PIECES OF THE PUZZLE

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

The Council of the Great City Schools

Fall 2011

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

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 led 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

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 2

Chapter 4

EXECUTIVE SUMMARY

Overview

This report summarizes preliminary and exploratory research conducted by the Council of the Great City Schools and the American Institutes for Research on urban school systems participating in the Trial Urban District Assessment (TUDA) of the National Assessment of Educational Progress (NAEP). The study is one of the first large-scale analyses of urban NAEP trends, and the first to examine local instructional and organizational practices alongside changes in NAEP scale scores in the participating cities. This report is also preliminary in the sense that it attempts to lay out a framework for how NAEP data on the TUDA districts might be analyzed in the future as the number of participating cities grows and the amount of data expands.

The purpose of this project was to identify urban school systems that are making academic progress and to examine possible factors in their improvement. The overarching goal was to identify variables that might be contributing to improvement in urban education across the nation and to explore what might be needed to accelerate those gains. The report also discusses broad lessons for the implementation of the common core state standards.

Summary of Methodology

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 math at the fourth- and eighth-grade levels, both overall and at differing points across the distribution of student achievement scale scores?

- Which of the TUDA districts outperformed the others on NAEP, controlling for relevant student background characteristics?

- Which of the TUDA districts have made significant and consistent gains on NAEP in reading and math at the fourth- and eighth-grade levels among student groups defined by race/ethnicity, language, and other factors?

- How have selected TUDA districts scored on NAEP subscales in reading, math, and science? What were their relative strengths and weaknesses across the subscales?

- What was the degree of alignment between (1) the NAEP frameworks in place between 2003 and 2007 in reading, math, and science and (2) selected districts' respective state standards? What was the relationship between that alignment and district performance or improvement on NAEP during those years?

- What instructional conditions and practices were present in districts that made significant and consistent gains on 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?

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, the Council of the Great City Schools and the American Institutes for Research used data from NAEP spanning 2003 to 2007, the latest year available when this project started. The report also summarizes reported scale scores from 2003 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 used data from 2003, 2005, and 2007 on fourth- and eighth-grade reading and math achievement. Because science results were available only on 2005 testing when the analysis for this report was conducted, we could not examine trends in science scale scores.[1] However, one-year science results are presented. All data were analyzed using both reported results and scale scores that account for differences in exclusion rates, known as full population estimates (FPE). For some analyses, scale scores also were also 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, all 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 only occur at the higher or lower end of the distribution?).

We used these processes to select two districts showing significant and consistent improvements in reading and math, as well as one district that did not show 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 using both reported NAEP results and full population estimates and determined that the selection of districts did not depend on the kind of analysis we conducted.

Fourth, we analyzed NAEP trends by student group for each of the TUDA city school systems to ensure that the study districts were not showing gains at the expense of one group or another. The

[1] The 2009 NAEP Science Assessment, which was released in February 2011, used a different framework than used in 2005, so these assessment results cannot be compared to one another to show changes in achievement over this time period.

analysis included trends by race/ethnicity, gender, eligibility for the National School Lunch Program (NSLP), disability, and language status.

Fifth, to determine whether there were any discernable strengths and weaknesses in reading, math, 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" 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 average 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 each subject area—reading, math, and science—and comparing them to state (and district) standards that were in place in reading and math in 2007 and in science in 2005. 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 codes were determined by at least three 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 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.

Major Findings

The analysis yielded a number of important and unique results that improve our understanding of how and why urban school districts show progress on NAEP and that enhance our ability to boost student achievement in the future. All results refer to reported scale scores unless otherwise indicated as being results from full population estimates. Major findings include the following—

Overall Urban Trends

- Public schools in the large cities[2] showed statistically significant gains between 2003 and 2009 in fourth and eighth grade reading and fourth- and eighth-grade mathematics.

[2] The bullets in this section refer to the large-city (LC) school sample and not to results on individual TUDA districts.

- Between 2003 and 2009, the public schools in the large cities showed significantly larger gains statistically than the national public school sample in both fourth and eighth grades, in both reading and math.

- Large-city schools and TUDA districts, on average, scored below the nation on NAEP in reading, math, and science in both fourth and eighth grades.

- Between 2003 and 2007, there were statistically significant gains in reading using FPEs among large-city fourth graders in the second, third, and fourth quintiles—or the achievement bands representing students between the 21st and 80th percentiles. In contrast, the nation showed a statistically significant improvement across all quintiles. In the eighth grade, the large-city schools showed no appreciable movement in reading in any quintile, while the nation showed statistically significant declines in the lowest and the two highest quintiles.

- Between 2003 and 2007, there were statistically significant gains in mathematics using FPEs in large-city schools across the achievement distribution, although there were exceptions. In fourth grade, large cities showed statistically significant improvements in average scale 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 math at every quintile, as did the national sample.

- Between 2003 and 2007, the large-city schools made more average gains using FPEs (both overall and at selected ranges of the achievement distribution) in mathematics (across the five quintiles) at both fourth- and eighth-grade levels than they made in reading at both grades.

Cities Showing or Failing to Show Significant and Consistent Gains

- Of the 11 TUDA districts, the Atlanta Public Schools made significant—as well as the most consistent[3] —improvements in reading between 2003 and 2007 at both the fourth- and eighth-grade levels, even after adjusting for test-exclusion rates (FPE).[4]

- Of the 11 TUDA districts, the Boston Public Schools made significant—as well as the most consistent—gains in mathematics between 2003 and 2007 at both fourth- and eighth-grade levels, after adjusting for test-exclusion rates (FPE).

- Other cities posted significant gains between 2003 and 2007, but the progress was often seen in only one subject and one grade level, rather than being as uniform as the improvements in Atlanta and Boston.

---

[3] By "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.

[4] 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.

- The Charlotte-Mecklenburg Public Schools outperformed all other TUDA districts in reading and math at both grade levels, after controlling for relevant student background characteristics and test exclusion rates (FPE). The district also scored as high as or higher than the national averages and showed student group performance that was higher than peer-group performance nationwide.

- 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 (FPE).

Districts Showing Higher or Lower Performance than Expected Statistically

- In grade four reading, Austin, Boston, Charlotte, New York City, and San Diego scored higher in 2007 than would be expected statistically among the 11 TUDA districts, given their student background characteristics. Chicago, Cleveland, the District of Columbia, and Los Angeles scored lower. Results were not different from what was predicted in Atlanta and Houston.

- In grade eight reading, Austin, Boston, Charlotte, Chicago, and Houston scored higher in 2007 than would be expected statistically among the 11 TUDA districts given their student background characteristics. District of Columbia and Los Angeles scored lower. Results were not different from what was predicted in Atlanta, Cleveland, New York City, and San Diego.

- In grade four math, Austin, Boston, Charlotte, Houston, and New York City scored higher in 2007 than would be predicted statistically among the 11 TUDA districts given their student background characteristics. Atlanta, Chicago, Cleveland, the District of Columbia, and Los Angeles were lower. Results were the same as predicted in San Diego.

- In grade eight math, Austin, Boston, Charlotte, Houston, and New York City scored higher in 2007 than would be predicted statistically among the 11 TUDA districts given their student background characteristics. The District of Columbia and Los Angeles were lower. Results were the same as predicted in Atlanta, Chicago, Cleveland, and San Diego.

Gains among Student Groups

- Atlanta, Boston, the District of Columbia, Houston and New York City posted significant reading gains between 2003 and 2009 at the fourth-grade level among African American students. Austin made significant reading gains among fourth graders between 2005, when they were first tested, and 2009. Atlanta showed significant reading gains among African American eighth graders. Atlanta, Boston, Chicago, the District of Columbia, and New York City posted significant math gains among fourth-grade African American students, and Atlanta, Boston, Charlotte, Chicago, the District of Columbia, Houston, Los Angeles, New York City, and San Diego posted significant math gains among African American eighth graders. And Austin made significant math gains among eighth graders between 2005 and 2009.

- Boston and the District of Columbia saw significant increases between 2003 and 2009 in reading scale scores among Hispanic fourth graders, and Houston and Los Angeles showed significant increases among Hispanic eighth graders in reading. Boston, Chicago, the District of Columbia, Houston, Los Angeles, New York City, and San Diego showed significant increases among Hispanic fourth graders in math. Boston, Chicago, the District of Columbia, Houston, Los Angeles, and San Diego made significant math gains among Hispanic eighth graders. Austin made significant math gains among eighth grade Hispanics between 2005 and 2009.

- Atlanta, Boston, Charlotte, Chicago, the District of Columbia, Houston, and New York City had significant increases between 2003 and 2009 in reading among National School Lunch Program (NSLP)-eligible fourth graders, and Atlanta, Boston, Houston, and Los Angeles made significant reading gains among NSLP-eligible eighth graders. In math, Atlanta, Boston, Chicago, the District of Columbia, Houston, Los Angeles, New York City, and San Diego made significant gains among NSLP-eligible fourth graders, and Atlanta, Boston, Charlotte, Chicago, the District of Columbia, Houston, Los Angeles, New York City, and San Diego made significant math gains among eighth graders. Austin made significant gains between 2005 and 2009 in eighth grade math.

- The District of Columbia and Houston made significant reading gains among limited English proficient (LEP) fourth graders. No one made significant reading gains among LEP eighth graders between 2003 and 2009. Boston, the District of Columbia, Houston, New York City, and San Diego made significant gains in math among LEP fourth graders, and Chicago, Houston, and San Diego made significant gains in math among eighth graders.[5] Austin made significant gains between 2005 and 2009 among eighth graders in reading and math.

Academic Strengths and Weaknesses

- 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 in 2007 tend to be strongly correlated, i.e., students who score well on one subscale tend to do well on others. In general, however, fourth graders in TUDA districts appeared to do somewhat better at reading for literary experience than at reading for information. There was considerable variation from city to city in the eighth grade, but it appeared that students in the 11 districts were more likely to do better in reading for literary experience than in reading for information or reading to perform a task.[6]

- In math, NAEP tests students in number properties and operations (“number” for short), measurement, geometry, data analysis and probability, and algebra. The analysis of 2007 TUDA results indicated considerable variation from city to city, but in general, fourth graders in TUDA districts appeared to score better in geometry and number and less well in measurement, algebra, and data. At the eighth-grade level, TUDA students appeared to do better in geometry and algebra and less well in number, data, and measurement.

- NAEP also assesses students at the fourth- and eighth-grade levels on their knowledge in the areas of earth science, physical science, and life science. The analysis of 2005 data indicated results that were low across the board among participating cities. While there

---

[5] Please note that in this report, the terms limited English proficient (LEP) and English language learners are sometimes used interchangeably.
[6] Tests of statistical significance were not performed on these subscale differences in reading, math, or science.

was considerable variation from one city to another in strengths and weaknesses, in general, fourth graders in TUDA districts appeared to do somewhat better in life sciences than in earth science and physical science, while eighth graders appeared to do equally well in all three fields of science.

Alignment Gaps

- The extent of content alignment between NAEP specifications and the respective district/state standards of the four selected TUDA cities ranged from a complete match or partial match of 48 percent to 80 percent in fourth-grade reading, 41 percent to 65 percent in eighth-grade reading, 66 percent to 72 percent in fourth-grade math, 51 percent to 84 percent in eighth-grade math, 19 percent to 57 percent in fourth-grade science, and 25 percent to 48 percent in eighth-grade science.

- Complete content matches in fourth- and eighth-grade reading and math were characterized as low to moderate, while fourth- and eighth-grade science matches were characterized as low.

- There was no apparent relationship between student performance or gains on NAEP and the degree of complete content alignment between NAEP specifications and state/district standards, although the sample size was too small to be definitive.

- The cognitive demand in the state and district standards was often similar to NAEP among specifications that were completely matched.

Differences in Practice and Results

- While the study was not designed to determine causality, it appears that instructional practices at the district level were more important in a school system's ability to improve on NAEP than whether their state standards were aligned with NAEP frameworks. The results of this study suggest that some districts made significant improvements on NAEP even when their state standards were not well-aligned with NAEP. Conversely, high alignment did not guarantee better results or more gains.

- Despite their differences, there were a number of traits and themes common among the improving and 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 set clear, systemwide goals and held staff members accountable for results, creating a culture of shared responsibility for student achievement.

  Curriculum and Instruction. The three improving and 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 with well-defined professional development or coaching tied to instructional programming 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 improving or high-performing 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 improving or high-performing districts had regular assessments of student learning and used these assessment data and other measures to gauge student learning, modify practice, and target resources and support.

- Among the TUDA districts, Atlanta showed significant and the most consistent overall gains in reading between 2003 and 2007. The district's literacy program, unlike those in many cities, did not use a single commercial reading program. Begun in 2001, the literacy initiative was instituted across the curriculum and remained largely unchanged during the intervening years. The district had strong leadership and program staff with deep content knowledge at each level of the organization and created a strong accountability system that emphasized growth across multiple performance levels, rather than gains only in the number of students reaching proficiency. This situation may have contributed to the district's improvement in reading across its achievement distribution (i.e., at every quintile). The district also devised a unique organizational structure that provided focused technical assistance and capacity-building to its schools, based on the use of detailed data on student achievement, a shared sense of mission, and staff teams with strong pedagogical knowledge. Finally, the district instituted a universal and sustained professional development effort that emphasized reading for information in fourth grade and reading to perform a task in eighth, areas in which the district showed the greatest gains.

- Boston showed significant and the most consistent gains in math on NAEP. Like Atlanta, Boston had strong and stable leadership at the school board, superintendent, and program-director levels. Boston's math leadership team began implementing a common, challenging, concept-rich math program 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. Also, like Atlanta with reading, Boston kept its math program in place for many years, supporting it with extensive and sustained professional development and coaching assistance for teachers. Unlike Atlanta, Boston had a "softer" accountability system, but the district was able to create a strong culture in support of results that served many of the same purposes as Atlanta's more formalized system.

- While Charlotte did not demonstrate the same gains as Atlanta or Boston in NAEP reading and math over the study period, the district maintained consistently high performance at or above the national averages from 2003 to 2007. Charlotte was selected for study because, after controlling for student background characteristics such as poverty and English language learner status, it out-performed all other TUDA districts in reading and math in 2007. Charlotte-Mecklenburg was one of the first districts in the nation to develop and institute academic standards and was a leader in pioneering an instructional management theory of action, something it has moved away from since 2007 in favor of less centralized instructional control. During the study period, Charlotte-Mecklenburg had (1) a highly defined curriculum and tiered interventions, (2) formal accountability systems with bonuses for improved student achievement, (3) regular assessments of student progress throughout the school year, (4) well-developed data systems that informed instruction and the management of instruction, and (5) expert central-office teams capable of intervening in schools if and when they fell behind.

- Cleveland—the district that showed few gains on NAEP between 2003 and 2007—had reasonably high alignment between its standards and NAEP. Yet until 2006, there was no functional curriculum 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. Ironically, according to school system officials, the district used the same math program as Boston but never expanded its use when it showed results. The district also lacked a system for holding its staff and schools accountable for student progress in ways that other 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. The district also endured substantial budget cuts in 2005 that resulted in the dramatic bumping of teachers by seniority, a move that left many working in subjects and grades for which they were unprepared, and there were few central office staff members who could help. By 2007, the district had fewer teachers for a school system of its enrollment than any of the other TUDA districts.

Conclusions

The report concludes with a discussion of findings and implications for improving urban education. In particular, the findings that student improvement on the NAEP was related less to content alignment than to the strength or weakness of a district's instructional programming has significant implications for the new Common Core State Standards. Many educators—and the public in general—assume that putting into place more demanding standards alone will result in better student achievement, but this study suggests 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, integrated, and consistent with the standards, and the standards themselves are not well-implemented.

This finding also has implications for a variety of high profile reform strategies and governance models. The city school systems studied for this project included a mixture of governance models, ranging from mayor-controlled systems to more traditional district structures. Yet what appears to matter in these differing organizational models has less to do with who controls the system than with what they do to improve student achievement. The same dynamic may also apply to various choice, labor, and funding models. We did not explicitly study the relationship between NAEP scale 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. The broader lesson is that these structural reforms are not likely to improve student achievement unless *they directly serve the instructional program*. We believe this is an important lesson for all large-city school systems to heed, because so often it is the governance, funding, choice, and other efforts and initiatives that attract the most attention, sometimes to the detriment of instructional improvements.

What may have also emerged from this study is further evidence that progress is possible when districts act at scale and systemically rather than trying solely to improve one school at a time. Moreover, it was clear from our study that districts making consistent progress in either reading or math 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 academic achievement.

Finally, each city school system had its own history with reforms, and each one had differing cultures, politics, and personalities that shape the sometimes erratic nature of urban school reform. It was apparent that a district's ability to accurately and objectively gauge where it is in the reform process, what its capacities are, 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 report wraps up with a short list of recommendations to urban school districts about what they might put into place based on the findings of this report and a set of conclusions about next steps.

CHAPTER I
INTRODUCTION AND ORGANIZATION OF REPORT

Purpose

America's urban schools are under more pressure to improve than any other institution—public or private—in the nation. They are being told to produce results or get out of the way. They are being told to improve or see the public go somewhere else. They are being told to be accountable for what they do or let someone else do it.

Some of this criticism is justified. Some of it is not. Either way, the nation's urban schools are being challenged in the court of public opinion and by history to improve student achievement to levels that the nation has never asked of them before.

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—and there are plenty of both on which to draw—and aggressively pursuing reforms that will boost academic performance.

There is fresh evidence that the efforts of these urban school systems are beginning to pay off. Academic achievement among urban students—the subject of this report—shows signs of improving. The gains have not muted the criticism or eased the pressure, but preliminary trend lines suggest that urban public education may be heading in the right direction. Still, urban schoolchildren lag behind their peers nationwide.

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 the National Assessment of Educational Progress (NAEP) reading and math assessments at grades four and eight. 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 also 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 the results, and what may be needed to further improve urban public schooling nationwide. The study also provides an initial framework for how future analyses might be conducted as more city school systems participate in the Trial Urban District Assessment (TUDA).

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?
- 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?

The Council of the Great City Schools tried to answer these questions with a number of initiatives that began in 2000. First, the Council persuaded the National Assessment Governing Board (NAGB) and Congress to oversample big-city school districts during the regular administrations of NAEP. The districts that volunteered for 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 member districts 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 reforms in ways that the current 50-state assessment system did not allow. NAGB and Congress granted the Council's request and the project grew from an initial cohort of six cities in 2002 to 11 cities in 2007 to 18 cities in 2009 and to 21 cities—or about one-third of the Council's membership—in 2011.

The Council's second initiative in 2000 was to launch a research project that asked some of the same questions that this report asks. Were urban schools getting better? Which urban districts were showing the largest gains and why? At the time, the organization only had state assessment results to work with, but the report that came out of the effort—*Foundations for Success: Case Studies of How Urban School Systems Improve Student Achievement*—was the first national study of big-city schools and hinted strongly at what lay behind the gains in some urban school districts.

Based on its research on why some districts improved and others did not, the Council launched its third initiative: A one-of-a-kind effort to provide technical assistance directly to urban school systems to improve academic performance. The Strategic Support Teams that the Council created to do the work crafted their proposals for improvement around the research and the experience of big-city school systems that were already showing gains, and found that city school systems that followed the blueprints could often make important progress. In turn, the efforts of the teams continue to inform the research on what is working and what isn't.

The Council's fourth initiative involved asking questions about the academic improvement of critical student groups, starting with English language learners. The questions were similar to the broader queries that the organization was asking about urban school districts, but in this case the questions concerned the student groups with whom our schools most needed to make progress. Were we making headway academically with English learners, for example? Which urban districts were making the most progress? What were these districts doing, and how did their practices differ from districts not making headway? Preliminary answers were proffered in the report, *Succeeding with English Language Learners: Lessons Learned from the Great City Schools.* Similar efforts are now underway with African American males after the Council's publication of *A Call for Change: The Social and Educational Factors Contributing to the Outcomes of Black Males in Urban Schools*.

Finally, with this new report, the Council has returned to some of the original questions that prompted cities to participate in TUDA in the first place. 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 and to start identifying possible driving forces behind these patterns. Specifically, this report examines—

- The academic performance and trends of large-city schools in the aggregate, compared with the nation at large.

- The performance of individual urban school systems that showed significant and consistent gains on NAEP reading and math from 2003 to 2007—on average and at all points along the distribution of achievement scale scores (i.e., with the highest and lowest performers and those students in between).

- The performance of individual urban school districts that appeared to outperform other urban school systems on the NAEP, after adjusting for relevant student background characteristics.

- The academic progress that urban school systems have made with critical student groups, including African Americans, Hispanics, poor students, students with disabilities,[1] and English language learners.

- The specific academic components and areas in reading, math, and science where urban students showed particular strengths and weaknesses on the NAEP.

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. This part of the study was intended to inform districts about the possibility that progress might be enhanced by better alignment. The issue was important because it addressed the concern that urban school instructional programs had become so tightly aligned with their state standards that they were undermining the ability of districts to make larger achievement gains as measured by NAEP.

Finally, the report investigates 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, prepared by the Council of the Great City Schools in collaboration with the American Institutes for Research and with funding from The Bill & Melinda Gates Foundation, presents the results from those inquiries.

Organization of Report

In addition to the executive summary and this introductory chapter, the chapters of this report are organized as follows:

Chapter 2 summarizes the demographic context of the large-city (LC) schools and TUDA districts, as well as the reported NAEP scale scores in reading and math of the 11 urban school districts that were taking part in TUDA between 2003 and 2009. It also highlights the gains of individual cities, compared with gains in large-city school and national samples.

[1] Sometimes students with disabilities (SD) are referred to as IEP students or students with Individualized Education Plans (IEPS).

Chapter 3 includes a detailed description of the methodology that was used in the study to conduct more fine-grained analyses of NAEP data from 2003 to 2007. (Not all 2009 data were available for detailed analysis in time for this study, although results are included in this report's addendum.) The chapter describes the statistical testing used in the technical analysis and contains the research team's methods for conducting statistical significance testing of results and treatment of both regular-sample results and full population estimates.

Moreover, the chapter describes the methods for conducting subscale analyses and analyzing the differences between the NAEP specifications and the respective district/state standards. The chapter also presents the methodology used for narrowing the analysis from the original 11 districts to four selected districts that were studied in greater depth, including the methodology for adjusting results for student and school background characteristics, and the results of that analysis. Finally, the chapter describes the methodology used for conducting the site visits in each of the four in-depth study districts.

Chapter 4 summarizes findings in reading, math, and science for the four selected districts in three sections, 4a (reading), 4b (math), and 4c (science). Each subject-area is divided into three sections. The first part presents the results of the detailed analysis of NAEP subscale data on the four study districts, compared with the nation and large-city schools, and examines each district's strengths and weaknesses. The second part reports the results of the alignment analysis between NAEP reading, math, or science specifications and district/state standards. This includes detailed analysis of the degree of match in content, grade level, and cognitive demand between the NAEP specifications and the state standards. In reading, additional information is presented comparing the NAEP test to each district's respective state tests in terms of item types and passage lengths. The third part of the chapter lays out the linkages between the NAEP results and the instructional context and practices in each study city.

Chapter 5 provides information on some of the shared strategies and general characteristics of each city school district that we studied in depth, background information on the context for each city's reforms, and key district instructional and other strategies that the districts used to improve student achievement. Information is presented on each district's instructional program, professional development, data and assessment systems, and accountability systems.

Chapter 6 presents a discussion of the overarching themes and patterns observed in the four districts, implications of the data and what they mean for the improvement of urban education nationwide, a series of recommendations, and a conclusion.

The report is rounded out with appendices containing more detailed information that may be of interest to the reader. Appendix A contains a brief description of how the National Assessment of Educational Progress (NAEP) is administered. Appendix B contains demographic data, city-by-city NAEP scale scores and trends, and funding and staff-level statistics on the TUDA districts. Appendix C presents specific information on the NAEP analysis methodology, including statistical adjustments, background, school and family variables, estimation formulas, and group average scale scores and standard errors. Scores are also reported in appendix C, using both reported scale scores and full-population estimates.

Appendix D presents the materials that were used to conduct the alignment analyses, including state and district standards documents by city and state, and a description of the detailed procedures and decision rules used for coding the matches and conducting the analysis. Appendix E describes the methodology used on the site visits to obtain information on the study districts' instructional programs. Also included in the appendix is the case study protocol used by the site visit teams. Appendices F through H contain case studies of the Atlanta (appendix F), Boston (appendix G), and Charlotte-Mecklenburg (appendix H) school systems and their reforms between 2003 and 2007. We did not write a separate case study on the contrasting district. Appendix I lists the individuals interviewed on each of the site visits and the materials reviewed as part of the site visits. Finally, appendix J lists the research advisory group for this project and the research teams.

CHAPTER 2

DEMOGRAPHICS AND ACHIEVEMENT IN LARGE-CITY SCHOOLS AND TUDA DISTRICTS

Introduction

This chapter begins by laying out the demographic characteristics of the large-city (LC) schools and Trial Urban District Assessment (TUDA) districts, and describing how this context differs from the nation as a whole.

The chapter then examines NAEP reading and math scale scores between 2003 and 2009 in the 11 urban school districts that were participating in TUDA in 2007 and compares the trends with those of LC and the nation. Subsequent chapters of this report present more detailed statistical analyses of the TUDA data from 2003 through 2007 only, because not all 2009 data were available for analysis when this report was being prepared.[1]

The TUDA initiative not only allows individual city school systems to participate in NAEP testing in a way that yields city-specific scale scores, but it also created a new NAEP variable—large-city (LC) schools—that permits tracking of the overall reading and math progress of public schools in the nation's major urban areas. [2, 3]

In addition, results for the 2009 NAEP science assessment allow us to compare science achievement among large-city schools to the nation.[4]

The data in this chapter examine LC and TUDA results between 2003 and 2009 in four ways:

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

[1] See the Addendum to this report for a detailed statistical analysis of NAEP scale scores on TUDA districts between 2007 and 2009.
[2] NAEP does not yield individual student scale scores, so individual student analyses are not part of this report.
[3] The LC variable includes public schools—both regular and charter—located in the urbanized areas of cities with populations of 250,000 or more. The sample is not district-specific like TUDA, but it includes schools in TUDA sites even when some of the schools in these districts are not classified as large-city schools.
[4] The NAEP science assessment was also administered in 2005, but the two tests are not comparable and therefore cannot yield any trend data. Chapter 4 contains detailed analyses of 2005 science results.

Demographics of Large-City Schools and TUDA Districts

The large-city (LC) schools that are the subject, in part, of this report are substantially different from the national public school sample that comprises the state NAEP. In general, the LC sample is composed of public schools located in the urbanized areas of cities with populations of at least 250,000 people. The national sample, on the other hand, is a random selection of students nationwide, state by state.

In 2009, the LC sample of fourth graders in reading was about 29 percent African American, 20 percent white, 42 percent Hispanic, and seven percent Asian/Pacific Islander, compared with the national public school sample that was 16 percent African American, 54 percent white, 21 percent Hispanic, and five percent Asian/Pacific Islander. The exact percentages in the sample differ somewhat between the fourth and eighth grades and between reading and math, and they may differ somewhat from actual enrollment figures in order to ensure a reliable sample. (See tables 2.1 and 2.2, and appendix B, tables B.2 and B.3.)

Table 2.1 Percentages of public school students in large-city schools and the national public sample in grades 4 and 8 on the NAEP reading assessment, by selected characteristics, 2003-2009[5]

Reading	Grade 4				Grade 8			
	2003	2005	2007	2009	2003	2005	2007	2009
African American								
National Public	17	17	17	16	17	17	17	16
Large City	35	32	31	29	36	32	31	27
White								
National Public	59	57	56	54	61	60	58	57
Large City	22	21	21	20	23	24	23	22
Hispanic								
National Public	18	19	20	21	15	17	18	20
Large City	34	38	38	42	32	36	37	41
Asian/Pacific Islander								
National Public	4	4	5	5	4	4	5	5
Large City	7	7	7	7	8	7	8	8
NSLP-eligible								
National Public	44	45	45	47	36	39	40	43
Large City	69	71	70	71	61	63	64	65
Students with Disabilities								
National Public	10	10	10	10	10	9	9	10
Large City	9	9	9	10	10	9	10	10
English Language Learners								
National Public	8	9	9	9	5	5	6	5
Large City	17	19	20	18	10	11	11	11

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.

[5] These percentages vary somewhat from the actual percents of students attending schools in large cities and in the selected TUDA districts. For reporting purposes weights are applied to both selected schools and students. The weights permit valid inferences to be drawn from the student samples about the respective populations from which they were drawn and, most importantly, ensure that the results of the assessments are fully representative of the target populations.

In addition, the fourth-grade reading sample in the large-city schools was 71 percent National School Lunch Program (NSLP)-eligible, 18 percent English language learners, and ten percent students with disabilities, compared with the national public school sample that was 47 percent NSLP-eligible, nine percent English language learners, and 10 percent students with disabilities. Again, the exact percentages differed somewhat between fourth and eighth grades and between reading and math, and may differ from actual enrollment figures.

Table 2.2 Percentages of public school students in large-city schools and the national public sample in grades 4 and 8 on the NAEP mathematics assessment, by selected characteristics, 2003-2009[6]

Mathematics	Grade 4				Grade 8			
	2003	2005	2007	2009	2003	2005	2007	2009
African American								
National Public	17	17	17	16	17	17	17	16
Large City	34	32	31	29	35	32	30	27
White								
National Public	58	57	55	54	62	60	58	56
Large City	22	21	20	20	24	24	23	21
Hispanic								
National Public	19	20	21	22	15	17	19	21
Large City	36	39	40	42	33	36	38	42
Asian/Pacific Islander								
National Public	4	4	5	5	4	5	5	5
Large City	7	6	7	7	8	8	8	8
NSLP-eligible								
National Public	44	46	46	48	36	39	41	43
Large City	69	71	71	71	60	62	65	66
Students with Disabilities								
National Public	11	12	11	12	11	11	9	10
Large City	10	11	11	11	11	10	9	11
English Language Learners								
National Public	9	10	10	10	5	6	6	6
Large City	19	20	21	20	12	12	12	12

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.

Unlike the LC variable, the 11 school districts that participated in the Trial Urban District Assessment (TUDA)[7] are individual school-system participants in NAEP that are included in their entirety in the LC sample. They are also the subject of this report. These 11 districts had a total enrollment of 2,892,269 students in 2008-09. The districts ranged in size in 2008-09 from New York City with 1,038,741 students to the District of Columbia with 44,331 students. (See appendix B, table B.1 for city-by-city enrollment and demographic data by year.) The 11 districts

[6] These percentages vary somewhat from the actual percents of students attending schools in large cities and in the selected TUDA districts. For reporting purposes weights are applied to both selected schools and students. The weights permit valid inferences to be drawn from the student samples about the respective populations from which they were drawn and, most importantly, ensure that the results of the assessments are fully representative of the target populations.

[7] The 11 urban school districts that participated in NAEP in 2007 and are the basis for this report were Atlanta, Austin, Boston, Charlotte, Chicago, Cleveland, the District of Columbia, Houston, Los Angeles, New York City, and San Diego. Seven additional city school districts participated in the TUDA in 2009and are included in the addendum to this report.

enrolled approximately 3,057,144 students in 2002-03, the baseline year for this study. The total enrollment of the 11 districts declined 5.4 percent between 2002-03 and 2008-09. In addition, the TUDA districts had fourth-grade reading samples in 2009 of NSLP-eligible students ranging from 100 percent in Cleveland (which is a Universal Meals district) to 47 percent in Charlotte-Mecklenburg. English language learners ranged from 41 percent in Los Angeles to one percent in Atlanta. The district samples also ranged from seven percent African American in Los Angeles to 80 percent in Atlanta and from five percent Hispanic in Atlanta to 77 percent Los Angeles. (See appendix B, table B.2)

NAEP Achievement in Large-City Schools and TUDA Districts

Reading[8]

Reported 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 testing in 2009 at both grades four and eight. Between 2003 and 2009, reported NAEP scale scores in reading rose in LC from an average or mean of 204 to 210 among fourth graders and increased from 249 to 252 among eighth graders. During the same period, reported NAEP scale scores in reading nationwide (a measure that includes students in large-city schools) moved from 216 to 220 among fourth graders and from 261 to 262 among eighth graders. (See table 2.3, appendix B, table B.4.)

Table 2.3 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 City	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.
Note: Changes in scale scores and tests of significance are based on differences between unrounded scores.
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.[9] (See table 2.3) 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 also narrowed from 12 points to 10 points over the same period.

[8] 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/ltt/bridge_study.asp for details.

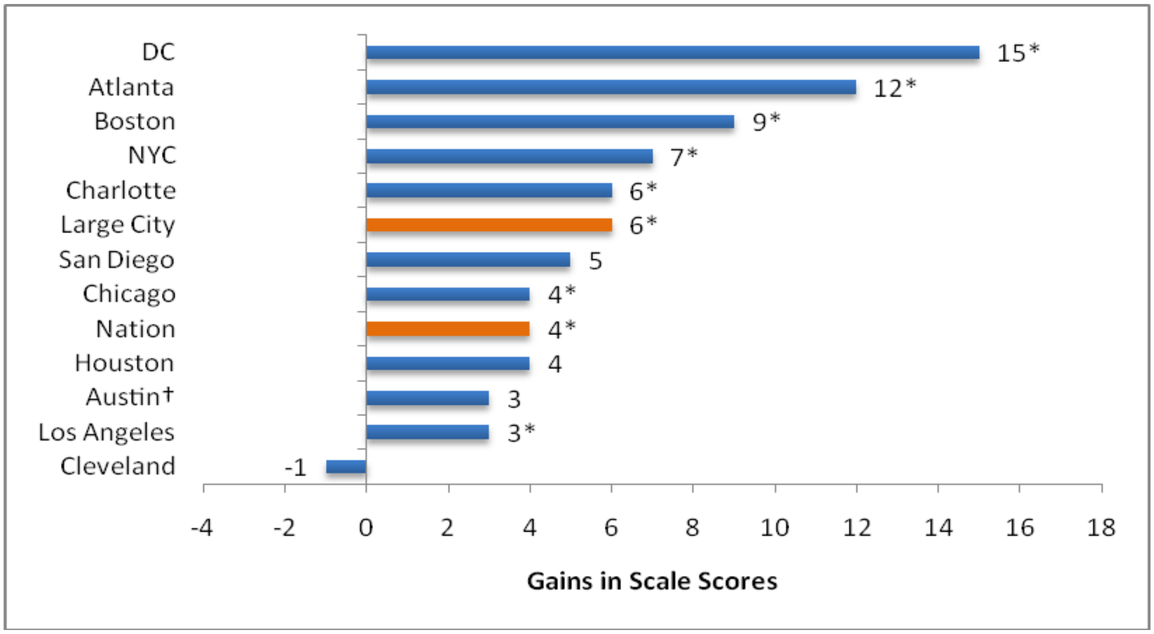
[9] Difference between rates of gain between 2003 and 2009 in fourth grade equals three scale score points, p<.05. Difference between rates of gain between 2003 and 2009 in eighth grade equals three scale score points, p<.05. All comparisons were independent tests for multiple pair-wise comparisons according to the False Discovery Rate procedure. (Tests of significance were conducted on unrounded scale scores.)

Moreover, the percentage of large-city fourth graders reading at or above basic levels of attainment 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.[10] (See appendix B, table B.8.)

The percentage of fourth graders nationwide reading at or above basic levels of attainment 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. (See appendix B, table B.8.)

In addition, the reported NAEP reading scale scores on individual TUDA cities showed significant gains in many cities. Significant reading gains among fourth graders between 2003 and 2009 were seen in Atlanta, Boston, Charlotte, Chicago, the District of Columbia (DC), Los Angeles, and New York City (NYC). (See figure 2.1.) Significant reading gains among eighth graders between 2003 and 2009 were seen in Atlanta, Boston, Houston, and Los Angeles. (See figure 2.2.)

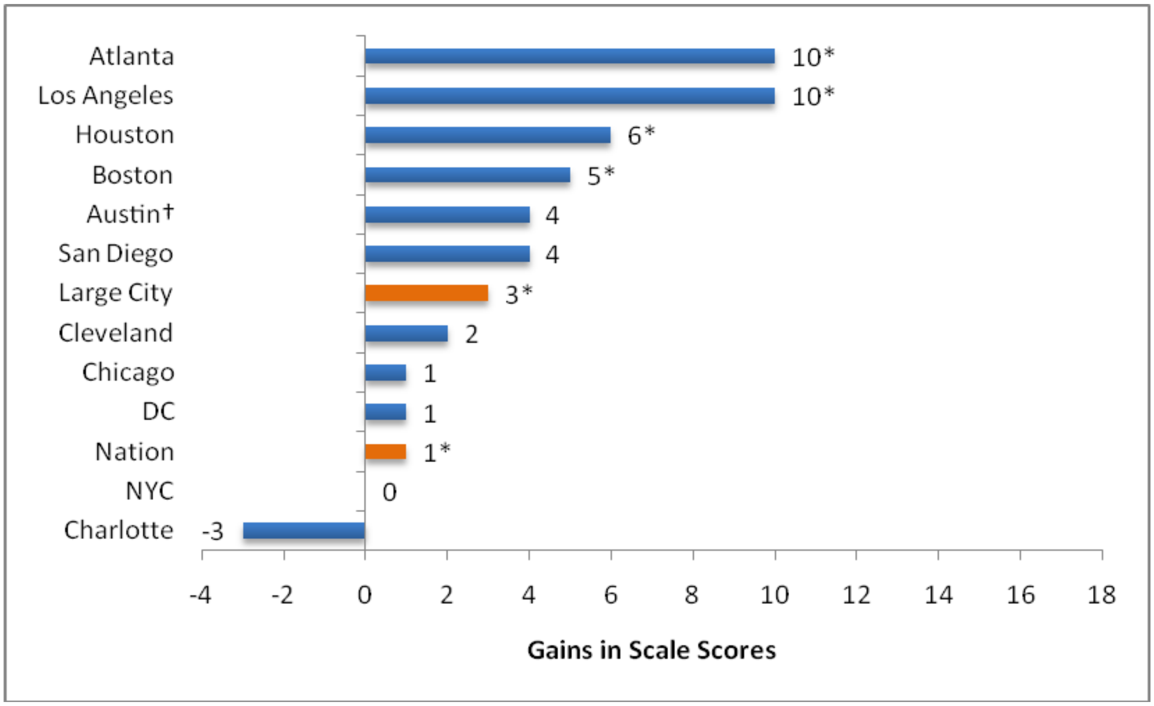
Figure 2.1 NAEP 4th-grade reading scale score increases in TUDA cities between 2003 and 2009, compared with large-city and national samples



† 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.

[10] 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.

Figure 2.2 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.

Overall, more TUDA districts saw increased reading scale scores among fourth graders than among eighth graders.[11] Looking at scale scores for 2009, Austin, Boston, and Charlotte—despite its eighth grade declines from 2003—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. (See appendix B, table B.4.)

Mathematics

Public schools in large cities also showed statistically significant gains between 2003 and 2009 in mathematics at both grades four and eight. Over that period, the reported NAEP scale scores of the large cities in mathematics increased from 224 to 231 among fourth graders and from 262 to 271 among eighth graders. (See table 2.4.)

During the same period, reported NAEP scale scores in math nationwide (which includes students in large-city schools) increased from 234 to 239 among fourth graders and from 276 to 282 among eighth graders. Both sets of gains were statistically significant.

[11] All references to gains or increases in NAEP scale scores are statistically significant at the p <.05 level.

Table 2.4 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 City	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.
Note: Changes in scale scores and tests of significance are based on differences between unrounded scores.
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.[12] (See table 2.4.)

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.[13]

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 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.[14] (See appendix B, table B.9.)

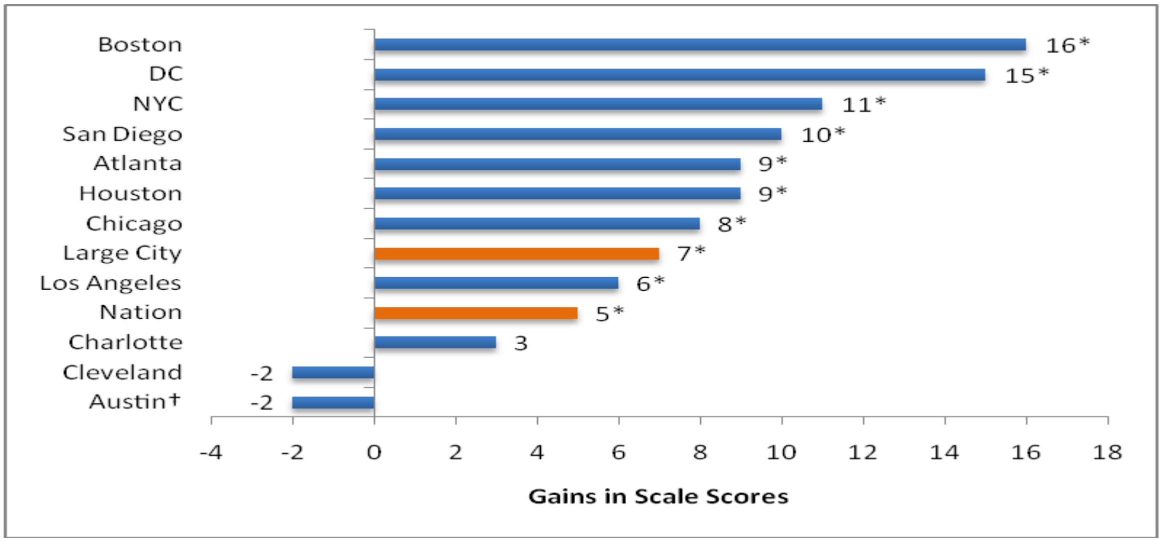
The percentage of fourth graders nationwide scoring at or above basic levels of attainment in math increased from 76 percent in 2003 to 81 percent in 2009, and those at or above proficient levels increased from 31 percent to 38 percent. The percentage of eighth graders scoring at or above basic levels 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. (See appendix B, table B.9.)

In addition, the reported NAEP math data on individual TUDA cities showed significant gains in many cities. Significant math gains among fourth graders between 2003 and 2009 were seen in Atlanta, Boston, Chicago, the District of Columbia (DC), Houston, Los Angeles, New York City (NYC), and San Diego. (See figure 2.3.) Significant math gains among eighth graders between 2003 and 2009 were seen in every TUDA city except Cleveland. (See figure 2.4.)

---

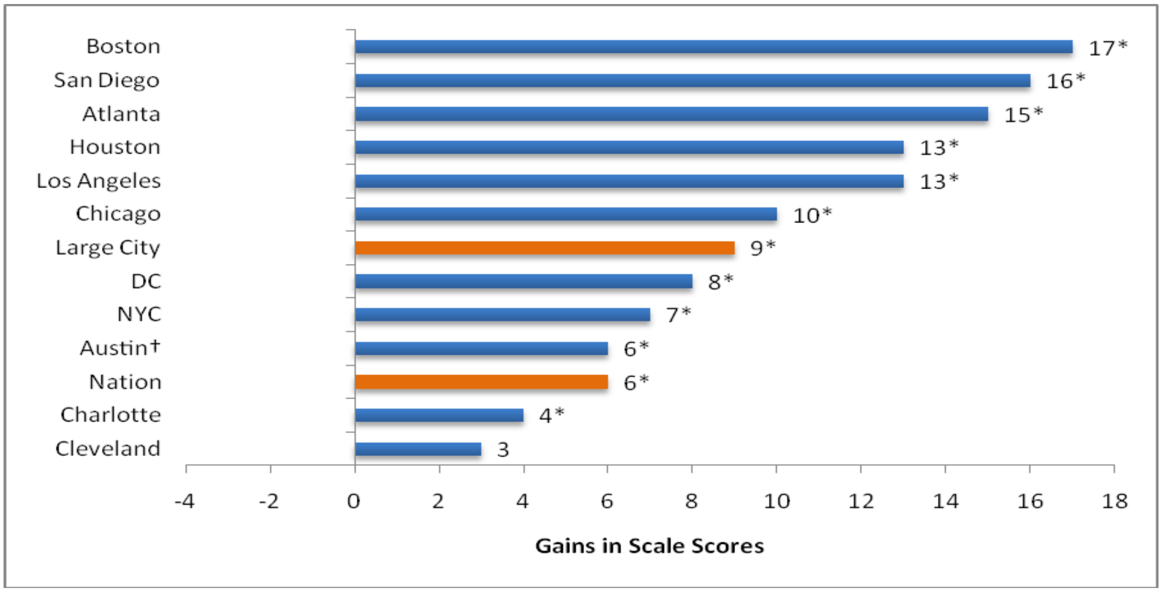
[12] Difference between rates of gain between 2003 and 2009 in fourth grade equals two scale score points, p<.05. Difference between rates of gain between 2003 and 2009 in eighth grade equals three scale score points, p<.05. (Tests of significance were conducted on unrounded scale scores.)
[13] Differences between numbers in the text and numbers in the accompanying tables are due to rounding.
[14] 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.

Figure 2.3 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.

Figure 2.4 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.

In contrast to reading, more TUDA districts increased reported math scale scores among eighth graders than among fourth graders. Austin, Boston, Charlotte, Houston, New York City, and San Diego outperformed their large-city peers in math 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. (See appendix B, table B.5.)

Student Groups

In addition to these overall trends, NAEP data show that over the study period, large-city districts generally improved the reading and math scores of key student groups. (See tables 2.5 and 2.6.)

Table 2.5. 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 City	193	196	199	201**	8***	241	240	240	243**	2***
White										
National Public	227	228	230	229*	2***	270	269	270	271	1***
Large City	226	228	231	233**	7***	268	270	271	272	4***
Hispanic										
National Public	199	201	204	204*	5***	244	245	246	248*	4***
Large City	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 City	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 City	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.
Note: Changes in scale scores and tests of significance are based on differences between unrounded scores.
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 2.6. 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				
	2003	2005	2007	2009	Δ	2003	2005	2007	2009	Δ
Overall										
National Public	234	237	239	239*	5***	276	278	280	282*	6***
Large City	224	228	230	231**	7***	262	265	269	271**	9***
African American										
National Public	216	220	222	222*	6***	252	254	259	260*	8***
Large City	212	217	219	219**	7***	247	250	254	256**	9***
White										
National Public	243	246	248	248*	5***	287	288	290	292	5***

Large City	243	247	249	250**	7***	285	288	292	294	9***
Hispanic										
National Public	221	225	227	227	6***	258	261	264	266	8***
Large City	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 City	246	247	251	253	7	281	289	291	299	18***
NSLP-eligible										
National Public	222	225	227	228*	6***	258	261	265	266*	8***
Large City	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.
Note: Changes in scale scores and tests of significance are based on differences between unrounded scores.
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 math at both fourth- and eighth-grade levels. 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 significant in either reading or mathematics.

Table 2.7 TUDA districts showing statistically significant reading gains or losses on NAEP by student group between 2003 and 2009

Reading	Black		Hispanic		Asian		White		NSLP		LEP		SPED	
City/Grade	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.

In fact, between 2003 and 2009, a majority of TUDA districts improved the scale scores of many of their student groups. (See tables 2.7 and 2.8 and appendix B, tables B.4 and B.5 for detailed city-by-city data by student group and subject area.)

Table 2.8 TUDA districts showing significant mathematics gains or losses on NAEP by student group between 2003 and 2009

Math	Black		Hispanic		Asian		White		NSLP		LEP		SPED	
City/Grade	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.

In addition, African American fourth-grade students in Austin, Boston, Charlotte, Houston, and New York City had significantly higher average reading scale scores than African American students in public schools throughout the country in 2009, and in Charlotte, this trend continued at the eighth-grade level. Moreover, Hispanic fourth-grade students in Boston and Charlotte outperformed their peers nationally in reading. (See appendix B, tables B.4 and B.5.)

In math, African American students in Boston, Charlotte, Houston, and New York City had significantly higher average math scale scores than their African American peers nationally at the fourth grade level, as did African American eighth graders in Austin, Boston, Charlotte, and Houston. Moreover, Hispanic fourth-grade students in Boston outperformed their peers nationally in math, while Hispanic students in Austin, Charlotte, and Houston outperformed their peers nationally in math at both the fourth- and eighth-grade levels in 2009.

In addition, poor students in Austin, Boston, Houston, and New York City had higher average math scale scores in 2009 at both the fourth- and eighth-grade levels than poor students nationwide. Finally, LEP students in a number of cities scored higher in reading and math than did their language peers nationwide.

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.

Science

Results on NAEP science assessments are also available for 2005 and 2009, although the two tests are not comparable. Data on the reported results from 2009 are included in this chapter, while the more detailed analysis in later chapters of this report cover the 2005 testing.

At both the fourth and eighth grade levels, large-city students scored lower than their national peers on the 2009 NAEP science assessment. The average scale scores for large cities were 135 among fourth graders and 134 among eighth graders, compared to an average scale score of 149 for both fourth and eighth graders in the national sample—a gap of 14 to 15 points, respectively.

However, looking at specific student groups, the gap was somewhat smaller for African American students in large cities—who scored lower than their national peers by five points at both the fourth and eighth grade levels--and large-city Hispanic students, who scored only three points lower than Hispanic students in the national sample at the fourth grade level, and four points lower than Hispanic students nationwide at the eighth grade level. There was no statistically significant difference between the scale scores of white students in large cities and nationwide at either the fourth- or eighth-grade level. (See table 2.9.)

Table 2.9 Average NAEP science scale scores of public school students nationwide and large-city public school students in grades 4 and 8, 2009

Science	Grade 4	Grade 8
	2009	2009
Overall		
National Public	149*	149*
Large City	135**	134**
African American		
National Public	127*	125*
Large City	122**	120**
White		
National Public	162	161
Large City	163	159
Hispanic		
National Public	130*	131*
Large City	127**	127**
Asian/Pacific Islander		
National Public	160*	159*
Large City	152**	152**
NSLP-eligible		
National Public	134*	133*
Large City	126**	125**

* Statistically different from large cities; ** Statistically different from national public.
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 the same year, 71 percent of fourth graders nationwide (which includes students in large-city schools) scored at the basic level or above in science and 62 percent of eighth graders scored at this level. The percentage of fourth graders nationwide scoring at or above proficient levels was 32 percent, while the percentage of eighth graders scoring at or above proficient levels was 29 percent in 2009. (See appendix B, table B.10.) The percentage of large-city fourth graders scoring at or above basic levels of attainment was 56 percent in 2009, and those at or above proficient levels was 20 percent the same year. The percentage of large-city eighth graders scoring at or above basic levels was 44 percent in 2009, while those at or above proficient levels was 17 percent that year.[15] (See appendix B, table B.10.) Finally, in 2009, science scale scores among fourth graders in Austin, Charlotte, and Jefferson County were not significantly different from their same-grade peers nationwide; while science scale scores among eighth graders in Austin were not significantly different from eighth graders nationwide.

Summary

In summary, fourth- and eighth-grade students attending public schools in large cities generally made statistically significant gains in reported reading and mathematics scale scores on NAEP between 2003 and 2009. Although trend data are not available for the NAEP science assessment, the data show that at both fourth- and eighth-grade levels, large-city students scored lower than their peers on the 2009 NAEP science assessment. However, the gaps in science achievement were somewhat smaller for African American and Hispanic students in large-city schools compared with their public school peers nationwide.

The data also show that the overall reading and mathematics gains among the large cities in both fourth and eighth grades were significantly larger than the gains seen nationwide between 2003 and 2009 in both subjects and grades. Large-city schools and the 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.

In addition, the NAEP data indicate that a number of cities are making significant reading and math gains among African American, Hispanic, poor, and LEP students.

The next chapter describes the methodology used to analyze the NAEP scale scores from 2003 to 2007 and select which districts to study in greater depth in order to determine what might be behind the gains in some urban schools and districts.

[15] Source: 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 2011-452), 2011.

CHAPTER 3

METHODOLOGY AND ANALYSIS OF TUDA DATA

Introduction

The primary goal of this research was to better understand the factors behind urban school achievement on NAEP. To achieve that goal, our analysis was designed to identify and select three types of districts for further case study:

1. Districts that made significant and consistent gains on the NAEP over the three administrations of the reading and mathematics assessments from 2003 to 2007.

