and knowledge base for teaching reading, as well as laying out “26 best practices” in literacy instruction. The CORE training continued until 2006, when district staff and coaches assumed responsibility for providing the professional development to new teachers, as well as refresher courses for others. As we saw in the previous chapter, some of the largest reading gains in Atlanta came on subscales that were a strong focus of CORE training, particularly reading for information.

Likewise, Boston provided professional development for teachers that was designed specifically to support implementation of TERC *Investigations* and *Connected Math*, providing math teachers with extensive training in math content as well as the workshop model of pedagogy. Professional development included, for example, on-site training, grade-level teams, math coaches focusing on unit preparation and student work, monthly professional development with principals, and training for coaches around data. Subject and topic-specific professional development in the pacing of classroom instruction was rolled out in advance of upcoming areas. This multi-faceted approach to professional development in Boston was

designed, moreover, to augment the limited number of formal professional development days provided for in the collective bargaining agreement.

In addition, the district's professional development not only covered important mathematical concepts at each grade level but also covered how they lined up with state and district standards, how they were infused in particular activities and lessons, and how they were reflected in the state assessments administered by the district. For instance, math coaches were trained to address claims by teachers, principals, and parents that the new program did not cover specific ideas and concepts. For example, many teachers claimed, at least initially, that the materials did not address "place value." What some teachers meant by this was that there were no place-value charts. But students were decomposing and recomposing numbers according to place value on a regular basis as they explored alternative algorithms. Many teachers, however, did not recognize this initially as place value.

Boston also provided extensive professional development to math coaches, who were placed in every school pursuant to the district's math plan. (Some of the math coaches came from the original pilot schools that had used *Investigations* and *Connected Math*.) Most coaches came to their work with strong expertise at a particular grade level, but this expertise had to be broadened so they could address entire grade-spans and beyond, since they needed to address how elementary math content connected to middle school and high school mathematics. In fact, coaches often set up structured opportunities for teachers to meet and talk across grade level in order to bolster a shared commitment to improving math instruction as a school. This practice included looking at student work across multiple grades in order to be clear on expectations for each grade level, as well as setting up opportunities for structured classroom visits across grades. The district's scope and sequence pacing guide was helpful in this process because it was organized so that teachers across grade levels were working on about the same mathematical strands at about the same time, making cross-grade-level work possible.

Another critical layer of this professional development was the extensive training provided to all principals on math instruction and on how to be instructional leaders accountable for advancing student achievement at their schools. The professional development for principals also covered the use of "learning walk" procedures, and math concepts used in the new materials.

In Charlotte-Mecklenburg Public Schools, professional development for teachers was defined around student assessment results and district instructional priorities. Courses followed the train-the-trainer model wherein curriculum and development coordinators were key instruction providers. At the high school level, the professional development department used a coaching model where highly qualified coaches were selected to work with struggling schools. These coaches were supervised by curriculum specialists in the central office.

In order to evaluate and determine the effectiveness of professional development, the district distributed surveys to teachers and analyzed student data against professional development offerings. The surveys looked at the instructional goals set by teachers, and the classroom data allowed the department to review growth based on the training. Teachers received five days of mandatory professional development before school started, but because each school had some autonomy, schools could provide additional training as needed. Teachers were also encouraged to become National Board Certified, and the professional development department recruited teachers and provided support to those who wanted to go through the process. Teachers were not penalized if they chose not to attend professional development sessions.

Cleveland also had a comprehensive professional development plan during the study period to accompany its instructional programming, but in contrast with the other three districts, it was largely designed around the attainment of credits for continuing education units rather than around the instructional priorities of the school district, state reading or math standards, or program implementation. While there was a highly

developed professional development tracking system at the time according to the Council's 2005 report on the district, the system was largely tracking staff participation and hours, rather than being used to evaluate the effect of the professional development on student achievement or teacher practice.

In addition, staff in the district at the time indicated to study team interviewers that schools were often left to define the nature of the professional development on their own, using their Title I set-aside dollars, a practice that contributed to a lack of focus and consistency in what was offered. Professional development during this time, therefore, remained voluntary, often unpaid or held after school or on weekends, and it was insufficient to train or prepare teachers for the new grades they were teaching when budget cuts and grade reconfigurations resulted in layoffs and staff redeployments. Finally, after the district implemented its new basal reading series (Harcourt's *Trophies*) as part of its 2003 Reading First grant, it did not have the resources to provide the necessary training for teachers on its use as the materials were adopted in later elementary grades.

The reader should be cautious about the team's findings on professional development, given that the research is quite mixed on the effects of professional development. Drawing causal links between the professional development offered by the selected districts and increases in NAEP results should be done with care. Professional development can be highly effective if designed in a way that it builds teacher capacity and used by teachers to enhance the student skills that NAEP is assessing. But the reader should not presume that any and all professional development is likely to produce substantial results if it is not directly used by teachers or connected to student learning.

Support for Implementation and Monitoring of Progress

In all three improving or high achieving districts, there was a strategy or mechanism in place for rolling out and supporting classroom implementation of districtwide reforms. This support came from a variety of policies, practices, and structures. Each district made a practice of monitoring, supporting, and refining programs over time rather than constantly replacing them. And each district strategically deployed staff to support its instructional programming at the school and classroom levels. This led to greater consistency and depth in program development and implementation districtwide.

For example, the Atlanta Public Schools based its initial reforms in 2000 around a series of individual school audits involving classroom observations in order to (1) to determine the quality of instruction provided at the beginning of the reform period, (2) to shape the nature of the professional development offered by the CSRDs and CORE, (3) and to determine how to differentiate professional development. These audits have continued to this day.

In 2000 and 2001, the district also developed and implemented a system of regionally based School Reform Teams (SRTs), headed by executive directors with deep knowledge of instructional practice and staffed by central-office content specialists to support and serve schools in their efforts to meet performance targets. The five SRTs, which were lead by executive directors, who evaluated their principals largely on student achievement, served about seven to fourteen schools each, and provided a critical mechanism for the district to receive feedback on the successes and challenges schools were facing, as well as what was needed to advance quality programming in real time.