2. Districts that consistently failed to make gains or that posted losses on the NAEP over the three administrations of the reading and mathematics assessments from 2003 to 2007.

3. Districts that outperformed other TUDA districts on the most recent administration of NAEP, controlling for relevant student background characteristics.

The first part of this chapter (1) details the statistical analysis methods used to examine reading and math achievement on the NAEP and (2) describes the process used to narrow the full TUDA sample down to a smaller set of districts for in-depth study.

The second part of this chapter (1) reports the process used to determine the degree of alignment between the NAEP and various state assessment programs in the selected districts, (2) summarizes the process of examining subscale data to determine district strengths and weaknesses, and (3) describes the site visit procedures and protocols used to examine the instructional programs that might have contributed to the NAEP results in these districts.

Part 1. Statistical Analysis of NAEP Data

This section describes the methodology used to analyze NAEP data on the TUDA districts, including analysis of the *reporting sample* and *full population estimates* (FPE), and the estimation of average or mean and gain scores across quintiles.[1]

Districts Included in the Analyses

Our analysis of student achievement focused on urban school districts participating in NAEP's Trial Urban District Assessment between 2003 and 2007. Detailed data on 2009 were not yet available in time for this analysis, so it is not included in this chapter but can be found in the addendum. (Reported data through 2009 are included in chapter 2.) Our goal in the 2003 to 2007 analysis was to compare results across the 11 TUDA districts in order to determine which ones showed significant and consistent gains and to select sites for in-depth case studies. To maximize the number of "time points" and districts for comparison, we chose 2003 as the starting point for all analyses. This decision allowed us to examine trends over three administrations of NAEP

[1] The *reporting sample* is the sample of NAEP students who actually participate in NAEP. It is the sample of students and data used in the computation of results reported in the major NAEP publications. *Full population estimates* (FPE) are the results for the full population by combining the actual performance of students who were assessed with the imputed performance of sampled students who did not participate in NAEP.

(2003, 2005, and 2007) in all TUDA districts, except Austin, which joined the project in 2005 and had only two data points.

Because our ultimate goal was to select districts with either significant and consistent improvement or consistent lack of improvement, we focused on NAEP reading and math scale scores. NAEP science testing has been conducted only once since 2003 (in 2005), so no trend data in science are available. Table 3.1 shows all 11 TUDA districts that participated in NAEP between 2002, the first year of the TUDA project, and 2007 by subject area tested.

Table 3.1 NAEP administrations and TUDA participation, by district, 2002-2007

	2002			2003			2005			2007		
Districts	R	M	S	R	M	S	R	M	S	R	M	S
Atlanta	√			√	√		√	√	√	√	√	
Austin							√	√	√	√	√	
Boston				√	√		√	√	√	√	√	
Charlotte				√	√		√	√	√	√	√	
Chicago	√			√	√		√	√	√	√	√	
Cleveland				√	√		√	√	√	√	√	
District of Columbia	√			√	√		√	√	√	√	√	
Houston	√			√	√		√	√	√	√	√	
Los Angeles	√			√	√		√	√	√	√	√	
New York City	√			√	√		√	√	√	√	√	
San Diego				√	√		√	√	√	√	√	

Note: R = Reading, M = Mathematics, S = Science
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, and 2007 Assessments.

Reporting Sample and Full Population Estimates

Not all students in jurisdictions sampled by NAEP actually participate in NAEP. Some students with disabilities and English language learners are excluded from the assessment in accord with state policies if their teachers or local administrators believe they are unable to meaningfully participate in NAEP. The ***reporting sample*** is the sample of students who actually participate in NAEP. ***Full population estimates*** (FPE) incorporate imputed scale scores for the sampled students who did not participate in NAEP. We analyzed NAEP data using both methodologies.

Although excluded students amount to a small fraction of the student population (i.e., just three percent in 2007), exclusion rates vary substantially between jurisdictions and can have significant effects on achievement trends in some districts. Because excluded students may have lower scale scores than the general population, statistically significant gains are sometimes reported in jurisdictions where the gain may be the result of increased exclusions. Conversely, significant gains are sometimes missed in jurisdictions where exclusions were reduced.

FPEs minimize the effects of differing participation and exclusion rates among students with disabilities or English language learners across jurisdictions. The analysis team relied on FPEs when comparing the 11 districts and selecting sites for case studies, while double-checking the results against the reporting sample.

Additional information on the FPE methodology is provided in appendix C.

Where possible, our analyses were performed using both the reporting sample and the FPEs, but there were two notable exceptions:

1. **Analyses controlling for student background characteristics.** Some covariates for these analyses are derived from the NAEP student background questionnaire. Data on background characteristics for students not participating in NAEP are not available. Therefore, the calculation of scale scores controlling for background characteristics was conducted on the reporting sample only, not the FPE.

2. **Subscale analyses.** Imputed composite scale scores are estimated for students who do not participate in NAEP, but FPEs for subscale scores are not available. Therefore, subscale analyses were performed using only the reporting sample, not the FPE.

Quintile Scores

We also examined trends over time by looking at various points in the distribution of achievement scale scores and calculating average or mean scores and gains at each quintile. Change in achievement is often measured by comparing the overall differences in average scale scores between two periods. Instead of aggregating all students into one average scale score, however, we disaggregated the data into equally weighted quintiles, which yielded five separate achievement groups. We then calculated average scale scores and gains for each quintile using the following procedures.

1. For each group in each time period, we ranked student scale scores from lowest to highest.

2. For each group, we partitioned students into five equally weighted quintiles with the lowest-scoring students in the lowest quintile, the second lowest-scoring student group in the next quintile, and so on.

3. We computed the average scale score for students in each weighted quintile for each time period—2003, 2005, and 2007.

Process for District Selection

This section describes the methodology used to narrow the number of TUDA districts for more in-depth study.

Selection of Districts Based on Gains and Losses Across Years

Defining what it meant for a district to significantly and consistently make gains on NAEP was of critical importance to the study. Accordingly, we measured gains using changes in scale scores. We documented gains in reading and mathematics at grades four and eight using changes both in scale-score averages for the overall district sample and at each quintile for the time periods 2003 to 2005, 2005 to 2007, and 2003 to 2007. This method yielded 792 statistics.[2]

To make the district site selections for the subsequent case studies, we implemented the following steps:

[2] Six statistics for each of the 11 districts in the three time periods in each subject-grade combination, except for one district (Austin), which has data on only time period 2005 to 2007.

1. Tested for statistically significant gains or losses in each time period: 2003 to 2005, 2005 to 2007, and 2003 to 2007.

2. Marked the number of times a district made statistically significant gains across the three time periods.

3. Rank-ordered the districts according to the number of marks they received.

4. Determined the districts with the highest number of marks.

5. Repeated these steps for each analysis, i.e., fourth-grade mathematics, fourth-grade reading, eighth-grade mathematics, and eighth-grade reading.

6. Compiled the results from the five steps above.

Using this process, we identified the districts with the most frequent gains on NAEP across the different analyses as those with the most consistent gains for the purposes of site selection. Steps 2 to 5 above were also used for marking districts with consistent losses. Tables 3.2 through 3.5 present the detailed changes in scale scores for the 11 participating districts across all testing periods, the large-city (LC) schools, and the national public sample.

Then, we looked at districts that consistently made gains in math and those that consistently made gains in reading. Because of the large amount of data and the different ways we could define and measure the idea of *consistent gains*, we aggregated the data and analyzed the results a number of ways to inform the site selection process. This was necessary because a number of districts were similar to one another, and most districts showed at least some gains. In the sections below, we summarize this analysis of the gains among the districts.

Identification of districts consistently making gains on NAEP

Table 3.2 summarizes the number of significant gains in average scale scores across all comparison periods (2003 to 2005, 2005 to 2007, and 2003 to 2007) by subject and grade using full population estimates. For every time period in which a district displayed a significant gain or loss at the overall average, the district received one mark. That is, we counted the number of times, out of a possible three, in which each district made a significant gain or loss. Thus, with the exception of Austin, which did not participate in the 2003 assessments, the maximum number of marks for any district is three. The *top four* districts are highlighted as those making consistent gains overall.

Table 3.2 shows that Atlanta made the most statistically significant gains in reading at the fourth-grade level, showing gains across two time intervals. Further, Atlanta made statistically significant gains on overall average scores across all three time intervals in fourth-grade mathematics, as did the District of Columbia and New York City. In eighth-grade reading, Houston and Los Angeles made gains across two intervals. In eighth-grade mathematics, Boston, Houston, and Los Angeles all made statistically significant gains across all three time intervals.

The last column of the table shows the number of times each district was among the top four districts making significant gains in each subject and grade. Atlanta, Houston, and Los Angeles each appeared three times among the top four districts. Similar counts were produced for changes at each quintile for each grade level and subject. For every time period in which a district displayed a significant gain or loss at each quintile, the district received one mark. With the exception of Austin, which did not participate in the 2003 assessments, the maximum number of marks for any district is 15, one for each quintile in each time period.

Table 3.2 Number of statistically significant gains based on the full population estimates of average scale scores in reading and mathematics in grades 4 and 8, and the number of times a district is among the top four with significant gains, by district

District	Reading 4	Reading 8	Mathematics 4	Mathematics 8	Number of times in top four
Atlanta	2	0	3	2	3
Houston	0	2	2	3	3
Los Angeles	0	2	2	3	3
Boston	0	0	2	3	2
District of Columbia	1	0	3	1	2
San Diego	0	0	2	2	2
Chicago	0	0	1	2	1
New York City	0	0	3	0	1
Austin	0	0	0	1	0
Charlotte	0	0	0	1	0
Cleveland	0	0	1	0	0

Note: (1) Top four districts are highlighted in yellow. In cases of ties, districts with the same number of points as the fourth-ranked district are also highlighted. In cases where there are not four districts with one or more points, only those with points are highlighted. (2) Districts not in the top four in any category are listed alphabetically. (3) Austin did not participate in 2003.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Reading and Mathematics Assessments: Full Population Estimates.

Table 3.3 Number of statistically significant gains at each quintile based on the full population estimates of average scale scores in reading and mathematics in grades 4 and 8, and the number of times a district is among the top four with significant gains, by district

District	Reading 4	Reading 8	Mathematics 4	Mathematics 8	Number of times in top four
Atlanta	6	1	11	10	4
Los Angeles	2	6	7	13	3
Boston	0	0	12	13	2
District of Columbia	7	0	12	1	2
Houston	0	6	8	11	2
New York City	1	0	13	1	2
Austin	0	0	0	1	0
Charlotte	0	0	1	3	0
Chicago	0	0	6	3	0
Cleveland	0	0	3	2	0
San Diego	0	0	9	8	0

Note: (1) Top four districts are highlighted in yellow. In cases of ties, districts with the same number of points as the fourth-ranked district are also highlighted. In cases where there are not four districts with one or more points, only those with points are highlighted. (2) Districts not in the top four in any category are listed alphabetically. (3) Austin did not participate in 2003.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Reading and Mathematics Assessments: Full Population Estimates.

The results are shown in table 3.3 where, again, the *top four* districts are highlighted.

We saw from table 3.2 that a number of districts were repeatedly in the top four in terms of overall gains. But the results of table 3.3, which examines gains across quintiles, narrow the potential list of study districts to Atlanta, Boston, the District of Columbia, and Los Angeles.[3] Houston showed gains in many areas, but it also showed a period of significant loss in reading in fourth grade and was therefore dropped from consideration as a district that consistently made gains. We selected Atlanta for deeper study because it showed the greatest relative strength in reading and consistently showed gains on mean scores across the quintiles in all but one column.

In math, while a number of districts showed some consistency in gains, the project team used the same methodology as described earlier to select Boston for the relative strength and consistency of gains across the most quintiles.

Identification of districts not making consistent gains or posting losses on NAEP

As noted, the analyses of NAEP trends were designed to identify two types of districts for further study and comparison: districts that *consistently made gains* and *districts that consistently did not make gains* or even posted losses. Following the same procedure outlined above to identify districts making gains, losses on NAEP were examined by looking at both overall district averages and the average of each quintile of the scale score distribution.

Tables 3.4 and 3.5 show the number of significant losses in average scale scores overall and at each quintile across all comparison periods by subject and grade.

Although our goal was to identify two districts for the gain category and two for the loss category, only Cleveland among all the TUDA participants consistently posted losses and/or consistently failed to make gains across grade levels, subjects, and years through 2007.

Selection of Districts Based on Performance after Adjusting for Student Characteristics

In addition to the districts selected on the basis of gains/losses, we analyzed data to select the district(s) that outperformed others on the basis of their overall 2007 results in reading and mathematics after student background characteristics were taken into account. Regression analyses were conducted for this purpose.

For these analyses, we needed to determine what background variables to include as covariates. Because no NAEP document describes which background variables are the most reliable, valid, or predictive of NAEP scale scores, we conducted a literature search on the use of NAEP background variables and concluded that key variables included student race/ethnicity, parents' education, NSLP eligibility, and reading materials in the home. We also conducted a literature review to ensure consistency with previous NAEP analyses that had used background characteristics as controls. One recent report by the National Center for Education Statistics, Braun, Jenkins, and Grigg (2006a), examined differences in average NAEP reading and mathematics scale scores between public and private schools after selected characteristics of students and/or schools were taken into account. As expected, student characteristics included gender, race/ethnicity, disability status, and identification as an English language learner. Another NCES report, Braun, Jenkins, and Grigg (2006b), compared charter schools with public schools using the same approach.

[3] Detailed quintile data using full population estimates can be found in appendix B, tables B.12-B.39.

Table 3.4 Number of statistically significant losses based on the full population estimates of average scale scores in reading and mathematics in grades 4 and 8, and the number of times a district is among the top four with significant losses, by district

District	Reading 4	Reading 8	Mathematics 4	Mathematics 8	Number of times in top four
Cleveland	0	0	1	0	1
Houston	1	0	0	0	1
Atlanta	0	0	0	0	0
Austin	0	0	0	0	0
Boston	0	0	0	0	0
Charlotte	0	0	0	0	0
Chicago	0	0	0	0	0
District of Columbia	0	0	0	0	0
Los Angeles	0	0	0	0	0
New York City	0	0	0	0	0
San Diego	0	0	0	0	0

Note: (1) Top four districts are highlighted in yellow. In cases where there are not four districts with one or more points, only those with points are highlighted. (2) Districts not in the top four in any category are listed alphabetically. (3) Austin did not participate in 2003.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Reading and Mathematics Assessments: Full Population Estimates.

Table 3.5 Number of statistically significant losses at each quintile based on the full population estimates of average scale scores in reading and mathematics in grades 4 and 8, and the number of times a district is among the top four with significant losses, by district

District	Reading 4	Reading 8	Mathematics 4	Mathematics 8	Number of times in top four
Cleveland	0	0	5	1	2
San Diego	0	1	2	0	2
Charlotte	0	1	0	0	1
Chicago	0	1	0	0	1
Houston	1	0	0	0	1
New York City	1	0	0	0	1
Atlanta	0	0	0	0	0
Austin	0	0	0	0	0
Boston	0	0	0	0	0
District of Columbia	0	0	0	0	0
Los Angeles	0	0	0	0	0

Note: (1) Top four districts are highlighted in yellow. In cases of ties, districts with the same number of points as the fourth-ranked district are also highlighted. In cases where there are not four districts with one or more points, only those with points are highlighted. (2) Districts not in the top four in any category are listed alphabetically. (3) Austin did not participate in the 2003 assessment.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Reading and Mathematics Assessments: Full Population Estimates.

Based on this literature review, our study used the following control variables: race/ethnicity, special education status, English language learner (ELL) status, and eligibility under the National School Lunch Program (NSLP). The analysis did not include a specific examination of gender. The analysis also accounted for the highest level of education attained by either parent and information on the availability of literacy materials and computers in students' homes. The last three indicators were based on student responses to the NAEP student background questionnaire.[4]

The regression analyses controlling for these student-background variables estimated the relative performance a district would have if the average demographics in that district were the same as the average demographics across the districts. For these analyses, we used 2007 scale scores in both reading and math at grades four and eight.

The results indicated that, depending on the grade level and subject, three to six districts outperformed other TUDA districts in 2007 when accounting for student background characteristics.

- In fourth-grade reading, Austin, Boston, and Charlotte outperformed the other TUDA districts (see table 3.6).
- In eighth-grade reading, there was little difference among the TUDA districts. Six districts (Austin, Boston, Charlotte, Chicago, Cleveland, and Houston) scored higher than the remaining five. These six were not significantly different from one another (see table 3.7).
- In fourth-grade math, the districts that scored higher than the rest of the TUDA districts (and were not significantly different from one another) were Austin, Boston, Charlotte, and Houston (see table 3.8).
- In eighth-grade math, three districts (Austin, Boston, and Charlotte) scored higher than the other TUDA districts (see table 3.9).

The results across grade levels and subjects indicated that three districts consistently appeared among the top-scoring districts after accounting for student background characteristics: Charlotte, Boston, and Austin. Boston, however, had already been selected on the basis of consistent gains over time, particularly in math.

The project team initially selected both Charlotte and Austin on the basis of the adjusted results. To ensure that the findings were consistent with the results for key populations of students, we conducted an additional analysis of the unadjusted student groups. In these analyses, we computed the average NAEP scale scores of students who were African American or Hispanic, students classified as low income (i.e., eligible for the National School Lunch Program), students with disabilities, and English language learners.[5] We then compared the performance of each student group across TUDA districts. These analyses of student groups in 2007 favored the selection of Charlotte. With only one exception, Charlotte was either the highest-scoring district or among the highest-scoring districts across all student populations, grades, and subjects. The only exception was among Hispanic students in eighth-grade math, where Charlotte ranked sixth among TUDA districts but still was not significantly different from the higher-ranking districts.

---

[4] See appendix C.2 for information about how the variables we used in the regression analyses were operationally defined.

[5] See appendix B, tables B.12 through B.39 for data on changes between 2003 to 2007 across student groups by city and quintile using the full population estimates.

We found weaker support for the selection of Austin. Austin was frequently among the highest-scoring districts for each student group by grade and subject, particularly in reading, but other results were more mixed. Austin was generally in the middle of the rankings of the 11 TUDA districts when we looked specifically at student groups. Based on this combination of findings, the project team selected Charlotte as the fourth district for further study.

Finally, we examined the same regression data and asked the question about district performance in 2007 after adjusting for student background variables in a slightly different way, although we did not use the results specifically to select which districts to study. Instead, we asked which districts—given the previous analysis—were performing higher or lower than what might be expected statistically based on the student background characteristics previously described. Positive effects would indicate the district was performing higher among the 11 TUDA participants than expected statistically; negative effects would indicate that the district was performing lower than expected. 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 (see table 3.10).[6]

- In grade four reading, the results indicated that the district effects in 2007 were positive and significant in Austin, Boston, Charlotte, New York City, and San Diego, 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 Houston.

- In grade eight reading, the results indicated that the district effects were positive and significant in Austin, Boston, Charlotte, Chicago, 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, Cleveland, New York City, and San Diego.

- In grade four math, the results indicated that the district effects were positive and significant in Austin, Boston, Charlotte, Houston, and New York City, and were negative and significant in Atlanta, Chicago, Cleveland, the District of Columbia, and Los Angeles. Results were the same as predicted in San Diego.

- In grade eight math, the results were positive and significant in Austin, Boston, Charlotte, Houston, and New York City, and were negative and significant in the District of Columbia and Los Angeles. Results were the same as predicted in Atlanta, Chicago, Cleveland, and San Diego.

[6] 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 3.6 Average NAEP scores in grade 4 reading, adjusted for student background characteristics, by district, 2007

District	Adjusted Mean	Standard Error	Boston	Charlotte	Austin	New York City	San Diego	Houston	Atlanta	Chicago	Los Angeles	Cleveland	District of Columbia
Boston	215	1.4	215										
Charlotte	212	1.2		212									
Austin	210	1.4			210								
New York City	210	0.9				210							
San Diego	208	1.1					208						
Houston	208	1.0						208					
Atlanta	206	1.2							206				
Chicago	202	1.3								202			
Los Angeles	201	1.2									201		
Cleveland	197	1.4										197	
District of Columbia	197	0.8											197

Note 1: 1. The green and gray colors in the boxes of this graph indicate whether or not a lower average or mean score of one district is statistically different from the higher average score of another district.

Example: The green box in the column labeled Atlanta in the row labeled Charlotte means that Atlanta's average score of 206 is statistically different from Charlotte's average score of 212.

Example: The gray box in the column labeled Atlanta in the row labeled San Diego means that Atlanta's average score of 206 is **not** statistically different from San Diego's average score of 208.

Blocks of gray represent districts whose average adjusted scores are similar to one another and distinctly different from all others. In grade 4 reading, the top four districts make up an imperfect gray cluster because New York City's standard error is small and hence the difference between New York City and Boston is statistically significant even though its score is very similar to Austin's, which is not statistically different from Boston's. Points where the green hits the diagonal represent a statistically significant drop in scores.

Yellow indicates a set of districts whose average adjusted scores are not statistically different from the top-scoring district.

2. Control variables for this analysis include race/ethnicity, special education status, ELL status, NSLP-eligibility, and a composite literacy scale that includes the presence at home of newspapers, magazines, a computer, and more than 25 books.

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

Table 3.7 Average NAEP scores in grade 8 reading, adjusted for student background characteristics, by district, 2007

District	Adjusted Mean	Standard Error	Boston	Chicago	Houston	Austin	Charlotte	Cleveland	Atlanta	New York City	San Diego	Los Angeles	District of Columbia
Boston	253	1.6	253										
Chicago	252	1.2		252									
Houston	252	1.1			252								
Austin	252	1.9				252							
Charlotte	251	1.2					251						
Cleveland	250	1.5						250					
Atlanta	248	1.3							248				
New York City	246	1.3								246			
San Diego	245	1.1									245		
Los Angeles	245	0.8										245	
District of Columbia	243	0.7											243

Notes: 1. The green and gray colors in the boxes of this graph indicate whether or not a lower average or mean score of one district is statistically different from the higher average score of another district.

Example: The green box in the column labeled San Diego in the row labeled Cleveland means that San Diego's average score of 245 is statistically different from Cleveland's score of 250.

Example: The gray box in the column labeled Atlanta in the row labeled Austin means that Atlanta's average score of 248 is **not** statistically different from Austin's average score of 252. Blocks of gray represent districts whose average adjusted scores are similar to each other and distinctly different from all others. Points where the green hits the diagonal represent a statistically significant drop in scores. In grade 8 reading, the gray above the diagonal indicates that there is not a large spread in scores. Therefore, we don't see any distinct blocks of similar scores, but rather a *stepwise* band of similar scores. As you move down, from the *red* square to the *blue* square for example, one district drops out of the cluster of similar scores, but another moves in.

Yellow indicates a set of districts whose average adjusted scores are not statistically different from the top-scoring district.

2. Control variables for this analysis include race/ethnicity, special education status, English language learner status, NSLP eligibility, and a composite literacy scale that includes the presence at home of newspapers, magazines, a computer, and more than 25 books. In grade 8, an indicator of whether one parent has a college degree was included in the analysis.

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

Table 3.8 Average NAEP scores in grade 4 mathematics, adjusted for student background characteristics, by district, 2007

District	Adjusted Mean	Standard Error	Houston	Boston	Charlotte	Austin	New York City	San Diego	Atlanta	Los Angeles	Chicago	Cleveland	District of Columbia
Houston	238	1.0	238										
Boston	237	1.0		237									
Charlotte	237	0.9			237								
Austin	237	0.7				237							
New York City	233	0.8					233						
San Diego	229	1.2						229					
Atlanta	226	0.8							226				
Los Angeles	224	0.7								224			
Chicago	222	0.8									222		
Cleveland	217	1.2										217	
District of Columbia	217	0.7											217

Notes: 1. The green and gray colors in the boxes of this graph indicate whether or not a lower average or mean score of one district is statistically different from the higher average score of another district.

Example: The green box in the column labeled Los Angeles in the row labeled Atlanta means that Los Angeles' average score of 224 is statistically different from Atlanta's average score of 226.

Example: The gray box in the column labeled Atlanta in the row labeled San Diego means that Atlanta's average score of 226 is **not** statistically different from San Diego's average score of 229. Blocks of gray represent districts whose average adjusted scores are similar to one another and distinctly different from all others. In grade 4 mathematics, the top four districts make up such a cluster. Points where the green hits the diagonal represent a statistically significant drop in scores.

Yellow indicates a set of districts whose average adjusted scores are not statistically different from the top-scoring district.

2. Control variables for this analysis include race/ethnicity, special education status, ELL status, NSLP eligibility, and a composite literacy scale that includes the presence at home of newspapers, magazines, a computer, and more than 25 books.

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

Table 3.9 Average NAEP scores in grade 8 mathematics, adjusted for student background characteristics, by district, 2007

District	**Adjusted Mean**	**Standard Error**	**Austin**	**Charlotte**	**Boston**	**Houston**	**New York City**	**San Diego**	**Atlanta**	**Chicago**	**Cleveland**	**Los Angeles**	**District of Columbia**
Austin	277	0.9	277										
Charlotte	277	1.1		277									
Boston	275	0.9			275								
Houston	274	0.9				274							
New York City	267	1.5					267						
San Diego	265	1.5						265					
Atlanta	264	1.4							264				
Chicago	264	1.3								264			
Cleveland	264	1.7									264		
Los Angeles	260	0.9										260	
District of Columbia	255	0.8											255

Notes: 1. The green and gray colors in the boxes of this graph indicate whether or not a lower average or mean score of one district is statistically different from the higher average score of another district.

Example: The green box in the column labeled Los Angeles in the row labeled Chicago means that Los Angeles' average score of 260 is statistically different from Chicago's average score of 264. (Differences with Atlanta and Cleveland are not statistically significant due to variations in standard errors.)

Example: The gray box in the column labeled Chicago in the row labeled New York City means that Chicago's average score of 264 is **not** statistically different from New York City's average score of 267. Blocks of gray represent districts whose average adjusted scores are similar to one another and distinctly different from all others. In grade 8 mathematics, the top three districts make up such a cluster. Although Houston's adjusted average is not significantly different from Boston's, it is significantly lower than Austin's. The points where the green hits the diagonal represent a statistically significant drop in scores.

Yellow indicates a set of districts whose average adjusted scores are not statistically different from the top-scoring district.

2. Control variables for this analysis include race/ethnicity, special education status, ELL status, NSLP eligibility, and a composite literacy scale that includes the presence at home of newspapers, magazines, a computer, and more than 25 books. In grade 8, an indicator of whether one parent has a college degree was included in the analysis.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments.

Table 3.10 District effects by subject and grade after adjusting for student background characteristics, 2007*

	Reading Grade 4	Reading Grade 8	Math, Grade 4	Math Grade 8
Atlanta	0.1	0.1	-2.4*	-1.6
Austin	4.1*	4.1*	7.7*	11.8*
Boston	8.9*	5.9*	8.6*	9.3*
Charlotte	6.2*	3.8*	8.0*	11.0*
Chicago	-3.4*	4.8*	-6.9*	-1.8
Cleveland	-8.5*	2.7	-11.5*	-2.0
District of Columbia	-9.1*	-4.3*	-11.8*	-10.8*
Houston	2.2	4.4*	8.9*	8.2*
Los Angeles	-4.8*	-2.5*	-5.4*	-5.8*
New York City	3.9*	-1.3	4.5*	1.8*
San Diego	2.6*	-2.1	0.3	-0.7

* District effect is significantly different from zero.

The reader should note that 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 by 2007 (see next chapter) to levels of predicted performance as shown in table 3.10, while Cleveland continued to show 2007 performance below predicted levels, except in eighth grade reading.

Overall, the analysis of district effects showed 2007 performance of TUDA districts relative to one another after adjusting for student background characteristics and tended to confirm district selection decisions.

Summary of Selected Sites

Our analysis was conducted to identify districts that had demonstrated performance in specified ways on the NAEP over the period 2003 to 2007.[7] Our final selections were:

Districts that consistently made gains over time:

- Atlanta (particularly for gains in reading)[8]
- Boston (particularly for gains in mathematics)

[7] Additional analysis indicated that reading and math gains between 2003 and 2007 in Atlanta and Boston were not significantly different from gains by their respective states, except that Atlanta's NAEP math scores in grade 8 increased significantly more than its state did over the same period.

[8] 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.

District that consistently failed to make gains or posted losses:

- Cleveland

District that outperformed other districts adjusting for background characteristics:

- Charlotte

These four districts constituted our sample of districts for the alignment analyses and case studies of instructional programs and other contextual factors that may be associated with TUDA performance on NAEP.

Part 2. Alignment and Subscale Analyses and Site Visits

This section describes the methodology for analyzing the alignment between NAEP and the state and/or district standards in the four selected jurisdictions and for conducting the site visits to determine the nature of the instructional programs in the districts.

Alignment Analysis

The guiding research questions for this component of the project were as follows: In the four selected districts, to what degree were the state and district standards in place in 2006–2007 for reading and math in grades four and eight aligned with NAEP specifications in terms of content, sequence, and depth of cognitive demand? And to what degree were the state and district standards in place in 2004-2005 for science aligned with NAEP specifications in terms of content, sequence, and depth of cognitive demand?

This section describes the processes used to answer these questions, including the training, recording, and validation procedures used to match subject matter content by grade for states and districts, code the matched content, and summarize the codes.

Materials collected

To conduct the reading, mathematics, and science alignment analyses, the project team first assembled NAEP content specifications in place for 2003 through 2007. The specifications for NAEP reading in 1992 through 2003 were not public, so the core material for conducting the alignment included *The Reading Framework for the 2003 National Assessment of Educational Progress*, published by the National Assessment Governing Board (NAGB)[9]; its predecessor document, *Reading Framework for the National Assessment of Educational Progress: 1992-1998* (NAGB, no date); and internal documents provided by the Federal Statistics Program (FSP) housed at the American Institutes for Research (AIR). Of particular use was the 2003 *Reading Framework*, which provided examples of how the assessment had defined and measured "aspects of reading" within each "context for reading." This source provided material from which proxies for reading specification statements could be extrapolated. The team then developed most proxy specifications, with a brief note about the reading behaviors they covered.[10]

[9] Available at http://www.nagb.org/publications/frameworks/r_framework_05/toc.html

[10] The content leader for the reading component of the alignment work had coordinated the development of the 1990 NAEP reading assessment, served for four years on the technical review team for the 1992 NAEP reading assessment, and served at AIR, directing the project that developed *The Framework for the 2009*

The math alignment analysis used NAEP specifications in *The Mathematics Framework for the 2005 National Assessment of Educational Progress* published by NAGB.[11]

Like reading, the NAEP science "specifications" for 1994 through 2005 were not public. So, the research team used the *Science Assessment and Exercise Specifications for the 1994 National Assessment of Educational Progress*, published by NAGB and provided by the FSP housed at AIR. Of particular use was the *Grade-Level Specific Objectives and Ideas for Specific Items* in *Appendix A: Fields of Science Content Outlines Guide*. The science "framework" for 2005, however, is publicly available[12] and includes two appendices relevant to the alignment task: *Appendix B: Fields of Science*, which includes descriptions of major science topics and desired learning goals; and *Appendix D: Science Content Outlines (Excerpts)*, which includes excerpts from the science outlines that are fully detailed in the *Specifications* document.

Next, the project team assembled the state and/or district content standards in place during the 2006–2007 school year for reading and mathematics and the science content standards in place during the 2004–2005 school year. These documents were obtained primarily from state and district Websites or from curriculum leaders within the selected districts. (See appendix D.1 for a list of Websites and online documents used for the alignment analyses by district.)[13]

In science, only those standards that referenced specific science content were included in the alignment analysis. The standards that focused solely on science processes and/or skills were not matched to the NAEP content objectives. Science process standards were often written with the intention of serving as overarching statements that were applicable to multiple content areas and for this reason would match with too many objectives to be informative.[14]

In all three subjects—reading, math, and science, we examined materials in grades three and four and seven and eight to see if the content was likely to have been taught before NAEP testing. We also examined grade five materials to determine if NAEP grade four content was addressed after the assessment.

Standardized format

The project team developed a standardized chart for each content area in grade four and in grade eight to record all content objectives and alignment codes.[15] These charts included the actual NAEP specification language, the matching state and district content standards, and space for noting (1) the degree to which state and local content standards in the four selected sites matched

---

[11] Available at http://www.nagb.org/publications/frameworks/m_framework_05/toc.html

[12] Available at http://www.nagb.org/publications/frameworks/s_framework_05/761907-ScienceFramework.pdf

[13] For reading, information about state reading assessments was also obtained from state Websites. This information allowed a comparison of factors such as passage length and item type on NAEP reading assessment and on the state assessments, and it helped coders match the cognitive demand codings on each matched state and district standard.

[14] For example, the grade 3 scientific inquiry grade-level indicator from the Ohio Academic Content Standards is: "Read and interpret simple tables and graphs produced by self/others." The absence of specific content in this indicator requires that either (1) this indicator should be assumed appropriate for alignment to all NAEP content objectives or (2) the coders must make decisions about which content is most appropriate to the stated skill, a step that would significantly increase the subjectivity and decrease the reliability of the alignment task.

[15] The spreadsheet for reading was modified to accommodate the design of the assessment into "aspects for reading" and "contexts for reading."

the NAEP language, (2) cognitive demand codes, (3) content match codes, and (4) grade-level codes. (See appendix D.2 for a detailed list of the column labels for these charts.)

The team developed alignment charts for grades four and eight for each of the three disciplines and each of the four districts selected for deeper study. Column A indicated the NAEP content against which state and district standards were compared. The overall design of the reading alignment charts differed somewhat from the math and science charts in that those for reading were organized by "context for reading" (reading for literary experience, reading for information, and reading to perform a task) with proxy specifications for each "aspect of reading" listed under each "context." Each proxy specification was considered and coded independently.

Training

Two rounds of training were conducted to ensure reliability in coding. One involved content placement (matching) and the other involved content coding. For the four TUDA districts selected for deeper study, trainees received copies of state standards (and, where available, district standards) in reading, mathematics, and science for grades three, four, five, seven, and eight. At the start of training sessions, trainees were given copies of standardized charts for each grade. For each content area, the content leader trained two junior-level staff who were knowledgeable about the content area (referred to as "coders") to populate the spreadsheets. Training began with an orientation on the NAEP framework and specifications documents and on how to represent NAEP content in Column A. Coders also received an orientation on the standards documents for the states and districts that were to be studied. Moreover, training familiarized coders with chart formats, so that they could populate them uniformly for districts and states and participate in the verification process. Training included a complete discussion of the constructs of interest (grade-level match, content match, and cognitive demand), followed by exercises to ensure that the coders could complete their work reliably. Finally, content leaders discussed the exercises with the coders and provided retraining as needed.

Content matching processes

The process of matching state and district content to fourth-grade and eighth-grade NAEP content objectives entailed four steps:

1. **Orientation:** The content area leader reviewed the NAEP objectives for grades four and eight with two subject-area coders. This review included an orientation on the domains (e.g., numbers, geometry), the content[16] (i.e., the specific subject-matter specifications), the verbs used, and any elaborating language or details. Next, the content-area leader reviewed the organization of the state and district standards being aligned, with particular attention to grades three, four, and five for the grade four analyses and to grades seven and eight for the grade eight analyses. The leader then reviewed the format of the alignment charts and how responses would be recorded.

2. **The Content Matching Process and Practice:** The content area leader reviewed the "Process for Aligning NAEP and the State and District Standards," (see appendix D.3) responded to questions, reviewed the "Content-Matching Decision Rules"[17] (see appendix D.4) and then conducted practice matching exercises (see appendix D.5).

---

[16] In reading, the "content" was the proxy specifications that were arranged by "context" and "aspect" of reading. Otherwise, the content-matching processes were similar for reading, math, and science.

[17] Initial coding rules for reading were augmented as comparisons were made between the state and district standards and the NAEP proxy specifications.

3. **Content Matching:** Each coder then entered the appropriate state and district content for two districts into the grade four and grade eight alignment charts for a total of four charts per coder and eight charts in all.

4. **Verification of Content Placement:** Following the initial content matching, the content-area leader reviewed the "Content-Matching Verification Process" with the coders (see appendix D.6). Each coder then reviewed the other coder's charts, and all questions and concerns were noted. The two coders then discussed areas of uncertainty and reached agreement, whenever possible, to produce second drafts of the charts. The leader reviewed these draft charts and resolved any remaining disagreements. The resulting final drafts of the content-matching charts were then ready for coding.

Content coding processes

To analyze the degree of alignment between the NAEP content objectives and the district and state reading, math, and science standards in the TUDA districts, we assigned three types of codes:

- NAEP to state and district **grade-level match** codes
- NAEP to state and district **content match** codes
- NAEP, state, and district **cognitive demand** codes

Like the content-matching process, the process of completing this coding entailed four steps.

Step 1: Orientation: The content-area leader reviewed with the two content-area coders the three types of coding to be conducted at each grade level and for each state and district. This review included an orientation on the purposes of each code, the levels or categories of each code, and the proper placement of the following seven codes in the charts:

- NAEP to state grade-level match code—i.e., Is the content skill in the state standards assumed to be taught in the same grade level as tested on NAEP?
- NAEP to district grade-level match code—i.e., Is the content skill in the district standards assumed to be taught in the same grade level as tested on NAEP?
- NAEP to state content match code—i.e., Is the content of the state standards a complete, partial, or nonexistent match with the NAEP specifications?
- NAEP to district content match code—i.e., Is the content of the district standards a complete, partial, or nonexistent match with the NAEP specifications?
- NAEP cognitive demand code—i.e., What is the degree of rigor or complexity implied in the NAEP specifications?
- State standard cognitive demand code—i.e., What is the degree of rigor or complexity implied in the state standards?
- District standard cognitive demand code—i.e., What is the degree of rigor or complexity implied in the district standards?

Step 2: The Coding Process and Practice: The content area leader then reviewed the "Process for Coding Content Matches and Cognitive Demand," responded to questions, and administered and reviewed practice coding exercises with the coders.[18]

Step 3: Coding the Content: All three coders (content lead plus two content-area coders) then independently entered the appropriate seven codes into the grade four and grade eight alignment charts for two districts, producing a total of four charts per coder and eight charts in all.

Step 4: Reconciliation of Coding: When the three coders had completed their independent coding, they compared and discussed their ratings and attempted to resolve discrepancies through consensus. In cases where consensus was not reached, a majority opinion was used to complete the final District Chart.

Appendix D provides more detailed information on the process of coding content matches and cognitive demand (appendix D.7) and content coding exercises (appendix D.8). See appendix D.9 for detailed descriptions of Norman Webb's related Descriptors for Depth of Knowledge levels for mathematics and science, and see appendix D.10 for Karen Wixson's discussion of this topic in reading.

Summarizing the codes

The content-matching and content-coding results are summarized in the tables in chapter 4. The summaries provide data on the number and percentage of state and local standards that match NAEP specifications, by subscale and overall, for each of the four selected districts in grades four and eight.

Subscale Analysis

In addition to examining alignment issues, we conducted subscale analyses to identify strengths and weaknesses within each content area tested (reading, mathematics, and science) among the four selected districts. NAEP results are scaled separately by content area and grade. In addition, items in each content area are calibrated and scaled separately, and the composite scale is a weighted combination of those subscales. NAEP subscales are not reported on the same metric for all content areas. For example, an average math subscale score of 260 in geometry is not equivalent to an average subscale score of 260 in measurement. Therefore, average subscale scores or gains in average subscale scores are not directly comparable with one another. To examine district strengths and weaknesses within each content area without directly comparing average scale scores, we used the following approaches:

1. To compute the effect size corresponding to changes in subscale averages from 2003 to 2007 for reading and mathematics, we divided the change in subscale averages from 2003 to 2007 by the standard deviation of the subscale scores from 2003. We did this one subscale at a time. We also tested changes in average scores for statistical significance, again one subscale at a time.

2. We computed the percentile to which a given district's subscale average corresponded in the national public school sample. And we computed the changes in percentiles from 2003 to 2007 on the reading and mathematics assessments.

[18] For reading, the coding practice exercises involved independent coding by two junior team members, followed by checking and discussion with the senior coder or the team leader.

3. Finally, we computed the percentile to which a given district's *adjusted* subscale average corresponded in the national public school sample for reading and math in 2007 and for science in 2005 using the same method as above.

Site Visits

The purpose of the site visits was to gain a more detailed understanding, retrospectively, of the factors behind the NAEP achievement patterns in the four districts selected for more in-depth study. Each of the four districts received a three-day visit from an expert academic team composed of the Council of the Great City Schools' director of academic achievement, the director of research, and two or three other team members with specialized expertise in reading, math, or science. The Council's executive director participated in three of the four visits.

Prior to each visit, the four districts received a letter proposing a schedule of interviews and requesting an extensive list of materials for team review. A telephone conversation was held to clarify the list of interviewees and the documents needed. The requests focused on curriculum, professional development materials, data, assessment material, and strategic plans that governed the district's instructional programs during the 2003 to 2007 study period. District staff members were encouraged to search their archives for many of the documents.

Two days of individual and group interviews ranging from 30 to 90 minutes were scheduled with current and past district leadership, central office staff, principals, teachers, instructional coaches, and community members.

The team used a standardized protocol built around 10 key reform and improvement levers identified in *Foundations for Success,*[19] which compared and contrasted the characteristics of urban districts that were making notable student achievement gains on state assessments, and those that were making more modest gains, or failing to improve. Since 2003, the Council of the Great City Schools has used this research to guide instructional reviews it has conducted on numerous major city school systems. An expert advisory panel reviewed the protocol prior to its use, and the study team made modifications to the protocols based on the advisory panel's recommendations in keeping with the goals of the research project.

The teams began their interviews with a series of opening questions before delving into more detailed inquiries about how and why the instructional program of the district worked as it did during the study period. The detailed questions and follow-up were structured around the following components from *Foundations for Success* and from the Council's experience in reviewing the instructional programs in numerous big-city school districts: (1) political preconditions (context), (2) instructional goals and goal setting, (3) accountability, (4) curriculum and instruction, (5) program implementation, (6) professional development and teacher and principal quality, (7) assessments and data, (8) low-performing students and schools, (9) early childhood and elementary programs, and (10) secondary schooling. A copy of the case study protocol can be found in appendix E.

These program-component categories have been determined to be helpful in explaining the differences between urban school systems that show improvement and those that do not. Although questions from the site visit team were organized around these 10 broad categories, the teams did not follow a set script. Instead, drawing on their own expertise and building on responses provided during the interviews, the members of the site visit teams used the categories

---

[19] Source: Snipes, J. et al. *Foundations for Success: Case Studies of How Urban School Systems Improve Student Achievement*, MDRC for the Council of the Great City Schools, 2002.

to frame questions that needed to be answered to determine why each district achieved the NAEP results it did.

Based on the response to a set of initial questions, the team would focus on a series of more specific questions, such as:

- Why was this program developed?
- How was it developed?
- Describe the implementation process.
- How many schools/teachers/students were involved in the implementation?
- How was the level of implementation measured?
- How was progress monitored?
- How was success measured?
- Were there any modifications based on data? Can you provide an example?
- Is it still in place? To what degree? If it is no longer used, how was the decision made?

Following two days of interviews and document review, the team met for one day to synthesize findings and discuss emerging themes. The findings were summarized in the case studies for each district and incorporated into a chapter examining overarching themes and shared strategies.

What Was Not Examined

This research project looked at a considerable number of variables and contextual factors, 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. This continues to be a major problem in educational research in general. We did use data on changes in the overall student/teacher ratios and the percentage of total staff members who were teachers in each of the study districts, but we did not have data on such basic teacher background variables as undergraduate or graduate degrees, college major, teaching experience, teacher pay or incentives, and the like. 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 who 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.

Our analysis also did not examine what the results in these districts might look like if their teaching forces came from a higher echelon of college graduates as is the case in other higher performing countries. Moreover, the analysis did not include an examination of the effects of pay-for-performance initiatives in these cities.

The study looked at student background variables, as described earlier in this chapter, but we did not explicitly examine immigrant students (although the numbers of English language learners are used as part of the student background measures). Nor did the study attempt to measure the extent of casual reading among students, a variable that some research has correlated with reading scores. The study, moreover, did not look at discipline levels, tracking practices, or policies on retaining students in grade.

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.

We also did not explicitly examine such factors as class-size, school size, quantifiable measures of parent involvement, school choice, the use of early-childhood programs, extended-time initiatives or instructional time, community engagement measures, and other such variables. Also, we did not look explicitly at the role of wrap-around services in these city school systems or attempt to figure out what kind of effect they might have or not have on student achievement. The case-study teams often probed for evidence that school staff members, teachers, and others viewed these and similar variables as critical to their school system's movement on NAEP, but we did not explicitly attempt to measure them. We urge subsequent studies to begin considering them.

Finally, the study team did look at changes in overall resources available to the districts during the study period. This included a look at average per pupil expenditures between 2003 and 2007 and changes in the amount of funding devoted to instruction during the study period. Otherwise, we did not examine how districts deployed their financial resources. Where these factors were relevant to the results, they are mentioned.

CHAPTER 4
CONTENT, SUBSCALE, AND ALIGNMENT ANALYSIS ON THE SELECTED DISTRICTS

Introduction

The previous chapters presented NAEP data for the TUDA districts in the aggregate and by district. The methodology by which the project team analyzed the reported NAEP results and selected four districts for further study—Atlanta, Boston, Charlotte, and Cleveland—was also described. This chapter presents our detailed analysis of NAEP results in the four districts by content area: reading, mathematics, and science. For reading and math, we seek to understand district strengths and weaknesses among the four case-study school systems with data from 2003, 2005, and 2007. Science data were available only for 2005 when this analysis was conducted. The first questions this chapter addresses are:

- In which content strands are urban students in the selected districts showing the greatest gains in reading and math?
- In which content strands are urban students showing the greatest academic strengths and weaknesses in reading, math, and science?

The chapter also examines factors that might explain, in part, these content-specific trends and patterns. Specifically, we address two broad sets of questions about district performance on NAEP:

- 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 performance on the NAEP?
- What instructional practices were present in districts that may have contributed to the gains or high performance in each content strand on NAEP?

These four questions are addressed for reading in section 4a, mathematics in section 4b, and science in section 4c. The reading and mathematics sections summarize data on changes in subscale performance over time (2003-2007) in each of the four districts. Similar data do not exist for science because the science assessment was administered only once (in 2005). Additional data on subscales are presented in percentiles. Data are also presented on item omission rates by item type and rates of correct responses by item type.

For each subject, we also 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 four and grade eight NAEP specifications. As described in chapter 3, we examined alignment in terms of both content and cognitive demand. In addition, the reading section includes an analysis of test-item types.

Finally, each section concludes with a discussion of what was learned during the site visits regarding the instructional practices that might help explain the relationship between the alignment results and the NAEP data for each of the four districts.

*****

Part 1. District Performance on NAEP Reading Subscales

Content

The NAEP's grade four reading test assesses reading skills on two subscales or "contexts" (reading for a literary experience and reading for information). Each context is composed of a set of four "aspects" that indicate the cognitive tasks the items ask. Therefore, there are a total of eight aspects. The NAEP's grade eight reading test during the 2003-07 study period assessed three contexts—the same two tested in grade four, plus reading to perform a task. Again, each context has four aspects, for a total of 12 aspects. Table 4a.1 shows the distribution of NAEP test items by contexts for reading for grades four and eight.

Table 4a.1 Percentage of items by reading content area and grade level, 2007

Subscale	Grade 4	Grade 8
Reading for a Literary Experience	55%	40%
Reading for Information	45%	40%
Reading to Perform a Task	N/A	20%

Source: National Assessment Governing Board (2002). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, D.C.

Composite, Subscale and Item Analyses–Strengths and Weaknesses in Reading

Chapter 2 discussed reported NAEP achievement overall and by district. This first section focuses on district strengths and weaknesses in reading among the four case-study districts. The analysis includes data on the two fourth-grade and three eighth-grade reading contexts listed above (table 4a.1). Data by aspect are not available through NAEP.

As noted in chapter 3 (Methodology), NAEP subscales are not all reported on the same metric. Therefore, average or mean subscale scores or gains on average subscale scores are not directly comparable from one subscale to another. In order to estimate relative strengths and weaknesses among the districts in each content area, we examine subscale and item-level performance in several ways.

First, we display changes between 2003 and 2007 in subscale performance in the selected TUDA districts in terms of effect size and statistical significance. Second, we provide the percentile rankings of the average composite and subscale scores for each of the four districts, based on the distribution of scale scores from the national public school population, not the scale scores from the full population estimates. Third, we graphically display percentile rankings of average subscale scores for each district, adjusted for student background characteristics. Finally, we provide item-level information about omission rates and percentage of correct answers for each district.

Note that because these analyses used different methods to examine changes in scale scores between 2003 and 2007, they occasionally led to slightly different results. For example, a change in the percentile rankings of the average subscale scores for a given district may not correspond to a statistically significant change in average scale scores. Therefore, the reader is encouraged to look at the results as a complete package rather than one finding at a time. Taken together, the results provide rich information about the reading performance, strengths, and weaknesses of the four selected TUDA districts.

Changes in Subscale Performance from 2003 to 2007

As we reported in chapter 3, 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. 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-city (LC) schools and the national public school sample. Table 4a.2 shows the results for fourth-grade reading and table 4a.3 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 4a.2 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-city schools
↑ 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 4a.3 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

Note: The results presented in this table are based on average or mean scale scores of students in the reporting sample. The results displayed in chapter 4 utilized the full population estimates (FPEs). Those results differ from the composite reading scores in this table in certain cases: (1) Using FPEs, changes in average scale scores at the composite level from 2003 to 2007 are not significant for Atlanta or Cleveland, but these changes are statistically significant using reported scores. (2) According to the FPE-based results presented in chapter 3, average scale scores for the national public school sample declined significantly from 2003 to 2007; here the results show the change from 2003 to 2007 was not statistically significant.
Key: LC=Large-city schools.
↑ 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) schools and the national public samples. 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 eight 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.[1]

Percentile Measures by Subscale

In the next analyses, we made indirect, normative, within-district comparisons between subscales by noting the percentile of each district's average subscale in terms of the national public school sample. These indirect comparisons reflect technical issues that do not allow direct comparisons of one NAEP subscale to another. Again, the purpose was to estimate specific district strengths and weaknesses in reading. Tables 4a.4 through 4a.7 show the percentiles of each district's average reading scores (composite and subscales) by year at grades four and eight. The tables also show the changes in percentile points between 2003 and 2007 for each district, but the significance of the change was not tested because of the indirect nature of percentiles as a measure of achievement. Instead, the analysis relied on the use of effect sizes seen in the previous section to determine the significance of subscale change.

Atlanta

As shown in table 4a.4, the average fourth-grade performance in Atlanta on both the composite reading score and on all subscales was below the national public median (50th percentile) in 2003, 2005, and 2007.

Table 4a.4 *Atlanta's* average NAEP reading percentiles and changes in percentiles, by subscale and grade, 2003-2007

	Grade 4				Grade 8			
	Percentile of the mean scale score			Shift in percentile	Percentile of the mean scale score			Shift in percentile
	2003	2005	2007	2003-2007	2003	2005	2007	2003-2007
Composite	28	31	33	5	25	26	29[2]	4
Literary	29	31	35	6	26	28	30	3[3]
Information	29	31	33	4	26	26	31	5
Task	N/A	N/A	N/A	N/A	26	29	32	6

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

[1] Cleveland did not show significant reading gains in eighth grade when analyzed with full population estimates.
[2] Note that composite scores can be lower than the individual subscale scores due to how the subscales are weighted.
[3] Difference is due to rounding.

In 2007, the average fourth-grade student in Atlanta was at the 33rd percentile on the reading composite score (up from the 28th percentile in 2003), and at the 35th percentile in reading for literary experience and the 33rd percentile in reading for information (up from the 29th percentile in 2003 for both context subscales). (See table 4a.4).

As in grade four, the grade eight average performance in Atlanta on the composite score and on all three subscales was below the national public median in 2003, 2005, and 2007. But, in 2007, the average eighth-grade student in Atlanta scored at the 29th percentile on the composite measure—up from the 25th percentile in 2003, and at the 32nd percentile in reading to perform a task, the 31st percentile in reading for information, and the 30th percentile in reading for literary experience. The largest effect size gain in subscale reading scores among Atlanta's eighth graders was in reading to perform a task.

In addition, table 4a.4 shows that Atlanta's standing compared to the national public sample appears somewhat better at grade four than at grade eight on the composite measure and on all subscales.

Finally, the analysis of item responses in Atlanta did not identify any reading items on which the district's fourth graders were able to answer more readily than students in the national public sample. But, Atlanta's fourth graders had more difficulty than others with such reading tasks as—

- Determining what caused polar bears to lose weight, as discussed in a passage (information)
- Determining the meaning of the word "dismantle" (information)

At the eighth grade, Atlanta students found it easier than their peers nationwide to answer such reading tasks as—

- Describing why polar bears could go for months without eating (information)
- Determining the difference between antiquarians and archeologists (information)
- Describing similarities and differences between Jefferson and Schliemann, based on a passage (information)
- Describing how an author uses language (literary)

On the other hand, Atlanta's eighth graders had a more difficult time than their peers nationwide with such reading tasks as—

- Writing one's senator with a petition and argument (task)
- Answering questions about the protection of dolphins, based on a passage (information)
- Describing what to put into a time capsule and why (task)
- Taking the vantage point of a narrator in a story (literary)

Boston

Table 4a.5 displays the same information for Boston. At grade four, the average performance of Boston on the composite measure and both subscales was below the national public median in 2003, 2005, and 2007. In 2007, the average student in Boston was at the 36th percentile on the reading composite measure, the 38th percentile in reading for literary experience, and the 35th percentile in reading for information. There were no significant gains on either reading subscale between 2003 and 2007 at grade four.

In grade eight, the average performance of Boston's students on the composite reading score and on all three reading subscales was also below the national public median in 2003, 2005, and 2007. The average eighth grader scored at the 38th percentile on the composite measure. The highest subscale performance appeared to be in reading for information where Boston's eighth graders were at the 41st percentile in

2007. The city's eighth graders were at the 39th percentile in reading to perform a task in 2007 and at the 38th percentile in reading for literary experience.

Table 4a.5 also shows that, unlike Atlanta, Boston's standing, in terms of the national public sample, was slightly better at grade eight than at grade four on the composite measure and in reading for information.

Table 4a.5 *Boston's* average NAEP reading percentiles and changes in percentiles, by subscale and grade, 2003-2007

	Grade 4				Grade 8			
	Percentile of the mean scale score			Shift in percentile	Percentile of the mean scale score			Shift in percentile
	2003	2005	2007	2003-2007	2003	2005	2007	2003-2007
Composite	36	37	36	#	37	39	38	1
Literary	37	38	38	1	40	40	38	-2
Information	36	37	35	-1	38	41	41	3
Task	N/A	N/A	N/A	N/A	34	39	39	5

Rounds to zero
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005 and 2007 Reading Assessments.

Finally, the analysis of item responses in Boston found that fourth graders in the district were more able than their peers nationwide to answer such reading tasks as—

- Describing how a character in a story is like someone they know (literary)
- Describing how money matters in a family, from a story (literary)

Conversely, fourth graders in Boston found it more difficult than their peers nationwide to answer such reading tasks as—

- Defining the meaning of the word "dismantle" (information)
- Articulating what does a specified circle describes (information)

Eighth graders in Boston were more likely than their peers nationwide to do well on such tasks as—

- Writing one's senator with a petition and argument (task)
- Offering good advice with an explanation (literary)

On the other hand, Boston's eighth graders had more difficulty than their peers nationwide with such reading tasks as—

- Learning from technology (literary)
- Describing the role of charcoal makers, from a passage (literary)

Charlotte

Table 4a.6 shows the same information for Charlotte. At grade four, Charlotte's average performance on the composite measure and on both subscales was at or near the national public median in 2003, 2005, and 2007. In 2007, the average student in Charlotte was at the 50th percentile on the composite reading measure, the 49th percentile in reading for literary experience and 51st percentile in reading for information. There were no significant gains using the effect-size measures, except in reading for information.

At grade eight, the average performance for Charlotte's students on the composite measure and all three subscales was somewhat below the national public median in 2003, 2005, and 2007. In 2007, the average student in Charlotte was at the 45th percentile on the composite reading score, the 47th percentile in reading for literary experience, the 44th percentile in reading for information, and the 45th percentile in reading to perform a task. Charlotte showed no positive gains in the eighth grade from 2003 to 2007. The largest negative change was in reading to perform a task.

Table 4a.6 shows that Charlotte's standing in terms of the national public sample is slightly higher at grade four than at grade eight on the composite measure and the literary and information subscales.

Table 4a.6 *Charlotte's* average NAEP reading percentiles and changes in percentiles, by subscale and grade, 2003-2007

	Grade 4				Grade 8			
	Percentile of the mean scale score			Shift in percentile	Percentile of the mean scale score			Shift in percentile
	2003	2005	2007	2003-2007	2003	2005	2007	2003-2007
Composite	50	52	50	#	48	46	45	-3
Literary	50	52	49	#	50	45	47	-2[4]
Information	50	52	51	1	45	47	44	-1
Task	N/A	N/A	N/A	N/A	50	47	45	-5

Rounds to zero
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005 and 2007 Reading Assessments.

Finally, the analysis of item responses in Charlotte did not identify any reading items on which the district's fourth graders were able to answer more readily than students in the national public sample. Nor were there any items with which the district's fourth graders had more difficulty than others.

At the eighth grade, Charlotte students were more likely than their peers nationwide to do well on—

- Writing one's senator with a petition and argument (task)

There were no tasks where Charlotte's eighth graders had more difficulty than did the national sample.

Cleveland

Table 4a.7 shows the same information for Cleveland. At grade four, Cleveland's average performance on the composite reading measure and both subscales was below the national public median in 2003, 2005, and 2007. In 2007, the average fourth-grade student in Cleveland was at the 25th percentile on the composite measure, the 27th percentile in reading for literary experience, and the 24th percentile in reading for information. The reader will note that Cleveland and Atlanta had what appeared to be similar fourth-grade composite reading percentile scores in 2003 (27 and 28, respectively), but they appeared quite different in 2007 (25 and 33, respectively).

[4] Difference is due to rounding.

At grade eight, the average performance of Cleveland's students on the composite reading measure and all three subscales was below the national public median in 2003, 2005, and 2007. In 2007, the average student in Cleveland was at the 30th percentile on the composite measure, the 34th percentile in reading for literary experience, the 31st percentile on reading for information, and the 29th percentile on the "task" subscale. There were no significant changes in effect sizes on any of the eighth-grade reading subscales during the study period, despite the apparent changes in percentiles. Table 4a.7 shows that Cleveland's standing in terms of the national public sample appears to be slightly higher on the composite measure and the literary and information subscales at grade eight than at grade four.

Table 4a.7 *Cleveland's* average NAEP reading percentiles and changes in percentiles, by subscale and grade, 2003-2007

	Grade 4				Grade 8			
	Percentile of the mean scale score			Shift in percentile	Percentile of the mean scale score			Shift in percentile
	2003	2005	2007	2003-2007	2003	2005	2007	2003-2007
Composite	27	27	25	-2	26	26	30	4
Literary	28	29	27	-1	30	30	34	4
Information	27	28	24	-3	25	24	31	6
Task	N/A	N/A	N/A	N/A	24	28	29	4[5]

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

Finally, the analysis of item responses in Cleveland found that fourth graders in the district were more able than their peers nationwide to do better than their peers on such tasks as—

- Describing how a character in a story spent time talking to ducks (literary)

Cleveland's fourth graders were more likely than their peers nationwide to have trouble with such reading tasks as—

- Describing how an author presents information (information)
- Knowing what one does when finding a banded bird (information)
- Determining the meaning of the word "pleading" (literary)
- Describing why a character in a story feels proud (literary)

Among eighth graders, Cleveland students did better than their peers nationwide on such tasks as—

- Providing detail from a story and explaining it (literary)
- Describing why polar bears could go for months without eating (information)
- Writing one's senator with a petition and argument (task)

On the other hand, Cleveland's eighth graders had more difficulty than their peers nationwide with such reading tasks as—

- Describing what to put into a time capsule and why (task)
- Determining the meaning of the word "deciphering" (information)
- Describing what newspaper clippings would go into a time capsule and why (task)

[5] Difference is due to rounding.

Percentile Measures by Subscale, Adjusted for Student Background Characteristics

Table 4a.8 takes the adjusted subscale averages for fourth- and eighth-grade reading in 2007 for each study district and (1) shows them in terms of percentile position (based on the national public school sample) in the NAEP assessments, and (2) modifies them for the same demographic variables discussed in chapter 3 in order to compare district performance once background variables were taken into account. The table shows that the adjusted averages for both subscales in grade four in all four districts were below the adjusted national median. In addition, the percentiles of the adjusted reading subscales of each grade appear similar within each district.[6] At grade eight, the adjusted subscale averages were also below the national median for all four districts. In Atlanta, the score in reading to perform a task appears higher than other subscales. In Boston, the three subscales appear to be relatively close to each other. Reading for literary experience appears to be a strength in Charlotte and Cleveland.

Table 4a.8 Adjusted NAEP reading subscale averages scores in percentiles on the national public school sample, by district and grade, 2007

	Grade 4		Grade 8		
	Literary Experience	Information	Literary Experience	Information	Task
Atlanta	33	32	32	32	36
Boston	43	40	38	40	39
Charlotte	38	39	39	34	37
Cleveland	26	24	38	34	33

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

Percentage of Omitted Items by Item Type

In addition to conducting the subscale analyses, we examined the percentage of items left blank, i.e., omitted items, by item type. Table 4a.9 shows the average omission rates by item type in grades four and eight for the four selected districts. In considering omission rates on the NAEP reading assessment, one must remember that the passages students are asked to read are long (from about 250 to some 1,200 words), and that many students, especially at grade four, are not necessarily accustomed to reading such long passages in a timed situation.

Table 4a.9 Item omission rates on NAEP reading, by item type, grade, and district, 2007

	Grade 4		Grade 8	
	MC items	CR items	MC items	CR items
Atlanta	0.6	3.1	0.6	6.6
Boston	0.7	4.4	1.1	7.2
Charlotte	0.6	3.2	0.7	4.2
Cleveland	0.8	5.3	0.7	6.4
LC	0.6	3.7	0.5	5.7
National Public	0.5	3.1	0.3	3.7

Note: MC=multiple-choice, CR=constructed-response
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessment.

[6] Percentiles for all 11 TUDA districts are shown in appendix B, tables B.40-41. Results show that fourth graders appeared to do better at reading for literary experience than in reading for information. Eighth-grade students appeared to do better in reading for literary meaning than at reading for information or reading to perform a task.

At grade four, the omission rates on multiple-choice (MC) items ranged from 0.6 percent in Atlanta and Charlotte to 0.8 percent in Cleveland. The omission rates among fourth graders on constructed-response (CR) items ranged from 3.1 percent in Atlanta, which was similar to the national rate, to 5.3 percent in Cleveland.

The omission rates among fourth graders on multiple-choice items for large-city schools and the four selected districts appeared similar to or higher than the national average. The omission rates on constructed-response items in large cities and two of the four selected districts—Boston and Cleveland—were higher than the national average. Omission rates in Atlanta and Charlotte were similar to the national average.

In grade eight, the omission rates in the selected jurisdictions on multiple-choice (MC) items ranged from 0.6 percent in Atlanta to 1.1 percent in Boston. The omission rates on constructed-response (CR) items ranged from 4.2 percent in Charlotte to 7.2 percent in Boston.

Not surprisingly, the highest omission rates at both grades and in all four districts were on CR items. In addition, in all districts, omission rates for CR items were higher at grade eight than at grade four. The omission rates in all selected districts and large-city schools generally were higher than the national average.

Approximately half of the items on the NAEP reading assessment require written responses that ask students to explain and support their ideas. There are two types of CR items—short, which require one- or two-sentence answers, and extended responses, which require students to write a paragraph in response.

Percentage of Correct Items by Item Type

Finally, we examined the percentage of correct items by item type in each of the four TUDA districts and compared the percentages to the national public sample and the LC averages. Table 4a.10 displays the average percent-correct rates by item type in grades four and eight.

Every district—and the nation—has a higher rate of correct responses on multiple-choice than on constructed-response items. In grade four, the rates on multiple-choice items ranged from 58 percent (Cleveland) to 74 percent (Charlotte) and on constructed-response items, from 40 percent (Cleveland) to 52 percent (Charlotte).

Table 4a.10 Percent-correct rates on NAEP reading, by item type, grade, and district, 2007

	Grade 4		Grade 8	
	MC items	CR items	MC items	CR items
Atlanta	65	45	67	49
Boston	65	47	70	55
Charlotte	74	52	74	56
Cleveland	58	40	67	50
LC	66	46	69	52
National Public	72	51	75	57

Note: MC=multiple-choice, CR=constructed-response
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Reading Assessments.

In grade eight, the percent-correct rates on constructed-response items ranged from 49 percent (Atlanta) to 56 percent (Charlotte), and on multiple-choice items from 67 percent (Cleveland) to 74 percent (Charlotte).

Every district—and the nation—had a higher rate of correct responses on multiple-choice than on constructed-response items. In addition, the percent-correct rates for CR items in all four districts were somewhat higher in grade eight than in grade four. The largest difference was observed in Cleveland, where the percent-correct rate in CR items was 40 percent in grade four and 50 percent in grade eight.

Part 2. Potential Factors Behind Subscale Reading Trends

To help us further understand the reading results, we explored two hypotheses about what might be driving student reading performance overall and at the subscale levels. First, we examined the alignment of state and/or district reading standards, specifications, expectations, or indicators with the NAEP reading specifications by context and aspect. The purpose of this analysis was to determine the extent to which we could expect that students' reading instruction had prepared them for the kinds of reading materials and tasks that are included on the NAEP assessment.

Second, the research team conducted site visits to the four selected districts to see what the districts had done instructionally that would help explain the NAEP reading scores. The methodology for both parts of this chapter is described in chapter 3 and in appendices C and D.

Alignment of State and District Standards to NAEP Reading Specifications[7]

The purpose of this part of the analysis was to determine how well each state or district's reading content standards were aligned with the NAEP specifications and to see if there was any connection to how well a district did on NAEP. For three of the four selected TUDA districts, we compared the state reading standards to NAEP. For Boston, we conducted the alignment analysis against both the state and the district standards, which were distinct.

Degree of Content Match

Fourth-Grade Reading

Our analysis for grade four showed that only about half the time did the NAEP specifications completely or partially match most district and state standards in content (between 44 percent and 56 percent). The exception was in Charlotte/North Carolina, where NAEP specifications were completely or partially matched by their standards about 80 percent of the time. These results are shown in table 4a.11 and figure 4a.1. The details follow in the bullets below.

There were 54 NAEP specifications in fourth-grade reading. The pattern of overall content matching was different from jurisdiction to jurisdiction. (Districts in bold are those selected for significant and less significant gains in reading.)

- **Atlanta/Georgia's standards matched 26 (48 percent) of the 54 NAEP specifications, with 21 complete and five partial matches. Therefore, 39 percent of the 54 NAEP specifications were completely aligned with the Atlanta/Georgia standards.**

- Boston, which had slightly different standards than its state, matched 28 (52 percent) of the 54 NAEP specifications, with 21 complete and seven partial matches. Therefore, 39 percent of the 54 NAEP specifications were completely aligned with the Boston standards. The state of Massachusetts had 35 percent complete matches.

[7] The specifications used for these analyses were developed from internal NCES documents and the detailed descriptions of the NAEP reading aspects presented in the official 2003 NAEP Reading Framework.

- Charlotte/North Carolina's standards, the district/state with the highest overall alignment, matched 43 (80 percent) of the 54 specifications, with 36 complete and seven partial matches. Therefore, 67 percent of the 54 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland/Ohio's standards matched 30 (56 percent) of the 54 NAEP specifications with 21 complete and nine partial matches. Therefore, 39 percent of the 54 NAEP specifications were completely aligned with the Cleveland/Ohio standards.**

In general, the degree of complete and partial content matches in fourth-grade reading was modest at best, except in Charlotte/North Carolina where matches were relatively strong. Atlanta, Boston, and Cleveland all had the same percentage of complete matches (39 percent).

If we look at two of the reading contexts—reading for literary experience and reading for information—the patterns show a somewhat more complicated picture. There were 28 NAEP specifications in the subscale of *reading for literary experience* for the fourth grade.

- **Atlanta/Georgia matched 16 (57 percent) of the 28 subscale specifications, with 13 complete and three partial matches. Therefore, 46 percent of the 28 NAEP specifications were completely aligned with the Atlanta/Georgia standards.**

- Boston matched 17 (61 percent) of the 28 subscale specifications, with 10 complete and seven partial matches. Therefore, only 36 percent of the 28 NAEP specifications were completely aligned with the Boston's standards.

- Charlotte/North Carolina matched 23 (82 percent) of the 28 subscale specifications, with 19 complete and four partial matches. Therefore, 68 percent of the 28 NAEP specifications were completely aligned with the Charlotte/North Carolina standards. The level of complete matches in Charlotte was notably higher than the percentage of complete matches in the other selected districts.

- **Cleveland/ Ohio matched 20 (71 percent) of the 28 subscale specifications, with 13 complete and seven partial matches. Therefore, 46 percent of the 28 NAEP specifications were completely aligned with the Cleveland/Ohio standards.**

In the subscale on *reading for information* in fourth grade, there were 26 NAEP specifications.

- **Atlanta/Georgia matched 10 (38 percent) of those 26 subscale specifications, with eight complete and two partial matches. Therefore, 31 percent of the 26 NAEP specifications were completely aligned with the Atlanta/Georgia standards.**

- Boston matched 11 (42 percent) of the 26 subscale specifications, with all 11 being complete matches (no partial matches). Therefore, 42 percent of the 26 NAEP specifications were completely aligned with the Boston's standards.

- Charlotte/North Carolina matched 20 (77 percent) of the 26 subscale specifications, with 17 complete and three partial matches. Therefore, 65 percent of the 26 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland/Ohio matched 10 (38 percent) of the 26 subscale specifications, with eight complete and two partial matches. Therefore, like Atlanta, 31 percent of the 26 NAEP specifications were completely aligned with the Cleveland/Ohio standards.**

In addition, all districts had notable variations in the degree of content match across the eight aspects of reading in the fourth grade. (See table 4a.11.) Each *aspect* had between six and eight NAEP specifications in each subscale.[8]

- **Atlanta/Georgia's highest level of alignment in reading for literary experience was in the aspect of "forming a general understanding," where all seven NAEP specifications (100 percent) were completely matched by the Atlanta/Georgia standards. The lowest level of alignment in that subscale was in the aspect dealing with "examining content and structure," which had no matches. In the subscale related to reading for information, the highest level of match was in "developing an interpretation." Five of seven or 71 percent of NAEP aspects were completely matched by the Atlanta/Georgia standards, with four complete matches and one partial match. The lowest level of alignment in reading for information involved "making reader/text connections," where there were no matches, either complete or partial.**

- Boston's highest level of alignment in reading for literary experience was in the aspect of "forming a general understanding," matching all seven specifications (100 percent), with five complete and two partial matches. The lowest level of alignment in that subscale was in the aspect dealing with "making reader/text connections," which matched three of eight specifications (38 percent) with two complete matches and one was partial match. In the subscale related to reading for information, the highest level of match was in "forming a general understanding." This aspect matched on five of six (83 percent) NAEP specifications—all of which were complete. The lowest alignment in reading for information involved "examining content and structure," where there was only one match of seven specifications (14 percent), although it was a complete one.

- Charlotte/North Carolina, as expected, had the highest overall level of matching by aspect, with at least a 50 percent match in every aspect. The highest level of match in reading for literary experience was in the aspect of "developing an interpretation," with complete matches for all seven specifications (100 percent). The lowest match in that subscale was in the aspect dealing with "making reader/text connections," which matched five of eight specifications (63 percent), with all five being complete. In the subscale related to reading for information, the highest level of match was in "developing an interpretation," with all seven NAEP aspects (100 percent) completely matched by Charlotte/North Carolina standards. The lowest match in reading for information involved "making reader/text connections"—the same aspect that was lowest in the reading for literary experience subscale. Here, the aspect matched on only three of six (50 percent) specifications, with one complete and two partial matches.

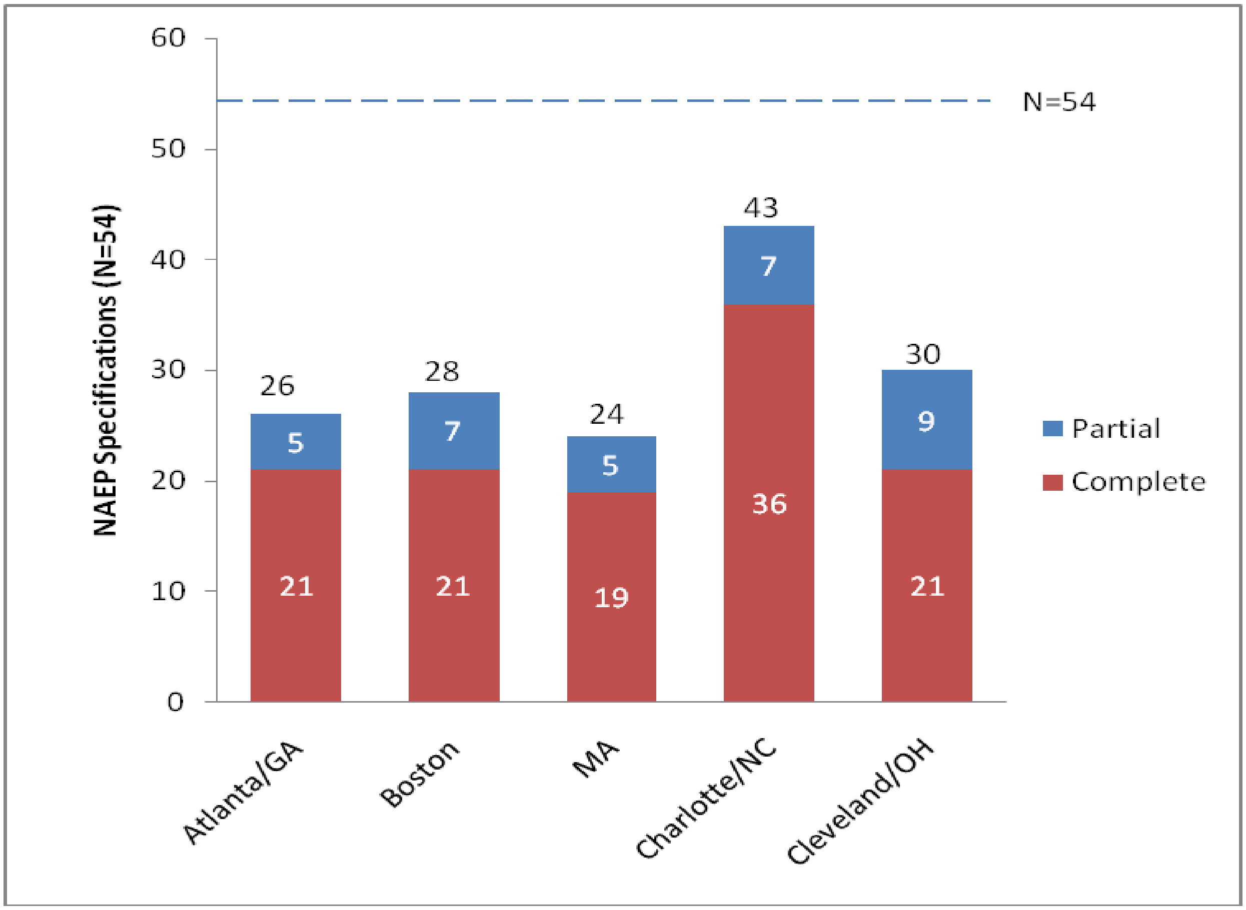
- **Cleveland/Ohio's highest level of alignment in reading for literary experience was in the aspect of "forming a general understanding," with all seven NAEP aspects (100 percent) being completely matched by the Cleveland/Ohio standards. The lowest match in that subscale was in the aspect of "making reader/text connections," matching four of eight (50 percent) specifications—all of which were partial matches. In the subscale related to reading for information, the highest level of match was in "forming a general**

[8] NAEP performance data are not disaggregated or reported by aspect.

understanding." Five of six (83 percent) of NAEP specifications were completely matched by the Cleveland/Ohio standards, all five of which were complete. The lowest match in reading for information involved "examining content and structure" and "making reader/text connections," where there were no matches of any kind on either aspect.

In general, the alignment in reading at the fourth-grade level was higher in reading for literary experience than in reading for information. In both subscales "forming a general understanding" tended to have the highest level of alignment, while "making reader/text connections" showed the lowest alignment. Finally, Charlotte/North Carolina showed the highest overall level of alignment (80 percent) between its standards and the NAEP specifications in fourth-grade reading. In contrast, Atlanta, Boston, and Cleveland showed similarly moderate or low complete and partial alignments overall, ranging from 48 percent (Atlanta) to 56 percent (Cleveland). When one examines complete matches only, then the overall alignment ranges from 39 percent in Atlanta, Boston, and Cleveland to 67 percent in Charlotte.

Figure 4a.1 Number of complete and partial matches with NAEP grade 4 reading specifications, by selected districts (*N* of NAEP specifications = 54), 2007*



*26 (48 percent) of Atlanta's grade 4 reading standards matched NAEP's 54 reading specifications either completely or partially; 28 (52 percent) of Boston's grade 4 reading standards matched NAEP's 54 reading specifications either completely or partially; 24 (44 percent) of Massachusetts's grade 4 reading standards matched NAEP's 54 reading specifications either completely or partially; 43 (80 percent) of Charlotte's grade 4 reading standards matched NAEP's 54 reading specifications either completely or partially; and 30 (56 percent) of Cleveland's grade 4 reading standards matched NAEP's 54 reading specifications either completely or partially.

Table 4a.11 Degree of match with NAEP grade 4 reading specifications/expectations/indicators, by subscale, aspect, and district, 2007

		Number of specifications in NAEP, by aspect																	
Subscale:		Reading for a Literary Experience								Reading for Information									
Aspect:		Forming a General Understanding		Developing an Interpretation		Making Reader/Text Connections		Examining Content and Structure		Forming a General Understanding		Developing an Interpretation		Making Reader/Text Connections		Examining Content and Structure		Total	
		7		7		8		6		6		7		6		7		54	
	Atlanta/ GA	7	100%	6	86%	3	38%	0	0%	4	67%	5	71%	0	0%	1	14%	26	48%
		C=7	P=0	C=5	P=1	C=1	P=2	C=0	P=0	C=4	P=0	C=4	P=1	C=0	P=0	C=0	P=1	C=21	P=5
	Boston	7	100%	4	57%	3	38%	3	50%	5	83%	4	57%	1	17%	1	14%	28	52%
		C=5	P=2	C=3	P=1	C=2	P=1	C=0	P=3	C=5	P=0	C=4	P=0	C=1	P=0	C=1	P=0	C=21	P=7
	MA	6	86%	5	71%	3	38%	1	17%	4	67%	4	57%	0	0%	1	14%	24	44%
		C=6	P=0	C=5	P=0	C=2	P=1	C=0	P=1	C=4	P=0	C=2	P=2	C=0	P=0	C=0	P=1	C=19	P=5
	Charlotte/ NC	6	86%	7	100%	5	63%	5	83%	4	67%	7	100%	3	50%	6	86%	43	80%
		C=4	P=2	C=7	P=0	C=5	P=0	C=3	P=2	C=4	P=0	C=7	P=0	C=1	P=2	C=5	P=1	C=36	P=7
	Cleveland/ OH	7	100%	5	71%	4	50%	4	67%	5	83%	5	71%	0	0%	0	0%	30	56%
		C=7	P=0	C=4	P=1	C=0	P=4	C=2	P=2	C=5	P=0	C=3	P=2	C=0	P=0	C=0	P=0	C=21	P=9

Note: How to read this table. Example: There are seven specifications in the context *Reading for a Literary Experience* under the aspect "Forming a General Understanding." Atlanta's standards matched the content on seven of those seven specifications or 100 percent. All seven were complete content matches.
C= complete match; P=partial match

Table 4a.12 summarizes the degree of complete alignment for each subscale and aspect in each of the selected jurisdictions. Matches were deemed to be high when at least 80 percent of NAEP specifications were completely matched by district/state objectives. Matches were deemed low when 50 percent or less of the NAEP specifications were completely matched by the district/state objectives for that subscale and aspect. Only seven of the 40 cells in table 4a.12 indicated high alignment and 20 of the 40 were low. Consequently, complete alignments in fourth-grade reading could be characterized as low to moderate. Overall, "forming a general understanding" in both reading for literary experience and reading for information tended to have the highest degree of complete alignment.

Table 4a.12 Degree of complete match of NAEP subscales with district/state standards in grade 4 reading, by subscale, aspect, and district, 2007*

Subscale/ Context	Aspect of Reading	Atlanta/ GA	Boston	MA	Charlotte/ NC	Cleveland/ OH
Reading for a Literary Experience	Forming a General Understanding	High	Moderate	High	Moderate	High
	Developing Interpretation	Moderate	Low	Moderate	High	Moderate
	Making Reader/ Text Connections	Low	Low	Low	Moderate	Low
	Examining Content and Structure	Low	Low	Low	Moderate	Low
Reading for Information	Forming a General Understanding	Moderate	High	Moderate	Moderate	High
	Developing Interpretation	Moderate	Moderate	Low	High	Low
	Making Reader/ Text Connections	Low	Low	Low	Low	Low
	Examining Content and Structure	Low	Low	Low	Moderate	Low

* High (80 percent or more) and low (50 percent or less)

Eighth-grade Reading

Our analysis for grade eight reading showed that between 37 percent (Massachusetts) and 65 percent (Cleveland/Ohio) of NAEP reading specifications were either completely or partially matched by district/state standards. These results are shown in table 4a.13 and figure 4a.2. The details follow in the bullets below.

There were 78 NAEP specifications in eighth-grade reading. In this case, Cleveland/Ohio standards had the highest overall complete and partial alignment. The pattern of matching differed substantially from jurisdiction to jurisdiction. (Districts in bold are the main comparison districts in reading.)

- **Atlanta/Georgia standards matched 33 (42 percent) of the 78 NAEP specifications, with 31 completely and two partially aligned. Therefore, 40 percent of the 78 NAEP specifications were completely aligned by the Atlanta/Georgia standards.**

- Boston, which had slightly different standards than its state, matched 32 (41 percent) of the 78 NAEP specifications, with 27 completely and five partially aligned. Therefore, some 35 percent of the 78 NAEP specifications were completely aligned with the city's standards. Boston and Massachusetts had the same number of standards completely aligned with NAEP specifications, although Boston had slightly more partial alignments than did the state.

- Charlotte/North Carolina standards matched 46 (59 percent) of the 78 NAEP specifications, with 43 complete and three partial matches. Therefore, 55 percent of the 78 NAEP specifications were completely aligned with Charlotte/North Carolina's standards.

- **Cleveland/Ohio matched 51 (65 percent) of the 78 specifications, with 44 complete and seven partial matches. Therefore, 56 percent of the 78 NAEP specifications were completely aligned with the Cleveland/Ohio standards.**

In general, the overall degree of complete and partial content matches in eighth-grade reading was modest at best.

If we look at the three reading strands in eighth grade—reading for literary experience, reading for information, and reading to perform a task—the alignment patterns showed a somewhat more complicated picture.

There were 29 NAEP specifications in the subscale on *reading for literary experience*.

- **Atlanta/Georgia matched 18 (62 percent) of the 29 subscale specifications, with 17 complete matches and one partial match. Therefore, 59 percent of the 29 NAEP specifications were completely aligned with the Atlanta/Georgia standards.**

- Boston matched 15 (52 percent) of the 29 subscale specifications, with 14 complete matches and one partial match. Therefore, 48 percent of the 29 NAEP specifications were completely aligned with the Boston standards.

- Charlotte/North Carolina matched 15 (52 percent) of the 29 subscale specifications, with 14 complete matches and one partial match. Therefore, 48 percent of the 29 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland/Ohio matched 18 (62 percent) of the 29 subscale specifications, with 14 complete and four partial matches. Therefore, 48 percent of the 29 NAEP specifications were completely aligned with Cleveland/Ohio standards.**

In the subscale on *reading for information* in eighth grade, there were 27 NAEP specifications.

- **Atlanta/Georgia matched 10 (37 percent) of the 27 subscale specifications, with nine complete matches and one partial match. Therefore, 33 percent of the 27 NAEP specifications were completely aligned with Atlanta/Georgia standards**.

- Boston matched 11 (41 percent) of the 27 subscale specifications, with nine complete and two partial matches. Therefore, 33 percent of the 27 NAEP specifications were completely aligned with Boston standards.

- Charlotte/North Carolina matched 18 (67 percent) of the 27 subscale specifications, with 17 complete matches and one partial match. Therefore, 63 percent of the 27 NAEP specifications were completely aligned with Charlotte/North Carolina standards.

- **Cleveland/Ohio matched 19 (70 percent) of the 27 subscale specifications, with 18 complete matches and one partial match. Therefore, 67 percent of the 27 NAEP specifications were completely aligned with Cleveland/Ohio standards.**

In the subscale on *reading to perform a task* in eighth grade, there were 22 NAEP specifications.

- **Atlanta/Georgia matched five (23 percent) of the 22 subscale specifications, with all five being complete matches. Therefore, 23 percent of the 22 NAEP specifications were completely aligned with Atlanta/Georgia standards.**

- Boston matched six (27 percent) of the 22 subscale specifications, with four complete and two partial matches. Therefore, only 18 percent of the 22 NAEP specifications were completely aligned with Boston standards.

- Charlotte/North Carolina matched 13 (59 percent) of the 22 subscale specifications, with 12 complete matches and one partial match. Therefore, 55 percent of the 22 NAEP specifications were completely aligned with Charlotte/North Carolina standards.

- **Cleveland/Ohio matched 14 (64 percent) of the 22 subscale specifications, with 12 complete matches and two partial matches. Therefore, 55 percent of the 22 NAEP specifications were completely aligned with Cleveland/Ohio standards.**

In addition, all districts had notable variations in the degree of content match across the 12 *aspects* of eighth-grade reading. Each subscale had four aspects. (See table 4a.13.) Each aspect had between five and eight NAEP specifications in each subscale.

- **Atlanta/Georgia's highest level of alignment with NAEP specifications in reading for literary experience was in the aspect of "developing an interpretation," matching six of seven (86 percent) specifications, all complete matches. In the subscale related to reading for information, the highest level of match was in "forming a general understanding" matching four of six (67 percent) specifications, with three complete and one partial match. In the subscale on reading to perform a task, the highest match was in "developing an interpretation," matching three of six (50 percent) specifications—all complete matches. The lowest level of alignment in Atlanta/Georgia across all subscales was in "making reader/text connections," where matches (either complete or partial) ranged from 0 percent to 25 percent**.

- Boston's highest level of alignment with NAEP specifications in reading for literary experience was in the aspect of "examining content and structure," matching six of seven specifications (86 percent), with five complete matches and one partial match. In the subscale related to reading for information, the highest match was in "examining content and structure," matching only four of eight (50 percent) specifications, with three complete matches and one partial. In the subscale on reading to perform a task, the highest match related to "developing an interpretation," matching three of six (50 percent) specifications, with two complete matches and one partial. There were variations across the three subscales as to which aspect showed the least alignment to NAEP, ranging from 0 percent on "making reader/text connections" in the reading to perform a task subscale to 33 percent on both "forming a general understanding" and "making reader/text connections" in the reading for information subscale.

- Charlotte/North Carolina's highest levels of match in reading for literary experience were in two aspects: "developing an interpretation" and "examining content and structure," both with five matches out of seven specifications (71 percent). In "developing an interpretation," all matches were complete. In "examining content and structure," four matches were complete and one was partial. In the subscale of reading for information, the highest match was in "examining content and structure," matching six of eight specifications (75 percent)—all six being complete matches.

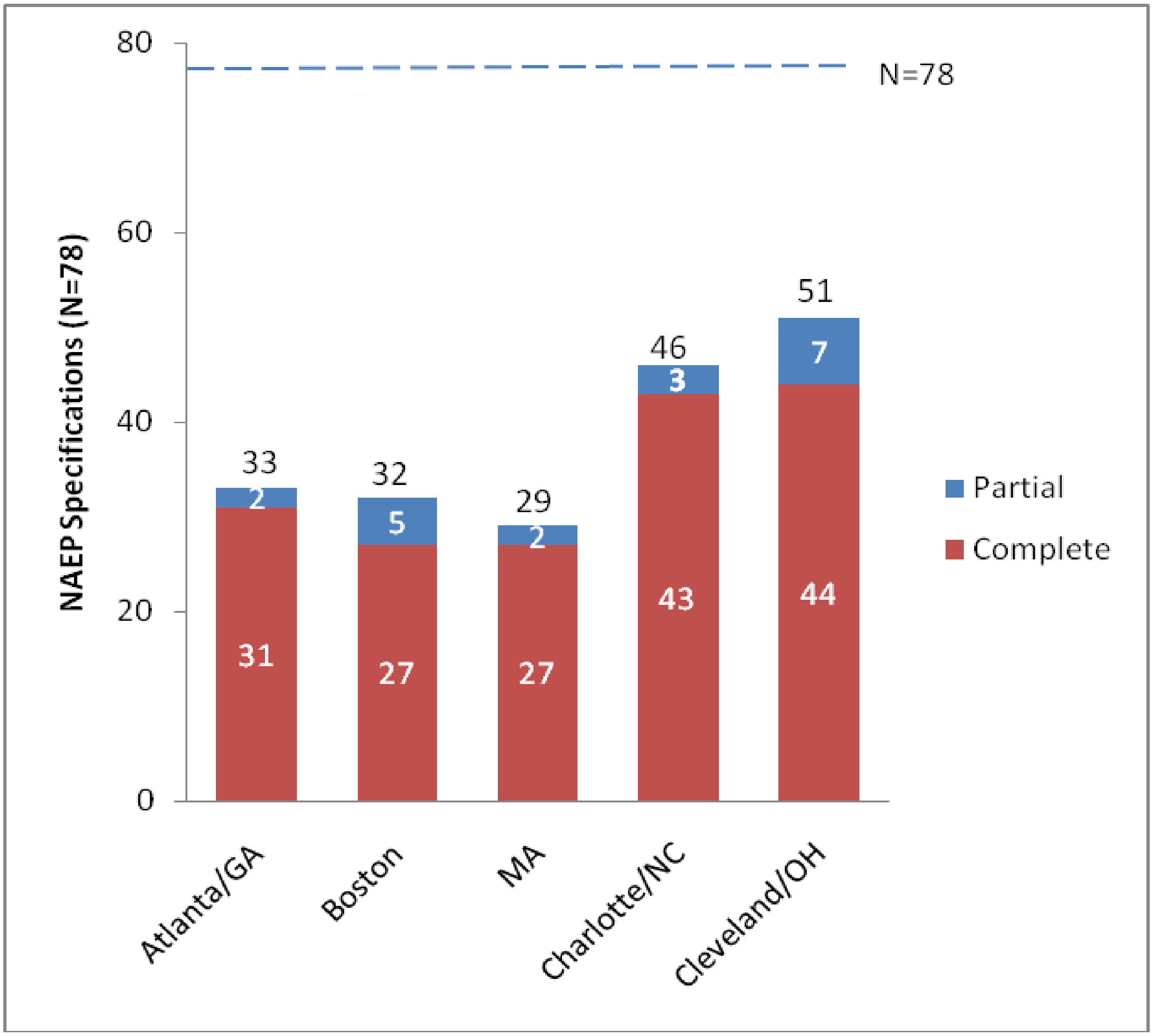
In the subscale on reading to perform a task, the highest matches related to "developing an interpretation" and "examining content and structure"—each of which was aligned on four of six (67 percent) specifications. In "developing an interpretation," three of the four matches were complete; in "examining for content and structure, all four matches were complete. "Making reader/text connections" showed the lowest alignment in two subscales—in reading for reading for information (50 percent) and in reading to perform a task (40 percent). In reading for literary experience, the lowest aspect was "forming a general understanding" (29 percent).

- **Cleveland/Ohio had the highest overall level of matching by aspect in grade eight reading with six aspects having 80 percent or more complete or partial matches. The highest match in reading for literary experience was in the aspect of "developing an interpretation," with all seven of the specifications matched (five complete matches and two partial). In the subscale of reading for information, the highest matches were in "forming a general understanding" and "developing an interpretation." Both aspects matched 100 percent either completely or partially. All matches on "forming a general understanding" were complete, while there was one partial match on "developing an interpretation." In the subscale on reading to perform a task, the highest match related to "developing an interpretation," which was aligned on five of the six (83 percent) specifications, with four complete matches and one partial. The lowest level of alignment in Cleveland/Ohio across all subscales related to "making reader/text connections," where matches (either complete or partial) ranged from 17 to 40 percent.**

In general, the alignment in reading at the eighth-grade level was lower than at fourth grade. In eighth grade, the greatest alignment was in the reading for literary experience subscale; the lowest level of alignment was in reading to perform a task. Across the three subscales, "developing an interpretation" *tended* to have the highest level of alignment, while "making reader/text connections" showed the lowest alignment—the same as in fourth grade.

Finally, Cleveland/Ohio showed the highest overall level of complete and partial alignment (65 percent) between the NAEP specifications and its standards in eighth-grade reading. In contrast, Atlanta, Boston, and Charlotte showed similarly moderate or low alignment overall, ranging from 41 percent (Boston) to 59 percent (Charlotte).

Figure 4a.2 Number of complete and partial matches with NAEP grade 8 reading specifications, by selected districts (*N* of NAEP specifications = 78), 2007*



*33 (42 percent) of Atlanta's grade 8 reading standards matched NAEP's 78 reading specifications either completely or partially; 32 (41 percent) of Boston's grade 8 reading standards matched NAEP's 78 reading specifications either completely or partially; 29 (37 percent) of Massachusetts's grade 8 reading standards matched NAEP's 78 reading specifications either completely or partially; 46 (59 percent) of Charlotte's grade 8 reading standards matched NAEP's 78 reading specifications either completely or partially; and 51 (65 percent) of Cleveland's grade 8 reading standards matched NAEP's 78 reading specifications either completely or partially.

Table 4a.13 Degree of match with NAEP grade 8 reading specifications/expectations/indicators, by subscale, aspect, and district, 2007

	Number of Specifications in NAEP, by aspect																									
Subscale	Reading for a Literary Experience								Reading for Information								Reading to Perform a Task									
Aspect:	Forming a General Understand-ing		Developing an Inter-pretation		Making Reader/Text Connections		Examining Content and Structure		Forming a General Understand-ing		Developing an Inter-pretation		Making Reader/Text Connections		Examining Content and Structure		Forming a General Understand-ing		Developing an Inter-pretation		Making Reader/Text Connections		Examining Content and Structure		Total	
	7		7		8		7		6		7		6		8		5		6		5		6		78	
Atlanta/ GA	5	71%	6	86%	2	25%	5	71%	4	67%	4	57%	0	0%	2	25%	1	20%	3	50%	0	0%	1	17 %	33	42%
	C=4	P=1	C= 6	P= 0	C=2	P=0	C=5	P=0	C=3	P=1	C=4	P=0	C=0	P=0	C=2	P=0	C=1	P=0	C=3	P=0	C=0	P=0	C=1	P= 0	C= 31	P= 2
Boston	1	14%	5	71%	3	38%	6	86%	2	33%	3	43%	2	33%	4	50%	1	20%	3	50%	0	0%	2	33 %	32	41%
	C=1	P=0	C= 5	P= 0	C=3	P=0	C=5	P=1	C=2	P=0	C=2	P=1	C=2	P=0	C=3	P=1	C=1	P=0	C=2	P=1	C=0	P=0	C=1	P= 1	C= 27	P=5
MA	6	86%	6	86%	2	25%	5	71%	2	33%	3	43%	0	0%	1	13%	2	40%	2	33%	0	0%	0	0%	29	37%
	C=6	P=0	C= 6	P=0	C=2	P=0	C=5	P=0	C=1	P=1	C=2	P=1	C=0	P=0	C=1	P=0	C=2	P=0	C=2	P=0	C=0	P=0	C=0	P= 0	C= 27	P=2
Char-lotte/ NC	2	29%	5	71%	3	38%	5	71%	4	67%	5	71%	3	50%	6	75%	3	60%	4	67%	2	40%	4	67 %	46	59%
	C=2	P=0	C= 5	P=0	C=3	P=0	C=4	P=1	C=4	P=0	C=5	P=0	C=2	P=1	C=6	P=0	C=3	P=0	C=3	P=1	C=2	P=0	C=4	P= 0	C= 43	P=3
Cleve-land/ OH	3	43%	7	100 %	2	25%	6	86%	6	100 %	7	100 %	1	17%	5	63%	4	80%	5	83%	2	40%	3	50 %	51	65%
	C=3	P=0	C= 5	P=2	C=1	P=1	C=5	P=1	C=6	P=0	C=6	P=1	C=1	P=0	C=5	P=0	C=3	P=1	C=4	P=1	C=2	P=0	C=3	P= 0	C= 44	P=7

Note: How to read this table. Example: There are seven specifications in the context *Reading for a Literary Experience* under the aspect "Forming a General Understanding." Atlanta standards matched the content of five of those seven specifications or 71 percent. Four of the matches were complete content matches and one was a partial content match.
C=complete match; P=partial match

Table 4a.14 summarizes the degree of complete alignment for each subscale and aspect in each of the selected jurisdictions. Matches were deemed to be high when at least 80 percent of NAEP specifications were completely aligned with district/state objectives. Matches were deemed low when 50 percent or fewer of NAEP specifications were completely matched by that district/state's objectives for that subscale and aspect. Only six of the 60 cells in table 4a.12 indicated high alignment, while 36 of the 60 were low. Consequently, complete alignments between NAEP specifications and local/state standards in eighth-grade reading again could be characterized as low to moderate.

Table 4a.14 Degree of complete match of NAEP subscales with district/state standards in grade 8 reading, by subscale, aspect, and district, 2007*

Subscale/ Context	Aspect of Reading	Atlanta/GA	Boston	MA	Charlotte/ NC	Cleveland/ OH
Reading for a Literary Experience	Forming a General Understanding	Moderate	Low	High	Low	Low
	Developing Interpretation	High	Moderate	High	Moderate	Moderate
	Making Reader/ Text Connections	Low	Low	Low	Low	Low
	Examining Content and Structure	Moderate	Moderate	Moderate	Moderate	Moderate
Reading for Information	Forming a General Understanding	Low	Low	Low	Moderate	High
	Developing Interpretation	Moderate	Low	Low	Moderate	High
	Making Reader/ Text Connections	Low	Low	Low	Low	Low
	Examining Content and Structure	Low	Low	Low	Moderate	Moderate
Reading to Perform a Task	Forming a General Understanding	Low	Low	Low	Moderate	Moderate
	Developing Interpretation	Moderate	Low	Low	Low	High
	Making Reader/ Text Connections	Low	Low	Low	Low	Low
	Examining Content and Structure	Low	Low	Low	Moderate	Low

***** High (80 percent or more) and low (50 percent or less)

Degree of Match in Cognitive Demand

In addition to determining the degree of content match between local/state standards and NAEP specifications, the research team examined how well those standards that matched completely corresponded in their cognitive demand or complexity to NAEP specifications. (See chapter 3 and appendices C and D for a detailed description of the methodology.) This process entailed examining the wording of the district/state standards and the NAEP specifications that matched completely to determine the cognitive demand or rigor in each statement and comparing the results.

When coding state and district standards, we focused on the verb in the standards showing the highest level of cognitive demand. If the vocabulary in a standard included "identifying, summarizing, and *analyzing* text," it was coded as **H** for high demand because of the rigor implied by the verb "analyzing." Although available documents such as curriculum guides were consulted, there was no way to determine how these standards translated into classroom instruction or which of the three verbs—identify, summarize, or analyze—actually received the most emphasis in teachers' instruction. This caution should frame and inform our discussion of the alignment of standards based on cognitive demand.[9]

One should also note that NAEP had a higher percentage of specifications with **low** cognitive demand in reading than did the district/state standards we reviewed in both grade four and grade eight. This may be because NAEP has a clear progression or sequence from relatively low-level specifications on two aspects--"forming a general understanding" and "developing an interpretation"--to relatively higher cognitive levels found in "making reader/text connections" and "examining content and structure" (National Assessment Governing Board, 2002). In grade four, 60 percent of the reading questions on NAEP were aligned to the two lower-level aspects, and at grade eight, 55 percent of the questions were aligned to those lower level aspects.[10] The assessment is structured this way to better measure what students actually know at the lower ends of the scale, while some states do not sequence the difficulty of their standards in the same way.

Tables 4a.15 and 4a.16 show the level of complete content match discussed in the previous section, along with the number and percentage of state and local standards that were classified as low, moderate, or high on cognitive demand in fourth and eighth-grade reading. Only those standards that matched NAEP specifications completely are included in the analysis, because it was nearly impossible to determine the nature and degree of the partial matches. This analysis of complete matches, however, gives the reader a sense of the rigor or complexity of state and local standards, but only for the portion of standards that completely match with NAEP.

Omitted from the cognitive demand codes were all standards that did not correspond completely to NAEP. When interpreting the percentage of state and local standards whose *cognitive demand* was moderate or high, one should note that the overall *content* match with NAEP was low to moderate.

The data in the tables indicate that the level of cognitive demand in the state and district standards appeared to be more closely aligned with NAEP reading in grade four than in grade eight. In grade four, most district/state standards and NAEP specifications had **moderate** cognitive demand. However, the four districts/states had standards with a greater percentage of **high** cognitive demand than the completely matched NAEP specifications. For instance, 38 percent of Atlanta's completely matched local/state reading standards were coded as high cognitive demand, while NAEP had 15 percent. On the other hand, 19 percent of Cleveland/Ohio's matched standards were coded as high cognitive demand.

In grade eight, we found that a higher percentage of completely matching district/state standards had **high** cognitive demand than did the NAEP specifications. For instance, 52 percent of Atlanta's completely matched local/state standards were coded as high cognitive demand, while NAEP had 17 percent. Cleveland/Ohio had 50 percent of its matched items coded at the high cognitive demand level. Charlotte had 79 percent.

The same two tables offer another way to look at cognitive demand. The tables present a weighted total and a weighted average for cognitive demand in each district/state. The weighted total is based on a system that awards one point for low, two points for moderate, and three points for high cognitive

[9] The readers should keep this caution in mind throughout the study because there are limitations to comparing specifications behind NAEP and the state standards that are intended to drive instruction. This study did not visit classrooms or check teacher assignments. In addition, state standards in areas like reading can be highly inclusive, e.g., "finding and evaluating a main idea or premise." It was impossible for the study to determine how thoroughly teachers implement a standard like this.

[10] Source: National Assessment Governing Board, 2002

demand. The weighted average is calculated in each jurisdiction by dividing the weighted total by the total number of complete matches with NAEP.

The analysis suggests that the degree of cognitive demand in grade four reading was similar to NAEP's weighted average (baseline of 1.9). Massachusetts and Charlotte/North Carolina had the highest cognitive demand (2.4) of all selected jurisdictions in grade 4 reading. At grade eight, the weighted averages among all jurisdictions were higher than in grade four and also similar to the NAEP baseline of 1.9 in all jurisdictions. Charlotte/North Carolina again had the highest cognitive demand level (2.8) of all selected jurisdictions.