This organizational structure was unique in that it moved a large number of district-level staff out of the central office and created a school-based, "direct-service model" of support that differed considerably from anything site-visit team members had seen before in other major urban school systems. This support structure not only reinforced teachers in the classroom with cross-functional experts who could provide comprehensive feedback on specific steps needed to improve literacy instruction, but it also worked to

free principals from the role of site management and operations, giving them the skills and knowledge to become instructional leaders of their schools.

Boston also utilized school-based staff and support structures to guide implementation of its new math programming. The process of implementing these new math programs was mounted in stages, starting with the naming of Math Leadership Teams of three to six teachers and principals in pilot schools and expanding to all remaining schools in spring 2001. The numbers of teachers on each team in each building increased over time, and the teams themselves were employed to oversee and conduct lesson planning, examine data, develop homework packets, and provide professional development one period a week.

All teachers received mathematics program materials in the fall of 2000, but the teachers in some schools began implementing the program faster than in others. The pace of the program phase-in was partly determined by the schools themselves. Some school principals and Math Leadership Teams wanted full implementation schoolwide as fast as possible. Other schools wanted to start the phase-in with team members only and then roll it out to other teachers later. And other schools wanted to get farther along in their literacy reforms before tackling the new math program. But after three years, all teachers were using the program and participating in professional development on the program's implementation, including ELL and special education teachers.

Once the program was rolled out districtwide, Boston developed a series of “walkthroughs” or “learning walks” in 2002 and 2003 to track math program implementation and gauge student engagement and then acted on the results. The process was initiated by the central office but was designed to help principals and others know what to pay attention to when they visited classrooms and looked at math instruction. In some cases, central-office instructional staff and math coaches were involved in the walks and offered principals direction on how to conduct them, depending on the school. The walkthrough rubrics contained detailed observations and follow-up questions to guide central-office staff, principal, and teacher reflections on what they observed.

The district also used its math coaching plan as a tool for supporting and monitoring program implementation, placing math coaches in every school to provide support to teachers beyond the limited professional development time allowed in the teacher contract. At least initially, coaches reported to the central office and served as “communicators” of all the curriculum materials and the links between the central office and school sites. Teachers reported that math coaching, which was done at all grade levels, was a key component of the school-based support they received, helping them adjust to the new math program and implement it properly, as well as giving them a sense of program ownership and more confidence in teaching math concepts.

These coaches—along with math teachers and principals—received extensive professional development on content, pedagogy, and the collaborative model of coaching, and met regularly to compare practices and results. In order to effectively support program fidelity, math coaches also needed to be prepared to discuss how a particular activity or lesson laid the groundwork for the development of an important math idea in subsequent years or even later in the year, given the tendency of some teachers to skip content with which they were not familiar or did not think was important.

In fact, this strategy of building buy-in through broad-based knowledge about the program extended to the district’s outreach efforts to parents. One of the unique facets of the math plan in Boston was that content instruction was offered to parents at libraries and afterschool tutorial sessions to help support student learning and drive full program implementation.

Like Atlanta and Boston, Charlotte also created extensive school-based support structures. Central-office staff and principals were expected to be out of their offices and in classrooms, supporting and overseeing instruction. Principals were included in training on district initiatives and given professional development on instructional management, walkthrough processes, and the use of balanced scorecards to ensure that, as the instructional leaders of schools, they were monitoring and supporting implementation of district programs in their buildings.

In addition, Charlotte deployed literacy and academic facilitators to elementary and middle schools to help principals develop school literacy plans (consistent with district goals), provide professional development for teachers, and provide support for parents. Like the math coaches in Boston, these literacy and academic facilitators in Charlotte provided a critical line of communications between schools and the district, closely monitoring literacy programs for quality assurance and meeting with district leadership monthly to discuss ways to better support the schools with which they were working.

Charlotte, moreover, provided intensive support to school sites through “Rapid Response Teams”—teams that were deployed to schools that were falling behind on district benchmark tests—in order to help them address areas of instructional weakness identified in the data. These Rapid Response Teams, which sometimes included the academic facilitators referred to previously, would remain on campus for two weeks or more to observe implementation of district initiatives and work with teachers by modeling or co-teaching lessons to promote district standards of instructional practice. Visits by these teams were then followed up by subsequent check-ins and monitoring to ensure improved performance. The presence of these teams, along with academic and literacy facilitators and other support staff in schools, not only helped schools and teachers improve, but also drove transparency and ownership for student achievement.

Throughout the study period, these support structures and lines of communication were reported to have helped Atlanta, Boston, and Charlotte make continuous adjustments to the curriculum and instructional materials based on feedback from school sites without constantly changing the underlying programs.

In Cleveland, however, support for program implementation and instructional capacity building was among the district's most notable areas of weakness. Unlike the other three districts, Cleveland lacked strong, school-based support structures or a cohesive plan for ensuring or monitoring quality instruction.

Whereas in other districts, principals, coaches, and other district staff became a very visible presence in schools and classrooms, there was not a culture of transparency or receptivity to classroom monitoring and support in Cleveland. In fact, principals and others (including coaches) had to be announced into classrooms if the visit was intended for any monitoring purposes. This hindered the ability of principals to oversee program implementation and take on the role of instructional leaders in their buildings. It also limited the role of coaches and dampened the likelihood that trust could be built between teachers and coaches.

Data and Assessments

In each district with significant and consistent gains or high performance, student assessment data were integral to driving the work of the central office and the schools. By and large, these data systems were built around regular diagnostic measures of student learning or benchmark assessments that were used by the central office as a monitoring system to inform placement of interventions or address specific professional development needs.

Each district also worked to create a “data culture,” providing teachers and principals with training in the use of data and developing protocols to help with interpretation and use of test results. Interviews with school level staff in all three districts revealed a strong familiarity with the use of data to inform

instruction and identify students' academic strengths and weaknesses. Staff members from all three districts could—without prompting—cite data to make their points. It was clear from the site visits that, in order to meet both individualized and systemwide objectives, every central-office member, principal, and teacher was expected to consistently review data and use them to make informed decisions about instruction and planning.

Atlanta, Boston, and Charlotte all used data aggressively to identify schools with low performance or growth in reading and mathematics in order to target resources and to refine and supplement the curriculum based on student and school-specific needs. In Atlanta, district staff at the most senior levels had regular meetings to drill down into school data to inform decisions about program refinements and school progress on explicit growth targets. Atlanta also modified its twice-a-year formative assessments to include NAEP-like questions, since the state test used only multiple-choice items.

All three districts, in fact, developed formative assessments to help gauge both program implementation and student progress toward their state standards.

In Boston, interviewees cited the rise of the “data principal” during the study period, and principals reported that their increased understanding of the use of data to inform instruction rather than just monitoring progress helped them gain a clear picture of progress at their school sites and of how to target extra support and professional development. The district also implemented its own interim assessments during the study period using released items from the state test (not NAEP), which research staff indicated helped focus instructional strategies around results. And the system designed and built its own data system (MY BPS) during 2002-2003 that contained student data for teacher use.

Principals and academic facilitators in Charlotte also reported using data to help target support and professional development in order to ensure that their teachers were equipped to meet student needs. Charlotte, in fact, was among the first school systems in the nation to establish locally developed quarterly exams and mini-assessments to track student progress throughout the year. The district also pioneered the use of balanced scorecards to track goals, implementation, and results through explicit assignment of responsibilities, detailed action plans, and measurable objectives for improved student achievement. The central office was charged with monitoring the results of all these data tools. In addition, common planning periods in Charlotte were devoted to sharing and analyzing student test results, and teachers reported relying on student data to create lesson plans, determine students' strengths and weaknesses, and identify areas of concern.

In contrast, although school-level staff members in Cleveland referred to being “data driven,” they were often unable to cite examples of how data were used during the study period to modify instructional practice or professional development, as could staff in the other three districts.

At the outset of the study period, there was little districtwide training in Cleveland on the interpretation and use of benchmark data and no evidence that these student data were used to reform curriculum or professional development. The district has become more data focused in more recent years, but it was much more narrowly attuned to state-test score results, particularly results from the Ohio Proficiency Tests (OPT) during the 2003 to 2007 period. In fact, the district used OPT-released items to write its own short-cycle tests and conduct extensive test-prep even after the test was phased out and the more rigorous Ohio Achievement Test (OAT) was put into place.

Moreover, benchmark tests in Cleveland were not approached as actionable, and low performance did not trigger interventions, additional support, professional development, or program adjustments as they did in the other districts during the study period.

Again, the reader should be cautious about drawing causal inferences about the effects of benchmark or formative assessments on student NAEP results in the selected districts. There is a school of thought that suggests that formative assessments might improve student achievement if they were used in a way that was directly linked with the curriculum and that yielded timely, accessible data, thereby encouraging greater teacher use of the data. At present, however, the research is sparse and links between formative assessments and increased student achievement are not always convincing.²³

Summary and Discussion

Each of the three districts showing gains or high performance on NAEP during the study period pursued reform in differing ways-- particularly at the program level and in how they put all the pieces of reform together to form a coherent strategy. Yet there was a set of common themes observable in their strategies and experiences. All three districts benefited from skillful, consistent, and sustained leadership and a focus on instruction. These leadership teams were unified in their vision for improved student achievement, setting clear, systemwide goals and creating a culture of accountability for meeting those goals. While they did not necessarily employ common programs or materials districtwide, there was a clear, uniform definition of what good teaching and learning would look like. That vision was communicated throughout the district, and a strategy for supporting high-quality instruction and program implementation through tailored, focused, and sustained professional development was aggressively pursued. And each of the districts used assessment data to monitor progress and to help drive these implementation and support strategies, ensuring that instructional reforms reached every school and every student.

Most importantly, these common themes seemed to work in tandem to produce an overall culture of reform in each of the three improving or high-performing districts. Each factor was critical, but it is unlikely that, taken in isolation, any one of these positive steps could have resulted in higher student achievement. Certainly, Cleveland shared some characteristics with the other three study districts, evidencing strong leadership and undergoing a substantial instructional overhaul during the study period. Yet the district lacked the combined force of all these other elements working together to promote instructional excellence. And it was the combined force of these reforms and how they locked together that appeared to make a difference in student achievement.

²³ This project also includes an extensive analysis of the effects of use of formative test data on student achievement. Results will be available in late 2011.

TABLE 16. SUMMARY OF KEY CHARACTERISTICS OF IMPROVING AND HIGH PERFORMING DISTRICTS VERSUS DISTRICTS NOT MAKING GAINS ON NAEP