Table 4a.15 Degree of match in cognitive demand for specifications with complete alignment on NAEP grade 4 reading, by district, 2007

	NAEP		Atlanta/ GA		Boston		MA		Charlotte/ NC		Cleveland/ OH/	
% of Complete Content Match	100%		39%		39%		35%		67%		39%	
Cognitive Levels												
Low	12	22%	5	24%	4	19%	1	5%	0	0%	4	19%
Moderate	34	63%	8	38%	10	48%	9	47%	21	58%	13	62%
High	8	15%	8	38%	7	33%	9	47%	15	42%	4	19%
Total	54	100%	21	100%	21	100%	19	100%	36	100%	21	100%
Weighted Total	104		45		45		46		87		42	
Weighted Mean		1.9*		2.1		2.1		2.4		2.4		2.0

* Number represents the balance among NAEP standards that were determined to be high, moderate, or low cognitive demand. 1=low cognitive demand, 2=moderate cognitive demand, and 3=high cognitive demand.

Table 4a.16 Degree of match in cognitive demand for specifications with complete alignment on NAEP grade 8 reading, by district, 2007

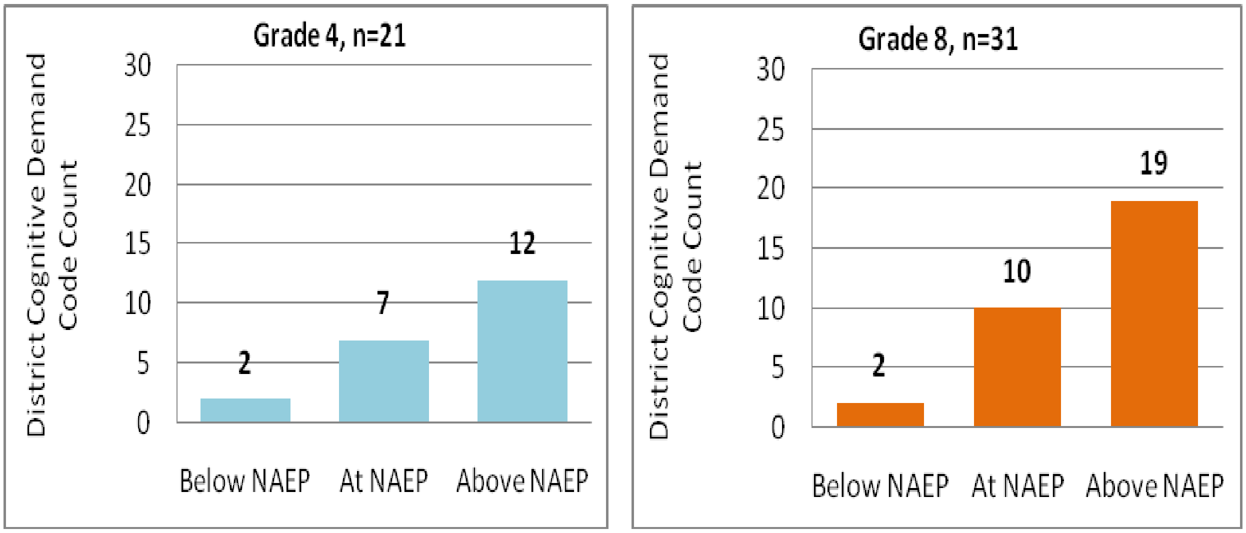
	NAEP		Atlanta/ GA		Boston		MA		Charlotte/ NC		Cleveland/ OH	
% of Complete Content Match	100%		40%		35%		35%		55%		56%	
Cognitive Levels												
Low	18	23%	3	10%	2	7%	1	4%	0	0%	8	18%
Moderate	47	60%	12	39%	9	33%	8	30%	9	21%	14	32%
High	13	17%	16	52%	16	59%	18	67%	34	79%	22	50%
Total	78	100%	31	100%	27	100 %	27	100%	43	100%	44	100%
Weighted Total	151		75		68		71		120		102	
Weighted Mean		1.9*		2.4		2.5		2.6		2.8		2.3

* Number represents the balance among NAEP standards that were determined to be high, moderate, or low cognitive demand. 1=low cognitive demand, 2=moderate cognitive demand, and 3=high cognitive demand.

Another way to capture the degree of alignment in cognitive demand between NAEP and the local/state standards is to directly compare each completely matching district/state standards with its corresponding

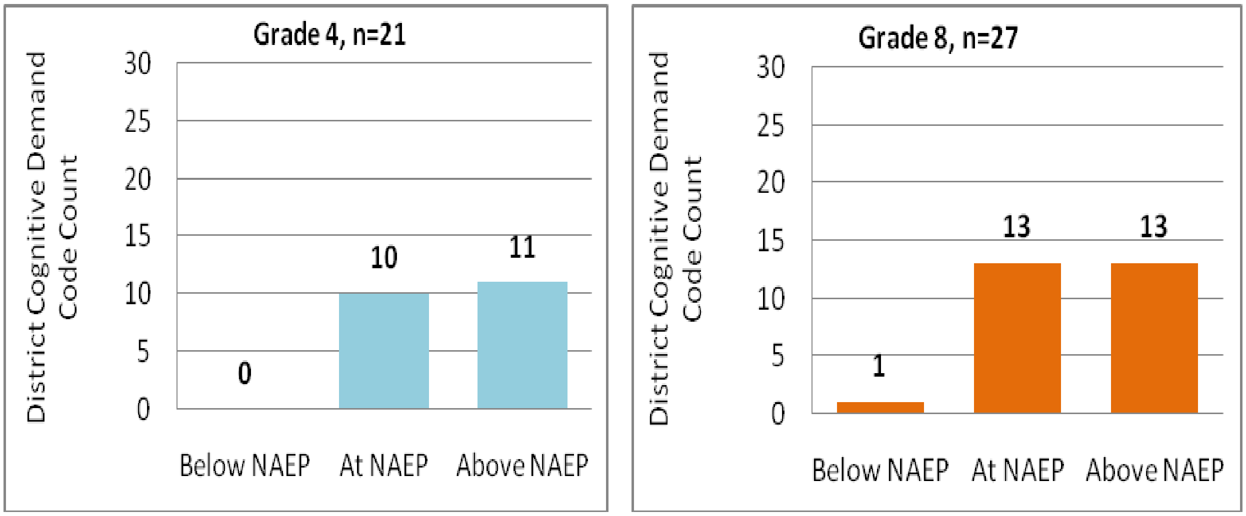
NAEP specification. Figures 4a.3 through 4a.12 show these data for grades four and eight in Atlanta, Boston, Charlotte, Cleveland, and Massachusetts, respectively. These graphs show that the cognitive demand codes of the completely matched standards generally were similar to NAEP specifications. In general, the analysis suggests that the matched standards in these jurisdictions were at least as high as the cognitive demand level in NAEP.

Figures 4a.3 and 4a.4 *Atlanta's c*omplete matches at grades 4 and 8 reading in cognitive demand compared to NAEP, 2007*



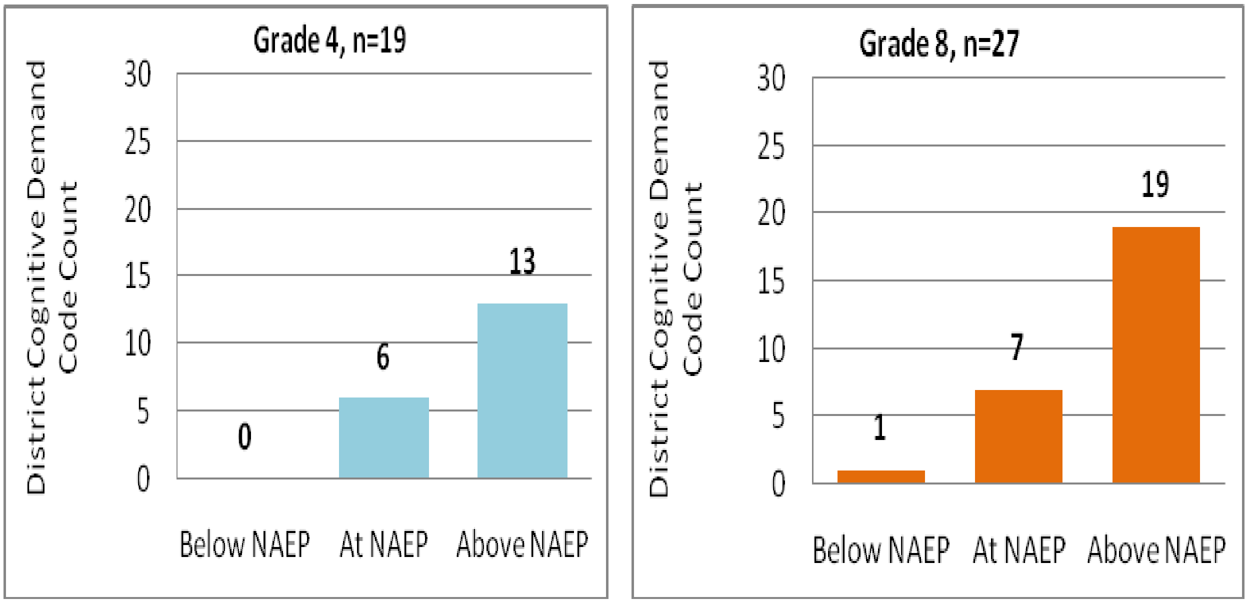
* 21 of Atlanta's grade 4 standards completely matched the 54 NAEP reading specifications (39 percent). Two of those 21 completely matched standards had a cognitive demand level below NAEP, seven were at the NAEP level, and 12 were above NAEP. Similarly, 31 of Atlanta's eighth grade standards completely matched the 78 NAEP reading specifications (40 percent). Two of those 31 completely matched standards had a cognitive demand level below NAEP, 10 were at the NAEP level, and 19 were above NAEP.

Figures 4a.5 and 4a.6 *Boston's* complete matches at grades 4 and 8 reading in cognitive demand compared to NAEP, 2007*



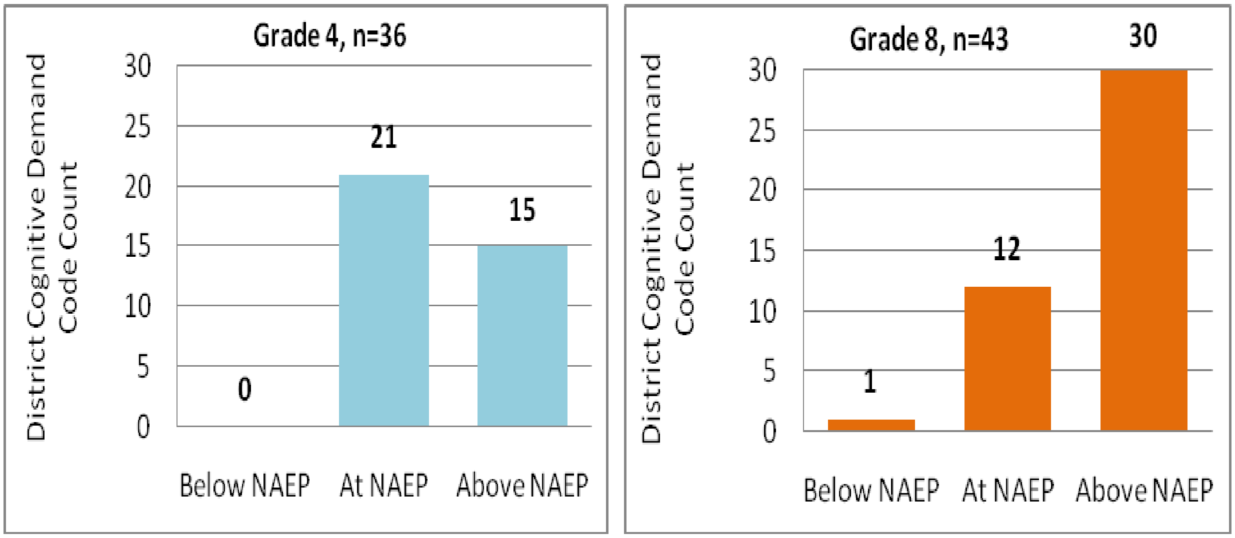
* 21 of Boston's grade 4 standards completely matched the 54 NAEP specifications (39 percent). None of those 21 completely matched standards had a cognitive demand level below NAEP, 10 were at the NAEP level, and 11 were above NAEP. Similarly, 27 of Boston's eighth grade standards completely matched the 78 NAEP reading specifications (35 percent). One of those 27 completely matched standards had a cognitive demand level below NAEP, 13 were at the NAEP level, and 13 were above NAEP.

Figures 4a.7 and 4a.8 *Massachusetts's* complete matches at grades 4 and 8 reading in cognitive demand compared to NAEP, 2007*



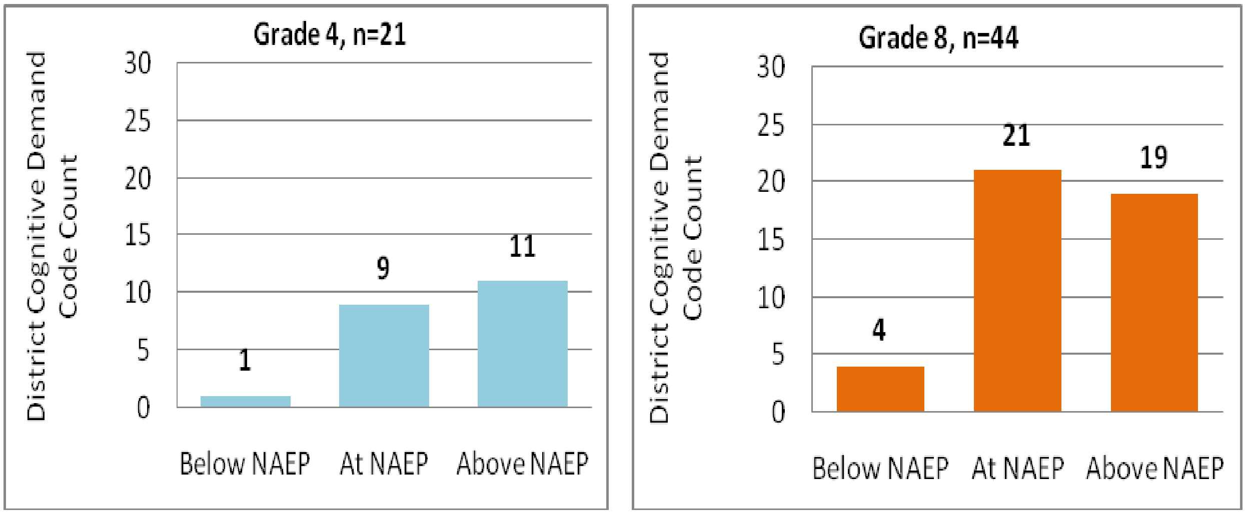
* 19 of Massachusetts's grade 4 standards completely matched the 54 NAEP specifications (35 percent). None of those 19 completely matched standards had a cognitive demand level below NAEP, six were at the NAEP level, and 13 were above NAEP. Similarly, 27 of Massachusetts's eighth grade standards completely matched the 78 NAEP reading specifications (35 percent). One of those 27 completely matched standards had a cognitive demand level below NAEP, seven were at the NAEP level, and 19 were above NAEP.

Figures 4a.9 and 4a.10 *Charlotte's* complete matches at grades 4 and 8 reading in cognitive demand compared to NAEP, 2007*



*****36 of Charlotte's grade 4 standards completely matched the 54 NAEP reading specifications (67 percent). None of those 36 completely matched standards had a cognitive demand level below NAEP, 21 were at the NAEP level, and 15 were above NAEP. Similarly, 43 of Charlotte's eighth grade standards completely matched the 78 NAEP reading specifications (55 percent). One of those 43 completely matched standards had a cognitive demand level below NAEP, 12 were at the NAEP level, and 30 were above NAEP.

Figures 4a.11 and 4a.12 *Cleveland's* complete matches at grades 4 and 8 reading in cognitive demand compared to NAEP, 2007*



* 21 of Cleveland's grade 4 standards completely matched the 54 NAEP reading specifications (39 percent). One of those 21 completely matched standards had a cognitive demand level below NAEP, nine were at the NAEP level, and 11 were above NAEP. Similarly, 44 of Charlotte's eighth grade standards completely matched the 78 NAEP reading specifications (56 percent). Four of those 44 completely matched standards had a cognitive demand level below NAEP, 21 were at the NAEP level, and 19 were above NAEP.

NAEP vs. State Tests – Item Type and Passage Length

We also looked at the NAEP reading assessment and at the individual state tests in reading in the four jurisdictions to further understand how students were tested on the state exams and how those testing experiences might affect NAEP performance. Student performance on any reading test usually reflects more than the content and cognitive demand of the standards and the specifications from which the test is constructed.

Other factors that may affect performance are students' familiarity with the content of the passages they read, passage and test length, and the concept and vocabulary loads of the passages. These factors are relevant to determining the overall cognitive demand of reading tests, along with the number and format of the items students must answer and how they must respond. Information on these factors on NAEP reading assessments and state reading tests provides further context for interpreting the comparisons of the cognitive demands between NAEP and the state and district standards.

Table 4a.17 contrasts several features of the NAEP assessments that were developed from the 2003 *Framework* and the state reading assessments that were used between 2003 and 2007. Information about the state reading assessments was drawn from postings about the tests on state education agency Websites and was purely descriptive. The extent of information available on the Websites varied. In some cases, for example, the content in the table was estimated by counting the number of words in released passages and averaging them.

The reader will note that NAEP is comprised of at least 50 percent constructed items at each grade, while tests in Georgia and North Carolina have none. In addition, NAEP uses reading passages of between 200 and 800 words in fourth grade, while Georgia never exceeds 400 words and Ohio uses between 300 and 700 words per passage.

Table 4a.17 Comparison of characteristics of NAEP and state reading assessments in grades 4 and 8, by state, 2007

Factor	NAEP	Georgia	Massachusetts	North Carolina	Ohio
Focus	Reading comprehension	Reading and English language arts	Reading and Composition	Reading	Reading
Balance of Item Types					
Multiple-choice	Fewer than 50% at each grade	100% of test	Majority of test*	100% of test	Majority of test*
Constructed Response	At least 50% at each grade	No	Yes (minimal)	No	Yes (minimal)
Balance of Text Types					
Literary	At least 55% of passages	Could not determine**	~ 60%	Could not determine**	~ 40% - 50%
Informational/ Procedural	At least 45%	Could not determine**	~ 40%	Could not determine**	~ 40 – 50%
Passage Length, in words	200 – 800 at fourth grade 700 – 1000 at eighth grade	~ 200 – 400*	640 – 885 at fourth grade* 440 – 1300 at eighth grade*	Could not determine**	300 – 700 at fourth grade 450 – 1000 at eighth grade

Notes: Information about the 2003 NAEP Reading Assessment was taken from the *Reading Framework.* Information on state reading assessments at grades 4 and 8 was taken from state education agencies' public Websites. In some cases, information on the assessments for years of interest in this study was limited or not available.
Entries marked by an asterisk (*) were estimated using material on state education agencies' public Websites.
Information for entries marked by a double asterisk (**) could not be found on state educational agencies public Websites.
Information about the Ohio tests was available from AIR's Assessment Program, which develops the tests.

One should be cautious about drawing conclusions about the information in table 4a.17 that compares NAEP and the state reading tests, however. The cognitive demand on NAEP reading may best be judged by examining actual "item sets," that is, the passages students must read and their accompanying multiple-choice and constructed-response items. NAEP reading passages are usually followed by eight to 10 questions or items. The cognitive demand of the proxy specifications, which were used for comparative purposes, provides only partial information about how demanding the assessments actually were. Longer passages were often more cognitively demanding to read because more information was included and because questions often required test takers to make connections across longer stretches of text.

In addition, constructed-response items can be more demanding than multiple-choice items because test takers may be required to synthesize or analyze information or stand back from what they have read and express their own evaluations or judgments in their writing. This is especially true for extended constructed response items that require students to write three or more connected sentences that are scored on a four-point scale.

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 four and eight and state and district standards in Atlanta, Boston, Charlotte, and Cleveland.

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

Generally, the greatest degree of complete and partial alignment was in reading for literary experience in grade four. In grade eight, 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. In addition, the analysis indicated that making "reader/text connections" was the least aligned aspect across all reading subscales in both grades.

Finally, there is 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 4a.18 and 4a.19.)

Table 4a.18 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 that saw a significant increase in reading, yet it had the same percentage of complete content matches with NAEP as Boston and Cleveland (39 percent), two districts that saw no significant increase in NAEP reading scale 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 average.

In eighth grade, Atlanta and Cleveland saw significant increases in reading scale 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 increases. 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 average.

Table 4a.19 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	55%	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

Site Visits and Linkages to Reading Results

The research team conducted site visits to the four selected districts to examine instructional and organizational reforms that took place during the study period that might help explain trends in NAEP reading scale scores. A description of the methodology and the protocols used during these site visits is included in chapter 3 and in appendix E. Teachers, staff, and community members were interviewed, and instructional materials used in the four districts during the study period were reviewed. (See appendix I.) This section examines what was learned on these site visits about the instructional programs in each city in order to inform the reading results, particularly the subscale results, presented in this chapter. A broader synthesis of the site visits is presented in the next chapter.

In discussing reading-related findings from the site visits, we paid particular attention to Atlanta and Cleveland, because Atlanta was found to have significant and consistent gains in reading on NAEP, while Cleveland had weaker and less consistent improvements. We also devote some attention to Boston's reading initiative in this section and why it may not have produced the same kinds of gains that its math program did.

The data in this and the previous chapter indicated that Atlanta made statistically significant progress in reading on NAEP scale scores. Specifically, the data showed that Atlanta's fourth graders made significant gains on the NAEP composite reading score and on the subscale of reading for information between 2003 and 2007. Over that same period, eighth graders showed significant improvements on the composite score and in reading to perform a task. There was no significant progress in reading for literary experience in either grade or in reading for information at the eighth-grade level.

The information obtained during the Atlanta site visit helped us understand why these NAEP reading patterns existed in the district. As will be described in greater detail in chapter 5 and in the case study in appendix F, Atlanta pursued an aggressive and sustained set of literacy reforms since 2000, after Beverly Hall became superintendent of the district's schools. In general, according to the site visit team, the reforms included a consistent and forward-looking vision for improving literacy across the curriculum, highly targeted professional development, detailed use of data to inform instruction, and assertive technical assistance to schools and teachers through an unusual organizational structure involving a series of School Reform Teams (SRT) that placed a strong emphasis on direct, site-based service and support. Moreover, the district's gains at the subscale level, particularly in reading for information and reading to perform a task, might be due, in part, to the district's nearly universal Consortium on Reading Excellence (CORE) training program for staff, principals, literacy coaches, and teachers that was conducted between 2001 and 2006. This training placed a strong emphasis on instructional approaches such as questioning, use of graphic organizers, and comprehensive monitoring to help students access narrative and informational texts. Beginning around 2003, the district also put strong emphasis on reading and writing across the curriculum, which was designed to bolster reading skills in multiple content and informational areas, and may have helped Atlanta's fourth and eighth graders do about as well as large cities generally on NAEP's constructed-response items even though Georgia's state test in fourth grade consisted solely of multiple-choice items.

In addition, the district used a series of Comprehensive School Reform Demonstration (CSRD) models, like Success for All and Direct Instruction, that emphasized instruction in the foundational reading skills that are likely to be absent among students of a low-performing school district at the outset of reform. Not all CSRDs have demonstrated effectiveness in raising student achievement, but the ones used in Atlanta show strong evidence of effectiveness.

Finally, the district's accountability system, which held staff answerable for student improvement across multiple performance levels and created a sense of shared ownership for student performance, may have been partly responsible for the reading gains observed in Atlanta across the achievement distribution (i.e., across quintiles).

Boston, on the other hand, did not see the gains in reading that it saw in math. During the 2003 to 2007 study period, Boston used a Reading and Writing Workshop (RWW) model for its literacy program, but the approach appeared to have had an uneven rollout and implementation. The RWW model is grounded in a "literature" or a "learning by discovery" approach that advocates the teaching of reading in the context of literature rather than in the systematic and explicit way that was recommended by the National Reading Panel Report.[11] When reading instruction is dependent almost entirely on the use of literature in the way it is with RWW, it is possible that children with little or no exposure to or instruction in reading tasks represented by the specifications on the NAEP subscales of "reading for information" and "reading to perform a task" may not do well on the national assessment. In addition, mismatches in the standards might have exacerbated this situation.

For example, the NAEP frameworks at fourth and eighth grades do not match the Massachusetts standards in content or grade level on such areas as understanding text organization and structure, understanding literary devices, and deepening understanding of text by attention to vocabulary. Boston's performance in areas such as these might have been related to the fact that they may not have been explicitly included in the workshop approach. In general, this program required the district to build strong conceptual knowledge and instructional capacity, but the district apparently did not do this to the same extent it did in implementing its math program. (See math section.)

In Cleveland, fourth graders posted no composite reading gains or gains on any of the individual subscales between 2003 and 2007. Eighth graders, however, showed a statistically significant gain on the

[11] National Institutes of Child Health and Human Development (2000). *Report of the National Panel: Teaching children to read: An evidence-based assessment of the specific research literature on reading and its importance for reading instruction* (NIH Publication No. 004769). Washington, D.C.: Author.

reported composite score over the same period[12] but made no significant subscale gains (although subscale percentiles appeared to show some movement). Other than the higher degree of content and cognitive alignment between NAEP and the state standards in grade eight than in grade four, the research team could find no reason why the reported eighth-grade scores in Cleveland went up (although the full population estimates did not). There were no special programs or initiatives in place and no change in instructional practice that would have prompted the gains in reported scale scores, although it is possible that by grade eight, students reaped the cumulative benefit of having been exposed to the district's standards-based program for their entire school careers. The district, moreover, showed no particular strength in any of the reading subscales at either grade, except for a small tendency to do somewhat better in reading for literary experience.

Overall, Cleveland's instructional program was weak and highly fractured, as will be described in greater depth in the next chapter. Academic initiatives in the district, though present, appeared not to have been strong enough to produce NAEP gains.

[12] This improvement was observed using the reported NAEP scores. Using full population estimates, the change in reading composite scores from 2003 to 2007 at the eighth-grade level was not statistically significant.

Part 1. District Performance on NAEP Math Subscales

Content

The framework and specifications used to guide the NAEP math assessments between 2003 and 2007 are anchored in five broad areas of mathematical content in grades four and eight: number properties and operations ("number" for short), measurement, geometry, data analysis and probability, and algebra. Table 4b.1 shows the percentage of items in each content area by tested grade.

Table 4b.1 Percentage of items by mathematics content area and grade level, 2007

Content Area	Grade 4	Grade 8
Number Properties and Operations	40%	20%
Measurement	20%	15%
Geometry	15%	20%
Data Analysis and Probability	10%	15%
Algebra	15%	30%

In 2007, the grade four assessment consisted of an item pool of 164 items, and the grade eight assessment consisted of an item pool of 167 items. Both tests contained multiple-choice, short constructed-response, and extended constructed-response items. At each grade level, students had access to a calculator for about one-third of the items. In addition, NAEP balanced the items among low-, moderate-, and high-complexity questions or prompts. The full specifications that governed the NAEP TUDA assessments can be found in *Mathematics Framework for the 2005 National Assessment of Educational Progress*.[13]

Composite, Subscale and Item Analyses–Strengths and Weaknesses in Math

The overall performance between 2003 and 2009 of the TUDA districts in mathematics (and reading) was discussed in chapters 2 and 3. In this chapter, we look at district strengths and weaknesses in math by examining the performance of the four districts in 2003, 2005, and 2007 on each of the five content-area math subscales. As noted in chapter 3 (Methodology), the NAEP subscales are not all reported on the same metric, so average subscale scores or gains in average subscale scores are not directly comparable across subscales. Therefore, we examine subscale and item-level performance in a number of ways to estimate district strengths and weaknesses within each content area.

First, we show changes from 2003 to 2007 in subscale performance for each TUDA district in terms of effect size and statistical significance. Second, we provide the percentile rankings for each TUDA district on the distribution of average subscale scores for the national public school population. Third, we graphically display the percentile rankings of average subscale scores for each TUDA district after adjusting for student background characteristics. Finally, we provide item-level information about omission rates and the percentage of items correct by item type for each TUDA district. Taken together, the results in this section provide a rich source of information about the math performance, strengths, and weaknesses of the four selected TUDA districts.

[13] This document is published by the National Assessment Governing Board.

Changes in Subscale Performance from 2003 to 2007

As we reported in chapter 3, Atlanta, Boston, Charlotte, and Cleveland were selected for deeper study. Boston was selected for the significance and consistency of its math achievement gains, and Charlotte was picked for its overall math (and reading) performance. The analysis begins with an examination of changes in subscale performance between 2003 and 2007 in the four districts and compares them to subscale results for the large-city (LC) schools and the national public school samples. Table 4b.2 shows the results for fourth-grade math and table 4b.3 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.

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 four with effect sizes that were two to three times larger than those of both the large-city (LC) schools and the national sample. Charlotte saw a significant gain only in geometry and did not see significant 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.

Table 4b.2 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 4b.3 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.

In grade eight 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. Average scores in Charlotte did not change significantly in any of the five content areas between 2003 and 2007, but 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 grades in Atlanta and Boston, effect sizes on the composite measure and individual subscales were generally greater than those of either the LC or the national public school sample.

Percentile Measures by Subscale

In the next analyses, we made indirect, normative comparisons between subscales (within a district) by looking at the percentile (on the national public school sample) to which a given district's subscale average corresponds. Again, the purpose was to estimate district strengths and weaknesses in math. Tables 4b.4 through 4b.7 (for Atlanta, Boston, Charlotte, and Cleveland, respectively) show the percentiles to which each district's averages correspond in composite scores and subscales by year in grades four and eight. The tables also show changes (gain or loss) in percentile points between 2003 and 2007, although statistical tests of significance were not performed because of the indirect way percentiles measure performance.

Atlanta

As shown in table 4b.4, the average performance of Atlanta on the composite math measure and all math subscales at grade four was below the national public school median in 2003, 2005, and 2007. In grade four math, the average student in Atlanta was at the 28th percentile in 2007, but the effect size analysis indicated that the gain over the study period was significant. Fourth graders in 2007 scored at the 30th percentile in number, the 24th percentile in measurement, the 32nd percentile in geometry, and the 29th percentile in both data and algebra. The overall fourth-grade math performance in the district was tightly clustered by subscale around the 30th percentile, except for measurement—the lowest of the five subscales. The effect size analysis indicates that gains between 2003 and 2007 were seen in all subscales, except measurement.

Table 4b.4 *Atlanta's* average NAEP mathematics percentiles and changes in percentiles, by subscale and grade, 2003-2007 (National Public School median=50)

	Grade 4				Grade 8			
	Percentile of the mean scale score			Shift in percentile	Percentile of the mean scale score			Shift in percentile
	2003	2005	2007	2003–2007	2003	2005	2007	2003–2007
Composite	26	27	28	2	19	18	25	6
Number	29	30	30	1	21	19	25	4
Measurement	23	25	24	1	15	16	26	11
Geometry	24	27	32	7*	18	18	24	6
Data	27	31	29	3*	22	18	27	5
Algebra	26	30	29	4*	22	23	26	4

* Difference is due to rounding.

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

In grade eight math, Atlanta's average performance on the composite measure and all subscales was also below the national public school median in 2003, 2005, and 2007. In 2007, the average eighth grader in Atlanta was at the 25th percentile on the composite math measure, the 25th percentile in number, the 26th percentile in measurement, the 24th percentile in geometry, the 27th percentile in data, and the 26th percentile in algebra. The effect size analysis, however, showed that gains between 2003 and 2007 were significant on the composite measure and all subscales except geometry. Table 4b.4 also shows that the relative standing of Atlanta on the composite measure and most subscales appeared to be somewhat better at grade four than at grade eight.

Finally, the analysis of item responses in Atlanta found that fourth graders in the district were more able than their peers nationwide to answer such math questions as—

- Adding three fractions with like denominators (number)
- Multiplying two two-digit whole numbers (number)
- Circling numbers with a factor or four (number)
- Finding distance between centers of two adjacent squares (geometry)
- Discerning the pattern of fractions (algebra)

Conversely, fourth graders in Atlanta had more difficulty than their peers nationwide with such math items as—

- Designating the number represented on a line (number)
- Determining the temperature on a thermometer (measurement)
- Drawing data on a graph (data analysis)

Atlanta's eighth graders were more able than their peers nationwide to correctly answer such math items as—

- Identifying or writing a number with 6 in the hundreds place (number)
- Recognizing a unit of volume (measurement)
- Identifying perpendicular streets (geometry)
- Using an average to solve a problem (data analysis)
- Determining an equation given a point and a slope (algebra)

Conversely, Atlanta's eighth graders had more trouble than their peers nationwide with such math items as—

- Writing the sum of fractions as a decimal (number)
- Identifying the image of a figure after its rotation (geometry)
- Comparing consumer price indices over two years (data analysis)

Boston

Table 4b.5 shows the same kind of information for Boston. In grade four, Boston's average math composite and subscale percentiles were below the national public median in 2003, 2005, and 2007. In 2007, the average fourth-grade student in Boston was at the 39th percentile on the composite math measure, the 42nd percentile on the number subscale, the 38th percentile on measurement, the 40th percentile on geometry, the 34th percentile on data, and the 37th percentile on algebra. However, the district posted significant gains in effect sizes between 2003 and 2007 on the composite measure and all subscales. In fact, the effect size on the composite measure in Boston in grade four was more than twice as large as that for the national sample and the large-city sample.

In grade eight math, Boston's average performance on the composite math measure and on all subscales was also below the national public school median in 2003, 2005, and 2007. In 2007, the average eighth-grade student in Boston was at the 44th percentile on the composite measure in 2007, the 43rd percentile on the number subscale, the 44th percentile on measurement, the 45th percentile in geometry, the 43rd percentile on data, and the 46th percentile in algebra. The effect size analysis, however, indicated that gains between 2003 and 2007 on the composite measure and all subscales were significant. The eighth-grade composite effect size, in fact, was over twice as large as that of the national sample. Table 4b.5 also shows that, in contrast to Atlanta, the relative standing of Boston was slightly better at grade eight than at grade four on the composite math measure and all math subscales.

Table 4b.5 *Boston's* average NAEP mathematics percentiles and changes in percentiles, by subscale and grade, 2003-2007 (National Public School median=50)

	Grade 4				Grade 8			
	Percentile of the mean scale score			Shift in percentile	Percentile of the mean scale score			Shift in percentile
	2003	2005	2007	2003–2007	2003	2005	2007	2003–2007
Composite	30	37	39	9	33	40	44	11
Number	32	37	42	10	34	39	43	9
Measurement	29	41	38	9	33	42	44	11
Geometry	30	39	40	10	34	41	45	11
Data	29	39	34	5	34	41	43	10*
Algebra	30	37	37	6*	33	40	46	13

* Difference is due to rounding.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Mathematics Assessments.

Finally, the analysis of item responses in Boston found that fourth graders in the district were more able than their peers nationwide to answer such math questions as—

- Identifying factors of a number (number)
- Circling numbers with a factor of four (number)
- Assembling pieces to cover a designated shape (geometry)
- Choosing the best graph to describe data (data analysis)

Conversely, fourth graders in Boston had more difficulty than their peers nationwide with such math items as—

- Identifying missing information (number)
- Recognizing the best measurement unit (measurement)
- Identifying a shape from a fold (geometry)
- Reading and interpreting a line-graph (data analysis)

Among eighth graders, Boston students were more likely than their peers nationwide to correctly answer such math items as—

- Applying the Pythagorean theorem (geometry)

- Analyzing relationship between eating fish and test scores (data analysis)
- Determining an equation given a point and slope (algebra)
- Interpreting a line equation in context (algebra)

On the other hand, Boston's eighth graders had a more difficult time than their peers nationwide in correctly answering such math items as—

- Identifying a number with 6 in the hundredths place (number)
- Recognizing a unit of volume (measurement)
- Determining the number of vertices of a box (geometry)

Charlotte

Table 4b.6 shows the same information for Charlotte. In contrast to Atlanta and Boston, Charlotte's average performance in grade four on the composite math measure and all math subscales was at or above the national public school median in 2003, 2005, and 2007. In 2007 the average fourth-grade student in Charlotte was at the 54th percentile on the composite math measure, the 54th percentile on the number subscale, the 49th percentile on measurement, the 59th percentile on geometry, the 52nd percentile on data, and the 59th percentile in algebra. From 2003 to 2007, the only effect-size gain the district saw, however, was in geometry, where the subscale moved from the 54th percentile to the 59th percentile.

In grade eight math, Charlotte's average performance was near the national public-school median on the composite math measure and most math subscales in 2003, 2005, and 2007. In 2007, the average eighth-grade student in Charlotte was at the 51st percentile on the composite math measure in 2007, the 46th percentile on the number subscale, the 49th percentile in measurement, the 54th percentile on geometry, the 49th percentile in data, and the 54th percentile on algebra. The only significant gain between 2003 and 2007 was on the composite math measure. Relative to the national sample, Charlotte did somewhat better in fourth grade than in eighth grade.

Table 4b.6 *Charlotte's* average NAEP mathematics percentiles and changes in percentiles, by subscale and grade, 2003-2007 (National Public School median=50)

	Grade 4				Grade 8			
	Percentile of the mean scale score			Shift in percentile	Percentile of the mean scale score			Shift in percentile
	2003	2005	2007	2003–2007	2003	2005	2007	2003–2007
Composite	59	59	54	–5	51	52	51	1
Number	60	58	54	–6	46	46	46	#
Measurement	57	57	49	–8	46	49	49	3
Geometry	54	62	59	6*	55	55	54	–1
Data	60	56	52	–8	49	50	49	1*
Algebra	62	58	59	–3	56	56	54	–2

* Difference is due to rounding.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Mathematics Assessments.

Finally, the analysis of item responses in Charlotte found that fourth graders in the district were more able than their peers nationwide to answer such math items as—

- Multiplying two two-digit whole numbers (number)
- Dividing numbers (number)
- Knowing that a triangle can be formed using three or more points (geometry)
- Identifying expressions (algebra)

Conversely, fourth graders in Charlotte had more difficulty than their peers with such math items as—

- Adding three fractions with like denominators (number)
- Measuring the length of an object (measurement)
- Determining the distance around a triangle (measurement)

At eighth grade, Charlotte students were more able than their peers nationwide to answer such math items as—

- Determining coordinates (geometry)
- Finding the equation of a line (algebra)
- Recognizing the effect of signs on operations (algebra)

On the other hand, Charlotte eighth graders had more difficulty with such math items as—

- Measuring an angle (geometry)
- Comparing the areas of two shapes (measurement)
- Recognizing a unit of volume (measurement)

Cleveland

Finally, table 4b.7 shows the same information for Cleveland. At both grade four and grade eight, Cleveland's average performance on the composite math measure and all subscales was below the national public school median in 2003, 2005, and 2007. In 2007, the average fourth-grade student in Cleveland was at the 20th percentile in 2007 on the composite math measure, the 20th percentile on both the number subscale and measurement, the 22nd percentile on geometry, the 23rd percentile on data, and the 21st percentile on algebra. There were no significant effect-size gains on the composite measure or subscales.

Table 4b.7 *Cleveland's* average NAEP mathematics percentiles and changes in percentiles by subscale and grade, 2003-2007 (National Public School median=50)

	Grade 4				Grade 8			
	Percentile of the mean scale score			Shift in percentile	Percentile of the mean scale score			Shift in percentile
	2003	2005	2007	2003–2007	2003	2005	2007	2003–2007
Composite	25	26	20	–5	25	21	25	#
Number	25	25	20	–5	27	21	22	–5
Measurement	22	27	20	–2	27	24	26	–1
Geometry	29	29	22	–7	28	27	28	#
Data	28	31	23	–6*	25	23	24	–1
Algebra	27	30	21	–6	23	19	27	4

* Difference is due to rounding.

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

In 2007, the average eighth-grade student in Cleveland was at the 25th percentile on the composite math measure in 2007, the 22nd percentile on the number subscale, the 26th percentile on measurement, the 28th percentile on geometry, the 24th percentile on data, and the 27th percentile on algebra. Cleveland showed a positive effect-size change only on the algebra subscale between 2003 and 2007 in grade eight. In general, relative to the national sample, eighth graders in Cleveland did somewhat better than fourth graders.

Cleveland's fourth graders were more likely than their peers nationally to correctly answer such items as—

- Adding three fractions with like denominators (number)
- Working with units of liquid measurement (measurement)
- Recognizing completed shapes (geometry)

On the other hand, Cleveland's fourth graders had more difficulty than their peers nationwide with such math items as—

- Finding the height of a puppy (measurement)
- Knowing units of measurement (measurement)
- Drawing a pictograph (data analysis)
- Determining which scales would balance (algebra)

Cleveland's eighth graders were likely to do better than their peers nationwide on such math items as—

- Identifying which measurement instruments to use for a particular task (measurement)
- Using similar triangles (geometry)

Conversely, Cleveland's eighth graders had more difficulty than their peers nationwide with such math items as—

- Writing the sum of fractions as a decimal (number)
- Determining the total weight of two apples (measurement)
- Evaluating an algebraic expression (algebra)

***

In summary, between 2003 and 2007, Boston posted the largest overall gain in effect sizes on the composite measure and on all five math subscales in both grades four and eight, even though it remained below the national median in all subscales at both grades. Atlanta also scored below the national median on all subscales at both grades, but it showed substantial and positive shifts on composite scores and most subscales at both grades four and eight between 2003 and 2007.

Cleveland's performance was below the national median on all subscales at both grades and showed a gain in only one subscale in one grade (algebra in eighth grade). Among the four districts, only Charlotte's performance was close to or above the national public school median in all subscales at both grades. However, Charlotte showed no positive or negative shifts between 2003 and 2007 in either grade, except for the composite score in grade eight and the geometry subscale in grade four.

Percentile Measures by Subscale, Adjusted for Student Background Characteristics

Figures 4b.1 and 4b.2 show another way of capturing the relative performance of the districts. These "radar graphs" show the percentile (on the national public-school sample) to which a given district's adjusted subscale average corresponded on the 2007 NAEP math assessment. The averages were adjusted for the same demographic variables discussed in chapter 3.[14]

For example, the 32nd percentile for the algebra subscale in grade four would mean that 68 percent of students in the nation performed better in algebra than the average fourth grader in that district after adjusting for differences in background variables. Therefore, the closer the graph is to the center, the weaker the performance; the pentagon vertices farthest from the center signify relative subscale strength.

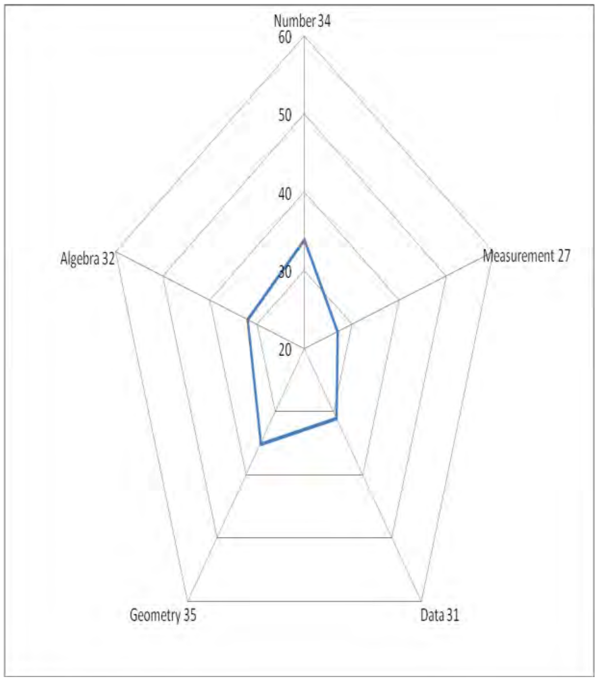
The figures (figures 4b.1 and 4b.2) show that the adjusted averages for all subscales in grade four were below the national median in all four districts, except Charlotte in geometry. Algebra was also a strength for Charlotte, compared to its other subscales. Geometry was a comparative strength for Atlanta, as were number, geometry, and measurement for Boston. In Cleveland, the percentiles in each of the five subscales were low and relatively close to one another.

In grade eight, the adjusted averages on all subscales were below the national median in all four districts. Charlotte was the only district where the adjusted averages were close to the national public school median. In Atlanta and Boston, the percentiles on the five subscales appeared to be relatively close to one another—ranging from the 31st to the 33rd percentile in Atlanta and from 42nd to 44th percentile in Boston. In Charlotte, algebra and geometry—both at the 48th percentile—appeared to be relative strengths for the district, compared with the other three subscales. Number was a relative weakness in Cleveland at the 28th percentile, while geometry and algebra were relative strengths (35th and 34th percentiles, respectively).

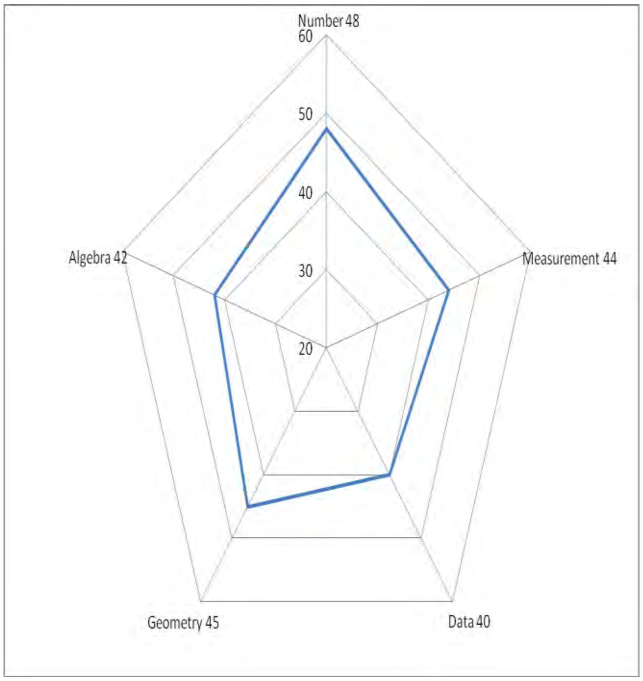
[14] Percentiles for all 11 TUDA districts are shown in appendix B, tables B.42-43. Results show that fourth graders appeared to be strongest in geometry, algebra, and number and weakest in measurement and data. At the eighth-grade level, urban students appeared to do better in geometry and algebra and less well in number.

Figure 4b.1 Percentile on national distribution to which each district's average adjusted NAEP grade 4 mathematics scores correspond, by district and subscale, 2007

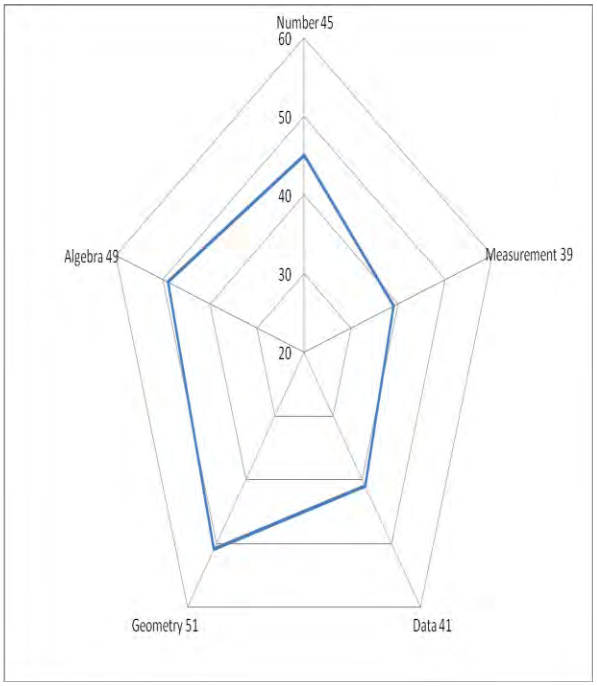
Atlanta



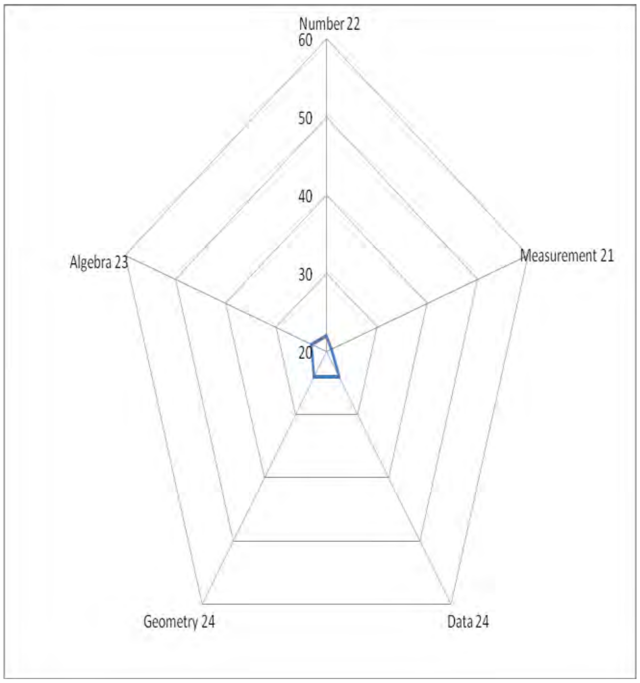
Boston



Charlotte



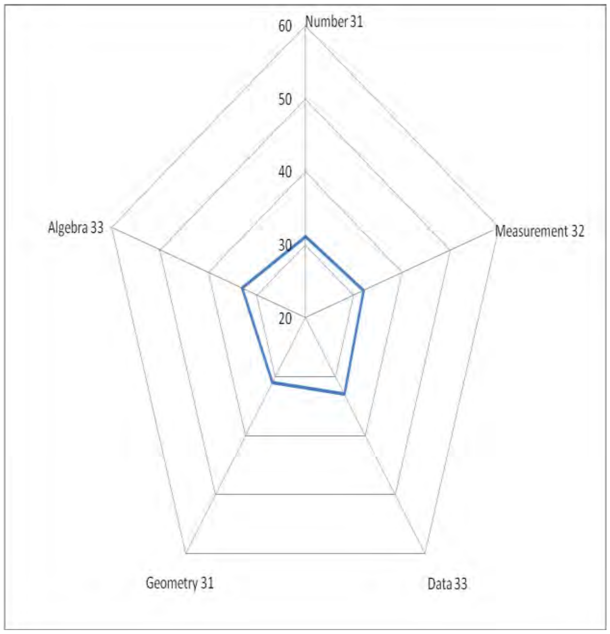
Cleveland



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

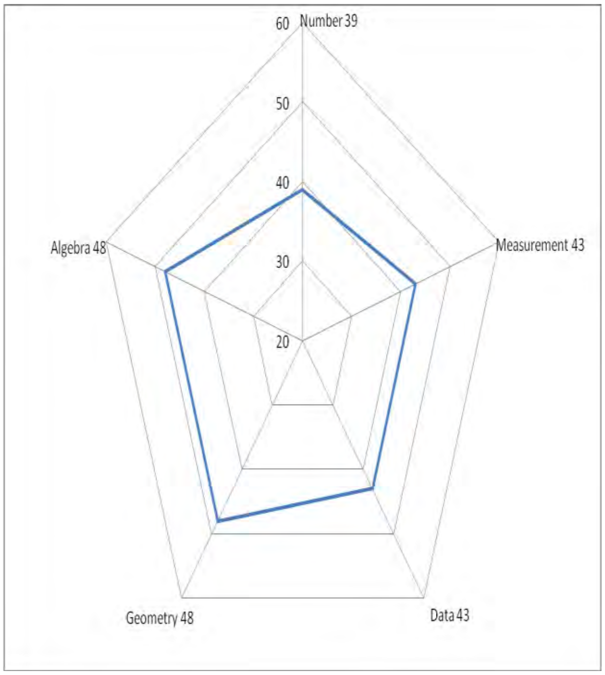
Figure 4b.2 Percentile on national distribution to which each district's average adjusted NAEP grade 8 mathematics scores correspond, by district and subscale, 2007

Atlanta

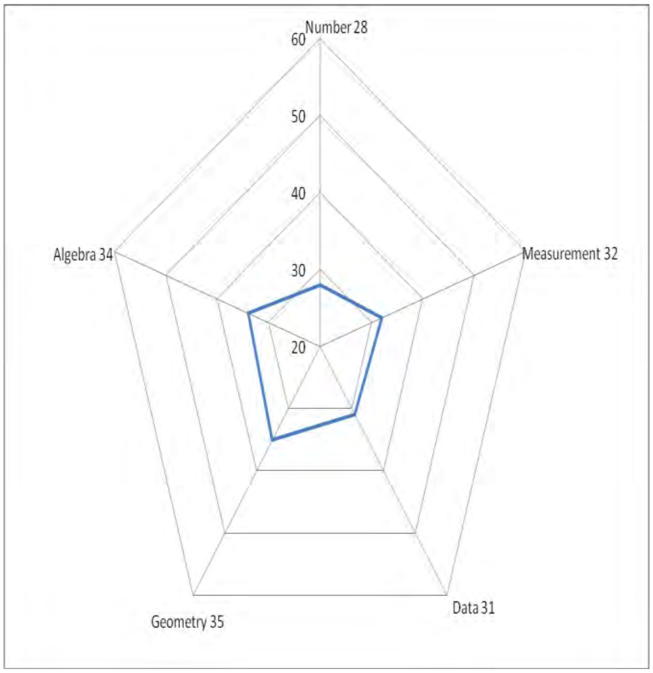


Boston

Charlotte



Cleveland



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

Percentage of Omitted Items by Item Type and Complexity

In addition to conducting the subscale-level analyses, we examined the percentage of items that were left blank—i.e., omitted items—by item type. Tables 4b.8 and 4b.9 show the average percentage of omission rates by item type in grades four and eight. At grade four, the omission rates on multiple-choice (MC) items ranged from 1.4 percent in Charlotte to 2.2 percent in Boston. The omission rates on constructed-response (CR) items in grade four ranged from 2.5 percent in Charlotte to 4.0 percent in Boston. The omission rates among fourth graders on multiple-choice items appeared similar to large-city schools and the nation, with the exception of Boston, which was higher. Omission rates on constructed-response items were typically higher in the selected districts and large-city schools than the national averages, with the exception of Charlotte, where rates appeared generally lower.