CHARACTERISTIC/STRATEGY	IMPROVING/HIGH PERFORMING DISTRICTS	STAGNANT/LOW PERFORMING DISTRICTS
Leadership	Strong, consistent focus on improving teaching and learning.	Despite a reform-minded CEO, financial challenges diverted the focus of reform away from the core elements of teaching and learning.
	The school board, superintendent, and central-office staff were able to unify the district behind a shared vision for instructional reform and sustain these reforms over a number of years, despite initial pushback.	The district lacked a coherent approach to instructional reform, and principals were left to shape their school's instructional efforts over the study period with little guidance, oversight, or technical assistance from the central office.
Goal-setting	Leadership remained stable over a relatively long period of time, by urban school district standards, and superintendent led districts on new strategies.	The tenure of the superintendent was stable over the study period, but the CEO was unable to build momentum behind instructional reforms.
	Each district articulated systemwide goals for improvement that went beyond state and federal targets, and were clear, measurable, and communicated throughout the district.	Goal-setting did not go much beyond meeting NCLB safe-harbor targets.
Accountability	While accountability systems varied in terms of explicitness, each district enacted systems for holding school and district staff accountable for meeting achievement goals and standards of performance.	There was little support or monitoring of progress at school sites, and school and district staff members were evaluated only minimally on academic gains.
	The transparency of improvement targets and the district's efforts to create buy-in for new programs helped create a culture of ownership for student achievement.	Staff throughout the organization demonstrated little confidence in or ownership of reforms.
Curriculum and Instruction*	Each district defined curriculum and learning objectives and laid out the knowledge and skills students were expected to have at various grade levels.	The district lacked a coherent, fully-articulated program of instruction, leaving schools to depend on textbook adoptions and state standards as the main arbiters of what students should learn.
	While specific programs sometimes varied from school to school, a common curriculum was deliberately rolled out and helped to create coherent instructional programming throughout the district.	Without guidance or oversight from the central office, schools used a wide range of materials to implement state standards, which resulted in poor cohesion of instructional programs overall.
Professional Development	District leadership was clear about defining what quality instruction looked like, and putting support structures in place to build staff capacity to deliver it. These support structures included pedagogical and content training, training for principals, coaching, and professional learning communities.	While there was a professional development plan in place, schools were often left to define the nature of this professional development themselves, leading to a lack of focus and consistency throughout the district.
	Professional development was generally perceived by school staff as "high quality," and was used to support curricula and programs.	The district's professional development plan was designed largely around the attainment of credits for continuing education, rather than around the instructional priorities of the school district or program implementation. Moreover, training was insufficient to prepare teachers for the new grades they were teaching when budget cuts resulted in layoffs and staff redeployment.
Support for Implementation	Each district employed a comprehensive strategy for rolling out and providing support and oversight for districtwide reforms, allowing them to monitor and refine programs over time rather than constantly replacing them.	The district lacked a strategy for supporting or overseeing instructional programming at the school level.
	Support came from a variety of policies, practices, and structures, and often involved the strategic deployment of school-based support staff.	There was no culture of transparency or receptivity to classroom monitoring and support during the study period. This limited the role of coaches and the ability of principals to oversee program implementation.
Use of Data and Assessments	All three districts employed data systems to monitor program implementation, identify low performing schools and target resources and interventions, identify professional development needs, and refine or supplement the curriculum.	During the study period, data from benchmark tests were not generally viewed as "actionable," and low performance did not trigger interventions, additional support, professional development, or program adjustments.
	Each district worked to create a "data culture," providing teachers and principals with training and protocols for the use of data and promoting the use of data to identify student needs and inform instruction.	There was little training on the interpretation and use of data. While staff referred to being "data driven," they were often unable to cite examples of how data were used during the study period to modify instructional practice or professional development.

* This applies to programming at the elementary and middle School levels, not at the secondary level for any of the districts studied.

CHAPTER 5
RECOMMENDATIONS
AND CONCLUSIONS

Discussion

The results of this exploratory study are encouraging because they indicate that urban schools are making significant academic progress in reading and mathematics and may be catching up with national averages. Our findings have special import because they suggest some reasons for this progress and the steps that might be required to accelerate this headway, particularly as the new common core standards are being implemented.

This section synthesizes our findings and observations around broad themes that we think warrant additional discussion and research as the nation's urban school districts move forward. Debate continues, of course, about what separates urban school systems that make major progress from those making more incremental or no gains. And sometimes that debate confuses what are perceived to be bold reforms with what actually improves student achievement. This chapter draws on the findings of our study to sort through some of the main issues.

Alignment of Standards and Programming

The research team working on this study hypothesized that we would find a close relationship between the alignment of NAEP reading and math specifications and state standards, on the one hand, and the ability to make significant gains on NAEP on the other. What we found was far more complex than what we had originally anticipated.

While the reader should keep in mind the limitations to the alignment analysis that we point out in the main report, the analysis found that the content alignment or match in reading and mathematics between the NAEP frameworks and state standards in the four study districts was generally low or moderate. North Carolina appeared to have the most consistently aligned standards in reading and fourth grade mathematics, and it also had the highest overall performance, but it is difficult to draw a causal relationship between alignment and performance because of the small sample. In sum, it appeared that alignment on its own was insufficient to affect significant movement on student NAEP scores in the four city school systems.

It was clear from the results of this analysis, moreover, that student improvement on the NAEP was related less to content alignment than to the strength or weakness of a district's overall instructional programming. Two of the districts with significant and consistent gains on NAEP—Atlanta and Boston—appeared to overcome the lack of content alignment with coherent, focused, high quality instructional and professional development programs. Conversely, Cleveland was unable to boost its student achievement even though Ohio's standards were as well aligned to NAEP specifications as those of Georgia and Massachusetts. In other words, it was clear that unaligned standards were not fatal to a district's ability to raise achievement. What seemed more important was the ability of the district to articulate a clear direction and implement a seamless set of academic reforms that were focused, high quality, and defined by high expectations.

This finding has significant implications for the new Common Core State Standards that some 45 states have now adopted, because many educators—and the public in general—assume that putting into place more demanding standards alone will result in better student achievement. The results of this study suggest that this is not necessarily the case.

In fact, the findings suggest that the higher rigor embedded in the new standards is likely to be squandered, with little effect on student achievement if the content of the curriculum, instructional materials, professional development, and classroom instruction are not high quality and well implemented and coordinated. Moreover, our findings strongly suggest that the manner in which the common core standards are put into practice in America's classrooms is likely to be the most important factor in their potential to raise academic performance.

The Pursuit of Reform at Scale

What may have also emerged from this study is further evidence that progress in large urban school districts is possible when they act at scale and systemically rather than trying to improve one school at a time.

The education reform movement has been grounded for years on the supposition that progress was attainable only at the school level and that considering school districts as major units of large-scale change was a waste of time. However, this study found that each of the districts that showed consistent gains did so by working to improve the entire instructional system. The districts were able to define and pursue a suite of strategies simultaneously and lock them together in a way that was seamless and mutually reinforcing.

At the same time, even these systemwide efforts left a number of chronically low-performing schools in place. But it may be the case that these districts are now in a better position to devote more focused attention on these few failing schools than districts that have not developed the same kind of systemic capacity.

To be fair, our contrasting district—Cleveland—also appeared to act at scale. Yet, Cleveland was more inclined to grant staff, principals, and teachers instructional autonomy, lacking the capacity to provide support to schools and teachers on a consistent, districtwide basis. In fact, part of the lesson from this study was that what sometimes passes as systemic reform is unlikely to produce results if the broad-scale reforms are poorly defined and executed.

The Interplay of Strategic vs. Tactical Reforms

It was also clear from our study that districts making consistent progress in either reading or mathematics undertook convincing reforms at both the strategic level—as a result of strong, consistent leadership and goal-setting—and the tactical level, with the programs and practices adopted in the pursuit of higher student achievement. There is little other way to explain why some districts saw larger gains in one subject or another when their strategic-level reforms looked very much alike.

At first glance, it may seem that it was the adoption of specific reading or math programs that produced the differing results in each city, but that is not the case. The successful tactical reforms were not program-specific. The Atlanta school system, for example, achieved significant gains in reading, although it did not actually use a single reading program. Instead, it used a series of comprehensive school reform demonstration models that have shown little effect in other major cities. And the math program used in

the Boston school system, which saw substantial gains in mathematics, was the same one used in the Cleveland school system, which saw little math gain.

What allowed these programs to work was a series of decisions regarding how to implement them with consistency and fidelity, how to leverage the expertise and focus of district reading and math directors and teachers, and how to thoughtfully and continuously refine the programs, based on what performance data suggested. These tactical efforts were clearly the main factors driving the patterns of gains that the study team observed in Atlanta, where growth in reading outpaced growth in mathematics, and in Boston, where growth in math outpaced growth in reading.

At the same time, it seems implausible that these tactical changes by themselves could have sustained the gains in either reading or math without having broader strategic reforms in place. Instead, it was the combined force of tactical decisions made in the name of well-defined, strategic efforts that seemed to yield the largest gains in achievement. In fact, although the district contexts differed, there was often more commonality across districts at the strategic level than at the tactical level. While the programs and approaches they chose may have varied, the success of reforms in Atlanta, Boston, and Charlotte was driven by stable, longstanding leadership teams and the ability of these leaders to translate a vision for improvement into definable goals and to hold people responsible for attaining these goals. These strategic factors served to define a broad set of expectations and preconditions for the tactical reforms under them.

Phases of Reform

The reader will note from the data in the main report that the study districts did not start their reforms at the same level of student proficiency and staff capacity. In addition, each city school system had its own history with reform, and each one had differing cultures, politics, and personalities that shape the sometimes erratic nature of urban school reform. And the reader should keep in mind that the starting point for reform was not necessarily 2003, the date we used to benchmark NAEP results.

Charlotte, for instance, had been pursuing standards-based reforms since the early 1990s. Its work in defining and implementing standards pre-dated that of most states, including North Carolina. The length of time that standards were in place, how comparatively well aligned they were to NAEP, the consistency and focus of their instructional program, the general consistency of the district's leadership, and the school system's lack of concentrated poverty relative to other cities may explain—in part—why Charlotte performed at or above national averages, even after adjusting for student background characteristics. If this is true, then it suggests that more time may be needed to attain something close to the same results in other cities.

At the same time, it is interesting that Charlotte did not see appreciable gains on NAEP during the study period. It is possible that what brought Charlotte to the national averages is not what it needed to move beyond this high level of achievement. It might have been the case that, in order for the district to see NAEP gains, Charlotte needed to move away from the kinds of prescriptive instructional programs that it was using in the 1990s and early 2000s toward programs that stressed more conceptual, higher-level understanding of academic content. And it may also be the case that the district's standing near the national average simply makes it hard to move beyond that level.

With Charlotte under new leadership, and having begun to move in new directions over the last several years, it will be interesting to see whether the reorientation of Charlotte's instructional program and theory of action will produce NAEP gains on the 2011 testing.

The strategies that Atlanta was using, on the other hand, were similar in intensity to what one sees in historically low-performing urban school districts that are working to create capacity, direction, and accountability from square one. The district and its leadership outlined a vision for reform and tightly defined and implemented it in a way that was necessary to break a culture of complacency. It was not likely that Atlanta could have seen substantial gains on NAEP without the clarity, direction, and discipline that defined its reform agenda over the last decade.

Likewise, Boston appears to have accurately gauged its overall performance levels and staff capacity to design a program in mathematics with a strong conceptual base, looser form of accountability, and strong overall leadership.

In other words, what may work at one stage of reform may not work at another. Recent analyses of data from the Program for International Student Assessment (PISA) suggest that the strategies used to move a district from poor to fair may be significantly different from those needed to move from good to great.²⁴ That finding is on ample display in this report. It is apparent that where one starts in the reform process matters in a district's ability to stage its efforts effectively over the years. A district's ability to accurately and objectively gauge where it is in the reform process and when and how to transition to new approaches or theories of action is critical to whether the district will see continuous improvement in student achievement or whether it will stall or even reverse its own progress.

The Role of Governance and Structural Change

The city school districts studied for this project included a mixture of governance structures. Some operated under the aegis of their mayors, and some had traditionally elected school boards. And while sample sizes were small, there was little reason to conclude that these structures of governance had a direct effect on NAEP gains, for high-achieving and improving districts and districts showing little gain were represented by governance structures of all types. Atlanta, which saw significant reading gains, and Charlotte, which had high performance, both had traditionally elected school boards; Boston, which saw significant math gains, and Cleveland, which saw few gains, were under mayoral control with appointed school boards.

To be sure, governance certainly has a role to play in district reform. For instance, Atlanta, which started its reforms with a traditionally elected but very fractious school board and a mayor who played little direct role in the school system, underwent a significant shift, with the business community playing a strong role in recruiting school board members who would constructively support the superintendent and her reforms. With this school board support, the Atlanta superintendent was able to push for a series of organizational changes to the system and spearhead the strategic reforms we referred to earlier that led to a decade of instructional change and growth on NAEP.

Yet what appears to matter in these differing governance and organizational models had less to do with who controlled the system than with what they did to improve student achievement. If the governance or organizational structure allows the district to focus on and support instruction in ways that it was not able to do under a more traditional structure, then it was likely to improve academic results—and to show greater gains than a traditional structure that did not focus on instructional improvement. Conversely, if

²⁴ Sources: Asia Society and the Council of Chief State School Officers, 2010. *International Perspectives on U.S. Education Policy and Practice: What Can We Learn from High-Performing Nations?* and Mourshed, M., Chijioke, C., and M. Barber (2010). *How the World's Most Improved School Systems Keep Getting Better*. Washington, D.C.: McKinsey & Company.

the structure—traditional or nontraditional—does not allow instructional changes to happen rather quickly or it does not focus on instruction, then it probably will not show much academic progress.