In grade eight, the omission rates on multiple-choice items ranged from 1.3 percent in Charlotte to 2.8 percent in Boston. The omission rates on constructed-response items in grade eight ranged from 4.8 percent in Charlotte to 9.0 percent in Cleveland. The omission rates among eighth graders on multiple-choice and constructed-response items were larger for the selected districts and large cities than for the nation.

Table 4b.8 Item omission rates on NAEP grade 4 mathematics, by item type, complexity, and district, 2007

	MC items	CR items	Low Complexity	Moderate Complexity	High Complexity
Atlanta	1.5	3.7	2.0	2.4	3.8
Boston	2.2	4.0	2.5	3.0	6.5
Charlotte	1.4	2.5	1.6	2.0	3.0
Cleveland	1.7	3.7	2.1	2.6	4.4
LC	1.6	3.5	1.9	2.4	4.4
National Public	1.5	2.9	1.7	2.1	3.4

Note: MC=multiple-choice, CR=constructed-response

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

Table 4b.9 Item omission rates on NAEP grade 8 mathematics, by item type, complexity, and district, 2007

	MC items	CR items	Low Complexity	Moderate Complexity	High Complexity
Atlanta	1.8	8.7	2.6	4.8	9.8
Boston	2.8	8.1	3.2	5.3	12.1
Charlotte	1.3	4.8	1.5	3.0	7.5
Cleveland	2.0	9.0	2.5	5.4	11.4
LC	1.6	6.7	2.1	4.0	9.2
National Public	1.2	4.5	1.5	2.8	6.1

Note: MC=multiple-choice, CR=constructed-response

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

Percentage of Correct Items by Item Type and Complexity

Finally, we examined the percentage of correct items by item type and difficulty in each of the four districts and compared the results to the LC averages and the national public school sample. Tables 4b.10 and 4b.11 show these results in grades four and eight, respectively. In grade four, the percent-correct rates ranged from 18 percent (high-complexity items in Atlanta) to 61 percent (low-complexity items in Charlotte).

In grade eight, the percent-correct rates ranged from 22 percent (high-complexity items in Cleveland) to 64 percent (low-complexity items in Charlotte). As expected, in all four districts and at both grade levels, multiple-choice and low-complexity items were the easiest and high-complexity and constructed-response items were the most difficult.

Table 4b.10 Percent-correct rates on NAEP grade 4 mathematics, by item type, complexity, and district, 2007

	MC items	CR items	Low Complexity	Moderate Complexity	High Complexity
Atlanta	42	31	47	30	18
Boston	52	43	56	42	26
Charlotte	55	44	61	42	28
Cleveland	41	32	47	29	19
LC	48	38	53	36	24
National Public	54	44	58	41	31

Note: MC=multiple-choice, CR=constructed-response
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments.

Table 4b.11 Percent-correct rates on NAEP grade 8 mathematics, by item type, complexity, and district, 2007

	MC items	CR items	Low Complexity	Moderate Complexity	High Complexity
Atlanta	51	35	54	35	25
Boston	55	42	57	42	26
Charlotte	61	49	64	48	38
Cleveland	45	31	48	30	22
LC	54	40	57	39	27
National Public	59	46	62	44	36

Note: MC=multiple-choice, CR=constructed-response
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments.

Part 2. Potential Factors Behind Subscale Math Trends

To help us further understand the math results, we explored two hypotheses concerning student NAEP math performance overall and at the subscale levels.

First, we examined the alignment of state and/or district math standards with the NAEP math specifications by subscale.

Second, the research team conducted site visits to the four selected districts to see what they were doing instructionally that would help explain the NAEP math scale scores. The methodology for both parts of this chapter is described in chapter 3 and in appendices C and D.

Alignment of State and District Standards to NAEP Mathematics Specifications

The purpose of this part of the analysis was to determine how well each state's or district's math content standards were aligned with the NAEP specifications and to see if there was any connection between the degree of alignment and how well a district did on NAEP. This work was done using the math specifications found in (1) the *Mathematics Framework for the 2005 National Assessment of Educational Progress* published by National Assessment Governing Board,[15] (2) the state math standards, and (3) in the case of Boston and Cleveland, the district math standards in place during the 2006–2007 school year.

Degree of Content Match

Fourth-grade Mathematics

Our analysis on grade four math showed that between 66 percent and 72 percent of NAEP specifications were either completely or partially matched by the local/state standards in the four jurisdictions. The highest overall matches appeared to be in Boston. These results are shown in table 4b.12 and figure 4b.3. The details follow in the bullets below. (Districts in bold are the main comparison districts in math.)

There were 65 NAEP specifications in fourth-grade math. All jurisdictions showed similar patterns of overall matching.

- Atlanta/Georgia standards matched 44 (68 percent) of the 65 NAEP specifications with 25 complete and 19 partial matches. Therefore, some 38 percent of the 65 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- **Boston, which had slightly different standards than its state, matched 47 (72 percent) of the 65 NAEP specifications, with 25 complete and 22 partial matches. Therefore, 38 percent of the 65 NAEP specifications were completely aligned with the Boston standards. The state's degree of complete match was 19 percentage points higher, at 57 percent.**

- Charlotte/North Carolina's standards matched 43 (66 percent) of the 65 specifications, with 30 complete and 13 partial matches. Therefore, 46 percent of the 65 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland standards matched 43 (66 percent) of the 65 NAEP specifications, with 26 complete and 17 partial matches. Therefore, some 40 percent of the 65 NAEP specifications**

[15] Available at http://www.nagb.org/publications/frameworks/m_framework_05/toc.html

were completely aligned with the Cleveland standards. The state's degree of complete match was 22 percentage points higher, at 62 percent.

In general, the overall degree of complete and partial content matches in fourth-grade math was modest, but the matches were generally higher in math than the content matches in fourth-grade reading.

If we look at the five math strands—number, measurement, geometry, data, and algebra—the patterns showed a more complex picture.

There were 20 NAEP specifications in the *number* subscale in fourth grade.

- Atlanta/Georgia matched 14 (70 percent) of the 20 subscale specifications, with 10 complete and four partial matches. Therefore, 50 percent of the 20 NAEP specifications were completely aligned with the Atlanta/Georgia standards.
- **Boston matched 15 (75 percent) of the 20 subscale specifications, with eight complete and seven partial matches. Therefore, only 40 percent of the 20 NAEP specifications were completely aligned with the Boston standards.**
- Charlotte/North Carolina matched 13 (65 percent) of the 20 subscale specifications, with 10 complete and three partial matches. Therefore, 50 percent the 20 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.
- **Cleveland matched 16 (80 percent) of the 20 subscale specifications, with 10 complete and six partial matches. Therefore, 50 percent of the 20 NAEP specifications were completely aligned with the Cleveland standards.**

In the subscale on *measurement* in fourth grade, there were 10 NAEP specifications.

- Atlanta/Georgia matched seven (70 percent) of the 10 subscale specifications, with four complete and three partial matches. Therefore, 40 percent of the 10 NAEP specifications were completely aligned with the Atlanta/Georgia standards.
- **Boston matched 10 (100 percent) of the 10 subscale specifications, with six complete and four partial matches. Therefore, 60 percent of the 10 NAEP specifications were completely aligned with the Boston standards.**
- Charlotte/North Carolina matched six (60 percent) of the 10 subscale specifications, with three complete and three partial matches. Therefore, only 30 percent of the 10 NAEP specifications were completely aligned with Charlotte/North Carolina standards.
- **Cleveland matched six (60 percent) of the 10 subscale specifications, with three complete and three partial matches. Therefore, only 30 percent of the 10 NAEP specifications were completely aligned with Cleveland standards.**

In the subscale on *geometry* in fourth grade, there are 15 NAEP specifications.

- Atlanta/Georgia matched 11 (73 percent) of the 15 subscale specifications, with two complete and nine partial matches. Therefore, only 13 percent of the 15 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- **Boston matched 11 (73 percent) of the 15 subscale specifications, with five complete and six partial matches. Therefore, 33 percent of the 15 NAEP specifications were completely aligned with the Boston standards.**

- Charlotte/North Carolina matched seven (47 percent) of the 15 subscale specifications, with five complete and two partial matches. Therefore, only 33 percent of the 15 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland matched eight (53 percent) of the 15 subscale specifications, with five complete and three partial matches. Therefore, only 33 percent of the 15 NAEP specifications were completely aligned with the Cleveland standards.**

In the subscale on *data* in fourth grade, there were nine NAEP specifications.

- Atlanta/Georgia matched four (44 percent) of the nine subscale specifications, three complete and one partial match. Therefore, 33 percent of the nine NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- **Boston matched nine (100 percent) of the nine subscale specifications, with six complete and three partial matches. Therefore, 67 percent of the nine NAEP specifications were completely aligned with the Boston standards**.

- Charlotte/North Carolina matched six (67 percent) of the nine subscale specifications, with three complete and three partial matches. Therefore, only 33 percent of the nine NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland matched five (56 percent) of the nine subscale specifications, with four complete and one partial match. Therefore, 44 percent of the nine NAEP specifications were completely aligned with the Cleveland standards.**

In the subscale on *algebra* in fourth grade, there were 11 NAEP specifications.

- Atlanta/Georgia matched eight (73 percent) of the 11 subscale specifications, with six complete and two partial matches. Therefore, 55 percent of the 11 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

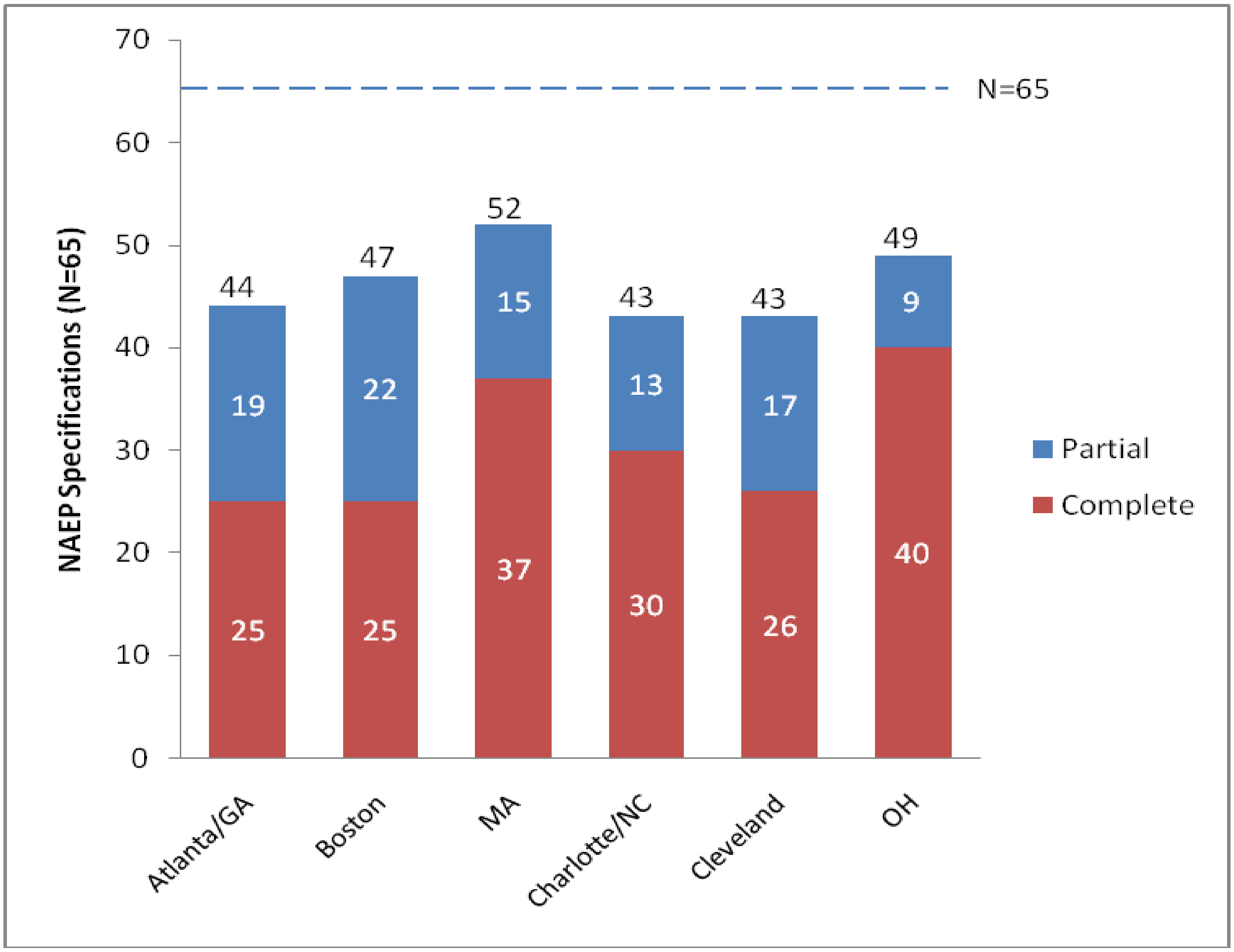
- **Boston matched two (18 percent) of the 11 subscale specifications; both were partial matches. Therefore, none of the 11 NAEP specifications were completely aligned with the Boston standards.**

- Charlotte/North Carolina matched 11 (100 percent) of the 11 subscale specifications, nine complete and two partial matches. Therefore, 82 percent of the 11 NAEP specifications were completely aligned with Charlotte/North Carolina standards.

- **Cleveland matched eight (73 percent) of the 11 subscale specifications, with four complete and four partial matches. Therefore, only 36 percent of the 11 NAEP specifications were completely aligned with the Cleveland standards.**

In general, the complete and partial alignment in algebra at the fourth-grade level was highest in Charlotte/North Carolina, but in Boston this was the lowest matched strand. The alignment in geometry was the lowest strand in two jurisdictions—Charlotte/North Carolina and Cleveland. Finally, Boston showed the highest overall level of complete and partial alignment (72 percent) between its standards and the NAEP specifications in fourth-grade math. In contrast, Atlanta, Cleveland, and Charlotte/North Carolina showed similarly moderate alignment. By and large, however, complete alignments across the four jurisdictions were at or below 50 percent overall and in the subscales. (See table 4b.12 and figure 4b.3)

Figure 4b.3 Number of complete and partial matches with NAEP grade 4 mathematics specifications, by selected districts (*N* of NAEP specifications = 65), 2007*



*44 (68 percent) of Atlanta's grade 4 math standards matched NAEP's 65 math specifications either completely or partially; 47 (72 percent) of Boston's grade 4 math standards matched NAEP's 65 math specifications either completely or partially; 52 (80 percent) of Massachusetts's grade 4 math standards matched NAEP's 65 math specifications either completely or partially; 43 (66 percent) of Charlotte's grade 4 math standards matched NAEP's 65 math specifications either completely or partially; 43 (66 percent) of Cleveland's grade 4 math standards matched NAEP's 65 math specifications either completely or partially; and 49 (75 percent) of Ohio's grade 4 math standards matched NAEP's 65 specifications either completely or partially.

Table 4b.12 Degree of match with NAEP grade 4 mathematics specifications/expectations/indicators, by subscale and district, 2007

		Number of NAEP Specifications, by Strand											
	Subscale:	Number		Measurement		Geometry		Data		Algebra		Total	
		20		10		15		9		11		65	
State and District #/% of Matching Specifications, and Breakdown of Complete and Partial Matches	Atlanta/ GA	14	70%	7	70%	11	73%	4	44%	8	73%	44	68%
		C= 10	P= 4	C= 4	P= 3	C= 2	P= 9	C= 3	P= 1	C= 6	P= 2	C= 25	P= 19
	Boston	15	75%	10	100%	11	73%	9	100%	2	18%	47	72%
		C= 8	P= 7	C= 6	P= 4	C= 5	P= 6	C= 6	P= 3	C= 0	P= 2	C= 25	P= 22
	MA	16	80%	10	100%	14	93%	6	67%	6	55%	52	80%
		C= 10	P= 6	C= 9	P= 1	C= 8	P= 6	C= 5	P= 1	C= 5	P= 1	C= 37	P= 15
	Charlotte/ NC	13	65%	6	60%	7	47%	6	67%	11	100%	43	66%
		C= 10	P= 3	C= 3	P= 3	C= 5	P= 2	C= 3	P= 3	C= 9	P= 2	C= 30	P= 13
	Cleveland	16	80%	6	60%	8	53%	5	56%	8	73%	43	66%
		C= 10	P= 6	C= 3	P= 3	C= 5	P= 3	C= 4	P= 1	C= 4	P= 4	C= 26	P= 17
	OH	17	85%	8	80%	9	60%	7	78%	8	73%	49	75%
		C= 15	P= 2	C= 7	P= 1	C= 7	P= 2	C= 6	P= 1	C= 5	P= 3	C= 40	P= 9

Note: How to read this table. Example: There are 20 specifications in the subscale *Number*. Atlanta standards matched the content of 14 of those 20 specifications or 70 percent. Ten of the matches were complete content matches and four were a partial content match.
C=complete match; P=partial match

Eighth-grade Mathematics

Our analysis on grade eight mathematics showed that between 51 percent and 84 percent of the NAEP specifications were either completely or partially matched by the local/state standards in the four jurisdictions. The highest overall matches were in Cleveland and the lowest were in Charlotte/North Carolina. These results are shown in table 4b.13 and figure 4b.4. The details follow in the bullets below. (Districts in bold are the main comparison districts in math.)

There were 101 NAEP specifications in eighth-grade mathematics, and the analysis of matches showed variation in the selected districts.

- Atlanta/Georgia standards matched 54 (53 percent) of the 101 NAEP specifications, with 32 complete and 22 partial matches. Therefore, 32 percent of the 101 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- **Boston, which had slightly different standards than its state, matched 71 (70 percent) of the 101 specifications, with 45 complete and 26 partial matches. Therefore, some 45 percent of the 101 NAEP specifications were completely aligned with the Boston standards (six percentage points higher than the state's complete matches [39 percent]).**

- Charlotte/North Carolina's standards matched 52 (51 percent) of the NAEP specifications, with 24 complete and 28 partial matches. Therefore, some 24 percent of the 101 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland standards matched 85 (84 percent) of the 101 NAEP specifications, with 57 complete and 28 partial matches. Therefore, 56 percent of the 101 NAEP specifications were completely aligned with the Cleveland standards, four percentage points higher than the state's degree of complete match (52 percent).**

In general, the overall degree of complete and partial content matches in eighth-grade mathematics was modest, but the matches were generally higher in math than the content matches in eighth-grade reading, except in Cleveland.

If we look at the five math strands in grade eight—number, measurement, geometry, data, and algebra—the patterns showed a more complex picture.

There were 27 NAEP specifications in the *number* subscale in eighth grade.

- Atlanta/Georgia matched 10 (37 percent) of those 27 subscale specifications, with eight complete and two partial matches. Therefore, 30 percent of the 27 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- **Boston matched 20 (74 percent) of the 27 subscale specifications, with 12 complete and eight partial matches. Therefore, only 44 percent of the 27 NAEP specifications were completely aligned with the Boston standards.**

- Charlotte/North Carolina matched 10 (37 percent) of the 27 subscale specifications, with four complete and six partial matches. Therefore, 15 percent of the 27 NAEP specifications were completely aligned with Charlotte/North Carolina standards.

- **Cleveland matched 21 (78 percent) of the 27 subscale specifications, with 14 complete and seven partial matches. Therefore, 52 percent of the 27 NAEP specifications were completely aligned with the Cleveland standards.**

There were 13 NAEP specifications in the *measurement* subscale in eighth grade.

- Atlanta/Georgia matched one (8 percent) of the 13 subscale specifications, with a complete match. Therefore, only 8 percent of the 13 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- **Boston matched 11 (85 percent) of the 13 subscale specifications, with nine complete and two partial matches. Therefore, 69 percent of the 13 NAEP specifications were completely aligned with the Boston standards.**

- Charlotte/North Carolina/ matched four (31 percent) of the 13 subscale specifications, with three complete and one partial match. Therefore, only 23 percent of the 13 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland matched 11 (85 percent) of the 13 subscale specifications, with nine complete and two partial matches. Therefore, only 69 percent of the 13 NAEP specifications were completely aligned with the Cleveland standards.**

There were 21 NAEP specifications in the subscale on *geometry* in eighth grade.

- Atlanta/Georgia matched 14 (67 percent) of the 21 subscale specifications, with eight complete and six partial matches. Therefore, only 38 percent of 21 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- **Boston matched 13 (62 percent) of the 21 subscale specifications, with five complete and eight partial. Therefore, 24 percent of the 21 NAEP specifications were completely aligned with the Boston standards.**

- Charlotte/North Carolina matched 11 (52 percent) of the 21 subscale specifications, with five complete and six partial matches. Therefore, only 24 percent of 21 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland matched 17 (81 percent) of the 21 subscale specifications, with 12 complete and five partial matches. Therefore, only 57 percent of the 21 NAEP specifications were completely aligned with the Cleveland standards.**

In the subscale on *data* in eighth grade, there were 22 NAEP specifications.

- Atlanta/Georgia matched 13 (59 percent) of the 22 subscale specifications, with six complete and seven partial matches. Therefore, 27 percent of the 22 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- **Boston matched 13 (59 percent) of the 22 subscale specifications, with nine complete and four partial matches. Therefore, 41 percent of the 22 NAEP specifications were completely aligned with the Boston standards.**

- Charlotte/North Carolina matched 10 (45 percent) of the 22 subscale specifications, with five complete and five partial matches. Therefore, only 23 percent of the 22 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland matched 19 (86 percent) of the 22 subscale specifications, with 14 complete and five partial matches. Therefore, 64 percent of the 22 NAEP specifications were completely aligned with the Cleveland standards.**

In the subscale on *algebra* in eighth grade, there were 18 NAEP specifications.

- Atlanta/Georgia matched 16 (89 percent) of the 18 subscale specifications, with nine complete and seven partial matches. Therefore, 50 percent of the 18 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- **Boston matched 14 (78 percent) of the 18 specifications, with 10 complete and four partial matches. Therefore, 56 percent of the 18 NAEP specifications were completely aligned with the Boston standards.**

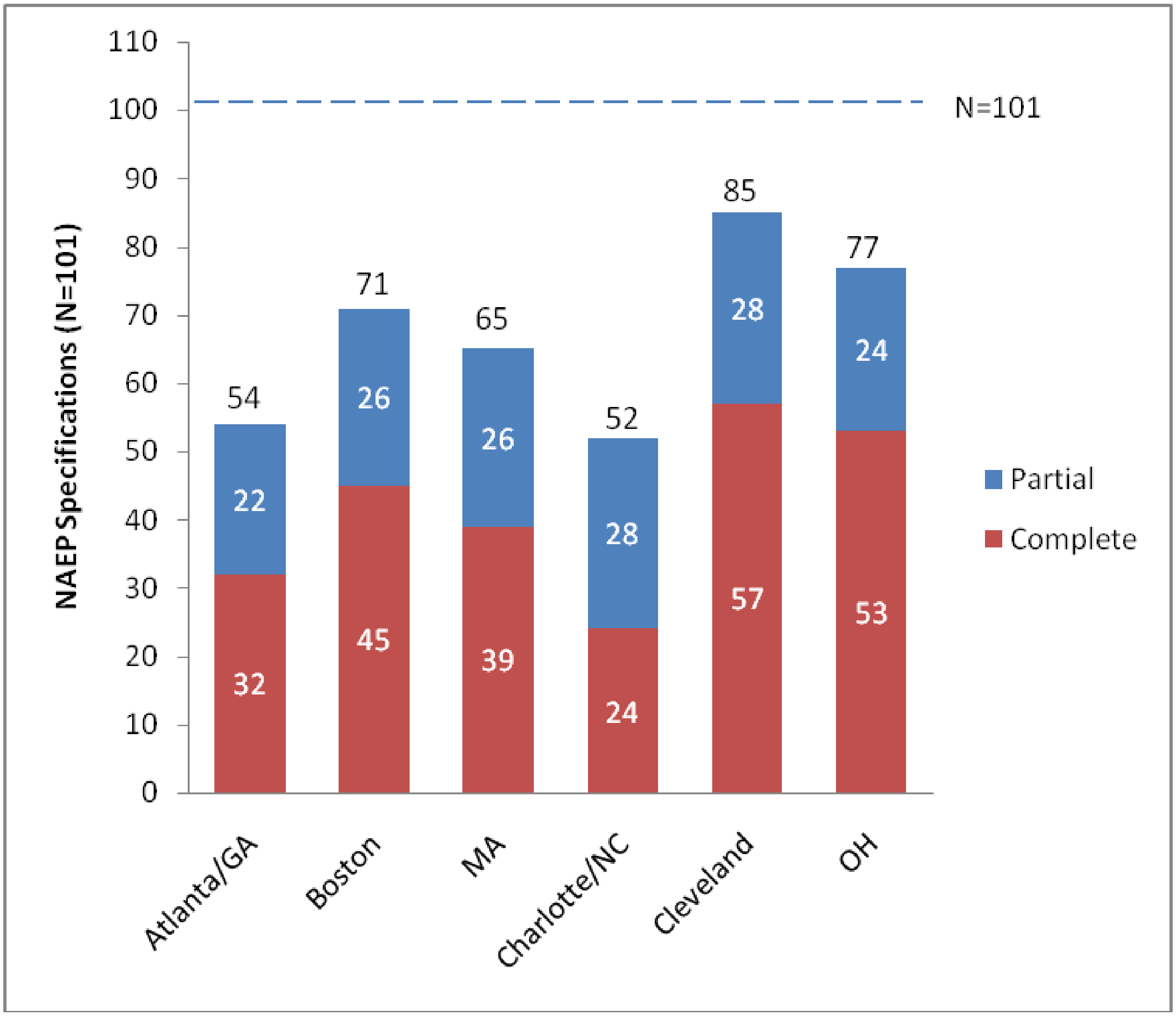
- Charlotte/North Carolina matched 17 (94 percent) of the 18 specifications, with seven complete and 10 partial matches. Therefore, 39 percent of the 18 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- **Cleveland matched 17 (94 percent) of the 18 specifications, with eight complete and nine partial matches. Therefore, only 44 percent of the 18 NAEP specifications were completely aligned with the Cleveland standards.**

In general, the complete and partial alignment in algebra at the eighth-grade level was the highest area matched across the four jurisdictions. No other patterns emerged in the degree of match in grade eight on complete and partial alignments on any other strand. Cleveland's standards showed the highest overall level of complete and partial alignment (84 percent) with the NAEP specifications in eighth-grade math, and Cleveland had the highest overall complete matches as well (56 percent).

In contrast, Charlotte/North Carolina and Atlanta/Georgia showed the lowest overall complete and partial alignment of around 52 percent. By and large, however, complete alignments across the four jurisdictions were at or below 50 percent overall and in the subscales.

Figure 4b.4 Number of complete and partial matches with NAEP grade 8 mathematics specifications, by selected districts (*N* of NAEP specifications = 101), 2007*



*54 (53 percent) of Atlanta's grade 8 math standards matched NAEP's 101 math specifications either completely or partially; 71 (70 percent) of Boston's grade 8 math standards matched NAEP's 101 math specifications either completely or partially; 65 (64 percent) of Massachusetts's grade 8 math standards matched NAEP's 101 math specifications either completely or partially; 52 (51 percent) of Charlotte's grade 8 math standards matched NAEP's 101 math specifications either completely or partially; 85 (84 percent) of Cleveland's grade 8 math standards matched NAEP's 101 math specifications either completely or partially; and 77 (76 percent) of Ohio's grade 8 math standards matched NAEP's 101 specifications either completely or partially.

Table 4b.13 Degree of match with NAEP grade 8 mathematics specifications/expectations/indicators, by subscale and district, 2007

		Number of NAEP Specifications, by Strand											
	Subscale:	Number		Measurement		Geometry		Data		Algebra		Total	
		27		13		21		22		18		101	
	Atlanta/ GA	10	37%	1	8%	14	67%	13	59%	16	89%	54	53%
		C= 8	P= 2	C= 1	P= 0	C= 8	P= 6	C= 6	P= 7	C= 9	P= 7	C= 32	P= 22
State and District #/% of Matching Specifications, and Breakdown of Complete and Partial Matches	Boston	20	74%	11	85%	13	62%	13	59%	14	78%	71	70%
		C= 12	P= 8	C= 9	P= 2	C= 5	P= 8	C= 9	P= 4	C= 10	P= 4	C= 45	P= 26
	MA	19	70%	10	77%	11	52%	12	55%	13	72%	65	64%
		C= 9	P= 10	C= 7	P= 3	C= 5	P= 6	C= 8	P= 4	C= 10	P= 3	C= 39	P= 26
	Charlotte/ NC	10	37%	4	31%	11	52%	10	45%	17	94%	52	51%
		C= 4	P= 6	C= 3	P= 1	C= 5	P= 6	C= 5	P= 5	C= 7	P= 10	C= 24	P= 28
	Cleveland	21	78%	11	85%	17	81%	19	86%	17	94%	85	84%
		C= 14	P= 7	C= 9	P= 2	C= 12	P= 5	C= 14	P= 5	C= 8	P= 9	C= 57	P= 28
	OH	17	63%	9	69%	13	62%	20	91%	18	100%	77	76%
		C= 7	P= 10	C= 8	P= 1	C= 9	P= 4	C= 17	P= 3	C= 12	P= 6	C= 53	P= 24

Note: How to read this table. Example: There are 27 specifications in the subscale *Number*. Atlanta standards matched the content of 10 of those 27 specifications or 37 percent. Eight of the matches were complete content matches and two were a partial content match.
C=complete match; P=partial match

Tables 4b.14 and 4b.15 summarize the degree of complete match with NAEP specifications in fourth- and eighth-grade mathematics. Matches of 80 percent or more were deemed high, while matches 50 percent or below were deemed low. Of the 30 cells in table 4b.14, only two were high—measurement in Massachusetts and algebra in Charlotte—and 20 (67 percent) were low. Of the 30 cells in grade eight, none of the complete matches were deemed high, while 19 (63 percent) were considered low.

Table 4b.14 Degree of complete match of NAEP subscales with district/state standards in grade 4 mathematics, by subscale and district, 2007*

	District/State					
Strand	Atlanta/GA	Boston	MA	Charlotte/NC	Cleveland	OH
Number	Low	Low	Low	Low	Low	Moderate
Measurement	Low	Moderate	High	Low	Low	Moderate
Geometry	Low	Low	Moderate	Low	Low	Low
Data	Low	Moderate	Moderate	Low	Low	Moderate
Algebra	Moderate	Low	Low	High	Low	Low

***** High (80 percent or more) and low (50 percent or less)

Table 4b.15 Degree of complete match of NAEP subscales with district/state standards in grade 8 mathematics, by subscale and district, 2007*

	District/State					
Strand	Atlanta/GA	Boston	MA	Charlotte/NC	Cleveland	OH
Number	Low	Low	Low	Low	Moderate	Low
Measurement	Low	Moderate	Moderate	Low	Moderate	Moderate
Geometry	Low	Low	Low	Low	Moderate	Low
Data	Low	Low	Low	Low	Moderate	Moderate
Algebra	Low	Moderate	Moderate	Low	Low	Moderate

***** High (80 percent or more) and low (50 percent or less)

Degree of Match in Cognitive Demand

In addition to determining the degree of content match between local/state standards and NAEP specifications, the research team examined how well those completely matched standards corresponded in their cognitive demand or complexity to NAEP specifications. (See chapter 3 and appendices C and D for a detailed description of the methodology.) The analysis entailed examining the wording of district/state standards and NAEP specifications to determine the cognitive demand or rigor in each statement and comparing the results of the two.

Tables 4b.16 and 4b.17 show the level of complete content match discussed in the previous section of this chapter along with the number and percentage of state and local standards that were classified as low, moderate, or high on cognitive demand in fourth- and eighth-grade math. Only those standards that matched NAEP specifications completely were included in the analysis. This gives the reader a sense of the rigor or complexity of state and local standards, but only for the portion of standards that match completely with NAEP. Omitted from the cognitive demand codes were all standards that did not correspond to NAEP. The data in the tables indicate that the level of cognitive demand in the state and district standards appeared to be closely aligned with NAEP in both grade four and grade eight. In fact, the cognitive demand of the completely matched standards in the four selected districts appeared often to be as high as the NAEP specifications.

Overall, most district/state standards and NAEP specifications had moderate **cognitive** demand. Tables 4b.16 and 4b.17 on grades four and eight, respectively, show that 66 percent of the grade four NAEP math specifications and 84 percent of the grade eight specifications were **moderate** in cognitive demand. Our analysis showed that the overwhelming majority of state and local standards that matched the NAEP specifications were also moderate in cognitive demand. In general, the cities and states had smaller percentages of standards written with low cognitive demand than NAEP and greater percentages of standards with moderate cognitive demand than NAEP.

To further quantify the degree of cognitive demand, the tables below show weighted total and weighted averages for each district. The total weight was based on assigning one point for low, two points for moderate, and three points for high cognitive demand. The weighted average was derived by dividing the weighted total by the total number of complete matches. The analysis suggests that, for completely matching standards, the degree of cognitive demand at grade four mathematics was as high as or higher in the four selected districts than on NAEP. For instance, Boston's weighted average was 2.0, a level that was somewhat higher than NAEP's 1.8 (the baseline).

Table 4b.16 Degree of match in cognitive demand for specifications with complete alignment to NAEP grade 4 mathematics, by district, 2007

	NAEP		Atlanta/ GA		Boston		MA		Charlotte/ NC		Cleveland		OH	
% of Complete Content Match	100%		38%		38%		57%		46%		40%		62%	
Cognitive Levels														
Low	19	29%	3	12%	1	4%	2	5%	2	7%	4	15%	7	18%
Moderate	43	66%	20	80%	24	96%	35	95%	26	87%	20	77%	30	75%
High	3	5%	2	8%	0	0%	0	0%	2	7%	2	8%	3	8%
Total	65	100%	25	100%	25	100%	37	100%	30	100%	26	100%	40	100%
Weighted Total	114		49		49		72		60		50		76	
Weighted Mean		1.8*		2.0		2.0		1.9		2.0		1.9		1.9

* Number represents the balance among NAEP standards that were determined to be high, moderate, or low cognitive demand. 1=low cognitive demand, 2=moderate cognitive demand, and 3=high cognitive demand.

Table 4b.17 Degree of match in cognitive demand for specifications with complete alignment to NAEP grade 8 mathematics, by district, 2007

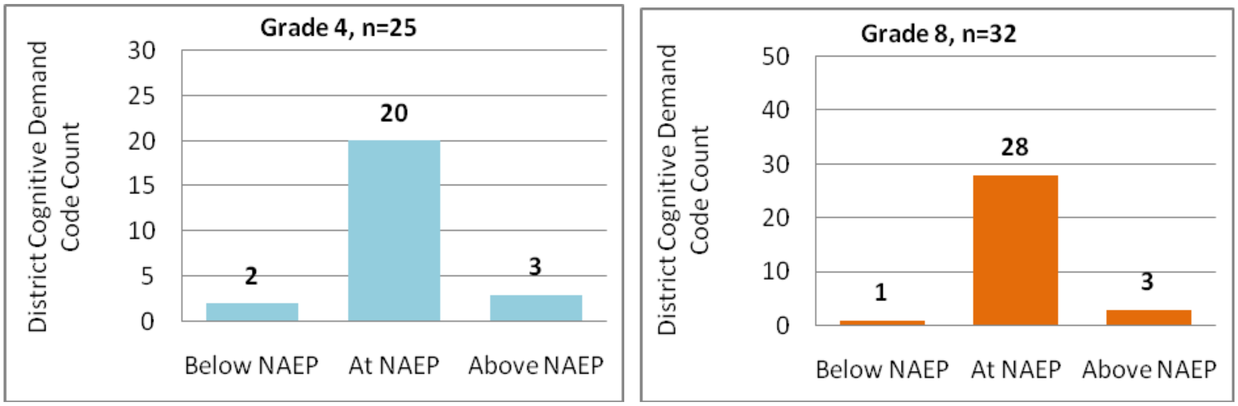
	NAEP		Atlanta/ GA		Boston		MA		Charlotte/ NC		Cleveland		OH	
% of Complete Content Match	100%		32%		45%		39%		24%		56%		52%	
Cognitive Levels														
Low	7	7%	0	0%	0	0%	0	0%	0	0%	1	2%	1	2%
Moderate	85	84%	30	94%	40	89%	35	90%	23	96%	52	91%	49	92%
High	9	9%	2	6%	5	11%	4	10%	1	4%	4	7%	3	6%
Total	101	100%	32	100%	45	100%	39	100%	24	100%	57	100%	53	100%
Weighted Total	204		66		95		82		49		117		108	
Weighted Mean		2.0*		2.1		2.1		2.1		2.0		2.1		2.0

* Number represents the balance among NAEP standards that were determined to be high, moderate, or low cognitive demand. 1=low cognitive demand, 2=moderate cognitive demand, and 3=high cognitive demand.

At grade eight, the cognitive demand of NAEP was again slightly lower than the weighted averages of all of the local/state standards because, as with reading, NAEP intentionally has a range of items from low to high in order to measure what students at the lowest end of the scale actually know.

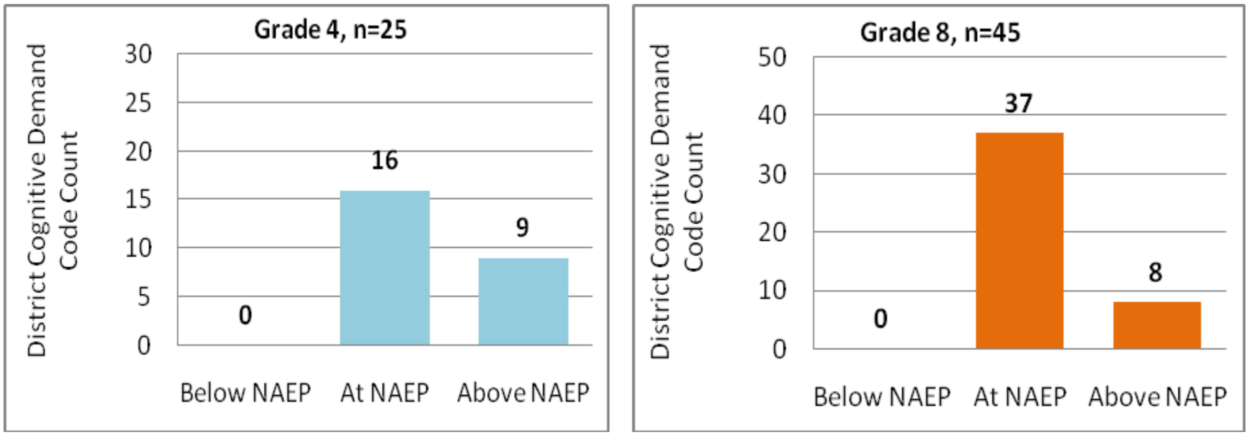
One additional way to capture the degree of alignment in cognitive demand is to directly compare each *completely matching* district/state standard with the corresponding NAEP specification. Figures 4b.5 through 4b.16 present this information at grades four and grade eight for Atlanta, Boston, Massachusetts, Charlotte, Cleveland, and Ohio, respectively. As with the prior analyses, these graphs showed that in all jurisdictions and for both grades, the majority of completely matched standards were at similar levels of cognitive demand as NAEP.

Figures 4b.5 and 4b.6 *Atlanta's c*omplete matches at grades 4 and 8 mathematics in cognitive demand compared to NAEP, 2007*



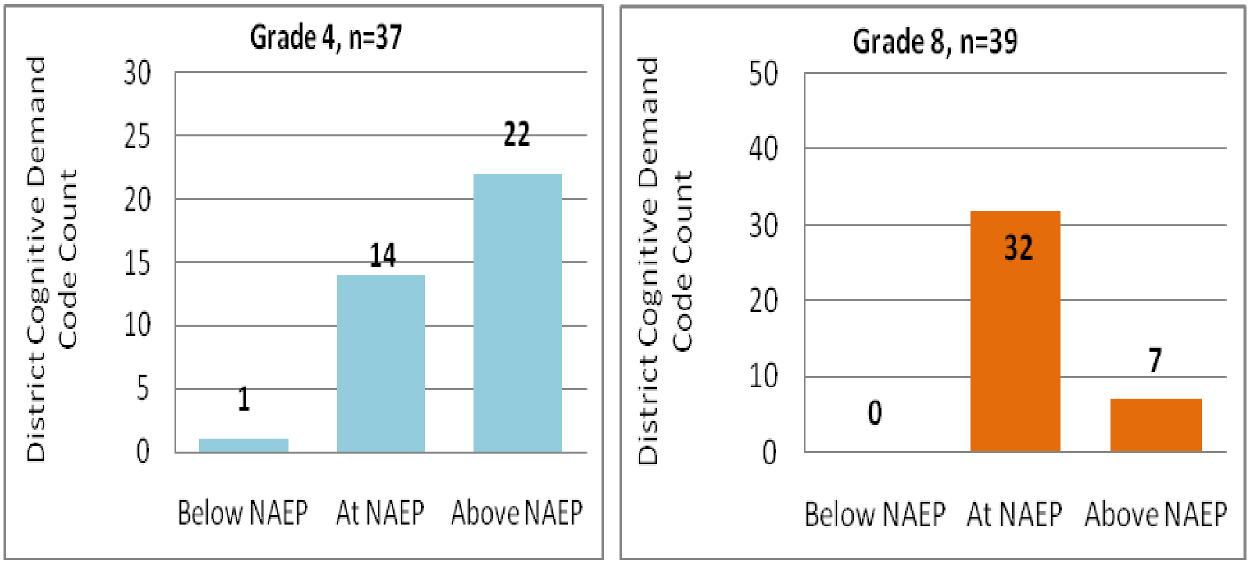
* 25 of Atlanta's grade 4 standards completely matched the 65 NAEP math specifications (38 percent). Two of those 25 completely matched standards had a cognitive demand level below NAEP, 20 were at the NAEP level, and three were above NAEP. Similarly, 32 of Atlanta's eighth grade standards completely matched the 101 NAEP math specifications (32 percent). One of those 32 completely matched standards had a cognitive demand level below NAEP, 28 were at the NAEP level, and three were above NAEP.

Figures 4b.7 and 4b.8 *Boston's c*omplete matches at grades 4 and 8 mathematics in cognitive demand compared to NAEP, 2007*



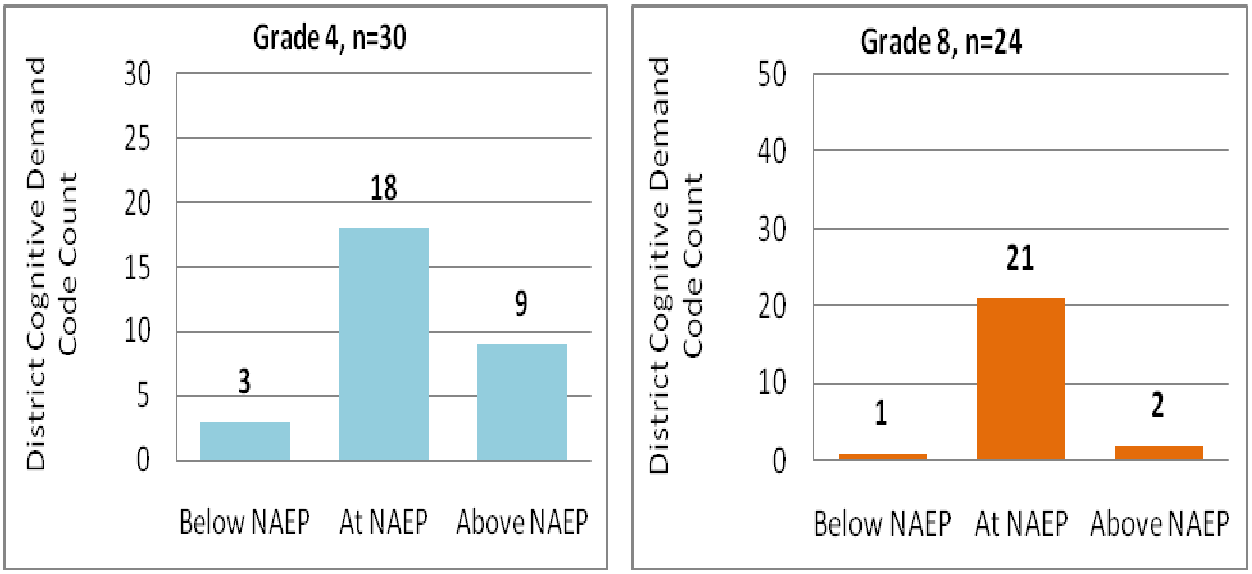
* 25 of Boston's grade 4 standards completely matched the 65 NAEP math specifications (38 percent). None of those 25 completely matched standards had a cognitive demand level below NAEP, 16 were at the NAEP level, and nine were above NAEP. Similarly, 45 of Boston's eighth grade standards completely matched the 101 NAEP math specifications (45 percent). None of those 45 completely matched standards had a cognitive demand level below NAEP, 37 were at the NAEP level, and eight were above NAEP.

Figures 4b.9 and 4b.10 *Massachusetts's* complete matches at grades 4 and 8 mathematics in cognitive demand compared to NAEP, 2007*



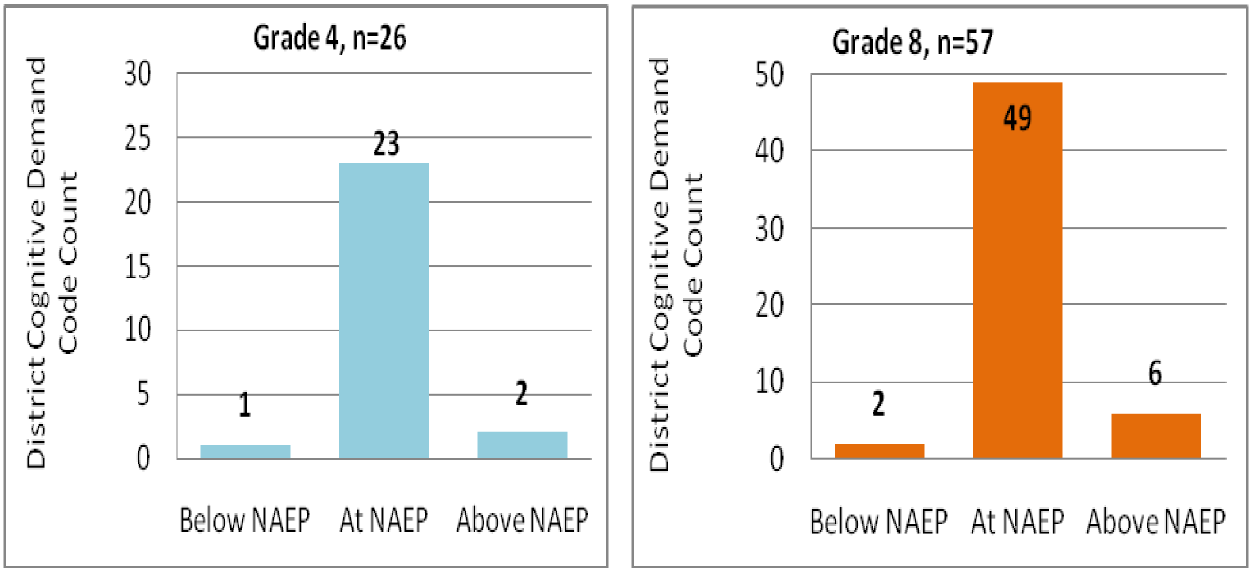
* 37 of Massachusetts's grade 4 standards completely matched the 65 NAEP math specifications (57 percent). One of those 37 completely matched standards had a cognitive demand level below NAEP, 14 were at the NAEP level, and 22 were above NAEP. Similarly, 39 of Massachusetts's eighth grade standards completely matched the 101 NAEP math specifications (39 percent). None of those 39 completely matched standards had a cognitive demand level below NAEP, 32 were at the NAEP level, and seven were above NAEP.

Figures 4b.11 and 4b.12 *Charlotte's c*omplete matches at grades 4 and 8 mathematics in cognitive demand compared to NAEP, 2007*



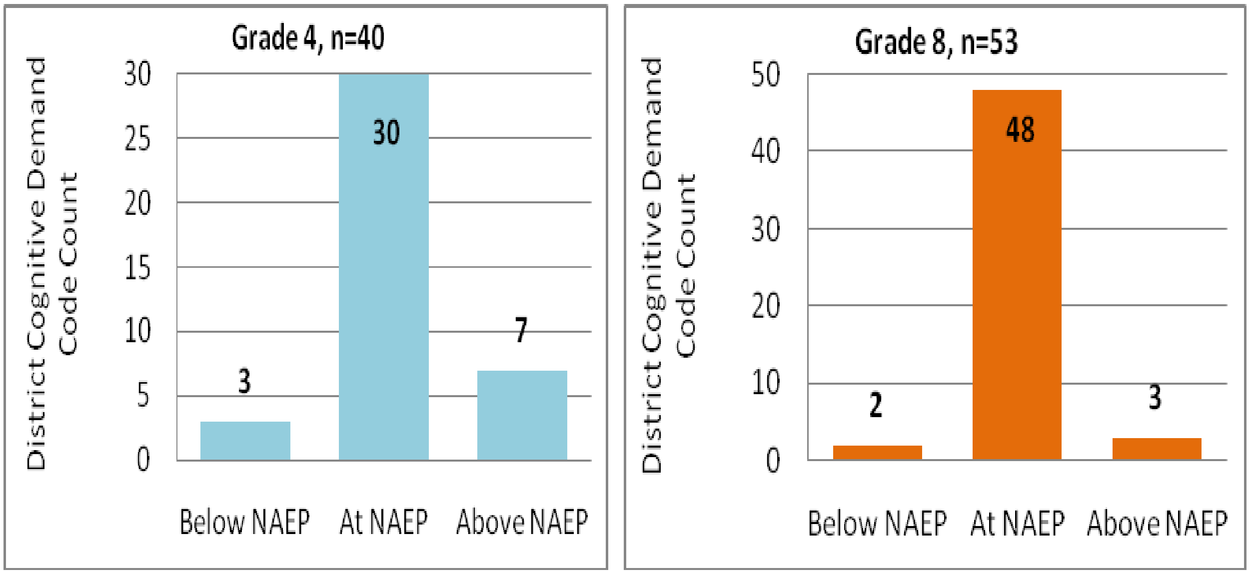
* 30 of Charlotte's grade 4 standards completely matched the 65 NAEP math specifications (46 percent). Three of those 30 completely matched standards had a cognitive demand level below NAEP, 18 were at the NAEP level, and nine were above NAEP. Similarly, 24 of Charlotte's eighth grade standards completely matched the 101 NAEP math specifications (24 percent). One of those 24 completely matched standards had a cognitive demand level below NAEP, 21 were at the NAEP level, and two were above NAEP.