The same dynamic may also apply to various choice, labor, and funding issues. We did not explicitly study the relationship between NAEP scores and charter schools, vouchers, collective bargaining, or funding levels. But we note that these factors were present to differing degrees in both improving and non-improving districts. Boston and Cleveland, for instance, were unionized districts; Atlanta and Charlotte were not. Cleveland had vouchers; the others did not. Boston had high funding levels, while Atlanta and Charlotte did not. And all had a wide range in the number of charter schools that operated in each jurisdiction. We cannot conclude with certainty that these factors do not matter, but we believe it would be difficult to argue based on the data we have that any of these were critical factors in the improvement or lack of improvement on NAEP in the study districts.

An example might help. It is likely that instructional quality is driving the results seen in studies of charter school effectiveness relative to other public schools. A large number of these studies find that students in charter schools perform at roughly the same levels of other public school students—a conclusion that is unsurprising if, despite differences in governance, instructional programming is actually similar in both settings. The more important comparison would involve charter schools with unusually high performance, a comparison that is likely to show differences from regular schools in focus, accountability, time-on-task, and instructional quality.

The broader lesson is that governance and structural reforms alone are not likely to improve student achievement unless they directly serve the instructional program. We believe that this is an important lesson for all large-city school systems to heed, because so often it is the governance, organizational, funding, choice, and other efforts and initiatives that attract public attention, sometimes to the detriment of instructional improvements. We think this point is bolstered by how closely student gains on various NAEP strands seemed to be associated with what the districts were doing instructionally. It is not plausible, for instance, that the reorganization of the Atlanta school district, in itself, could have improved students' ability to read for information. But teacher and principal professional development that focused on those skills and was implemented in the context of broader strategic reforms might well have brought about the improvement. In other words, part of Atlanta's success on NAEP is a function of how well it organizationally aligned itself to its instructional priorities. This also appears to be the case in Boston.

Implications for Implementing the Common Core State Standards

Building on this point about the centrality of instructional quality and reform, we think that the results of this study have important implications for the development and implementation of curriculum and for classroom instruction, particularly in light of the new common core standards:

1. The low degree of content matching described in this study suggests that even clearly written curriculum supported by professional development and coaching might not produce the results we want with the common core if our instructional efforts are not broadly consistent with the new standards in quality, rigor, and capacity. In other words, a significant challenge for urban school districts and others will be to reflect the rigorous thinking behind the standards and their progressions without getting bogged down in each individual standard.
2. The results of this study also imply that districts that are able to use the new common core state standards to improve student achievement are more likely to do so with curriculum that lays out clear expectations about student performance to all staff members, provides teachers with explicit examples of

student work illustrating varying levels of concept mastery, and differentiates instruction for students who bring special challenges to the classroom.

3. The new common core standards will compel classroom instruction that is more conceptual in its orientation than what most educators are used to. In mathematics, for example, the new common core state standards will require deeper understanding of math concepts and more rigorous application of them. Boston's experience in boosting subscale performance in both fourth and eighth grades using a more concept-based program will help our understanding of how to meet the demands of the common core in other major urban school systems.

Similar shifts may also be required in reading, as the common core state standards will emphasize far more reading for information than is currently the case in most classroom instruction, curricula, or textbooks. The data from this project suggest that urban school districts generally did less well in this area than they did in reading for literary experience. Over the long run, the growing emphasis on teaching concepts should result in students doing well academically regardless of the nature of the tests.

4. Finally, the implementation of the common core will depend heavily on the overall effectiveness and commitment of teachers and administrators alike, as well as the capacity of districts to support their teacher corps through a variety of strategies. Ultimately, the implementation of the common core standards should raise the overall quality of people who want to be in the education field in the first place, because the new standards will define a higher bar for what is required to ensure that students are academically prepared for a more complex future. Establishing the mechanisms by which this process works will be one of education's most substantial challenges as the new standards spread and the nation moves toward becoming more internationally competitive.

Recommendations

The Council of the Great City Schools and the American Institutes for Research make the following recommendations to urban school districts participating in the Trial Urban District Assessment of the National Assessment of Educational Progress (NAEP), as well as to other urban districts, on how they might increase or accelerate the academic progress that they have been making.

1. Devote considerable time and energy to articulating and building both a short-term and a long-term vision among city leaders, the school board (whether appointed or elected), the superintendent, key instructional staff members, and teachers for the direction and reform of the school system—and then sustain it over time, even when the individual actors change.
2. Take advantage of the development and implementation of the common core state standards to upgrade and align the district's curriculum (in scope, richness, and balance), materials, professional development, teacher and student supports and monitoring, assessments, communications, and community outreach efforts. It is clear from the results of this study that the common core is not likely to boost student achievement by itself, without high quality instructional programming consistent with the new higher standards and strong student supports.
3. Ensure that the school district has the right people in the right places to lead reforms, build coalitions, and oversee change management. Devote long-term strategic effort to building and enhancing the capacity of district personnel at both the central-office and school levels to deliver high-quality instruction and manage operations.

4. Continuously evaluate the effectiveness of instructional programs, professional development, personnel recruitment and deployment, data systems, and student supports and interventions—and make strategic and tactical changes as necessary based on that data.
5. Ensure that the implementation of reforms is monitored for fidelity, and that district accountability and personnel evaluation systems align with district academic goals and priorities.
6. Allow sufficient time for district reforms to take root, while using data to make necessary tactical changes. Our findings showed that persistence over a sustained period (more than five years) was critical to a district's ability to see long-term improvement, despite low initial buy-in and early results.
7. Be mindful of where your district stands in the reform process, and what approaches are appropriate and necessary to either kick start or sustain progress according to your current needs, levels of student achievement, and staff capacity.
8. Create multi-faceted internal outreach and communications systems, so staff members throughout the organization understand why they are doing what they are doing. Build a culture of ownership in both the work and the results.
9. Keep budget cuts away from the classroom as much as possible, so students are not affected by sudden changes, drops or shifts in personnel, or alterations in programs that have been producing results. If teachers have to be reassigned to grades or subjects they have not taught recently, ensure that they have adequate supports and professional development to enable them to adapt and deliver quality instruction in their new assignments.
10. Be transparent with your district's data, don't overstate your progress, and be your own toughest critic.

Conclusions

The purpose of this study was to answer a series of important questions about the degree and nature of urban school improvement and to determine what separates urban districts that have made progress from those who haven't. We tried to answer these questions by looking at the trends, standards, characteristics, and practices of big-city school systems with widely contrasting performance. These analyses have helped us draw lessons about the factors behind the improvement of urban school systems and the barriers that may slow down our progress.

This study affirms many of the conclusions that the Council of the Great City Schools made in its 2002 report with MDRC, *Foundations for Success*, and broadens our understanding of what spurs academic gains in urban school systems—or fails to do so—into such areas as standards, alignment, organizational structure, accountability, rigor and instructional focus and cohesion.

Over the long run, we will need to do more than explain *post hoc* why urban school systems improved or why they did not. We will need to be able to predict it. This study puts us a step closer to being able to predict which large-city school districts are likely to show progress on the NAEP and under what circumstances the gains are likely to occur.

The challenge, of course, is not to forecast improvement for its own sake, but to be more confident that we are looking at the right levers in raising student achievement in large-city school districts. If we are

not confident of that, then there may be reason to think that gains are coming for reasons that we have not been able to articulate and that large-city school systems may be pursuing the wrong reforms. As NAEP trend lines get longer, as more urban districts participate in the TUDA program, and as the research base grows, our ability to understand what is likely to spur better performance should improve.

This study also raises some interesting questions and avenues for future research. For example, the need for policies and programming designed to raise student achievement among our most vulnerable student groups has become imperative. In our examination of the patterns of achievement on NAEP, we found that the districts in which students in the aggregate made progress in reading and math saw academic improvement in these subjects among individual student groups as well. Among African American students nationally, for example, those in the Atlanta Public Schools tended to show some of the strongest gains in reading; and in the Boston Public Schools, African American, Hispanic, and poor students tended to show some of the most consistent gains in mathematics. In neither of these cases, however, were African American, Hispanic, poor, or other student groups targeted for special programming. The assumption in each of these cases appeared to be that good instruction for some students was good instruction for all students. However, this study leaves unanswered questions about the potential that specialized, targeted or differentiated programming and services might hold, or what strategies will be necessary to not only raise achievement across the board, but to eliminate achievement gaps based on poverty or race.

Another unanswered question arises from the nature and size of the gains documented in this study. While we may have succeeded in identifying characteristics and approaches of districts that have helped move the needle on student achievement, we are left to ponder what the effects on NAEP performance would be if any of these cities pursued the broader and more wholesale level of reforms seen in such high-performing nations as South Korea, Finland, and Singapore. It is also left for speculation what the effects on NAEP achievement might be if districts pursued reforms that are widely discussed in the public arena, i.e., performance pay, the alteration of seniority systems, more aggressive turnaround of troubled schools, and similar initiatives.

Whatever its unanswered questions, however, this study shows that there is increasing reason to be optimistic about the future of urban public education, not because big-city schools are making significant progress (which they are) but because the progress appears to be the result of purposeful and coherent reforms. This exploratory report was part of our larger effort to increase our performance as urban educators through knowledge and research.

Too much of the history of urban education has been defined around who is valuable in this society and who is not; for whom we have high hopes and for whom we have no hopes at all; for whom we have high standards and for whom we hold no great expectations. But our job in public education is not to reflect and perpetuate these inequities or to let them define us or hold us or our kids back. Our job is to overcome them. The great civil rights battles were not fought so that urban children could have access to mediocrity; they were fought over access to excellence and the resources to provide it. Our job is to create excellence. This project is one more step toward that goal, one more piece of the puzzle.

BIBLIOGRAPHY

Asia Society and the Council of Chief State School Officers, 2010. *International Perspectives on U.S. Education Policy and Practice: What Can We Learn from High-Performing Nations?* Washington, DC: Authors.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, 1*, 289–300.

Braun, H., Jenkins, F., and Grigg, W. (2006a). *Comparing private schools and public schools using hierarchical linear modeling* (NCES 2006-461). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Braun, H., Jenkins, F., and Grigg, W. (2006b). *A closer look at charter schools using hierarchical linear modeling* (NCES 2006-460). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Braun, H., Zhang, J., and Vezzu, S. (2006). *Evaluating the effectiveness of a full-population estimation method* (Unpublished paper). Princeton, NJ: Educational Testing Service.

Forgione Jr., P. D. (1999). *Issues surrounding the release of the 1998 NAEP Reading Report Card. Testimony to the Committee on Education and the Workforce*, U.S. House of Representatives, on May 27, 1999. Retrieved March 16, 2006, from http://www.house.gov/ed_workforce/hearings/106th/oi/naep52799/forgione.htm.

Horwitz, A., Uro, G., et al. (2009) *Succeeding with English language learners: Lessons learned from the Great City Schools*. Washington, D.C.: Council of the Great City Schools.

Mathematics 2009, Trial Urban District Assessment, Results at Grades 4 and 8. National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education (NCES 2010-452), 2010.

McLaughlin, D. H. (2000). *Protecting state NAEP trends from changes in SD/LEP inclusion rates* (Report to the National Institute of Statistical Sciences). Palo Alto, CA: American Institutes for Research.

McLaughlin, D. H. (2001). *Exclusions and accommodations affect state NAEP gain statistics: Mathematics, 1996 to 2000* (appendix to chapter 4 in the NAEP Validity Studies Report on Research Priorities). Palo Alto, CA: American Institutes for Research.

McLaughlin, D. H. (2003). *Full population estimates of reading gains between 1998 and 2002* (Report to NCES supporting inclusion of full population estimates in the report of the 2002 NAEP reading assessment). Palo Alto, CA: American Institutes for Research.

McLaughlin, D. H. (2005). *Properties of NAEP full population estimates* (Unpublished report). Palo Alto, CA: American Institutes for Research.

McLaughlin, D. H. (2005). *Achievement gap display study* (NAEP State Analysis Project Technical Report to NCES). Palo Alto, CA: American Institutes for Research.