Figures 4b.13 and 4b.14 *Cleveland's c*omplete matches at grades 4 and 8 mathematics in cognitive demand compared to NAEP, 2007*



* 26 of Cleveland's grade 4 standards completely matched the 65 NAEP math specifications (40 percent). One of those 26 completely matched standards had a cognitive demand level below NAEP, 23 were at the NAEP level, and two were above NAEP. Similarly, 57 of Cleveland's eighth grade standards completely matched the 101 NAEP math specifications (56 percent). Two of those 57 completely matched standards had a cognitive demand level below NAEP, 49 were at the NAEP level, and six were above NAEP.

Figures 4b.15 and 4b.16 *Ohio*'s complete matches at grades 4 and 8 mathematics in cognitive demand compared to NAEP, 2007*



* 40 of Ohio's grade 4 standards completely matched the 65 NAEP math specifications (62 percent). Three of those 40 completely matched standards had a cognitive demand level below NAEP, 30 were at the NAEP level, and seven were above NAEP. Similarly, 53 of Ohio's eighth grade standards completely matched the 101 NAEP math specifications (52 percent). Two of those 53 completely matched standards had a cognitive demand level below NAEP, 48 were at the NAEP level, and three were above NAEP.

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 4.b18 and 4.b19.) Overall, the content matches appeared similar in grade four and grade eight, although there was greater variability in grade eight. Although the complete and partial matches on the NAEP standards never fell below 50 percent in math, only at grade eight in Cleveland did the content match exceed 80 percent. However, analyses of the complete matches provided a different picture. At grade four, complete matches were at or below 50 percent in the four cities, and at grade eight none exceeded 56 percent.

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

Table 4b.18 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

Table 4b.19 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

In fourth grade, Atlanta and Boston were the only two of the four districts to see significant increases in math, yet both districts had lower complete content matches than Charlotte and Cleveland, which saw no significant increases in NAEP math scale scores. Moreover, the cognitive demand averages of all four districts appeared to be similar. As in reading, Charlotte had the highest percentile measure in math and what appeared to be the highest overall level of complete content matches.

In eighth grade, Atlanta, Boston, and Charlotte saw significant increases in math scale 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 (56 percent). 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 averages that were similar to the other districts.

Site Visits and Linkages to Mathematics Results

As was indicated previously, the research team conducted site visits to the four selected districts to examine practices and policies that could help explain the trends in NAEP math scale scores between 2003 and 2007. A description of the methodology and the protocols used during these site visits was presented in chapter 3. At each site, teachers, staff, and community members were interviewed and instructional materials used during the study period were reviewed. (See appendix I.) This section examines what the team learned on these visits about the instructional programs in each city that could inform the math results, particularly the subscale and strand results, presented earlier in this chapter. In the next chapters we examine the broader contextual features of the four districts and the particular instructional practices of the school systems, and we synthesize the results from this and earlier chapters into a more cohesive picture of why student achievement scores on NAEP may have improved or failed to improve.

In this section, we pay particular attention to the data on Boston and Cleveland because Boston had significant and the most consistent gains in math and Cleveland had weaker and less consistent improvements in math.

The data presented in this and the previous chapters indicated that Boston made statistically significant progress in mathematics on NAEP scale scores between 2003 and 2007. The data also showed that the district's math gains over this period in terms of effect sizes at the fourth- and eighth-grade levels were significant on the NAEP composite math scores and on every subscale. The data, moreover, suggests that the district saw the largest gains in fourth grade in the number and geometry subscales and in the eighth grade in algebra, measurement, and geometry.

The information gained during the Boston site visit helps us understand why these NAEP patterns exist. As will be described in greater detail in chapter 5 and in the case study in appendix G, Boston pursued an aggressive set of math reforms for the better part of 10 years starting around the year 2000. In general, Boston's gains, overall and at the subscale level, appear to be due, in part, to (1) the district's adoption of math programs with a strong emphasis on understanding math concepts and problem solving, (2) its alignment with state standards that were consistent with NAEP, (3) the extensive amount of professional development and math coaching received by school staff and teachers, (4) the gradual phase-in of the program that helped build capacity and ownership, (5) the convincing feed-back loops that the central office built into the program's implementation, (6) the careful monitoring of math achievement and its progress, and (7) a districtwide math plan.

In addition, district math program staff indicated during site visit interviews that the topic-specific professional development seminars most chosen by teachers was in number and geometry, which, in fourth grade, were the areas in which student NAEP scale scores improved. Conversely, teachers participated in less professional development in measurement, an area where students made less relative progress. Moreover, the district's gains on the algebra subscale in the eighth grade may have been partly due to the math program's strong emphasis on this area in the middle grades.

In Cleveland, fourth graders posted no composite math gains or gains on any of the individual subscales between 2003 and 2007. Eighth graders saw no composite gains but did post an increase in the algebra subscale over the study period. Cleveland's students did not show statistically significant improvements on any other subscale. In general, the district's percentile rankings also trended downward in every subscale and on the composite score in both grades, except in the algebra subscale at eighth grade. In general, the district showed no particular strengths in any subscale at either grade.

The information gained during the Cleveland site visit helped us understand this pattern. As noted in the reading section, Cleveland had an instructional program that was highly fractured (also described in greater depth in the next chapter). Interestingly, although Cleveland's composite score over the study period did not improve as it did in Boston, Cleveland used the same middle school math program as Boston and provided National Science Foundation funding for selected teachers to receive extensive training in math and science content. Our site visit data indicated that Cleveland's middle school math program, however, was limited to a set of pilot schools and never expanded. Although the program was in only a few schools, it is possible that it did contribute to the increase in algebra subscale scores among Cleveland's eighth graders.

Otherwise, Cleveland's math initiative appeared to have been too weak and too isolated to produce districtwide gains in NAEP math scale scores in either grade. The district did not collect or disseminate the detailed data that Boston did; nor did it have the same kind of program leadership, professional development, or monitoring that appeared to facilitate gains in Boston. In addition, the interviews and focus groups indicated that principals and teachers did not always use the district's scope and sequence guides and that student academic work was not the focus of classroom observations, where they were made at all. Student achievement data were not used to shape instructional modifications or enhancements or to inform professional development to the same extent as in Boston.

Finally, the structure of the general math program itself may have presented a problem. The 2006 math matrix developed by the district, called the Big Book of Math Standards, laid out a series of math standards grade by grade without any reference to specific instructional materials or time frames for when to teach what. Organizing instruction in this way can lead teachers to search for activities that have little overt connection to the standards. Research suggests that the result of this approach is often erosion in the integrity of the mathematics taught, a situation that may have been the case in Cleveland.

Part 1. District Performance on NAEP Science Subscales

Content

The framework and specifications used to guide NAEP's 2005 science assessments included content from three broad fields of science: earth science, physical science, and life science. The NAEP specifications present the science content in outlines that include grade-level objectives (at grades four, eight, and 12) along with ideas for exercises aligned with these objectives. Table 4c.1 shows how the item pool is distributed across the three fields in science (approximated as the proportion of the total amount of time that would be required if the entire pool were administered to a single individual).

Table 4c.1 Percentage of items by science content area and grade level, 2005

Field of Science	Grade 4	Grade 8
Earth Science	33%	30%
Physical Science	33%	30%
Life Science	33%	40%

In 2005, the grade four science assessment consisted of 157 specifications across three science subscales, and the grade eight assessment consisted of 222 specifications across the same three subscales. In addition to the specifications across fields of science, NAEP employs a cognitive-demand structure that incorporates low-, moderate-, and high-complexity items. The assessments at both grade levels contain multiple-choice, short constructed-response, and extended constructed-response items.

Also included are performance exercises that allow students to manipulate physical objects and to draw scientific understandings from those manipulations. The full set of specifications that governed the NAEP science assessments in 2005 can be found in *Science Assessment and Exercise Specifications for the 1994 National Assessment of Educational Progress*.[16]

Composite, Subscale, and Item analyses – Strengths and Weaknesses in Science

In this section, we compare and contrast the academic strengths and weaknesses of the four selected districts in each of the three science fields or subscales. As described in chapter 3 (Methodology), NAEP subscales are not all reported on the same metric, so the average subscale scores or gains in average subscale scores are not directly comparable from one subscale to another. So in order to estimate district strengths and weaknesses, we examine subscale and item-level performance in two ways.

First, we provide the percentile rankings for each selected TUDA district on the distribution of average subscale scores on the national public school sample, after adjusting for student background characteristics. Second, we provide item-level information about omission rates and the percentage of correct items by item type for each TUDA district in the three fields of science.

[16] Published by the National Assessment Governing Board.

Taken together, the results provide a picture of overall performance on the science 2005 NAEP assessment in the selected districts and of academic strengths and weaknesses in the three fields of science. (See table 4c.2.)

Table 4c.2 Average NAEP science percentiles by subscale and grade corresponding to the subscale score distribution of the national public school sample, 2005

	Atlanta		Boston		Charlotte		Cleveland	
	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8
Composite Science	29	20	29	31	42	42	25	24
Physical Science	28	18	29	29	40	40	25	22
Earth Science	32	22	29	33	43	45	25	23
Life Science	29	21	31	33	45	42	28	28

Percentile Measure by Subscale, Adjusted for Student Background Characteristics

Figures 4c.1 and 4c.2 show another way of capturing the relative performance of the districts. These "radar graphs" show the percentile (on the national public sample) to which a given district's adjusted subscale average corresponded on the 2005 NAEP science assessment. The averages were adjusted for the same demographic variables discussed in chapter 3.

For example, the 30th percentile for the life science subscale in grade four would mean that 70 percent of students in the nation performed better in life science than the average fourth grader in that district after adjusting for differences in background variables. Therefore, the closer the graph is to the center, the weaker performance; the vertices farthest from the center signify relative subscale strength.

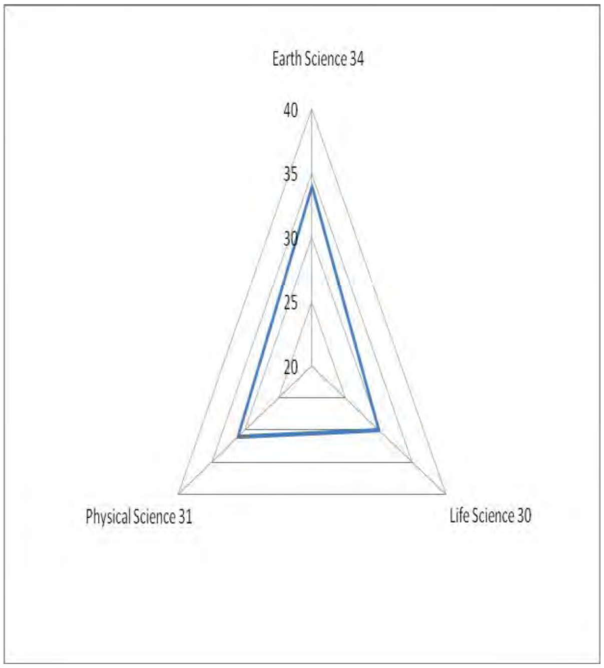
The figures show that, in grade four in 2005, the adjusted averages on all subscales in all four districts were below the national median, which would be at 50. All of the selected districts performed within the 28th to 34th percentile in science in fourth grade.[17] Earth science, however, appeared to be a relative strength for Atlanta, compared with other subscales in the same district. In Charlotte, physical science (29th percentile) appeared to be weaker than the other two subscales. In Boston and Cleveland, the percentiles on the three subscales appeared to be relatively close to one another (Boston's percentiles ranged from 31 to 33 and Cleveland from 28 to 31.)

In grade eight, the adjusted averages on all subscales were also below the national median in all four districts, but the range of performance (24th percentile to the 36th percentile) was somewhat wider than in grade four. Compared to other subscales within their own districts, earth science appeared to be a relative strength in Atlanta and in Charlotte, and physical science appeared to be a weakness in Atlanta, Boston, and Cleveland.[18]

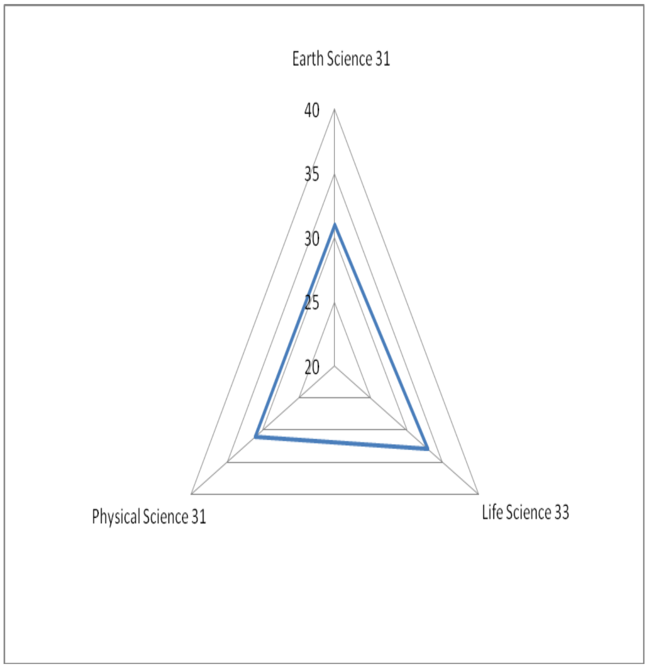
[17] Percentiles for all 11 TUDA districts, except the District of Columbia, are shown in tables B.44 and B.45. Results show that fourth graders appeared to do somewhat better in life sciences than in earth science and physical science, while eighth graders appeared to do about the same in all three fields of science.
[18] Subscale data on other TUDA districts is included in appendix C.

Figure 4c.1 Percentile on national distribution to which each district's average adjusted NAEP grade 4 science scores correspond, by district and subscale, 2005

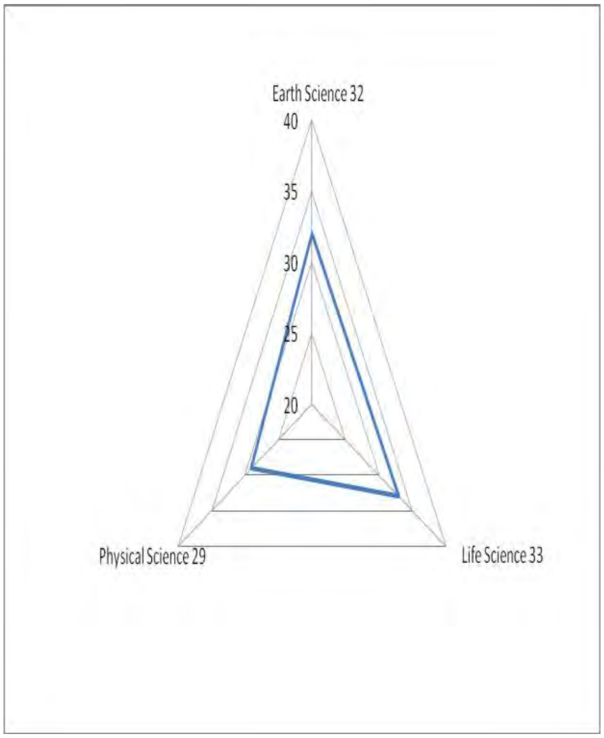
Atlanta



Boston



Charlotte



Cleveland

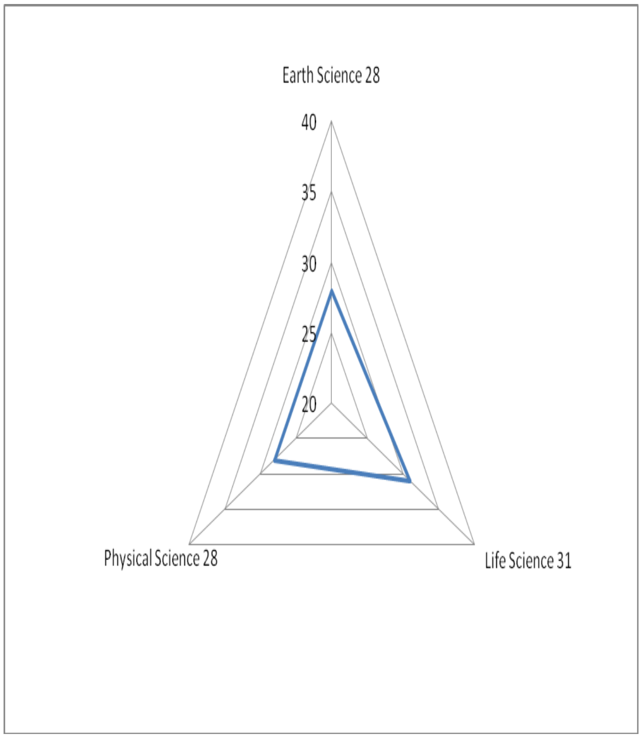
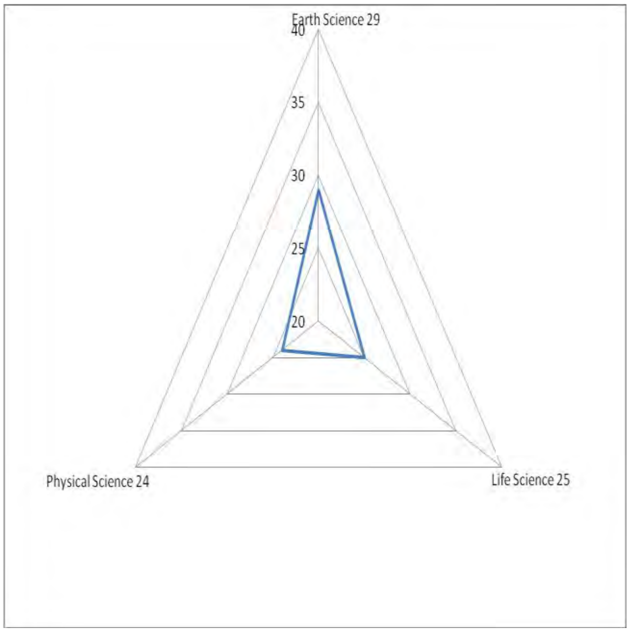


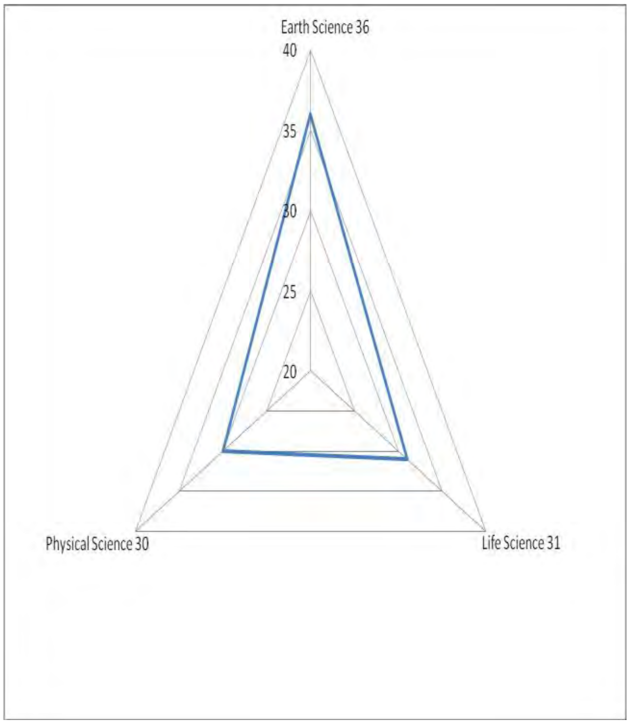
Figure 4c.2 Percentile on national distribution to which each district's average adjusted NAEP grade 8 science scores correspond, by district and subscale, 2005

Atlanta

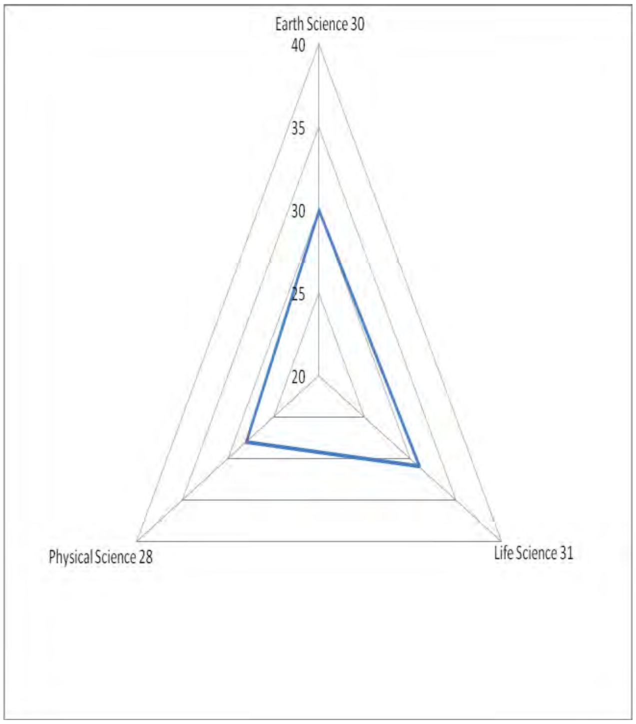


Boston

Charlotte



Cleveland



Percentage of Omitted Items by Item Type

The team also examined the percentage of items that were left blank—i.e., omitted items—by item type. Table 4c.3 shows the average percentage of omission rates by item type in grades four and eight. As was the case in reading and math, the omission rates in science were higher on constructed-response (CR) than on multiple-choice (MC) items. At grade four, the omission rates on MC items ranged from 0.8 percent in Atlanta to 1.8 percent in Boston. The omission rates on constructed-response items in grade four ranged from 3.2 percent in Charlotte to 5.6 percent in Cleveland. Among fourth graders in the four selected districts and in large-city (LC) schools, the omission rates on MC items were higher than the national public school rate, with the exception of Atlanta, where the omission rates were similar to the national rate. Omission rates on CR items were typically higher in the selected districts and large cities than national averages with the exception of Charlotte, which was lower and Atlanta, which appeared similar.

In grade eight, the omission rates on multiple-choice items ranged from 0.5 percent in Charlotte to 1.1 percent in Boston. The omission rates on constructed-response items in grade eight ranged from 4.6 percent in Charlotte to more than twice that rate in Cleveland, 9.6 percent. The omission rates among eighth graders on MC and CR items were greater in the selected districts and large-city schools than in the nation.

Table 4c.3 Item omission rates on NAEP science, by item type, grade, and district, 2005

	Grade 4		Grade 8	
	MC items	CR items	MC items	CR items
Atlanta	0.8	3.9	0.6	8.9
Boston	1.8	5.3	1.1	8.8
Charlotte	1.0	3.2	0.5	4.6
Cleveland	1.2	5.6	0.8	9.6
LC	1.1	4.7	0.5	6.5
National Public	0.8	3.9	0.4	4.5

Note: MC=multiple-choice, CR=constructed-response

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Science Assessment.

Percentage of Correct Items by Item Type

Next, we examined the average percentage of correct items by item type in each of the four selected districts and compared the percentages to the national public sample and the LC averages. Table 4c.4 displays the average percent-correct rates by item type in grades four and eight. In grade four, the percent-correct rates ranged from 45 percent (Cleveland) to 53 percent (Charlotte) on multiple-choice items and from 29 percent (Cleveland) to 36 percent (Charlotte) on constructed-response items. In each of the four selected districts and in the large-city schools, the percent-correct rate was lower than in the nation on both multiple-choice and constructed-response items. Additionally, each district—and the nation—had a higher rate of correct responses on MC than on CR items.

In grade eight, the percent-correct rates ranged from 44 percent (Atlanta and Cleveland) to 52 percent (Charlotte) on multiple-choice items and from 25 percent (Atlanta) to 34 percent (Charlotte) on constructed-response items. Every district—and the nation—had a higher rate of correct responses on MC than on CR items. However, the percent-correct rates for CR items in all four districts were somewhat higher in grade four than in grade eight. The largest difference was in Atlanta, where the percent-correct rate in CR items was 31 percent in grade four and 25 percent in grade eight. On both MC and CR items, the percent-correct rate in each of the four selected districts and in large cities was lower than in the nation.

Table 4c.4 Percent-correct rates on NAEP science, by item type, grade, and district, 2005

	Grade 4		Grade 8	
	MC items	CR items	MC items	CR items
Atlanta	47%	31%	44%	25%
Boston	46	31	48	30
Charlotte	53	36	52	34
Cleveland	45	29	44	26
LC	48	32	48	30
National Public	54	38	54	36

Note: MC=multiple-choice, CR=constructed-response
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2005 Science Assessments.

Part 2. Potential Factors Behind Subscale Science Trends

To help us further understand the science results, we explored two hypotheses on reasons for student science performance overall and at the subscale levels. First, we examined the alignment of state and/or district science standards with the NAEP science specifications by field.

Second, the research team conducted site visits to the four selected districts to see what they were doing instructionally that would help explain the NAEP science scale scores. The methodology for both parts of this chapter is described in chapter 3 and in appendices C and D.

Alignment of State and District Standards to NAEP Science Specifications

The purpose of this part of the analysis was to determine how well each state's or district's science content standards were aligned with the NAEP specifications. In other words, to what degree was the content encompassed by the NAEP specifications covered completely or partially by state or district standards? This analysis was done using the science-content specifications and item ideas found in the *Science Assessment and Exercise Specifications for the 1994 National Assessment of Educational Progress*, published by National Assessment Governing Board. The analysis team also examined the relevant state science standards and, in the case of Boston, district science standards, in place during the 2004-2005 school year.

Grade four science standards for Atlanta/Georgia were not available and therefore, not coded. In addition, data were not available for Boston on the life science strand and therefore not coded.

In addition, one should note that state science standards in Massachusetts and Ohio were written by grade bands. Thus, there is the potential that estimates of matching may be artificially high in these two states because it was not possible to determine what content was taught in grades three and four versus grade five. Where separate grade-level standards were available (i.e., North Carolina), we examined the degree of match in grade five but did not include them in the calculations of the percentage of overlap.[19]

[19] Another method for comparing the percentage of matches among these three states/districts at grade 4 could be to add the number of matching grade 5 standards to the current number for Charlotte/North Carolina. This action

Finally, the data on Boston, whose standards matched completely or partially at very low levels seems to suggest that the district's objectives were the least aligned of the four districts. Between fall 2003 and spring 2006, however, Boston phased in a new districtwide science curriculum that consisted of units produced by the Full Option Science System (FOSS) and Science and Technology for Children (STC). Only those units implemented during the 2003-2004 and 2004-2005 academic years, i.e., prior to the administration of NAEP, were included in the alignment study.

In addition, the curricular materials associated with the STC provided lists of content topics (many of which were aligned to the NAEP specifications), but the materials did not describe the topics in a way that allowed one to determine what content students were expected to learn by the end of the unit. These two factors led to the exclusion of multiple FOSS and STC curricular units from the alignment study and may have resulted in an underestimation of the actual overlap between the Boston curriculum and the NAEP specifications. It is important to note that the standards in Massachusetts showed a higher degree of complete and partial alignment (62 percent) than the other districts/states, so the findings in Boston may not provide an accurate picture of the district's taught curriculum. It was also impossible to determine whether teachers were using the state standards, the district's objectives, or some combination of the two—a caution that also applies to reading and math.

Degree of Content Match

Fourth-grade Science

Our analysis on grade four science showed that between 19 percent and 57 percent of the NAEP specifications were either completely or partially matched with local/state standards in three of the four jurisdictions (data were not available for Atlanta/Georgia at grade four). The highest overall matches appeared to be in Cleveland/Ohio. These results are shown in figure 4c.3 and table 4c.5 and described in detail in the bullets below.

There were 157 NAEP specifications in fourth-grade science, and matches varied considerably across jurisdictions.

- Boston, which had different standards than its state, matched 30 (19 percent) of the 157 NAEP specifications, with six complete and 24 partial matches. Therefore, some 4 percent of the 157 NAEP specifications were completely aligned with the Boston standards. However, 23 percent of the NAEP specifications aligned completely with the Massachusetts standards (19 percentage points above Boston's rate).

- Charlotte/North Carolina's standards matched 74 (47 percent) of the 157 specifications, with 15 complete and 59 partial matches. Therefore, some 10 percent of the 157 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- Cleveland/Ohio standards matched 90 (57 percent) of the 157 specifications, with 76 complete and 14 partial matches. Therefore, some 48 percent of 157 NAEP specifications were completely aligned with the Cleveland/Ohio standards.

---

would result in a much higher percentage of matches (47 percent + 24 percent matched in grade 5 = 71 percent of NAEP specifications matched). An overall match of 71 percent would result in Charlotte/North Carolina having the highest overlap with NAEP.

In general, the overall degree of complete and partial content matches in fourth-grade science for the three jurisdictions was low, except for Cleveland/Ohio, which was modest. The fourth-grade complete and partial content alignment for science was lower than the content matches for reading and math.

If we examine the three science strands—earth science, physical science, and life science—the patterns show a complex picture.

There were 63 NAEP specifications in the *earth science* subscale in fourth grade.

- Boston matched 10 (16 percent) of the 63 subscale specifications, with five complete and five partial matches. Therefore, only 8 percent of the 63 NAEP specifications were completely aligned with the Boston standards.
- Charlotte/North Carolina matched 22 (35 percent) of the 63 subscale specifications, with five complete and 17 partial matches. Therefore, 8 percent of the 63 NAEP specifications were completely aligned with the Charlotte/North Carolina standards—the same as Boston.
- Cleveland/Ohio matched 29 (46 percent) of the 63 specifications, with 24 complete and five partial matches. Therefore, 38 percent of the 63 NAEP specifications were completely aligned with the Cleveland standards.

There were 50 NAEP specifications in the subscale on *physical science* in fourth grade.

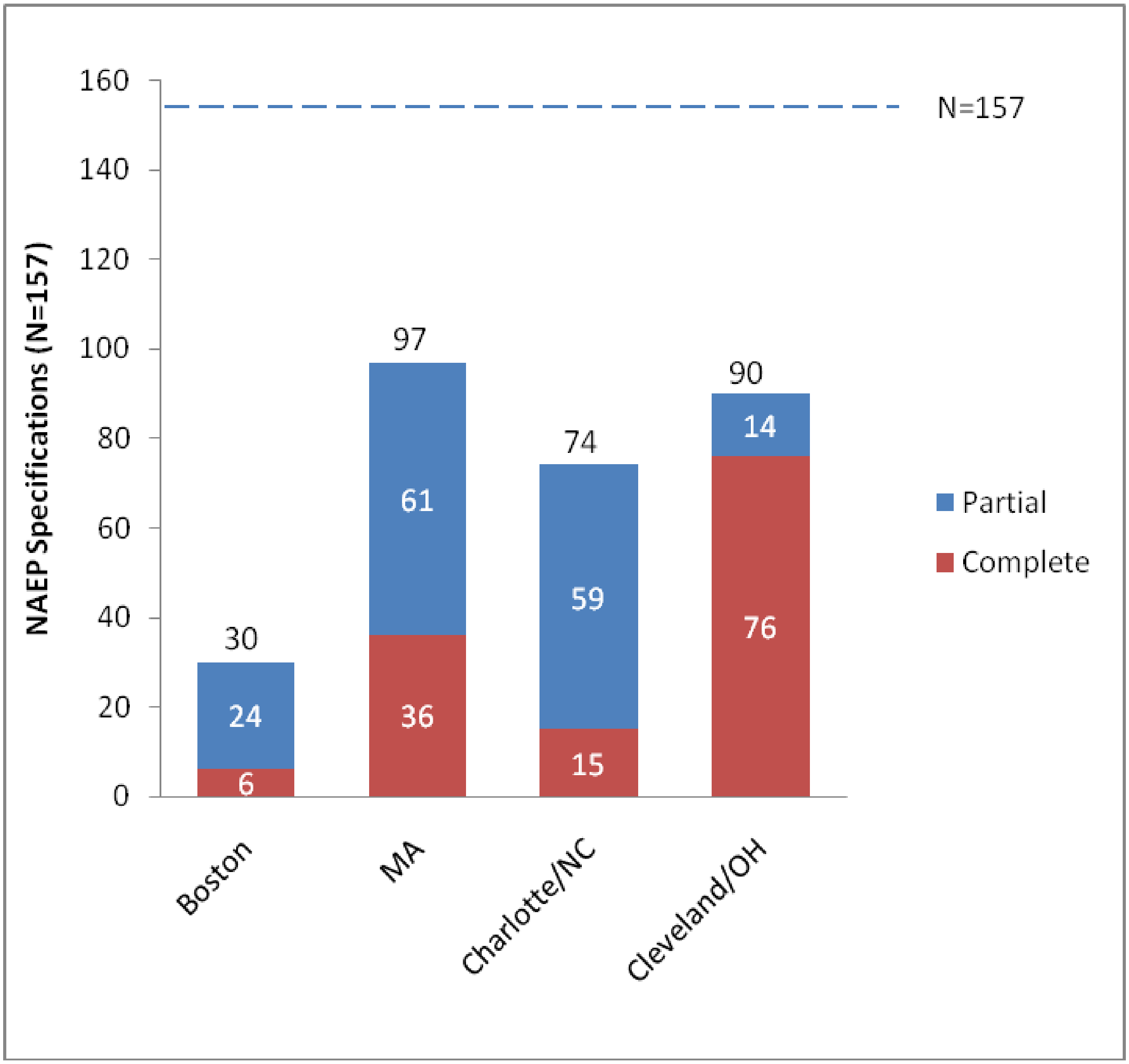
- Boston matched 20 (40 percent) of the 50 subscale specifications, with only one complete and 19 partial matches. Therefore, only 2 percent of the 50 NAEP specifications were completely aligned with the Boston standards.
- Charlotte/North Carolina matched 16 (32 percent) of the 50 subscale specifications, with only one complete and 15 partial matches. Therefore, only 2 percent of the 50 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.
- Cleveland/Ohio matched 26 (52 percent) of the 50 subscale specifications, with 19 complete and seven partial matches. Therefore, 38 percent of the 50 NAEP specifications were completely aligned with the Cleveland/Ohio standards—the same alignment that the state/city had on the earth science subscale.

There were 44 NAEP specifications in the subscale on *life science* in fourth grade, 19 fewer than in earth science. As explained previously, Boston's results on this subscale could not be coded.

- Charlotte/North Carolina matched 36 (82 percent) of the 44 subscale specifications, with nine complete and 27 partial matches. Therefore, only 20 percent of the 44 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.
- Cleveland/Ohio matched 35 (80 percent) of the 44 subscale specifications, with 33 complete and two partial matches. Therefore, 75 percent of the 44 NAEP specifications were completely aligned with the Cleveland standards.

In sum, Cleveland/Ohio showed the highest overall level of alignment (57 percent) between the NAEP specifications and its standards in fourth-grade science, and Boston had the lowest overall level of alignment (19 percent) for complete and partial matches. The alignment in earth science was highest in Cleveland/Ohio (46 percent) and the lowest in Boston (16 percent). The alignment in physical science was highest in Cleveland/Ohio (52 percent) and lowest in Charlotte/North Carolina (32 percent). And the alignment in life science appeared similar for both Charlotte/North Carolina (82 percent) and Cleveland/Ohio (80 percent). The degree of content match was highest in life science in the two districts measured. But overall in science, if one looks solely at complete matches, the alignment only exceeded 50 percent once, i.e., in life sciences in Cleveland/Ohio, and, in fact, was often in single digits.

Figure 4c.3 Number of complete and partial matches with NAEP grade 4 science specifications, by selected districts (*N* of NAEP specifications = 157), 2005*



*30 (19 percent) of Boston's grade 4 science standards matched NAEP's 157 science specifications either completely or partially; 97 (62 percent) of Massachusetts's grade 4 science standards matched NAEP's 157 science specifications either completely or partially; 74 (47 percent) of Charlotte's grade 4 science standards matched NAEP's 157 science specifications either completely or partially; and 90 (57 percent) of Cleveland's grade 4 science standards matched NAEP's 157 science specifications either completely or partially.

Table 4c.5 Degree of match with NAEP grade 4 science specifications/expectations/indicators, by subscale and district, 2005

State and District # / % of Matching Specifications, and Breakdown of Complete and Partial Matches	Subscale:	Earth Science		Physical Science		Life Science		Total	
		63		50		44		157	
	Atlanta/ GA*	– –	– –	– –	– –	– –	– –	– –	– –
		– –	– –	– –	– –	– –	– –	– –	– –
	Boston**	10	16%	20	40%	– –	– –	30	19%
		C = 5	P = 5	C = 1	P = 19	– –	– –	C = 6	P = 24
	MA	37	59%	29	58%	31	70%	97	62%
		C = 13	P = 24	C = 8	P = 21	C = 15	P = 16	C = 36	P = 61
	Charlotte/ NC	22	35%	16	32%	36	82%	74	47%
		C = 5	P = 17	C = 1	P = 15	C = 9	P = 27	C = 15	P = 59
	Cleveland/ OH	29	46%	26	52%	35	80%	90	57%
		C = 24	P = 5	C = 19	P = 7	C = 33	P = 2	C = 76	P = 14

* Grade 4 standards for Atlanta/GA were not available and not coded in science.

** For reasons described in the text, not all Boston standards were included in the analysis.

Note: How to read this table. Example: There are 63 specifications in *Earth Science*. Boston's standards matched the content of 10 of those 63 specifications or 16 percent. Five of the matches were complete content matches and five were a partial content match.
C=complete match; P=partial match

There was almost no variation among the three fields of science in the degree of complete content matching within and across districts. Only Cleveland had a complete match that was higher than "low" in any of the three fields of science in grade four, life science. (See table 4c.6.)

Table 4c.6 Degree of complete match of NAEP subscales with district/state standards in grade 4 science, by subscale and district, 2005*

	District/State			
Field of Science	Boston	MA	Charlotte/NC	Cleveland/OH
Earth Science	Low	Low	Low	Low
Physical Science	Low	Low	Low	Low
Life Science	– –	Low	Low	Moderate

***** High (80 percent or more) and low (50 percent or less)
Note: Data on Atlanta for 2005 were not available to the research team.

Eighth-grade Science

Our analysis of grade eight science showed that between 25 and 48 percent of the NAEP specifications were either completely or partially matched by the local/state standards in the four selected jurisdictions. There were 222 total NAEP science specifications in grade eight. These results are shown in figure 4c.4 and table 4c.7 and described in detail in the bullets below.

- Atlanta/Georgia standards matched 77 (35 percent) of the 222 NAEP specifications, with 28 complete and 49 partial matches. Therefore, some 13 percent of the 222 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- Boston, which had slightly different standards than its state, matched 55 (25 percent) of the 222 NAEP specifications, with only 14 complete and 41 partial matches. Therefore, some 6 percent of the 222 NAEP specifications were completely aligned with the Boston standards. Massachusetts, on the other hand, had a complete-match rate of 21 percent—15 percentage points higher than Boston. (Some of this lack of alignment may have been due to the fact that the physical science portion of the new curriculum was not implemented until the 2005-06 academic year and therefore not included in the analysis.)

- Charlotte/North Carolina's standards matched 104 (47 percent) of the 222 specifications—the same level of alignment as in grade four—with 47 complete and 57 partial matches. Therefore, some 21 percent of the 222 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- Cleveland/Ohio standards matched 106 (48 percent) of the 222 NAEP specifications, with 53 complete and 53 partial matches. Therefore, some 24 percent of the 222 NAEP specifications were completely aligned with the Cleveland/Ohio standards.

In general, the overall degree of complete and partial content matches in eighth-grade science for the four jurisdictions was low.

If we examine the three science strands—earth science, physical science, and life science—the patterns were more complex.

There were 116 NAEP specifications in the *earth science* subscale in eighth grade.

- Atlanta/Georgia matched 13 (11 percent) of the 116 subscale specifications, with only four complete and nine partial matches. Therefore, 3 percent of the 116 NAEP specifications were completely aligned with the Atlanta/Georgia standards.

- Boston matched 31 (27 percent) of the 116 subscale specifications with only four complete and 27 partial matches. Therefore, only 3 percent of the 116 NAEP specifications were completely aligned with the Boston standards.

- Charlotte/North Carolina matched 51 (44 percent) of the 116 subscale specifications, with 19 complete and 32 partial matches. Therefore, 16 percent of the 116 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- Cleveland/Ohio matched 59 (51 percent) of the 116 subscale specifications, with 31 complete and 28 partial matches. Therefore, 27 percent of 116 NAEP specifications were completely aligned with the Cleveland standards.

There were 62 NAEP specifications in the *physical science* subscale in eighth grade.

- Atlanta/Georgia matched 35 (56 percent) of the 62 subscale specifications, with 14 complete matches and 21 partial matches. Therefore, 23 percent of the 62 NAEP specifications were completely aligned with the Cleveland/Ohio standards.

- Charlotte/North Carolina matched 31 (50 percent) of the 62 subscale specifications, with 15 complete and 16 partial matches. Therefore, only 24 percent of the 62 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- Cleveland/Ohio matched 25 (40 percent) of the 62 subscale specifications, with 12 complete and 13 partial matches. Therefore, 19 percent of the 62 NAEP specifications were completely aligned with the Cleveland/Ohio standards.

There were 44 NAEP specifications in the *life science* subscale in eighth grade.

- Atlanta/Georgia matched 29 (66 percent) of the 44 subscale specifications, with 10 complete and 19 partial matches. Therefore, 23 percent of the 44 NAEP specifications were completely aligned with the Atlanta's standards.

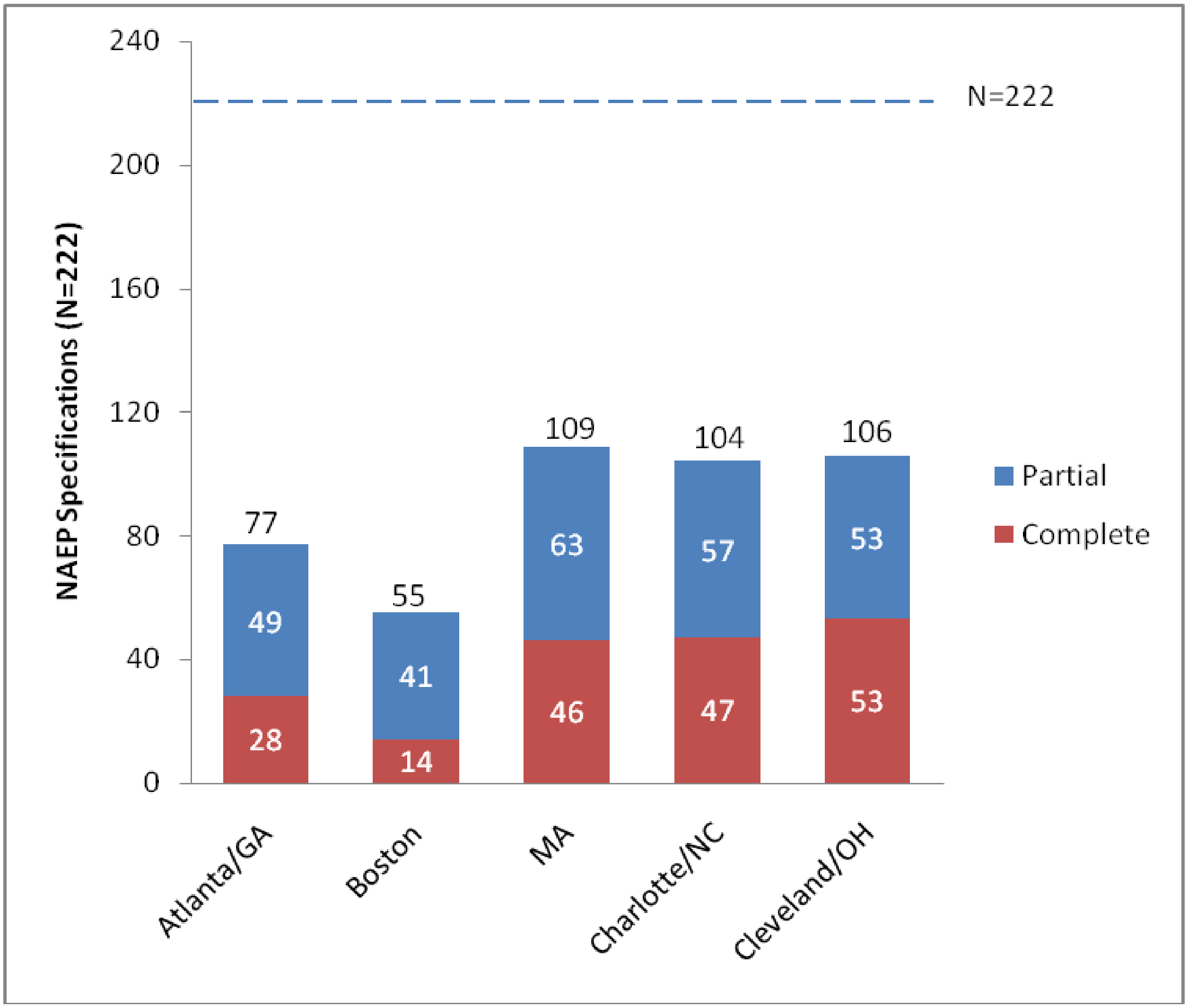
- Boston matched 24 (55 percent) of the 44 subscale specifications, with 10 complete and 14 partial matches. Therefore, 23 percent of the 44 NAEP life science specifications were completely aligned with the Cleveland standards.

- Charlotte/North Carolina matched 22 (50 percent) of the 44 subscale specifications, with 13 complete and nine partial matches. Therefore, 30 percent of the 44 NAEP specifications were completely aligned with the Charlotte/North Carolina standards.

- Cleveland/Ohio matched 22 (50 percent) of the 44 subscale specifications, with 10 complete and 12 partial matches. Therefore, 23 percent of the 44 NAEP specifications were completely aligned with the Cleveland standards.

The complete and partial alignment in life sciences at the eighth grade was highest in Atlanta/Georgia. The alignment in earth science was highest in Cleveland/Ohio and the lowest in Atlanta/Georgia. The alignment in physical science was highest in Atlanta/Georgia and lowest in Cleveland/Ohio. Cleveland/Ohio showed the highest overall level of complete and partial alignment (48 percent) between its standards and the NAEP specifications in eighth-grade science, and Boston had the lowest overall level of alignment (25 percent). Finally, complete matches between the NAEP specifications and the district/state standards on any subscale in grade eight never exceeded 30 percent and were as low as 3 percent.

Figure 4c.4 Number of complete and partial matches with NAEP grade 8 science specifications, by selected districts (*N* of NAEP specifications = 222), 2005*



*77 (35 percent) of Atlanta's grade 8 science standards matched NAEP's 222 science specifications either completely or partially; 55 (25 percent) of Boston's grade 8 science standards matched NAEP's 222 science specifications either completely or partially; 109 (49 percent) of Massachusetts's grade 8 science standards matched NAEP's 222 science specifications either completely or partially; 104 (47 percent) of Charlotte's grade 8 science standards matched NAEP's 222 science specifications either completely or partially; and 106 (48 percent) of Cleveland's grade 8 science standards matched NAEP's 222 science specifications either completely or partially.

Table 4c.7 Degree of match with NAEP grade 8 science specifications/expectations/indicators, by subscale and district, 2005

State and District # / % of Matching Specifications, and Breakdown of Complete and Partial Matches	Subscale:	Earth Science		Physical Science		Life Science		Total	
		116		62		44		222	
	Atlanta/GA	13	11%	35	56%	29	66%	77	35%
		C = 4	P = 9	C = 14	P = 21	C = 10	P = 19	C = 28	P = 49
	Boston*	31	27%	– –	– –	24	55%	55	25%
		C = 4	P = 27	– –	– –	C = 10	P = 14	C = 14	P = 41
	MA	56	48%	24	39%	29	66%	109	49%
		C = 22	P = 34	C = 9	P = 15	C = 15	P = 14	C = 46	P = 63
	Charlotte/ NC	51	44%	31	50%	22	50%	104	47%
		C = 19	P = 32	C = 15	P = 16	C = 13	P = 9	C = 47	P = 57
	Cleveland/ OH	59	51%	25	40%	22	50%	106	48%
		C = 31	P = 28	C = 12	P = 13	C = 10	P = 12	C = 53	P = 53

* In 2004-2005 Boston was in the midst of adopting a new districtwide curriculum. The physical science portion of the new curriculum was not implemented until 2005-2006.
Note: How to read this table. Example: There are 116 specifications in *Earth Science*. Atlanta's standards matched the content of 13 of those 116 specifications or 11 percent. Four of the matches were complete content matches and nine were a partial content match.
C=complete match; P=partial match

As shown in table 4c.8, no district or state had a high or moderate degree of complete content match with the NAEP specifications in grade eight. All matches were low.

Table 4c.8 Degree of complete match of NAEP subscales with district/state standards in grade 8 science, by subscale and district, 2005*

	District/state				
Field of Science	Atlanta/GA	Boston	MA	Charlotte/NC	Cleveland/OH
Earth Science	Low	Low	Low	Low	Low
Physical Science	Low	– –	Low	Low	Low
Life Science	Low	Low	Low	Low	Low

***** High (80 percent or more) and low (50 percent or less)

Degree of Match in Cognitive Demand

In addition to determining the degree of content match between local/state standards and NAEP specifications, the research team examined how well those completely matched standards corresponded in their cognitive demand or complexity to NAEP specifications. (See chapter 3 and appendices C and D for a detailed description of the methodology.) This entailed examining the wording of district/state standards and NAEP specifications to determine the cognitive demand or rigor in each statement and then comparing the results.

Tables 4c.9 and 4c.10 show the level of complete content match discussed in the previous section along with the number and percentage of state and local standards that were classified as low, moderate, or high on cognitive demand in fourth- and eighth-grade science. Only those standards that matched NAEP specifications completely were included in the analysis. This gives the reader a sense of the rigor or complexity of state and local standards but only for the portion of standards that match with NAEP. Omitted from the cognitive demand codes were all standards that did not correspond to NAEP.

First, the data in the tables indicate a range in the degree to which the level of cognitive demand in the state and district standards aligned with NAEP in both grades four and grade eight. Except in Boston, the cognitive demand of the completely matched standards in the selected districts appeared to be the same as or somewhat higher than the NAEP specifications in grades four and eight.

Tables 4c.9 and 4c.10 on grades four and eight, respectively, show that 49 percent of the grade four NAEP science specifications and 64 percent of the grade eight specifications were **moderate** in cognitive demand. Our analysis showed that the majority of the standards that matched the NAEP specifications were also **moderate** in cognitive demand—ranging from 0 percent (Boston) to 88 percent (Cleveland/Ohio) in grade 4, and from 29 percent (Boston) to 100 percent (Atlanta/Georgia) in grade eight. The degree of match for the selected districts at the **high** level of cognitive demand ranged in grade four from 0 percent (Boston and Massachusetts) to 47 percent (Charlotte/North Carolina) and in grade eight, from 0 percent (Atlanta/Georgia) to 47 percent (Charlotte/North Carolina).

Again, at both grade levels and in all jurisdictions, except Boston, the percentage of content-matched standards that were low in cognitive demand was smaller than the percentage of NAEP specifications.

To further quantify the degree of cognitive demand, the tables below also show weighted totals for each district based on assigning one point for low, two points for moderate, and three points for high cognitive demand. The weighted averages were derived by dividing the weighted total by the total number of completely matching specifications. This analysis suggests that the degree of cognitive demand in grade four science varied among the matching standards in the four selected districts from NAEP. For instance, Boston's weighted average in fourth grade was 1.0, a level that was lower than NAEP's 1.8 (the baseline).