Mourshed, M., Chijioke, C., and M. Barber (2010). *How the World's Most Improved School Systems Keep Getting Better*. Washington, D.C.: McKinsey & Company.

Reading 2009, Trial Urban District Assessment, Results at Grades 4 and 8. National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education (NCES 2010-459), 2010.

Science 2009, Trial Urban District Assessment, Results at Grades 4 and 8. National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education (NCES 2010-452), 2011.

Snipes, J., Dolittle, F., and Herlihy, C. (2002) *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, D.C.: Council of the Great City Schools.

Wise, L. L., Hoffman, R. G., and Becker, D. E. (2004). *Testing NAEP full population estimates for sensitivity to violation of assumptions* (Technical report TR-04-27). Alexandria, VA: Human Resources Research Organization.

U.S. Department of Education, National Center for Education Statistics, Common Core of Data, *Local Education Agency Universe Finance Survey 2008*.

APPENDIX.
**RESEARCH ADVISORY PANEL
AND RESEARCH TEAM**

Research Advisory Panel

Dr. Peter Afflerbach, Professor of Education
University of Maryland

Robin Hall, Principal
Atlanta Public Schools

Dr. Karen Hollweg, Director of K-12 Science Education (retired)
National Research Council

Dr. Andrew Porter, Dean
Graduate School of Education
University of Pennsylvania

Dr. Norman Webb, Senior Research Scientist
Wisconsin Center for Educational Research
National Institute for Science Education

Dr. Karen Wixson, Professor of Education
University of Michigan

Research Team

1. Council of the Great City Schools

Michael Casserly, Executive Director
Ricki Price-Baugh, Director of Academic Achievement
Sharon Lewis, Director of Research
Amanda Corcoran, Research Manager
Renata Uzzell, Research Manager
Candace Simon, Research Manager
Shirley Schwartz, Director of Special Projects

2. American Institutes for Research

Dr. Jessica Heppen, Senior Research Analyst
Steve Leinwand, Principal Research Analyst
Terry Salinger, Chief Scientist, Reading Research
Victor Bandeira de Mello, Principle Research Scientist
Enis Dogan, Senior Research Scientist
Mike Garet, Vice President Education, Human Development in the Workforce
Laura Novotny, Senior Research Analyst
Kerri Thomsen, Research Associate
Melissa Kutner, Research Assistant

Site Visit Teams

1. Atlanta

Michael Casserly, Executive Director
Council of the Great City Schools

Ricki Price-Baugh, Director of Academic Achievement
Council of the Great City Schools

Sharon Lewis, Director of Research
Council of the Great City Schools

Renata Uzzell, Research Manager
Council of the Great City Schools

Nancy Timmons, Chief Academic Officer (former)
Fort Worth Independent School District

Harry Pratt, Consultant
Science Associates
President of National Science Teachers Association

2. Boston

Michael Casserly, Executive Director
Council of the Great City Schools

Ricki Price-Baugh, Director of Academic Achievement
Council of the Great City Schools

Sharon Lewis, Director of Research
Council of the Great City Schools

Amanda Corcoran, Research Manager
Council of the Great City Schools

Nancy Timmons, Chief Academic Officer (former)
Fort Worth Independent School District

Norma Jost, Math Supervisor
Austin Independent School District

3. Charlotte-Mecklenburg

Ricki Price-Baugh, Director of Academic Achievement
Council of the Great City Schools

Sharon Lewis, Director of Research
Council of the Great City Schools

Candace Simon, Research Manager
Council of the Great City Schools

Nancy Timmons, Chief Academic Officer (former)
Fort Worth Independent School District

Maria Crenshaw, Director of Instruction
Richmond Public Schools

Harry Pratt, Consultant
Science Associates
President of National Science Teachers Association

4. Cleveland

Michael Casserly, Executive Director
Council of the Great City Schools

Ricki Price-Baugh, Director of Academic Achievement
Council of the Great City Schools

Sharon Lewis, Director of Research
Council of the Great City Schools

Candace Simon, Research Manager
Council of the Great City Schools

Nancy Timmons, Chief Academic Officer (former)
Fort Worth Independent School District

Linda Davenport, Director of Mathematics
Boston Public Schools

About the Council of the Great City Schools

The **Council of the Great City Schools** is a coalition of 65 of the nation's largest urban public school systems. The organization's Board of Directors is composed of the Superintendent, CEO or Chancellor of Schools, and one School Board member from each member city. An Executive Committee of 24 individuals, equally divided in number between Superintendents and School Board members, provides regular oversight of the 501(c)(3) organization. The composition of the organization makes it the only independent national group representing the governing and administrative leadership of urban education and the only association whose sole purpose revolves around urban schooling.

The mission of the Council is to advocate for urban public education and assist its members in their improvement and reform. The Council provides services to its members in the areas of legislation, research, communications, curriculum and instruction, and management. The group convenes two major conferences each year; conducts studies of urban school conditions and trends; and operates ongoing networks of senior school district managers with responsibilities for areas such as federal programs, operations, finance, personnel, communications, research, and technology. Finally, the organization informs the nation's policymakers, the media, and the public of the successes and challenges of schools in the nation's Great Cities. Urban school leaders from across the country use the organization as a source of information and an umbrella for their joint activities and concerns. The Council was founded in 1956 and incorporated in 1961, and has its headquarters in Washington, D.C.

Chair of the Board

Winston Brooks, Albuquerque Superintendent

Chair-elect of the Board

Candy Olson, Hillsborough County School Board

Secretary/Treasurer

Eugene White, Indianapolis Superintendent

Immediate-past Chair

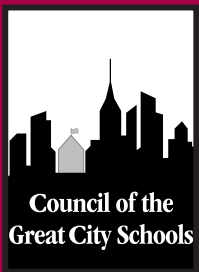
Carol Johnson, Boston Superintendent

Achievement Task Force Chairs

Eileen Cooper Reed, Cincinnati School Board

Carlos Garcia, San Francisco Superintendent

Michael Casserly, Executive Director



THE COUNCIL OF THE GREAT CITY SCHOOLS

1301 Pennsylvania Avenue, NW
Suite 702
Washington, DC 20004

202-393-2427
202-393-2400 (fax)
www.cgcs.org