At grade eight, the weighted averages indicated that the cognitive demand of NAEP was again lower than or equal to the weighted averages on all of the local/state standards, except Boston, which was 1.6. Charlotte's weighted average of 2.4 exceeded NAEP's baseline weighted average of 1.8.

Table 4c.9 Degree of match in cognitive demand for specifications with complete alignment on NAEP grade 4 science, by district, 2005

	NAEP		Boston		MA		Charlotte/ NC		Cleveland/ OH	
% of Complete Content Match	100%		4%		23%		10%		48%	
Cognitive Level										
Low	56	36%	6	100%	7	19%	1	7%	4	5%
Moderate	77	49%	0	0%	29	81%	7	47%	67	88%
High	24	15%	0	0	0	0%	7	47%	5	7%
Total	157	100%	6	100%	36	100%	15	100%	76	100%
Weighted Total	282		6		65		36		153	
Weighted Mean		1.8*		1.0		1.8		2.4		2.0

* Number represents the balance among NAEP standards that were determined to be high, moderate, or low cognitive demand. 1=low cognitive demand, 2=moderate cognitive demand, and 3=high cognitive demand.

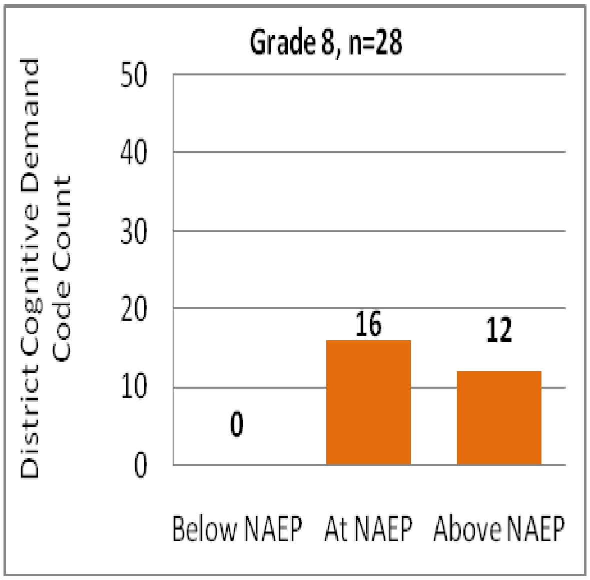
Table 4c.10 Degree of match in cognitive demand for specifications with complete alignment on NAEP grade 8 science, by district, 2005

	NAEP		Atlanta/ GA		Boston		MA		Charlotte/ NC		Cleveland/ OH	
% of Complete Content Match	100%		13%		6%		21%		21%		24%	
Cognitive Level												
Low	66	30%	0	0%	8	57%	10	22%	3	6%	1	2%
Moderate	142	64%	28	100%	4	29%	35	76%	22	47%	50	94%
High	14	6%	0	0%	2	14%	1	2%	22	47%	2	4%
Total	222	100%	28	100%	14	100%	46	100%	47	100%	53	100%
Weighted Total	392		56		22		83		113		107	
Weighted Mean		1.8*		2.0		1.6		1.8		2.4		2.0

* Number represents the balance among NAEP standards that were determined to be high, moderate, or low cognitive demand. 1=low cognitive demand, 2=moderate cognitive demand, and 3=high cognitive demand.

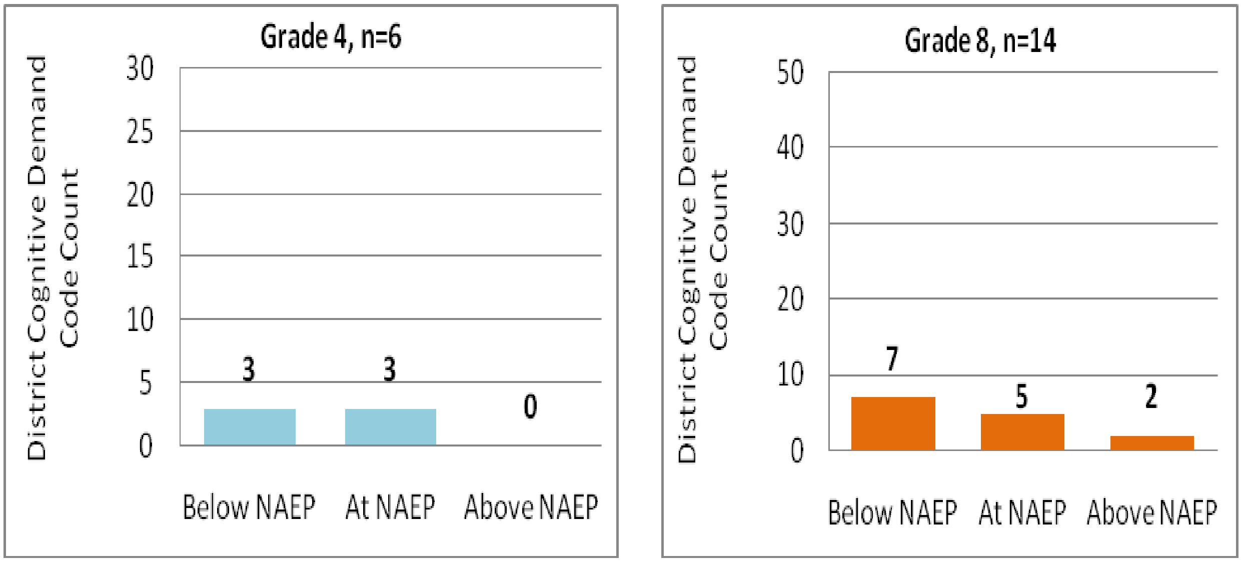
One additional way to capture the degree of alignment in cognitive demand was to directly compare the demand level of each *completely matched* district/state standard with that of the NAEP specification to which it was matched. Figures 4c.5 through 4c.13 present this information for grades four and eight in each of the four districts (only grade eight for Atlanta). In Charlotte, Cleveland, and Atlanta, most of the standards had a cognitive demand level that was similar to the matching NAEP specifications.

Figure 4c.5 *Atlanta's* complete matches at grade 8 science in cognitive demand compared to NAEP, 2005*



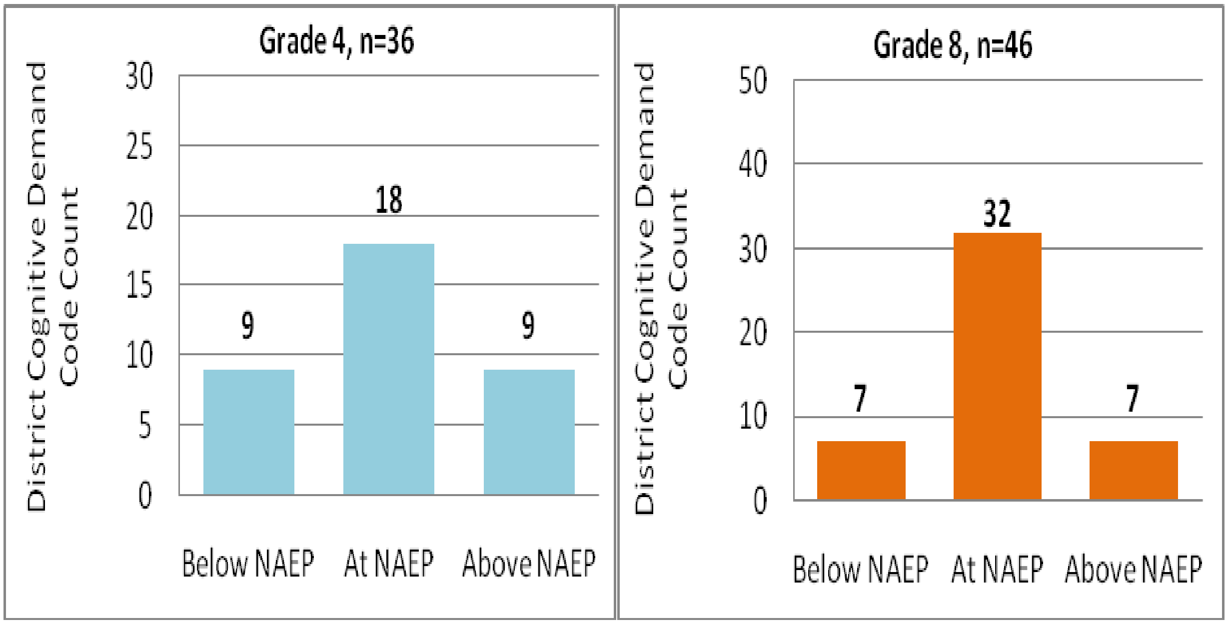
* 28 of Atlanta's grade 8 standards completely matched the 222 NAEP science specifications (13 percent). None of those 28 completely matched standards had a cognitive demand level below NAEP, 16 were at the NAEP level, and 12 were above NAEP.

Figures 4c.6 and 4c.7 *Boston's* complete matches at grades 4 and 8 science in cognitive demand compared to NAEP, 2005*



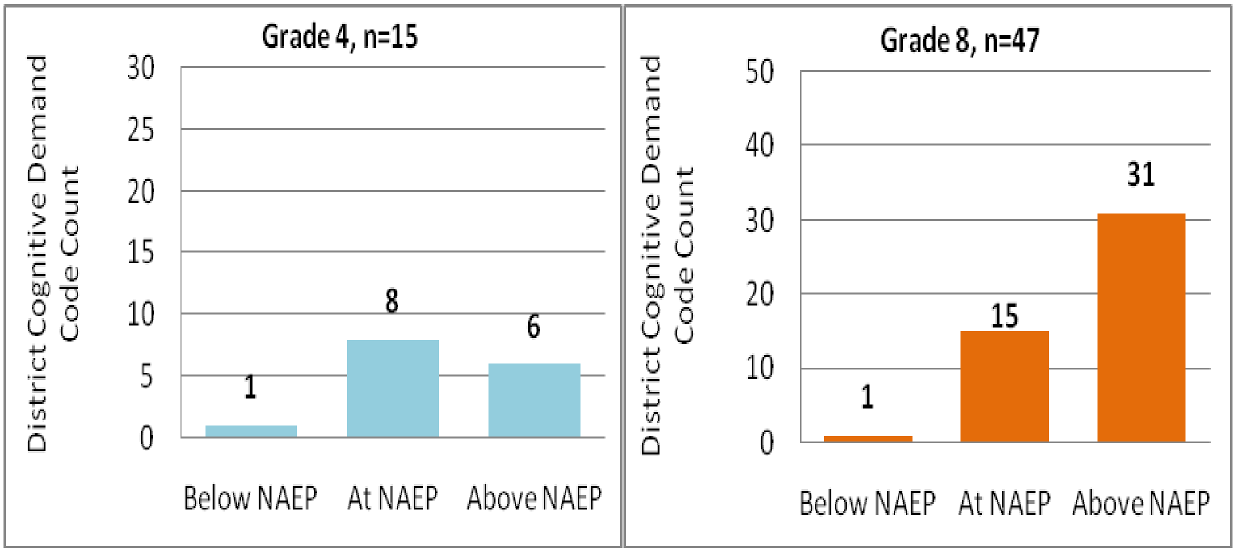
* 6 of Boston's grade 4 standards completely matched the 157 NAEP science specifications (4 percent). Three of those 6 completely matched standards had a cognitive demand level below NAEP, three were at the NAEP level, and none were above NAEP. Similarly, 14 of Boston's eighth grade standards completely matched the 222 NAEP science specifications (6 percent). Seven of those 14 completely matched standards had a cognitive demand level below NAEP, five were at the NAEP level, and two were above NAEP.

Figure 4c.8 and 4c.9 *Massachusetts's* complete matches at grades 4 and 8 science in cognitive demand compared to NAEP, 2005*



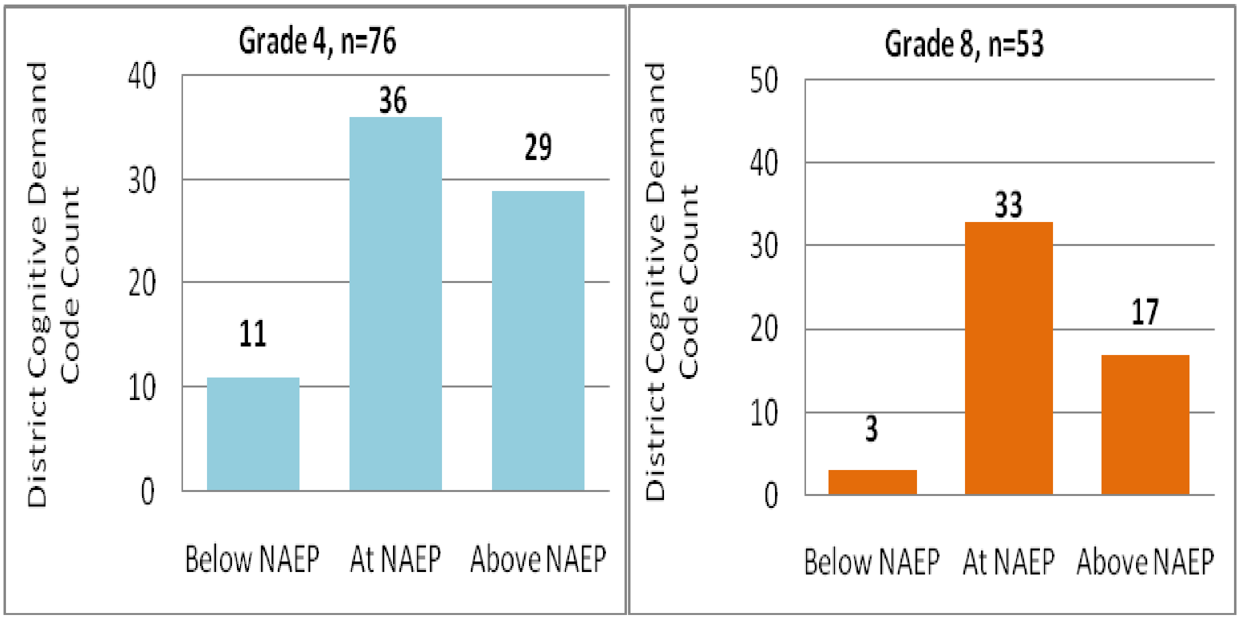
* 36 of Massachusetts's grade 4 standards completely matched the 157 NAEP science specifications (23 percent). Nine of those 36 completely matched standards had a cognitive demand level below NAEP, 18 were at the NAEP level, and nine were above NAEP. Similarly, 46 of Massachusetts's eighth grade standards completely matched the 222 NAEP science specifications (21 percent). Seven of those 46 completely matched standards had a cognitive demand level below NAEP, 32 were at the NAEP level, and seven were above NAEP.

Figures 4c.10 and 4c.11 *Charlotte's c*omplete matches at grades 4 and 8 science in cognitive demand compared to NAEP, 2005*



* 15 of Charlotte's grade 4 standards completely matched the 157 NAEP science specifications (10 percent). One of those 15 completely matched standards had a cognitive demand level below NAEP, eight were at the NAEP level, and six were above NAEP. Similarly, 47 of Charlotte's eighth grade standards completely matched the 222 NAEP science specifications (21 percent). One of those 47 completely matched standards had a cognitive demand level below NAEP, 15 were at the NAEP level, and 31 were above NAEP.

Figures 4c.12 and 4c.13 *Cleveland's* complete matches at grades 4 and 8 science in cognitive demand compared to NAEP, 2005*



* 76 of Cleveland's grade 4 standards completely matched the 157 NAEP science specifications (48 percent). 11 of those 76 completely matched standards had a cognitive demand level below NAEP, 36 were at the NAEP level, and 29 were above NAEP. Similarly, 53 of Cleveland's eighth grade standards completely matched the 222 NAEP science specifications (24 percent). Three of those 53 completely matched standards had a cognitive demand level below NAEP, 33 were at the NAEP level, and 17 were above NAEP.

Summary of Analysis of Science Standards Alignment and NAEP Results

In sum, we analyzed the degree of alignment between the NAEP grade four and eight science specifications and the state and district science standards for Atlanta, Boston, Charlotte and Cleveland on both content and cognitive demand. Our analysis showed varied results. The degree of complete and partial content match varied considerably across the states/districts at each grade level. Boston's standards seemed to have the lowest degree of content alignment with NAEP science at both grade levels, but this may have been partly because the standards associated with the STC curricular units were not fully included in the analysis. Moreover, the highest overall alignment was in Massachusetts, where state standards completely or partially match 62 percent of the NAEP science specifications at grade four.

While the Charlotte/North Carolina standards seemed to have low content match with NAEP, the matches specifically reflected grade three and four standards, while other states/districts and grade-band standards made it impossible to separate out grade five for the purpose of matching. In Massachusetts and Cleveland/Ohio, where the percentage of matches included the standards aligned at grades three, four, *and* five, our results may have overestimated the exposure of students to NAEP content prior to taking the assessment at grade four. Based on our analysis of the cognitive demand alignment, we saw that, in general, the districts did not appear to have standards that were significantly lower in cognitive demand than the NAEP specifications, with the possible exception of Boston, where a new science curriculum was being phased in during 2005 when the NAEP science exam was administered.

Finally, there appeared to be little relationship between the content and cognitive matches in science and the 2005 percentiles of each of the selected districts. (See tables 4c.11 and 4c.12.) Cleveland appeared to have the highest level of content match in fourth-grade science but had the lowest of the four districts on the overall science composite percentile. (Again, there were no trend data.) Charlotte appeared to have the highest science composite score and the highest overall cognitive demand relative to NAEP, but its level of complete content match with NAEP was only 10 percent in the fourth grade. The results in eighth grade were similarly unrelated. Charlotte and Cleveland appeared to have similar levels of complete content matches in science, but Charlotte's composite science score was substantially higher than Cleveland's.

Table 4c.11 Summary statistics on NAEP science in grade 4

Study District	2005 Unadjusted Composite Percentile	Percentage Complete Content Match with NAEP	Weighted Cognitive Demand Mean for Complete Content Matches (baseline 1.8)
Atlanta	29	NA	NA
Boston	29	4%	1.0
Charlotte	42	10%	2.4
Cleveland	25	48%	2.0
LC	--	---	--
National Sample	50	--	1.8

Table 4c.12 Summary statistics on NAEP science in grade 8

Study District	2005 Unadjusted Composite Percentile	Percentage Complete Content Match with NAEP	Weighted Cognitive Demand Mean for Complete Content Matches (baseline 1.8)
Atlanta	20	13%	2.0
Boston	31	6%	1.6
Charlotte	42	21%	2.4
Cleveland	24	24%	2.0
LC	--	--	--
National Sample	50	--	1.8

Site Visits and Linkages to Science Results

As indicated earlier, the research team conducted site visits to the four selected districts to examine practices and policies that could help explain NAEP science performance in grades four and eight in 2005. A description of the methodology and the protocols used during these site visits is included in chapter 3 and appendix E. Individuals interviewed and materials reviewed are listed in appendix I.

Unlike in reading and mathematics, examining the instructional programming of the selected districts in order to draw linkages to the NAEP science results, particularly the subscale and alignment results, presented a particular challenge. None of the selected districts, including Charlotte, which saw the highest overall science scale scores on NAEP in 2005, placed anywhere near the same emphasis on science instruction during the study period (2003 to 2007) as they did on reading and math, making it nearly impossible to draw any conclusions as to the instructional roots of the district's differing levels of science achievement.

Adding to the difficulty in drawing instructional linkages was the fact that science data on 2005 was the only information available when this analysis was conducted, so there was no way to gauge district progress. Data from the 2009 science testing were available in February 2011, but the results were not comparable to the 2005 data because each testing involved differing frameworks.

In the next chapters, we examine the broader contextual features of the four districts and the particular instructional practices of the school systems. We also synthesize the results from this and earlier chapters into a more cohesive picture of why student achievement scores on NAEP may have improved or failed to improve.

CHAPTER 5

POLICIES, PROGRAMS, AND PRACTICES OF THE SELECTED DISTRICTS

Introduction

The four TUDA districts 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—that 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.[1] (See table 5.1 for a summary of key characteristics of district reforms.) Detailed case studies of Atlanta, Boston, and Charlotte-Mecklenburg are provided in appendices F, G and H.

Atlanta

Atlanta showed significant and consistent gains in reading throughout the study period.[2] 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 (CSRD),[3] and bolstered by a system of regionally based School Reform Teams (SRT) 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

[1] 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.

[2] 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.

[3] The district used Success for All, Direct Instruction, America's Choice, Modern Red School House, Co-Nect, Middle Schools that Matter, High Schools that Matter, and IB. Some schools also used the Open Court reading program.

school system, creating a common vocabulary for reading instruction and providing extensive site-based and cross-functional support through literacy coaches and professional 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 the reading reforms did not benefit from the unanimity of approach observable in the district's later work in math. The district's literacy program, which was built around a Reading and Writing Workshop (RWW) model during the study period, appeared to be less well-defined and less focused than the district's math reforms. In addition, the study team noted from interviews with teachers and district leaders that philosophical differences at the central office level over approaches to literacy instruction contributed to a lack of coherence in reading instruction districtwide. 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 math, 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 math 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 even when initial results showed sparse 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 math 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, it out-performed all other TUDA districts in reading and math 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 math 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 2006, 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 that were common to 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.* Boston, Atlanta, 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.

In addition, the study team examined issues related to spending levels, governance, and staffing levels to determine whether these variables showed discernible patterns and might have been related to whether a district showed improvement on NAEP.

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 and energetic new superintendent, Beverly Hall, 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 (Elizabeth Reilinger and now Gregory Groover), who had strong working relations with the former and current superintendents—Tom Payzant and Carol Johnson, respectively. The board used its mandate for improvement to spearhead a comprehensive five-year plan in 1996 that focused on strengthening student achievement and advancing standards-based instructional practice. No doubt, the leadership of the district was also spurred by state action in 1998 to require students to pass the Massachusetts exams in order to graduate. Much of the original plan remains intact, though with substantial enhancements in reading, under the leadership of Superintendent Carol Johnson.

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 under superintendent John Murphy, and the district benefited subsequently from a series of strong superintendents—Eric Smith, James Pughsley, and Frances Haithcock, who focused on instructional issues even as the district was settling one of the nation's longest running court-ordered school desegregation cases. A new theory of action was pursued in the district under superintendent Peter Gorman.

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. Along the way, according to focus-group participants, teachers districtwide began to embrace the changes.

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, Barbara Byrd-Bennett. 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 as a whole 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 district staff members that did not seem 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 math 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 appeared to be at the heart of reforms in Atlanta, Boston, and Charlotte. These districts 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 commonly understood 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. It set in place 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—worked under 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 in the late 1990s 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 between district staff members 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 urban school districts that are very low-performing and just beginning their reforms to put 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, each demonstrated 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’s approach to goal setting and accountability did not go much beyond meeting NCLB school 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. And student academic gains figured minimally into principals’ and teachers’ evaluations 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 the senior staff’s analyzing and 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. Veteran 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, began its math reforms after its initial state test scores on which its students would need to pass in order to graduate came back and the district realized it needed to revamp its math programming. A study group was initiated and the small numbers of schools that were making gains were examined to determine reasons for improvements. A decision was made to apply for a National Science Foundation (NSF) grant and sessions with administrators and teachers were established to discuss why math instruction needed to change. Based on the gains in the limited number of schools and the NSF grant, the district 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 math 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 math professional development 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 did not develop in Cleveland during the study period, although the district put into place the Cleveland Literacy System and adopted the Harcourt *Trophies* reading program in selected grades. In fact, there was no published curriculum in place when Eugene Sanders 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. Although there was some use of textbook materials and lesson plans built around the standards, not everyone used them, and the new reading series was adopted initially only for grades k-3 because of 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 math and the state standards, which, as 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 on-going 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 teacher quality also played important roles 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 throughout the school system. The district, which allows 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 throughout the district, as well as laying out "26 best practices" to 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 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 for teachers who then trained other teachers at their schools. 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, its purpose was largely designed around monitoring teachers' 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 largely tracked 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 this study's 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 core 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 on a series of individual school audits involving classroom observations. The goals of these audits were (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 are 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 gave principals the skills and knowledge to become instructional leaders of their schools after freeing them from some of their responsibilities for site management and operations.

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 math 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 experts 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 seemed to be no 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 math 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 district research staff indicated helped focus instructional strategies around results. Moreover, Boston 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, data from benchmark tests in Cleveland were not viewed 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 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.[4]

Governance, Spending, and Staffing Levels

Finally, we examined a number of other features of the four study districts to see if there were discernible patterns in the governance structures, staffing levels, or spending amounts that might be related to district gains or lack of gains on NAEP.

The four districts had a variety of different governance structures. Atlanta and Charlotte both had traditionally elected school boards, while Boston and Cleveland had boards of education that were appointed by their respective cities' mayors. The mayors in the latter two cities played strong leadership roles in their schools, particularly Boston's Mayor Tom Menino, whose term spanned two superintendents in the city. Cleveland, on the other hand, had a number of mayors over the years, including during the study period of 2003 to 2007.

Moreover, two of the cities—Boston and Cleveland—have a large number of choice options for parents. During the study period, there were numerous charter schools in both cities, and Cleveland also had a court-approved private school voucher program that served several thousand students and might have provided competition to the traditional school system. Neither Atlanta nor Charlotte had large numbers of charter schools or voucher initiatives.

Three of the four cities, moreover, were financially independent—Atlanta, Charlotte, and Cleveland. Boston, on the other hand, was financially dependent on its general-purpose unit of government for its locally derived revenues. The average per-pupil expenditures (unadjusted for regional cost-of-living differences) of the districts also varied. In 2007, amounts ranged from $8,081 in Charlotte to $19,435 in Boston. Atlanta's per-pupil expenditure was $12,745 and Cleveland's was $11,383. (See appendix B, table B.46.) During the 2003 to 2007 study period, per pupil expenditures of the districts, in inflation-unadjusted dollars, rose between 11.5 percent (Atlanta) and 41.6 percent (Boston). Cleveland's per-pupil expenditures rose by 11.6 percent and Charlotte's by 12.4 percent in inflation-unadjusted dollars.

The percentage of those total per-pupil expenditures devoted to instruction also showed some variation, but all were at or above 50 percent:[5] 61.8 percent in Charlotte, 59.8 percent in Cleveland, 57.3 percent in Boston, and 54.4 percent in Atlanta. All districts showed increases in instructional spending between 2003 and 2007: 42.0 percent increase in Boston, 17.8 percent in Cleveland, 12.4 percent in Charlotte, and 7.7 percent in Atlanta.

The staffing levels of the four districts also showed variation over the study period. In 2007, the percentage of all staff members who were teachers in the four districts ranged from 43.3 percent in Cleveland to 60.8 percent in Boston. (See appendix B, table B.47.) In Atlanta, 53.5 percent of all staff members were teachers and in Charlotte, 53.2 percent. In addition, the teacher-to-pupil

[4] 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.

[5] Source: National Center for Educational Statistics. Instructional expenditures include "payments from all funds for salaries, employee benefits, supplies, materials, and contractual services for elementary/secondary instruction. It excludes capital outlay, debt service, and interfund transfers for elementary/secondary instruction. Instruction covers regular, special, and vocational programs offered in both the regular school year and summer school. It excludes instructional support activities, as well as adult education and community services. Instructional salaries include salaries for teachers and teacher aides and assistants."

ratios in the districts were very similar, except in Cleveland. The ratios were 13.7:1 in Atlanta and Charlotte and 13.2:1 in Boston. In Cleveland, however, the ratio was 15.8:1.

In 2003, teachers comprised 50.4 percent of the Cleveland school district's workforce, a level similar to that of other study districts, and the city had a pupil-to-teacher ratio of 10.7:1—the lowest among the four districts. However, during the 2003 to 2007 study period, while the other three districts were able to maintain or lower their pupil-to-teacher ratios, Cleveland made substantial budget cuts that affected both the size and the deployment of its teacher workforce often into unfamiliar grades, changes that may be related to the district's difficulty in raising its NAEP scores. These budget cuts also reportedly resulted in schools having to use their own funds to replace "consumables," something that was not always done.

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 was 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.

Atlanta had outstanding and long-serving leadership at the superintendent level that defined the overall academic direction of the school system, gave latitude over programming to schools in order to build ownership for reforms, strengthened staff capacity throughout the district, and glued the work together with data, consistency, and accountability for results.

Boston, on the other hand, started the study period farther ahead of Atlanta on its reforms and its NAEP scores. Boston's story was similar to Atlanta's at the broad strategic level in the sense that improvements were driven by strong and cohesive leadership and policy unity, the rise and use of data to inform instructional change, and professional development and coaching that built instructional capacity, consistency, and follow-up. At the tactical program level, however, Atlanta and Boston focused on different subjects, which may have contributed to why their reading and math results improved at differing rates.

Charlotte did not see significant gains in reading or math over the study period but maintained high achievement levels, even after adjusting for student background characteristics. Its instructional program was similar to both Atlanta's and Boston's.

In contrast, Cleveland had very low academic achievement on NAEP and did not show significant improvements in most grades and subjects during the 2003 to 2007 study period. The school district underwent substantial instructional change during the period, but it also saw significant budget cuts and grade and school reconfigurations that resulted in major personnel redeployments that may have undercut the positive effects the instructional reforms might otherwise have had.

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, for it was the joint force of these reforms and how they locked together in Atlanta, Boston, and Charlotte that appeared to make all the difference in better student achievement.

***

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.
	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.
Goal-setting	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 6

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. Moreover, our analysis indicated that gains among students in large-city schools were significantly larger than gains in the national sample, suggesting that urban schools may be catching up with national averages. Otherwise, we have shied away from characterizing the size of urban gains except to note that two of the study districts—Atlanta and Boston—had effect sizes between 2003 and 2007 in reading and math, respectively, that were several times larger than either the large-city school or national samples.

The findings in this report have special import because they suggest some reasons for these gains, although the reader is cautioned against assuming causal links in the results because of the limited number of study districts. The analysis also suggests steps that might be required to accelerate this progress, 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 schools move forward. Debate continues, of course, about what separates urban school systems that make major progress from those making more incremental gains 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. The reader should keep in mind the limitations to the alignment analysis that we pointed out in chapter 4, but what we found was far more complex than what we had originally anticipated.

Essentially, the analysis found that the content alignment or match in reading and math between the NAEP frameworks and state standards in the four study districts was low or moderate. (We did not define what good alignment was other than to designate a content match above 80 percent as high.) In general, North Carolina appeared to have the most consistently aligned standards in reading and grade four math, and it also had the highest overall performance, but it is difficult to draw a causal relationship between alignment and performance. In addition, there was no apparent relationship between the degree of content match and the likelihood that a district would see gains or losses on NAEP in either reading or math. It is possible, however, that the intersection of content and rigor may have greater import than either one alone. In all, it appeared that content alignment on its own was insufficient in the small sample we studied to affect movement on student NAEP scale scores in the four city school systems.

Moreover, it was clear from the results of this analysis that student improvement on NAEP was related less to content alignment than to the strength or weakness of a district's instructional programming. Two of the districts with significant and consistent gains on NAEP—Atlanta and Boston—were most likely able 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 or better 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 preliminary finding has significant implications for the new Common Core State Standards, which some 45 states have now adopted. Many educators—and the public in general—assume that putting into place more demanding standards alone will directly result in higher 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, integrated, consistent with the standards, and well coordinated. Moreover, our findings strongly suggest that the manner in which the common core standards are implemented and 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.

Social scientists have long puzzled over how to attain significant effects at scale. Many observers have concluded that it is pointless to try to affect social policy by developing innovations at small scale and then trying to ramp up isolated examples of excellence one project or school at a time. Yet, the education reform movement has been grounded for years on the supposition that progress was attainable mostly at the school level and that considering school districts as major units of large-scale change was largely a waste of time. However, this study suggests 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 attempted to act at scale. Yet, Cleveland was also more inclined to grant staff, principals, and teachers instructional autonomy, and lacked 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 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 math 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 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 most likely 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 math, 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 tactical decisions regarding how to implement the programs 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 math, 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.

For example, there was a striking contrast between the Atlanta and Boston school districts in how they handled their reading reforms. Both began their systemwide reforms with literacy, but Atlanta's reading initiative never wavered from its initial vision, even as it continued to refine its practices. Boston's reading program, on the other hand, was splintered over philosophical disagreements about the correct approach to literacy instruction—disagreements that had practical ramifications for programming and support at the school and classroom levels.

In contrast, the Boston school district's math initiatives were more like Atlanta's reading effort. Boston laid out a strong, unified vision for its improvements in math, rolled out its program in a deliberate fashion, and then sustained support for the initiative and its implementation, even in the face of pushback against it.

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, energetic leadership teams, and these leaders' vision for improvement, their skill in working collaboratively toward realizing that vision, their facility at using political opportunities to further reforms, their knack for holding people accountable for results, and their ability to manage change and sustain the implementation of the reforms over an extended number of years with instructional staff that were expert in their respective fields.

This leadership and its stability over a prolonged period were critical to each district's ability to build and maintain momentum behind their efforts.

It is worth emphasizing that the tenure of leadership in both Atlanta and Boston was remarkably long by the standards of large urban school districts, and this translated into reforms that were sustained, well-integrated, and cohesive. Beverly Hall served 12 years as the head of the Atlanta public schools before she stepped down at the end of the 2010-11 school year. Tom Payzant also served a dozen years as the superintendent of the Boston public schools. In both cases, their tenures were over three times longer than the average big city school superintendent nationwide. And their school boards benefited from substantial stability over the period, as well.

Charlotte also had stable leadership during the period, but the district's consistency was as much a product of seamless succession planning and the durability of its instructional agenda as the stability of its individual leaders. The district was very good at staying with the same set of instructional reforms, even as the people who carried the reforms forward changed.

Either way, the longevity of the leadership and the stability of the instructional reforms allowed the three districts—Atlanta, Boston, and Charlotte—to manage the process of change, to build capacity among teachers and staff to translate the vision and strategy for reform into action, and to develop momentum behind the reforms in ways that appeared to have cumulative effects over time.

However, we should note that longevity in superintendent leadership may be a necessary but not sufficient condition for improvement. Interestingly, the Cleveland superintendent remained the same during the study period, but there was substantial central-office staff turnover below her level during the study period that may have contributed to the district's lack of momentum behind a cohesive reform strategy.

We should also note that stability of leadership can also be a way of maintaining the status quo and resisting much-needed instructional reform, to the detriment of student achievement, although we did not see this circumstance in the selected districts.

In addition to strong, sustained leadership, it was clear to the study team that an important strategic element shared by the most consistently improving school districts was accountability—the ability to translate a vision for improvement into definable goals and to hold people responsible for attaining these goals. Sometimes these accountability systems were highly defined. The Atlanta school district, for instance, had what is sometimes referred to as "administrative accountability" because its infrastructure was centrally devised to create and institutionalize a culture of responsibility for results where it had not existed previously, model excellence where few people had seen it before, and to ignite more accelerated change. This approach is consistent with the successful reforms of school systems that are starting the process from low levels of student achievement.

In other places—such as Boston—accountability systems were looser and more dependent on persuasion and a culture of personal responsibility for results. This system is sometimes referred to as "professional accountability" because it relies on the proficiency, pride, and ownership of the individuals responsible for the results, and is more often seen in districts with longer histories of reform and capacity building. This approach is also consistent with districts with a longer track-record of reforms and whose performance and capacity is higher.

Also at the strategic level, consistently improving districts put into place sophisticated data systems that were used to drive accountability, monitor progress, and inform instructional practice. As a result of carefully orchestrated cultures of data use, staff members and teachers in the consistently improving districts were curious about their data, trained in how to interpret data, and skilled in using results to alter practice. In contrast, the lower-performing district we studied had poorer-quality data and low expectations for its use.

It was this set of strategic activities—leadership, goal setting, accountability, and data use—that defined a broad set of expectations and preconditions for the tactical reforms under them. Otherwise, it is impossible to attribute academic improvement to one tactical reform or another, because districts and schools often make informed programmatic decisions only to have them nullified by how they are implemented or applied. For instance, it is not unusual to see districts with great data systems but no one trained to use them. Or seemingly strong improvement targets that are actually defined around the minimal growth required to avoid sanctions.

Still, through this study, we tried to discern the net impact of both the strategic and various tactical-level actions a district pursued, as it appears that the districtwide strategic reforms may have been too generalized to produce the rapid gains that we saw in reading versus math or vice versa *without* these tactical changes.

Phases of Reform

The reader will note from the data in the earlier chapters that the study districts did not start their reforms with students at the same level of academic proficiency on NAEP or with the same staff capacities. 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, and Mecklenburg County'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 in student achievement on NAEP during the study period. It is possible that what brought Charlotte up 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. It may also be the case that the district's standing near the national average makes it hard to move beyond that level.

In other words, what may work at one stage of reform may not work at another. Charlotte succeeded in creating stability in programming, but it may have stayed with these efforts too long and outgrown what had been needed when its reforms were new. 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 or great to excellent.[1] That finding is on ample display in this report.

With Charlotte under new leadership over the last several years, and having begun to move in new directions, 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 indicative of what one sees in historically low-performing urban school districts that are working to move from poor to fair performance and 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. What was particularly interesting about the Atlanta approach, however, was that it was able to pair a relatively prescriptive instructional program with a site-based strategy that allowed individual schools to choose which program to adopt, building ownership and buy-in as a result. Yet, despite this element of flexibility, it was not likely that Atlanta could have seen substantial gains without the clarity, direction, and discipline that defined its reform agenda over the last decade.

Likewise, Boston's reforms also seemed appropriate to a city school system moving from fair to good performance academically, particularly in math. In fact, the district was much farther ahead of either Atlanta or Cleveland on NAEP in 2003 and was able to build on the many years of thinking, planning, false starts, and stakeholder buy-in that pre-dated our study period.

We have noted that Atlanta and Cleveland had similar NAEP reading and math scale scores in 2003, but Atlanta's reform efforts allowed it to outpace Cleveland over the study period.[2] Implicit in this observation is that Cleveland might have seen the same kinds of gains as Atlanta's if Cleveland had pursued many of the same instructional reforms over the same period and if it had not had to absorb large budget cuts with such disruptive staffing shifts.

In sum, 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. Certainly, more research on this question is necessary.

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 as well as districts showing little gain were represented by governance structures of all types. Atlanta, which saw significant reading gains on NAEP, and Charlotte, which had high performance, both

---

[1] 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.

[2] Atlanta made substantially greater reading gains on NAEP than did Cleveland, but it also appears to have moved from a reading performance level in 2003 that was significantly below what would have been expected statistically to a level that was not significantly different from what might be expected statistically in 2007. (See table 3.10.)

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. The business community began playing a strong role in recruiting and supporting 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 spearheaded the strategic reforms we referred to earlier that led to a decade of instructional change and growth on NAEP.

The Atlanta reforms also included transforming a traditional, regionally based district structure into five school-based units that provided hands-on technical assistance, coaching, professional development, and operational services directly to schools and teachers. The central office was downsized into these units—or school reform teams (SRTs)—that provided technical assistance, coaching, professional development, and operational services directly to their designated schools and teachers. The SRTs were staffed with some of the district's most talented people and evaluated on their ability to improve services and raise achievement in their schools. The configuration allowed the schools to pick their own academic programs—something that often does not work well in other city school districts because the multiple programs cannot be adequately supported centrally. But, in this case, the SRTs were specifically trained on the programs that the schools chose to implement, and the SRTs were given only a few schools to work with at any one time.

Therefore, Atlanta presented an interesting example of a school district that blended two seemingly irreconcilable organizational approaches—managed instruction and site-based decision-making—into a coherent approach to improving school-based staff and teacher capacity and raising student achievement.

What appears to matter in these differing governance and organizational models has less to do with who controls the system than with what they do to improve student achievement. In other words, part of Atlanta's success was a function of how well it organizationally aligned itself to its instructional priorities. But it is not plausible 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. Similarly, Boston did not make the same kinds of organizational changes as Atlanta did, but the district was able to provide some of the same kinds of hands-on support to schools through its coaching and math teams. In these situations, it appears that the process of the instructional reforms may matter more than their structure, or the "how" may be more important than the "what."

Still, 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 is likely to improve academic results—and to show greater gains than a traditional structure that does not focus on instructional improvement.[3] Conversely, if 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 Cleveland school district may be an example of this pattern.

[3] The District of Columbia Public Schools appears to be another example of this situation.

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 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 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 the most 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.

Implications for the Common Core State Standards

Building on this point about the centrality of instructional quality, 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 State Standards.

1. The results of this study imply that districts' ability to use the new common core standards to improve student achievement is likely to depend upon equally high levels of quality curriculum and instruction. The low degree of content matching described in this study suggests that even clearly written curriculum supported by professional development and coaching may 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, focus, and coherence. 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. A clearly defined set of curriculum objectives can be the foundation for assessing, choosing, using, and evaluating various textbooks and other instructional materials, as well as developing and providing professional development. The more successful districts described in this report were all effective in defining their curriculum around their respective state standards, even if the standards themselves did not align well with NAEP. These districts attempted to use their textbooks and programs in the way they were designed, while modifying the materials to fit the standards more completely. Charlotte, for instance, focused on faithful implementation of its adopted reading program and district curriculum yet expanded its focus on reading to include

writing across the curriculum. Atlanta insisted that vendors of reading textbooks worked with central office staff to address gaps in the published materials.

3. The new common core standards will compel classroom instruction that is more conceptual and more problem-solving in its orientation than what most educators are used to. School and district reliance on test-prep to meet minimal requirements will prove less effective as the assessments that emerge from the common core demand more sophisticated responses from students than the current multiple-choice formats require. At present, some of the state standards and assessments that we looked at for this project did not yet involve the kinds of short answer or extended constructed responses that NAEP requires. As the nation's two major assessment consortia finish their work on the next generation of tests linked to the common core, we are likely to see assessments with more complex, multi-part performance tasks that go well beyond NAEP. Over the long run, the growing emphasis on teaching concepts should result in students doing well academically regardless of the nature of the tests or the degree of alignment.

In addition, the new common core 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 suggested that urban school districts generally did less well in this area than they did in reading for literary experience. Similar shifts will be required in math instruction as the new standards require deeper understanding of math concepts and more rigorous application of them. At the same time, urban school districts need to be aware of research suggesting that more rigorous content and academic requirements alone can lead to student frustration without substantial monitoring and intensive supports of students by teachers and staff.

4. Finally, the implementation of the common core standards will depend heavily on the overall quality of teachers' and administrators' skills, as well as the capacity of districts to support their teacher corps through a variety of strategies. The districts we studied that made significant progress devoted substantial time and energy to improving the capacity of their teachers, principals, and staff and/or worked to recruit and retain new personnel at all levels of the organization. Most of the districts did not assume this could be done with professional development by itself. Instead, they designed multiple mechanisms to enhance the ability of teachers, principals, and staff to provide quality instruction, including coaching, technical assistance, professional learning communities, user-friendly data systems, accountability and monitoring, and team-building opportunities. And we should also note that the three more successful districts also shared one other characteristic that should serve all major urban school systems well in the implementation of the common core: they worked as collaboratively as possible with their state education departments in developing curriculum and professional development aligned with their instructional programs. 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 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 urban 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 the leaders of urban school districts participating in the Trial Urban District Assessment of NAEP, as well as to leaders in other urban districts. These recommendations suggest steps that these leaders might take to increase or accelerate the academic progress that their students 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 and Remaining Questions

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 distinguishes urban districts that have made major progress from those who have not. 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 also 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, rigor, organizational restructuring, accountability, and instructional focus and cohesion.

Over the long run, however, 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 improvements 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 math. 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 not only to raise achievement across the board, but also 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 may 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 the 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 reforms.

Whatever its unanswered questions, 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

References

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 A
HOW NAEP IS ADMINISTERED

The 2001 "No Child Left Behind" legislation that mandated participation in the National Assessment of Educational Progress (NAEP) for any state receiving Title I funds dramatically increased state participation in the assessment. As a result, the National Center for Education Statistics (NCES) faced the challenge of administering NAEP across an expanded participation base within a small testing window. To meet this goal, NCES assigned NAEP assessment activities to Westat, its data collection contractor, so that the burden on schools would be greatly reduced.

Since NAEP is conducted in partnership with states, each state has employed an NAEP State Coordinator to serve as the connection between the state education agency and schools selected for the sample. In general, the NAEP State Coordinator works with the schools, Westat, and NCES to ensure the quality of the state's NAEP results.

Individual schools participating in NAEP designate an in-school staff member to be the school coordinator. The school coordinator collaborates on assessment activities with the professional field staff trained by Westat. School coordinators carry out the following tasks with the help of Westat field staff and the NAEP State Coordinator:

- Schedule the assessment date.
- Upon request of the NAEP representative, provide a list of all eligible students.
- Inform teachers and students about the assessment.
- Inform parents about the assessment.
- Provide space for the assessment.
- Receive the school pre-assessment packet and conduct final preparations for the assessment.
- Distribute background questionnaires to the appropriate school staff and collect completed questionnaires.

The Westat field staff, who are responsible for all assessment day activities, are a national network of educators trained to collect and safeguard NAEP assessment data, to guarantee the accuracy and integrity of the data, and to provide support for the schools throughout the assessment process. In addition to assisting the school coordinator with his or her assigned tasks, the NAEP State Coordinator and Westat field staff are responsible for the following duties:

- Work with schools to set up the assessment dates.
- Provide the MySchool website to facilitate communications with schools and make available the NAEP Help Desk at 1-800-283-NAEP to respond to schools' questions.
- From the list of all eligible students, draw a random sample of students to be assessed.
- Provide schools with information about notifying parents/guardians.
- Prepare school, teacher, students with disabilities (SD), and English language learners (ELL) questionnaires for distribution.
- Send pre-assessment packets to the school coordinator.
- Provide all materials, including pencils and ancillary materials, for each assessment session.
- Conduct the assessment.
- Send the materials to the scoring facility.

View materials for NAEP administrators to learn more about the administration process.

[1] Source: http://nces.ed.gov/nationsreportcard/about/natadministered.asp

APPENDIX B

DISTRICT DEMOGRAPHICS, NAEP TRENDS, FUNDING, AND TEACHERS

Table B.1 General enrollment of TUDA districts by NAEP administration year, 2003-2009

	2002-03	2006-07	2008-09
Atlanta	54,946	50,631	49,032
Austin	78,608	82,140	83,319
Boston	61,552	56,388	55,923
Charlotte	109,767	128,789	134,060
Chicago	436,048	413,694	421,430
Cleveland	71,616	55,593	49,148
District of Columbia	67,522	56,943	44,331
Houston	212,099	202,936	200,252
Los Angeles	746,852	707,627	684,143
New York City	1,077,381	N/A	1,038,741
San Diego	140,753	130,983	131,890

Source: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, "Local Education Agency Universe Finance Survey."

Table B.2 Percentages of public school students in TUDA districts, large cities, and the national public sample in grades 4 and 8 on the NAEP reading assessment by selected characteristics, 2003-2009

Reading	Grade 4				Grade 8			
District	2003	2005	2007	2009	2003	2005	2007	2009
African American								
Atlanta	87	85	83	80	91	92	90	89
Austin	—	15	13	12	—	12	13	11
Boston	49	46	44	40	47	45	41	42
Charlotte	45	43	42	39	43	46	47	47
Chicago	53	48	49	46	52	46	49	47
Cleveland	73	70	66	70	78	75	75	72
District of Columbia	85	85	85	76	88	89	88	84
Houston	40	33	29	30	34	31	31	29
Los Angeles	12	10	11	7	13	11	10	9
New York City	37	35	29	29	38	34	33	32
San Diego	18	12	11	12	16	13	12	12
National Public	17	17	17	16	17	17	17	16
Large City	35	32	31	29	36	32	31	27
White								
Atlanta	10	11	14	13	5	4	6	7
Austin	—	30	28	29	—	35	31	31
Boston	11	12	13	14	16	15	16	15
Charlotte	42	40	36	37	46	40	35	32
Chicago	10	9	10	9	10	11	9	9
Cleveland	16	19	20	17	16	15	15	16
District of Columbia	5	4	6	9	3	3	3	5
Houston	10	12	7	8	8	9	9	9
Los Angeles	10	9	9	9	10	10	9	8
New York City	14	15	17	15	13	16	16	16
San Diego	22	22	24	28	24	25	26	28
National Public	59	57	56	54	61	60	58	57
Large City	22	21	21	20	23	24	23	22
Hispanic								
Atlanta	2	4	3	5	2	2	3	3
Austin	—	52	54	55	—	50	53	54
Boston	30	32	33	37	25	29	32	31
Charlotte	8	11	13	15	6	9	11	14
Chicago	35	41	39	42	34	39	39	40
Cleveland	7	9	9	10	5	9	8	10
District of Columbia	9	9	7	13	8	6	8	9
Houston	47	51	60	59	56	56	57	59
Los Angeles	72	74	75	77	69	72	74	75
New York City	37	38	39	39	33	37	37	37
San Diego	43	47	47	42	37	44	45	41
National Public	18	19	20	21	15	17	18	20
Large City	34	38	38	42	32	36	37	41
Asian/Pacific Islander								
Atlanta	#	1	#	1	1	1	#	#
Austin	—	3	4	4	—	4	3	3
Boston	9	10	9	7	11	10	11	11
Charlotte	4	3	4	4	4	4	4	4
Chicago	2	3	3	4	3	4	3	3

(Table B.2 Continued) Percentages of public school students in TUDA districts, large cities, and the national public sample in grades 4 and 8 on the NAEP reading assessment by selected characteristics, 2003-2009

Cleveland	1	#	2	1	1	#	1	1
District of Columbia	1	2	1	2	1	1	1	2
Houston	3	3	3	4	2	3	3	3
Los Angeles	6	7	5	7	8	7	7	7
New York City	11	12	14	16	16	12	15	14
San Diego	18	18	17	18	22	17	16	19
National Public	4	4	5	5	4	4	5	5
Large City	7	7	7	7	8	7	8	8
NSLP-eligible								
Atlanta	81	76	75	74	78	74	75	78
Austin	—	59	61	60	—	49	55	54
Boston	81	83	81	79	70	76	70	72
Charlotte	44	49	48	47	37	45	47	46
Chicago	85	84	86	87	88	81	85	86
Cleveland	100	100	100	100	100	100	100	100
District of Columbia	70	76	66	70	57	70	65	73
Houston	72	74	84	81	67	71	77	78
Los Angeles	83	84	77	84	67	78	76	82
New York City	89	86	85	87	85	84	85	79
San Diego	58	64	65	60	53	54	57	55
National Public	44	45	45	47	36	39	40	43
Large City	69	71	70	71	61	63	64	65
Students with Disabilities								
Atlanta	6	7	5	9	8	7	5	9
Austin	—	7	8	8	—	8	12	11
Boston	16	17	16	17	17	13	16	16
Charlotte	14	10	10	11	10	9	9	9
Chicago	10	9	9	12	12	14	15	14
Cleveland	5	5	4	6	9	7	6	11
District of Columbia	8	9	4	5	10	11	7	5
Houston	11	6	6	4	12	8	8	7
Los Angeles	9	7	9	9	10	9	9	9
New York City	12	12	13	15	13	9	14	13
San Diego	11	11	12	10	10	9	9	10
National Public	10	10	10	10	10	9	9	10
Large City	9	9	9	10	10	9	10	10
English Language Learners								
Atlanta	2	1	1	1	1	1	1	#
Austin	—	16	22	24	—	11	13	13
Boston	13	12	27	16	9	7	7	3
Charlotte	7	8	10	7	5	7	6	5
Chicago	16	14	18	10	4	3	4	5
Cleveland	2	3	4	3	1	2	3	4
District of Columbia	6	5	6	6	3	2	2	4
Houston	18	23	29	27	11	11	9	8
Los Angeles	54	54	47	41	31	33	28	22
New York City	6	8	15	14	7	7	7	7
San Diego	33	35	41	35	20	20	21	16
National Public	8	9	9	9	5	5	6	5
Large City	17	19	20	18	10	11	11	11

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

Table B.3 Percentages of public school students in TUDA districts, large cities, and the national public sample in grades 4 and 8 on the NAEP mathematics assessment by selected characteristics, 2003-2009

Math	Grade 4				Grade 8			
District	2003	2005	2007	2009	2003	2005	2007	2009
African American								
Atlanta	87	84	82	79	93	92	91	88
Austin	—	14	13	11	—	13	13	11
Boston	46	45	44	39	46	45	42	40
Charlotte	46	40	42	39	46	48	47	46
Chicago	52	46	46	45	51	45	47	48
Cleveland	76	70	66	68	72	70	74	71
District of Columbia	87	86	83	77	87	88	88	82
Houston	35	28	26	25	33	28	29	29
Los Angeles	10	10	10	7	12	13	11	10
New York City	35	35	29	28	36	35	33	32
San Diego	17	14	11	12	16	15	13	12
National Public	17	17	17	16	17	17	17	16
Large City	34	32	31	29	35	32	30	27
White								
Atlanta	10	11	12	13	5	5	4	7
Austin	—	28	26	25	—	33	31	31
Boston	12	13	12	14	16	16	17	14
Charlotte	41	41	36	36	42	38	34	32
Chicago	11	8	10	9	10	12	11	9
Cleveland	16	19	20	15	15	18	15	15
District of Columbia	4	4	6	9	3	4	3	5
Houston	7	10	6	7	8	10	9	8
Los Angeles	11	10	9	9	10	9	8	8
New York City	15	14	17	15	16	15	15	16
San Diego	23	23	23	27	27	26	23	28
National Public	58	57	55	54	62	60	58	56
Large City	22	21	20	20	24	24	23	21
Hispanic								
Atlanta	2	3	5	5	1	2	3	4
Austin	—	55	58	60	—	51	53	55
Boston	33	32	35	37	28	28	30	33
Charlotte	7	11	14	16	6	9	12	15
Chicago	34	42	41	42	36	38	39	40
Cleveland	6	7	11	13	11	10	10	12
District of Columbia	8	8	9	12	9	7	9	11
Houston	56	59	65	64	55	58	58	60
Los Angeles	73	74	75	77	71	72	74	75
New York City	37	39	40	40	34	38	38	39
San Diego	42	44	47	43	38	41	46	41
National Public	19	20	21	22	15	17	19	21
Large City	36	39	40	42	33	36	38	42
Asian/Pacific Islander								
Atlanta	#	1	#	1	#	#	#	#
Austin	—	3	3	3	—	3	2	3
Boston	8	9	8	8	9	11	10	11
Charlotte	4	5	4	5	5	4	5	4

(Table B.3 Continued) Percentages of public school students in TUDA districts, large cities, and the national public sample in grades 4 and 8 on the NAEP mathematics assessment by selected characteristics, 2003-2009

Chicago	3	3	3	4	4	4	3	3
Cleveland	1	1	1	1	1	1	1	1
District of Columbia	1	1	2	2	1	1	1	2
Houston	2	3	3	4	3	4	3	3
Los Angeles	6	6	5	7	7	6	7	7
New York City	12	12	13	16	14	13	13	14
San Diego	18	17	18	17	19	17	17	18
National Public	4	4	5	5	4	5	5	5
Large City	7	6	7	7	8	8	8	8
NSLP-eligible								
Atlanta	81	76	77	74	78	78	80	78
Austin	—	63	61	65	—	50	54	55
Boston	83	84	82	78	71	74	69	73
Charlotte	45	44	48	47	36	45	49	46
Chicago	85	87	86	87	88	81	84	86
Cleveland	100	99	100	100	100	100	100	100
District of Columbia	71	76	69	72	57	72	65	75
Houston	76	78	85	83	69	70	77	78
Los Angeles	83	86	77	84	65	77	76	82
New York City	88	84	87	87	83	84	86	79
San Diego	58	64	63	61	52	55	59	55
National Public	44	46	46	48	36	39	41	43
Large City	69	71	71	71	60	62	65	66
Students with Disabilities								
Atlanta	7	8	9	9	9	9	7	10
Austin	—	9	10	12	—	7	13	11
Boston	17	18	19	18	21	12	13	16
Charlotte	14	11	10	11	12	10	12	9
Chicago	11	10	11	12	13	15	14	14
Cleveland	7	9	6	11	9	10	9	13
District of Columbia	10	11	9	11	11	12	8	13
Houston	12	8	7	5	10	8	8	8
Los Angeles	9	9	10	10	11	10	9	10
New York City	12	12	15	18	14	11	12	14
San Diego	10	9	9	10	10	8	8	8
National Public	11	12	11	12	11	11	9	10
Large City	10	11	11	11	11	10	9	11
English Language Learners								
Atlanta	2	2	2	2	1	1	1	1
Austin	—	22	28	31	—	11	14	15
Boston	16	13	30	17	9	7	7	8
Charlotte	6	9	10	7	6	6	8	6
Chicago	17	17	19	10	5	5	5	5
Cleveland	3	4	6	6	5	3	4	6
District of Columbia	6	4	6	7	4	3	4	5
Houston	34	36	38	37	12	13	10	10
Los Angeles	55	53	48	41	32	33	27	23
New York City	7	10	15	15	10	9	10	9
San Diego	33	34	39	35	21	19	20	15
National Public	9	10	10	10	5	6	6	6
Large City	19	20	21	20	12	12	12	12

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

Table B.4 Average reported NAEP reading scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, and national public, 2003-2009

Reading	Grade 4					Grade 8				
District	2003	2005	2007	2009	Change	2003	2005	2007	2009	Change
Overall										
Atlanta	197	201	207	209**	12***	240	240	245	250**	10***
Austin	-	217	218	220*	3	-	257	257	261*	4
Boston	206	207	210	215*,**	9***	252	253	254	257*,**	5***
Charlotte	219	221	222	225*,**	6***	262	259	260	259*,**	-3
Chicago	198	198	201	202*,**	4***	248	249	250	249**	1
Cleveland	195	197	198	194*,**	-1	240	240	246	242*,**	2
District of Columbia	188	191	197	203*,**	15***	239	238	241	240*,**	1
Houston	207	211	206	211**	4	246	248	252	252**	6***
Los Angeles	194	196	196	197*,**	3***	234	239	240	244*,**	10***
New York City	210	213	213	217*	7***	252	251	249	252**	0
San Diego	208	208	210	213**	5	250	253	250	254**	4
National Public	216	217	220	220*	4***	261	260	261	262*	1***
Large City	204	206	208	210**	6***	249	250	250	252**	3***
African American										
Atlanta	191	194	200	201	10***	237	237	242	246	9***
Austin	-	200	201	211*,**	11***	-	242	238	247	5
Boston	202	203	204	212*,**	10***	245	244	250	248	3
Charlotte	205	206	206	211*,**	6	247	244	246	249*,**	2
Chicago	193	190	193	194*,**	1	243	240	240	243	0
Cleveland	191	193	192	189*,**	-2	238	236	243	239**	1
District of Columbia	184	187	192	195*,**	11***	236	235	238	235*,**	-1
Houston	201	207	205	210*,**	9***	244	242	249	243	-1
Los Angeles	187	187	196	195**	8	233	234	229	239	6
New York City	201	206	206	208*,**	7***	245	241	240	246	1
San Diego	196	198	199	206	10	236	242	240	239	3
National Public	197	199	203	204*	7***	244	242	244	245*	1***
Large City	193	196	199	201**	8***	241	240	240	243**	2***
White										
Atlanta	250	253	253	253*,**	3	‡	‡	‡	292*,**	
Austin	-	239	244	245*,**	6	-	279	284	282*,**	3
Boston	225	230	230	231	6	273	274	275	282*,**	9
Charlotte	237	240	244	243*,**	6	278	278	279	276	-2
Chicago	224	225	227	228	4	265	270	266	272	7
Cleveland	208	209	215	209*,**	1	250	255	262	258*,**	8
District of Columbia	254	252	258	257*,**	3	‡	301	‡	‡	
Houston	235	245	241	243**	8	270	280	281	280	10***
Los Angeles	217	229	228	222*	5	266	261	272	271	5
New York City	231	226	232	235	4	270	269	270	271	1
San Diego	231	226	234	236	5	269	273	271	273	4
National Public	227	228	230	229*	2***	270	269	270	271	1***
Large City	226	228	231	233**	7***	268	270	271	272	4***

(Table B.4 continued) Average reported NAEP reading scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, and national public, 2003-2009

Reading	Grade 4					Grade 8				
District	2003	2005	2007	2009	Change	2003	2005	2007	2009	Change
Hispanic										
Atlanta	‡	‡	‡	‡		‡	‡	‡	‡	
Austin	-	207	206	208*	1	-	243	244	251*	8
Boston	201	200	204	209*,**	8***	245	248	241	251*	6
Charlotte	202	209	207	212*,**	10	244	248	251	254	10
Chicago	196	201	201	203	7	249	251	255	249	0
Cleveland	201	201	200	200	-1	‡	248	249	237**	-11
District of Columbia	187	193	206	207	20***	240	247	249	249	9
Houston	203	203	200	206	3	242	245	246	250*	8***
Los Angeles	189	190	190	193*,**	4	228	235	236	239*,**	11***
New York City	205	207	203	208*	3	247	247	241	243	-4
San Diego	195	196	196	193*,**	-2	238	241	235	242	4
National Public	199	201	204	204*	5***	244	245	246	248*	4***
Large City	197	198	199	202**	5***	241	243	243	245**	4***
Asian/Pacific Islander										
Atlanta	‡	‡	‡	‡		‡	‡	‡	‡	
Austin	-	‡	236	‡		-	‡	‡	‡	
Boston	223	224	229	231	8	274	280	275	276	2
Charlotte	218	‡	235	233	15	‡	‡	‡	‡	
Chicago	‡	‡	237	232		268	277	‡	‡	
Cleveland	‡	‡	‡	‡		‡	‡	‡	‡	
District of Columbia	‡	‡	‡	‡		‡	‡	‡	‡	
Houston	‡	‡	231	240*		‡	‡	289	‡	
Los Angeles	218	223	219	220**	2	255	262	264	265**	10
New York City	227	235	230	235*	8***	264	271	268	270	6
San Diego	222	222	223	227	5	260	265	265	264**	4
National Public	225	227	231	234*	9***	268	270	269	273*	5***
Large City	223	223	228	228**	5	260	266	263	268**	8***
NSLP-eligible										
Atlanta	189	191	198	199**	10***	235	234	240	244**	9***
Austin	-	203	203	206	3	-	240	240	247	7
Boston	204	205	207	211*,**	7***	247	247	249	251*	4***
Charlotte	200	206	205	210*,**	10***	244	242	245	248*	4
Chicago	194	194	197	199*,**	5***	246	246	247	246	0
Cleveland	195	197	198	194*,**	-1	240	240	246	242**	2
District of Columbia	182	183	188	193*,**	11***	232	234	234	232*,**	0
Houston	201	202	201	206*	5***	241	243	247	246	5***
Los Angeles	189	190	191	193*,**	4	230	236	237	240*,**	10***
New York City	206	210	209	214*,**	8***	248	249	246	250*	2
San Diego	197	199	198	198**	1	240	243	236	242	2
National Public	201	203	205	206*	5***	246	247	247	249*	3***
Large City	196	198	200	202**	6***	241	243	242	244**	3***

(Table B.4 continued) Average reported NAEP reading scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, and national public, 2003-2009

Reading	Grade 4					Grade 8				
District	2003	2005	2007	2009	Change	2003	2005	2007	2009	Change
Limited English Proficiency										
Atlanta	‡	‡	‡	‡		‡	‡	‡	‡	
Austin	-	189	194	197*,**	8	-	213	210	223	9***
Boston	192	190	197	196*,**	4	215	217	210		
Charlotte	190	198	196	193	3	230	237	228	229*	-1
Chicago	176	175	182	176**	0	212	216	217	220	8
Cleveland	‡	‡	‡	‡		‡	‡	‡	‡	
District of Columbia	174	177	198	192	19***	231	‡	‡	‡	
Houston	186	192	186	196*,**	9***	214	216	209	219	6
Los Angeles	183	182	177	176*,**	-7***	205	213	212	206*,**	1
New York City	183	183	181	189	6	212	216	209	212	0
San Diego	186	188	189	186	0	220	219	209	211	-9***
National Public	186	187	188	188*	2	222	224	222	219	-4
Large Central Cities	185	184	183	184**	0	216	221	214	215	0
District	2003	2005	2007	2009	Change	2003	2005	2007	2009	Change
Students with Disabilities										
Atlanta	180	169	191	177	-3	208	203	‡	210**	2
Austin	-	184	190	194*	10	-	219	228	232*	13***
Boston	181	180	183	190*	9	217	220	223	234*	17***
Charlotte	191	194	187	196*	5	228	216	228	224	-4
Chicago	163	176	172	169**	6	215	210	213	216**	1
Cleveland	161	‡	‡	‡		208	‡	210	210**	2
District of Columbia	148	154	162	‡		199	199	210	‡	
Houston	183	187	174	178	-5	222	210	217	201*.**	-21***
Los Angeles	167	161	166	152*,**	-14***	195	201	200	206*,**	12***
New York City	181	183	181	189*	8	211	213	216	221**	10***
San Diego	185	180	171	167**	-18***	209	219	214	221	11
National Public	184	190	190	189*	5***	224	226	226	229*	5***
Large Central Cities	175	180	178	177**	2	212	213	214	217**	5***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.5 Average reported NAEP math scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, and national public, 2003-2009

Mathematics	Grade 4					Grade 8				
District	2003	2005	2007	2009	Change	2003	2005	2007	2009	Change
Overall										
Atlanta	216	221	224	225*,**	9***	244	245	256	259*,**	15***
Austin	-	242	241	240*	-2	-	281	283	287*,**	6***
Boston	220	229	233	236*,**	16***	262	270	276	279*	17***
Charlotte	242	244	244	245*,**	3	279	281	283	283*	4***
Chicago	214	216	220	222*,**	8***	254	258	260	264*,**	10***
Cleveland	215	220	215	213*,**	-2	253	249	257	256*,**	3
District of Columbia	205	211	214	220*,**	15***	243	245	248	251*,**	8***
Houston	227	233	234	236*,**	9***	264	267	273	277*,**	13***
Los Angeles	216	220	221	222*,**	6***	245	250	257	258*,**	13***
New York City	226	231	236	237*	11***	266	267	270	273**	7***
San Diego	226	232	234	236*	10***	264	270	272	280*	16***
National Public	234	237	239	239*	5***	276	278	280	282*	6***
Large City	224	228	230	231**	7***	262	265	269	271**	9***
African American										
Atlanta	211	215	217	218**	7***	241	242	253	255**	14***
Austin	-	228	226	226	-2	-	262	265	274*,**	12***
Boston	216	223	226	231*,**	15***	251	256	263	268*,**	17***
Charlotte	229	230	230	231*,**	2	258	264	267	270*,**	12***
Chicago	207	208	213	212*,**	5***	245	245	248	252**	7***
Cleveland	210	215	210	209*,**	-1	249	244	253	252*,**	3
District of Columbia	202	207	209	212*,**	10***	240	241	245	244*,**	4***
Houston	221	224	225	227*,**	6	259	257	265	266*,**	7***
Los Angeles	208	209	216	209*,**	1	234	239	245	247*,**	13***
New York City	219	222	227	227*,**	8***	253	257	258	261*	8***
San Diego	216	221	222	222	6	252	253	258	263	11***
National Public	216	220	222	222*	6***	252	254	259	260*	8***
Large City	212	217	219	219**	7***	247	250	254	256**	9***
White										
Atlanta	258	263	266	266*,**	8	298	‡	‡	‡	
Austin	-	262	263	262*,**	0	-	305	308	312*,**	7***
Boston	234	244	250	251	17***	289	299	305	311*,**	22***
Charlotte	257	261	261	263*,**	6***	301	304	308	304*,**	3
Chicago	235	243	244	242*	7	276	281	287	289	13***
Cleveland	233	233	233	228*,**	-5	269	265	269	275*,**	6
District of Columbia	262	266	262	270*,**	8***	‡	317	‡	‡	
Houston	254	262	263	260*,**	6	293	294	308	311*,**	18***
Los Angeles	241	247	247	245	4	277	280	285	287	10
New York City	244	245	249	254**	10***	289	286	289	295	6
San Diego	243	249	252	255**	12***	284	292	294	301*,**	17***
National Public	243	246	248	248*	5***	287	288	290	292	5***
Large City	243	247	249	250**	7***	285	288	292	294	9***

(Table B.5 continued) Average reported NAEP math scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, and national public, 2003-2009

Mathematics	Grade 4					Grade 8				
District	2003	2005	2007	2009	Change	2003	2005	2007	2009	Change
Hispanic										
Atlanta	‡	‡	223	222		‡	‡	‡	‡	
Austin	-	234	233	233*,**	-1	-	267	271	274*,**	7***
Boston	215	225	230	232*,**	17***	252	261	270	269*	17***
Charlotte	233	234	234	235*,**	2	262	262	264	272*,**	10
Chicago	217	217	219	226	9***	259	263	265	268	9***
Cleveland	220	224	215	217*,**	-3	249	251	258	250*,**	1
District of Columbia	205	215	220	227	22***	246	252	251	263	17***
Houston	226	232	234	235*,**	9***	261	265	270	275*,**	14***
Los Angeles	211	216	217	218*,**	7***	240	245	253	254*,**	14***
New York City	220	226	230	230*,**	10***	260	259	262	261**	1
San Diego	216	222	223	224	8***	248	258	259	265	17***
National Public	221	225	227	227	6***	258	261	264	266	8***
Large City	219	223	224	226	7***	256	258	261	264	8***
Asian/Pacific Islander										
Atlanta	‡	‡	‡	‡		‡	‡	‡	‡	
Austin	-	‡	268	‡		-	‡	‡	‡	
Boston	243	256	255	260	17***	300	309	305	312*,**	12***
Charlotte	252	256	263	257	5	293	‡	305	‡	
Chicago	‡	‡	249	255		286	292	‡	301	15***
Cleveland	‡	‡	‡	‡		‡	‡	‡	‡	
District of Columbia	‡	‡	‡	‡		‡	‡	‡	‡	
Houston	‡	‡	265	264*,**		‡	299	310	‡	
Los Angeles	241	246	246	248**	7	275	291	292	291**	16***
New York City	247	253	257	258	11***	286	295	299	309*,**	23***
San Diego	238	245	247	247**	9***	278	282	289	292**	14***
National Public	246	251	254	255	9***	289	294	296	300	11***
Large City	246	247	251	253	7	281	289	291	299	18***
NSLP-eligible										
Atlanta	209	213	216	216*,**	7***	239	240	251	253*,**	14***
Austin	-	232	229	231*,**	-1	-	261	267	271*,**	10***
Boston	218	227	231	233*,**	15***	256	264	271	273*,**	17***
Charlotte	229	230	231	232*,**	3	256	261	265	268*	12***
Chicago	212	212	216	219*,**	7***	252	254	257	261**	9***
Cleveland	215	220	215	213*,**	-2	253	249	257	256*,**	3
District of Columbia	200	206	207	210*,**	10***	235	241	243	243*,**	8***
Houston	223	228	231	233*,**	10***	259	262	268	271*,**	12***
Los Angeles	212	216	217	218*,**	6***	240	245	254	254*,**	14***
New York City	224	228	234	235*,**	11***	261	264	267	270*,**	9***
San Diego	217	225	224	224**	7***	252	258	260	268*	16***
National Public	222	225	227	228*	6***	258	261	265	266*	8***
Large City	217	221	223	225**	8***	252	256	260	262**	10***

(Table B.5 continued) Average reported NAEP math scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, and national public, 2003-2009

Mathematics	Grade 4					Grade 8				
District	2003	2005	2007	2009	Change	2003	2005	2007	2009	Change
Limited English Proficiency										
Atlanta	‡	‡	‡	‡		‡	‡	‡	‡	
Austin	-	225	226	229*,**	4	-	240	245	249*,**	9***
Boston	209	221	228	222*,**	13***	229	233	242	238	8
Charlotte	226	228	230	228*,**	2	258	252	252	256*,**	-2
Chicago	204	201	207	209*,**	5	228	235	240	241	13***
Cleveland	‡	‡	205	‡		‡	‡	‡	‡	
District of Columbia	200	206	209	217	17***	231	‡	226	‡	
Houston	221	228	229	231*,**	10***	240	245	241	247*	6***
Los Angeles	207	210	208	206*,**	-1	223	225	230	227*,**	4
New York City	203	211	216	219	16***	238	232	235	230**	-7
San Diego	211	217	217	217	5***	235	236	237	244	9***
National Public	214	216	217	218	4***	241	244	245	243*	1
Large Central Cities	212	214	214	216	5***	238	238	239	238**	0
District	**2003**	**2005**	**2007**	**2009**	**Change**	**2003**	**2005**	**2007**	**2009**	**Change**
Students with Disabilities										
Atlanta	200	198	207	202*,**	2	210	202	‡	228*,**	18***
Austin	-	227	226	222*	5	-	250	252	259*,**	9
Boston	201	210	214	219*	18***	227	233	247	247*	20***
Charlotte	225	228	222	226*	1	253	242	256	247*	-6
Chicago	194	198	196	200*,**	6	217	226	228	235**	17***
Cleveland	195	204	‡	193*,**	-2	223	216	222	227*,**	4
District of Columbia	177	188	188	194*,**	17***	204	208	211	204*,**	-1
Houston	216	214	214	209**	-7***	241	232	240	231**	-10***
Los Angeles	198	195	196	191*,**	-7	215	210	220	225*,**	10***
New York City	203	207	213	218*	15***	223	231	235	242**	18***
San Diego	210	214	201	205**	-5	228	234	234	246	17***
National Public	214	218	220	220*	6***	242	244	246	249*	7***
Large Central Cities	204	209	208	210**	6***	229	230	233	238**	9***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.6. Average reported NAEP science scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, national public, 2005‡

Science	4th Grade	8th Grade
District	2005	
Atlanta	133**	117*,**
Austin	147*	144*,**
Boston	133**	131**
Charlotte	145*,**	142*,**
Chicago	126*,**	124*,**
Cleveland	128*,**	122*,**
Houston	138**	130**
Los Angeles	126*,**	121*,**
New York City	134**	128**
San Diego	138**	136*,**
National Public	149*	147*
Large City	135**	132**

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; ‡ Data are not comparable to 2009 science results.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Science Assessments

Table B.7 Average reported NAEP science scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, national public, 2009

Science	Grade 4	Grade 8
District	2009	
Overall		
Nation	149*	149*
Large City	135**	134**
Atlanta	134**	127*,**
Austin	147*	149*
Baltimore City	117*,**	113*,**
Boston	139*,**	130*,**
Charlotte	150*	141*,**
Chicago	125*,**	121*,**
Cleveland	114*,**	121*,**
Detroit	111*,**	113*,**
Fresno	121*,**	124*,**
Houston	135**	138*,**
Jefferson County (KY)	150*	145*,**
Los Angeles	124*,**	123*,**
Miami-Dade	144*,**	137*,**
Milwaukee	126*,**	122*,**
New York City	135**	129*,**
Philadelphia	121*,**	119*,**
San Diego	144*,**	138**
African American		
Nation	127*	125*
Large City	122**	120**
Atlanta	126*	123
Austin	129	138*,**
Baltimore City	115*,**	110*,**
Boston	133*,**	120**
Charlotte	131*,**	126*
Chicago	113*,**	110*,**
Cleveland	109*,**	117**
Detroit	109*,**	113*,**
Fresno	110*,**	117
Houston	128*	128*
Jefferson County (KY)	129*	128*
Los Angeles	117**	113*,**
Miami-Dade	125	123
Milwaukee	115*,**	115*,**
New York City	125	119**
Philadelphia	115*,**	112*,**
San Diego	124	125

(Table B.7 continued) Average reported NAEP science scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, national public, 2009

White		
Nation	162	161
Large City	163	159
Atlanta	181*,**	
Austin	183*,**	178*,**
Baltimore City	143*,**	
Boston	161	160
Charlotte	174*,**	167*,**
Chicago	154	150*,**
Cleveland	136*,**	144*,**
Detroit	†	†
Fresno	144*,**	151*,**
Houston	174*,**	172*,**
Jefferson County (KY)	163	157**
Los Angeles	152*,**	152*,**
Miami-Dade	169*,**	159
Milwaukee	158	143*,**
New York City	159	151*,**
Philadelphia	141*,**	139*,**
San Diego	169**	158
Hispanic		
Nation	130*	131*
Large City	127**	127**
Atlanta	†	†
Austin	133*	134*,**
Baltimore City	†	†
Boston	134*	123**
Charlotte	136*	131
Chicago	128	125**
Cleveland	113*,**	122**
Detroit	122	117
Fresno	118*,**	119*,**
Houston	133*	137*,**
Jefferson County (KY)	138	†
Los Angeles	119*,**	118*,**
Miami-Dade	146*,**	138*,**
Milwaukee	132	127
New York City	127	120*,**
Philadelphia	120*,**	115*,**
San Diego	128	123**

(Table B.7 continued) Average reported NAEP science scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, national public, 2009

Asian/Pacific Islander		
Nation	160*	159*
Large City	152**	152**
Atlanta	†	†
Austin	†	†
Baltimore City	†	†
Boston	154	157
Charlotte	163	†
Chicago	159	†
Cleveland	†	†
Detroit	†	†
Fresno	123*,**	125*,**
Houston	160	166
Jefferson County (KY)	†	†
Los Angeles	151	156
Miami-Dade	†	†
Milwaukee	†	†
New York City	153	156
Philadelphia	141*,**	139**
San Diego	157	148**
NSLP-Eligible		
Nation	134*	133*
Large City	126**	125**
Atlanta	123*,**	120*,**
Austin	130*	130*
Baltimore City	114*,**	110*,**
Boston	134*	123**
Charlotte	132*	126**
Chicago	120*,**	118*,**
Cleveland	114*,**	121**
Detroit	108*,**	110*,**
Fresno	118*,**	119*,**
Houston	130*,**	133*
Jefferson County (KY)	136*	133*
Los Angeles	120*,**	119*,**
Miami-Dade	135*	130*
Milwaukee	120*,**	118*,**
New York City	132*	125**
Philadelphia	119*,**	115*,**
San Diego	128**	125**

(Table B.7 continued) Average reported NAEP science scale scores of public school students in grades 4 and 8, overall and by selected student characteristics, TUDA district, large city, national public, 2009

Limited English Proficiency		
Nation	114*	103*
Large City	111**	97**
Atlanta	†	†
Austin	120*,**	104
Baltimore City	†	†
Boston	119*	88**
Charlotte	127*,**	111*
Chicago	102*,**	99
Cleveland	†	†
Detroit	†	112
Fresno	105**	93**
Houston	124*,**	104
Jefferson County (KY)	†	†
Los Angeles	104*,**	88*,**
Miami-Dade	113	92**
Milwaukee	127*,**	†
New York City	110	95
Philadelphia	98*,**	97
San Diego	117*	93**
Students with Disabilities		
Nation	129*	122*
Large City	112**	103**
Atlanta	110**	98**
Austin	130*	124*
Baltimore City	111**	90*,**
Boston	121*,**	99**
Charlotte	130*	112**
Chicago	102*,**	96*,**
Cleveland	93*,**	97**
Detroit	88*,**	83*,**
Fresno	98*,**	91*,**
Houston	109**	97**
Jefferson County (KY)	126*	120*
Los Angeles	89*,**	88*,**
Miami-Dade	118**	112*,**
Milwaukee	102*,**	99**
New York City	117**	105**
Philadelphia	94*,**	92*,**
San Diego	115**	109**

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; ‡ Data are not comparable to 2009 science results. Data are not comparable to 2005 science results.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Science Assessments

Table B.8 Average reported NAEP reading performance levels of public school students in grades 4 and 8, overall and by TUDA district, large city, and national public, 2003-2009

Reading								
4th Grade	2003		2005		2007		2009	
District	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient
Atlanta	37***	14***	41***	17***	48	18***	50*,**	22**
Austin	—	—	61	28	62	30	65*	32*
Baltimore City	—	—	—	—	—	—	42*,**	12*,**
Boston	48***	16***	51***	16***	54***	20	61*,**	24**
Charlotte	64***	31	65	33	66	35	71*,**	36*
Chicago	40***	14	40	14	44	16	45*,**	16*,**
Cleveland	35	9	37	10	39	9	34*,**	8*,**
Detroit	—	—	—	—	—	—	27*,**	5*,**
District of Columbia	31***	10***	33***	11***	39***	14***	46*,**	18*,**
Fresno	—	—	—	—	—	—	40*,**	12*,**
Houston	48***	18	52	21	49***	17	55**	19**
Jefferson County	—	—	—	—	—	—	64*	30*
Los Angeles	35***	11	37	14	39	13	40*,**	13*,**
Miami-Dade	—	—	—	—	—	—	68*	31*
Milwaukee	—	—	—	—	—	—	39*,**	12*,**
New York City	53***	22***	57	22***	57***	25	62*,**	29*
Philadelphia	—	—	—	—	—	—	39*,**	11*,**
San Diego	51***	22***	51***	22***	55	25	59*,**	29*
National Public	62***	30***	62***	30***	66	32	66*	32*
Large City	47***	19***	49***	20***	53	22	54**	23**
8th Grade	2003		2005		2007		2009	
District	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient
Atlanta	47***	11***	46***	12***	53***	13	60**	17*,**
Austin	—	—	65***	27	66	28	71*	30*
Baltimore City	—	—	—	—	—	—	54*,**	10*,**
Boston	61***	22	61***	23	63	22	68**	23**
Charlotte	71	30	69	29	69	29	70*,**	28*
Chicago	59	15	60	17	61	17	60**	17*,**
Cleveland	48	10	49	10	56	11	52*,**	10*,**
Detroit	—	—	—	—	—	—	40*,**	7*,**
District of Columbia	47	10***	45	12	48	12	48*,**	14*,**
Fresno	—	—	—	—	—	—	48*,**	12*,**
Houston	55***	14***	59***	17	63	18	64**	18**
Jefferson County	—	—	—	—	—	—	68*,**	26*,**
Los Angeles	43***	11***	47***	13	50***	12	54*,**	15*,**
Miami-Dade	—	—	—	—	—	—	73*	28*
Milwaukee	—	—	—	—	—	—	51*,**	12*,**
New York City	62	22	61	20	59	20	62**	21**
Philadelphia	—	—	—	—	—	—	56*,**	15**
San Diego	60	20	63	23	60	23	65**	25
National Public	72***	30	71***	29***	73***	29***	74*	30*
Large City	58***	19***	60***	20	60***	20	63**	21**

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; ‡ Data are not comparable to 2009 science results.

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

Table B.9 Average reported NAEP mathematics performance levels of public school students in grades 4 and 8, overall and by TUDA district, large city, and national public, 2003-2009

Mathematics								
4th Grade	2003		2005		2007		2009	
District	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient
Atlanta	50***	13***	57***	17***	61	20	63*,**	21*,**
Austin	—	—	85	40	83	40	83*	38*
Baltimore City	—	—	—	—	—	—	64*,**	13*,**
Boston	59***	12***	72***	22***	77	27	81*	31**
Charlotte	84	41	86	44	85	44	86*,**	45*,**
Chicago	50***	10***	52***	13	58	16	62*,**	18*,**
Cleveland	51	10	60***	13***	53	10	51*,**	8*,**
Detroit	—	—	—	—	—	—	31*,**	3*,**
District of Columbia	36***	7***	45***	10***	49***	14***	57*,**	19*,**
Fresno	—	—	—	—	—	—	58*,**	14*,**
Houston	70***	18***	77	26	80	28	82*	30**
Jefferson County	—	—	—	—	—	—	72**	31**
Los Angeles	52***	13***	58	18	60	19	61*,**	19*,**
Miami-Dade	—	—	—	—	—	—	81*	33**
Milwaukee	—	—	—	—	—	—	59*,**	15*,**
New York City	67***	21***	73***	26***	79	34	79*	35*
Philadelphia	—	—	—	—	—	—	61*,**	16*,**
San Diego	66***	20***	74	29***	74	35	77*	36*
National Public	76***	31***	79***	35***	81	39	81*	38*
Large City	63***	20***	68***	24***	70***	28	72**	29**
8th Grade	2003		2005		2007		2009	
District	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient
Atlanta	30***	6***	31***	7***	41	11	46*,**	11*,**
Austin	—	—	68***	33***	72	34***	75*,**	39*,**
Baltimore City	—	—	—	—	—	—	43*,**	10*,**
Boston	48***	17***	58***	23***	65	27***	67*,**	31*
Charlotte	67***	32	69	33	70	34	72*	33*
Chicago	42***	9***	45***	11***	49	13	51*,**	15*,**
Cleveland	38	6	34***	6	45	7	42*,**	8*,**
Detroit	—	—	—	—	—	—	23*,**	4*,**
District of Columbia (DCPS)	29***	6***	31***	7***	34***	8***	38*,**	12*,**
Fresno	—	—	—	—	—	—	46*,**	15*,**
Houston	52***	12***	58***	16***	65	21	69*	24**
Jefferson County (KY)	—	—	—	—	—	—	60**	22**
Los Angeles	32***	7***	38***	11***	45	14	46*,**	13*,**
Miami-Dade	—	—	—	—	—	—	64*,**	22**
Milwaukee	—	—	—	—	—	—	37*,**	7*,**
New York City	54***	20***	54***	20	57	22	60**	26**
Philadelphia	—	—	—	—	—	—	52*,**	17*,**
San Diego	53***	18***	61***	22***	62***	24***	68*	32*
National Public	67***	27***	68***	28***	70***	31***	71*	33*
Large City	50***	16***	53***	19***	57***	22***	60**	24**

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; ‡ Data are not comparable to 2009 science results.

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

Table B.10. Average reported NAEP science performance levels of public school students in grades 4 and 8, overall and by TUDA district, large city, and national public, 2005‡

Science				
	4th Grade		8th Grade	
District	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient
Atlanta	42*	13	23*	7
Austin	60*	25	52*	27
Boston	43*	10	38	14
Charlotte	60*	23	51*	24
Chicago	34*	8	28*	9
Cleveland	37*	6	26*	5
Houston	48	15	29	12
Los Angeles	35*	9	29*	9
New York City	46	13	36	14
San Diego	52	19	43*	18
National Public	66	27	57	27
Large City	47	15	40	16

*Statistically different from large cities at $p < .05$; Test of significance were not available for comparisons between districts and the national public sample and at or above proficient levels; ‡ Data are not comparable to 2009 science results.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Science Assessments

Table B.11 Average reported NAEP science performance levels of public school students in grades 4 and 8, overall and by TUDA district, large city, and national public, 2009

	2009			
Science	4th Grade		8th Grade	
District	% At or Above Basic	% At or Above Proficient	% At or Above Basic	% At or Above Proficient
Atlanta	52*,**	19**	33*,**	10*,**
Austin	65*,**	31*	61*	33*,**
Baltimore City	31*,**	6*,**	20*,**	4*,**
Boston	62*,**	18**	39*,**	15**
Charlotte	70*	33*	52*,**	22*,**
Chicago	44*,**	12*,**	29*,**	7*,**
Cleveland	30*,**	4*,**	26*,**	6*,**
Detroit	26*,**	4*,**	20*,**	3*,**
Fresno	38*,**	8*,**	34*,**	9*,**
Houston	55**	16*,**	49*,**	17**
Jefferson County	70*	33*	57*,**	24*,**
Los Angeles	45*,**	11*,**	33*,**	10*,**
Miami-Dade	66*,**	25*,**	49*,**	18**
Milwaukee	44*,**	12*,**	28*,**	6*,**
New York City	56**	18**	38*,**	13*,**
Philadelphia	38*,**	8*,**	25*,**	6*,**
San Diego	65*,**	29*	49**	20*,**
National Public	71*	32*	62*	29*
Large City	56**	20**	44**	17**

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Data are not comparable to 2005 results.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Science Assessment

Table B.12 Changes in the average scale score of grade 4 African American public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	2.4	2.3	2.7	2.7	1.9	2.6
Austin	—	†	†	†	†	†
Boston	0.3	0.4	-0.1	0.4	1.1	-0.1
Charlotte	1.3	1.9	1.1	0.5	0.7	2.1
Chicago	-3.1	-1.7	-3.4	-3.3	-2.2	-4.7
Cleveland	3.2	8.3*	2.9	2.1	1.9	0.8
DC	1.4	2.1	1.4	0.9	1.2	1.2
Houston	4.0	0.4	4.0	4.1	4.4	7.2
Los Angeles	0.2	†	†	†	†	†
New York City	4.4	7.5	5.4	4.7	2.9	1.5
San Diego	-0.3	†	†	†	†	†
National Public	1.3*	3.2*	1.8*	1.2	0.6	-0.5
Large City	2.8*	4.9*	3.5*	2.7*	2.1	1.1
Changes 2005 to 2007						
Atlanta	4.1*	4.5*	6.5*	5.4*	4.0	0.0
Austin	0.8	†	†	†	†	†
Boston	3.0	1.6	3.8	3.4	2.9	3.2
Charlotte	-0.5	-2.7	1.7	0.6	-0.3	-1.9
Chicago	4.0	4.7	3.4	4.0	3.5	4.2
Cleveland	-7.8	-21.4	-5.6	-3.9	-3.5	-4.6
DC	2.2	0.3	2.3	3.0	2.4	3.0
Houston	-2.9	-6.4	-1.6	-1.0	-1.5	-3.8
Los Angeles	8.9	†	†	†	†	†
New York City	-0.1	-2.8	1.2	1.5	1.2	-1.8
San Diego	1.0	†	†	†	†	†
National Public	3.6*	2.3*	4.9*	4.6*	3.8*	2.5*
Large City	1.6	-1.4	2.8	2.8*	2.6*	1.4
Changes 2003 to 2007						
Atlanta	6.5*	6.8	9.2*	8.1*	5.9*	2.6
Austin	—	†	†	†	†	†
Boston	3.3	1.9	3.7	3.8	4.0	3.1
Charlotte	0.7	-0.8	2.8	1.0	0.4	0.2
Chicago	0.9	3.0	0.0	0.7	1.3	-0.5
Cleveland	-4.6	-13.1	-2.8	-1.8	-1.5	-3.8
DC	3.6*	2.5	3.7*	3.9*	3.6*	4.2*
Houston	1.2	-5.9	2.4	3.1	2.9	3.3
Los Angeles	9.1	10.8	†	†	†	†
New York City	4.2	4.7	6.6	6.2	4.0	-0.3
San Diego	0.7	†	†	†	†	†
National Public	4.9*	5.5*	6.7*	5.8*	4.4*	2.0*
Large City	4.5*	3.5	6.3*	5.5*	4.7*	2.4

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.13 Changes in the average scale score of grade 8 African American public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	-0.1	-1.1	-0.9	-0.7	-0.5	2.6
Austin	—	†	†	†	†	†
Boston	-1.5	-4.2	-0.9	-1.3	-1.2	0.2
Charlotte	-3.7	-11.4*	-3.8	-1.2	-1.7	-0.7
Chicago	-3.2	-11.3*	-3.8	-1.4	0.2	0.1
Cleveland	-3.7	-10.9*	-4.4	-2.3	-0.6	-0.4
DC	-0.7	0.1	-1.4	-1.9	-1.1	1.0
Houston	-1.2	-7.2	-1.0	0.9	1.4	0.0
Los Angeles	1.3	†	†	†	†	†
New York City	-3.5	-0.1	-4.4	-4.7	-3.9	-4.4
San Diego	3.9	†	†	†	†	†
National Public	-1.9*	-3.3*	-2.2*	-1.7*	-1.4	-1.0
Large City	-1.4	-4.1*	-1.7	-0.9	-0.3	0.1
Changes 2005 to 2007						
Atlanta	2.6	0.4	4.8	4.4*	3.5	0.1
Austin	1.2	†	†	†	†	†
Boston	5.0	5.6	6.0	5.4	4.7	3.1
Charlotte	2.4	5.1	2.3	1.2	1.8	1.6
Chicago	-1.3	-2.6	-0.8	-0.6	-1.0	-1.4
Cleveland	2.8	2.3	5.3	4.1	3.0	-0.7
DC	-0.6	-3.0	-0.8	0.1	0.5	0.3
Houston	3.0	1.8	5.9	4.1	2.2	1.0
Los Angeles	-4.2	†	†	†	†	†
New York City	-0.3	-3.4	0.4	0.4	-0.4	1.5
San Diego	-0.9	†	†	†	†	†
National Public	1.2*	0.0	2.4*	2.0*	1.4*	0.4
Large City	-0.5	-2.1	0.5	0.2*	-0.4*	-0.8
Changes 2003 to 2007						
Atlanta	2.5	-0.7	4.0	3.7	3.0	2.6
Austin	—	†	†	†	†	†
Boston	3.5	1.4	5.1	4.2	3.5	3.2
Charlotte	-1.4	-6.3	-1.5	0.0	0.2	0.8
Chicago	-4.5	-13.9*	-4.7	-2.0	-0.8	-1.3
Cleveland	-0.9	-8.7	0.9	1.8	2.4	-1.1
DC	-1.3	-2.9	-2.2	-1.7	-0.7	1.3
Houston	1.8	-5.4	4.9	5.0	3.6	1.0
Los Angeles	-2.9	†	†	†	†	†
New York City	-3.8	-3.5	-4.0	-4.2	-4.2	-2.9
San Diego	3.0	†	†	†	†	†
National Public	-0.7	-3.3*	0.2	0.3	0.0	-0.7
Large City	-1.9	-6.2*	-1.2	-0.7	-0.8	-0.7

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.14 Changes in the average scale score of grade 4 African American public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	3.7*	4.2	3.8*	3.6	3.8*	3.2
Austin	—	†	†	†	†	†
Boston	6.7*	3.5	5.2*	6.5*	8.3*	10.2*
Charlotte	0.9	-1.7	0.8	1.4	1.2	2.6
Chicago	0.7	-2.1	-1.4	-0.1	2.0	4.9
Cleveland	5.0*	3.9	4.8*	5.5*	5.4*	5.3*
DC	4.8*	5.7*	5.3*	4.8*	4.8*	3.3*
Houston	2.0	-0.1	2.1	2.8	3.0	2.1
Los Angeles	2.1	†	†	†	†	†
New York City	2.7	2.2	3.3	3.1	2.7	2.3
San Diego	2.2	†	†	†	†	†
National Public	3.8*	2.9*	3.8*	4.0*	4.1*	4.0*
Large City	4.5*	2.7*	4.6*	5.2*	5.2*	4.9*
Changes 2005 to 2007						
Atlanta	1.8	-0.4	1.9	2.4	2.4	2.8
Austin	-0.9	†	†	†	†	†
Boston	3.0	-0.4	3.6	4.1*	4.0	3.5
Charlotte	0.7	-0.6	0.4	1.2	2.0	0.3
Chicago	5.1*	4.8	6.2*	5.6*	5.1*	3.8
Cleveland	-8.9*	-15.9*	-8.7*	-7.1*	-6.5*	-6.2*
DC	1.3	-3.5	0.3	2.5	2.9*	4.5*
Houston	0.7	-3.1	1.2	1.9	1.7	1.7
Los Angeles	7.5*	†	†	†	†	†
New York City	5.6*	5.6	6.1*	5.8*	5.5*	4.9
San Diego	1.8	†	†	†	†	†
National Public	2.1*	0.6	2.4*	2.6*	2.4*	2.4*
Large City	1.7	0.1	2.0	2.2*	2.1	1.9
Changes 2003 to 2007						
Atlanta	5.6*	3.8	5.7*	6.1*	6.2*	6.0*
Austin	—	†	†	†	†	†
Boston	9.7*	3.1	8.8*	10.6*	12.4*	13.7*
Charlotte	1.5	-2.2	1.3	2.7	3.2	2.9
Chicago	5.8*	2.8	4.8	5.5*	7.1*	8.7*
Cleveland	-3.9	-12.0*	-3.9	-1.5	-1.1	-0.9
DC	6.1*	2.2	5.6*	7.3*	7.7*	7.9*
Houston	2.7	-3.1	3.3	4.7	4.7	3.8
Los Angeles	9.5*	†	†	†	†	†
New York City	8.3*	7.9*	9.4*	8.9*	8.2*	7.3*
San Diego	4.0	†	†	†	†	†
National Public	5.9*	3.5*	6.2*	6.6*	6.5*	6.4*
Large City	6.2*	2.9*	6.6*	7.4*	7.3*	6.9*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.15 Changes in the average scale score of grade 8 African American public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	0.9	-0.7	0.6	0.6	1.5	2.3
Austin	—	†	†	†	†	†
Boston	3.9	-2.4	4.0	5.3	6.3*	6.2
Charlotte	4.3*	1.5	4.4	5.9*	5.9*	3.5
Chicago	0.2	1.2	1.9	0.8	0.0	-2.8
Cleveland	-5.1*	-10.2*	-5.3*	-3.1	-2.4	-4.2
DC	2.2	5.1	3.0	1.4	0.6	0.8
Houston	-2.2	-8.0	-2.1	-0.4	-0.3	-0.1
Los Angeles	5.1	†	†	†	†	†
New York City	3.9	6.6	3.3	2.9	3.2	3.7
San Diego	-0.3	†	†	†	†	†
National Public	2.3*	2.0	2.3*	2.2*	2.4*	2.7*
Large City	2.2	1.7	1.5	1.9	2.2	3.6*
Changes 2005 to 2007						
Atlanta	10.3*	11.1*	11.8*	9.8*	8.5*	10.4*
Austin	5.8	†	†	†	†	†
Boston	6.4*	9.3	6.1	5.2	5.7*	5.4
Charlotte	4.4*	10.0*	6.4*	2.8	1.5	1.0
Chicago	2.5	-2.4	-0.5	3.2	4.6	7.5
Cleveland	4.7	-2.6	5.1	6.4*	7.3*	7.2*
DC	1.6	-3.2	0.9	2.0	3.2	4.8*
Houston	5.6*	0.2	4.8	7.1*	7.7*	8.4*
Los Angeles	5.1	†	†	†	†	†
New York City	1.2	3.3	1.4	1.4	1.0	-1.1
San Diego	4.1	†	†	†	†	†
National Public	3.9*	3.7*	4.2*	3.9*	4.0*	3.9*
Large City	3.7*	2.8	3.7*	3.9*	4.2*	4.0*
Changes 2003 to 2007						
Atlanta	11.2*	10.4*	12.4*	10.4*	10.0*	12.7*
Austin	—	†	†	†	†	†
Boston	10.2*	6.9	10.1*	10.5*	11.9*	11.6*
Charlotte	8.6*	11.6*	10.9*	8.7*	7.4*	4.5
Chicago	2.7	-1.2	1.4	4.0	4.7	4.7
Cleveland	-0.4	-12.8	-0.2	3.3	5.0*	3.0
DC	3.8*	2.0	3.9	3.4	3.8*	5.6*
Houston	3.5	-7.8	2.7	6.6*	7.5*	8.3*
Los Angeles	10.2*	†	†	†	†	†
New York City	5.2	9.9*	4.7	4.3	4.2	2.6
San Diego	3.8	†	†	†	†	†
National Public	6.3*	5.7*	6.5*	6.1*	6.4*	6.6*
Large City	5.9*	4.5*	5.2*	5.8*	6.4*	7.6*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.16 Changes in the average scale score of grade 4 White public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	1.8	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	3.1	†	†	†	†	†
Charlotte	3.3	2.3	1.9	2.8	3.8	5.7
Chicago	1.8	†	†	†	†	†
Cleveland	4.3	†	†	†	†	†
DC	-2.8	†	†	†	†	†
Houston	9.9*	†	†	†	†	†
Los Angeles	12.1*	†	†	†	†	†
New York City	-4.5	†	†	†	†	†
San Diego	-4.5	†	†	†	†	†
National Public	0.3	1.1	0.7	0.3	-0.2	-0.3
Large City	1.2	4.0	2.1	1.1	-0.1	-0.8
Changes 2005 to 2007						
Atlanta	0.5	†	†	†	†	†
Austin	5.1	5.5	7.8*	5.9	4.2	2.1
Boston	2.2	†	†	†	†	†
Charlotte	3.3	4.8	5.3	3.5	2.6	0.1
Chicago	1.0	†	†	†	†	†
Cleveland	2.5	†	†	†	†	†
DC	2.5	†	†	†	†	†
Houston	-5.3	†	†	†	†	†
Los Angeles	-0.9	†	†	†	†	†
New York City	6.6	†	†	†	†	†
San Diego	8.8	†	†	†	†	†
National Public	1.8*	1.6*	2.8*	2.0*	1.5*	1.2*
Large City	3.7*	4.0	4.8*	4.0*	3.3*	2.2
Changes 2003 to 2007						
Atlanta	2.3	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	5.2	†	†	†	†	†
Charlotte	6.6*	7.1	7.2*	6.3*	6.4*	5.8
Chicago	2.8	†	†	†	†	†
Cleveland	6.8	†	†	†	†	†
DC	-0.3	†	†	†	†	†
Houston	4.6	†	†	†	†	†
Los Angeles	11.2*	†	†	†	†	†
New York City	2.1	0.1	3.0	2.3	2.0	3.2
San Diego	4.3	4.4	3.2	4.6	5.3	4.2
National Public	2.1*	2.7*	3.4*	2.3*	1.4*	0.9*
Large City	4.9*	8.0*	6.9*	5.0*	3.2*	1.4

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.17 Changes in the average scale score of grade 8 White public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	2.8	†	†	†	†	†
Charlotte	1.1	-1.1	-0.4	0.5	1.9	4.4
Chicago	3.6	†	†	†	†	†
Cleveland	1.2	†	†	†	†	†
DC	2.7	†	†	†	†	†
Houston	9.8*	†	†	†	†	†
Los Angeles	-3.8	†	†	†	†	†
New York City	-0.2	†	†	†	†	†
San Diego	4.3	5.5	3.5	4.6	6.0	†
National Public	-1.3*	-2.3*	-1.6*	-1.3*	-0.9*	-0.4
Large City	1.7	0.9	1.9	1.6	1.6	2.5
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	4.8	7.8	6.8	5.0	3.9	0.7
Boston	0.8	†	†	†	†	†
Charlotte	0.2	-0.1	1.7	1.2	0.5	-2.2
Chicago	-1.3	†	†	†	†	†
Cleveland	8.6	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	1.1	†	†	†	†	†
Los Angeles	9.5	†	†	†	†	†
New York City	1.8	†	†	†	†	†
San Diego	-3.3	-8.4	-0.8	-1.5	-2.8	-3.2
National Public	0.6*	1.2	1.7*	0.9*	0.2	-0.8*
Large City	0.1	-0.8	0.8	0.5	0.4	-0.3
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	3.5	†	†	†	†	†
Charlotte	1.3	-1.2	1.3	1.7	2.3	2.2
Chicago	2.3	†	†	†	†	†
Cleveland	9.8*	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	10.9*	†	†	†	†	†
Los Angeles	5.7	†	†	†	†	†
New York City	1.6	†	†	†	†	†
San Diego	0.9	-2.9	2.7	3.1	3.3	†
National Public	-0.7*	-1.2	0.0	-0.4	-0.7*	-1.2*
Large City	1.8	0.1	2.8	2.1	1.9	2.1

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.18 Changes in the average scale score of grade 4 White public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	5.5	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	9.8*	†	†	†	†	†
Charlotte	4.3	3.6	3.2	4.3	4.8*	5.7*
Chicago	7.3	†	†	†	†	†
Cleveland	0.1	†	†	†	†	†
DC	5.1	†	†	†	†	†
Houston	9.1*	†	†	†	†	†
Los Angeles	6.2	†	†	†	†	†
New York City	1.4	†	†	†	†	†
San Diego	7.0*	11.6*	7.0*	6.1*	4.9	5.2
National Public	3.0*	2.8*	3.4*	3.2*	2.8*	2.8*
Large City	4.5*	2.5	5.0*	5.4*	5.1*	4.4*
Changes 2005 to 2007						
Atlanta	3.2	†	†	†	†	†
Austin	0.9	-2.9	1.9	3.2	2.1	0.3
Boston	4.9	†	†	†	†	†
Charlotte	-0.8	0.0	0.9	-0.9	-1.5	-2.4
Chicago	2.2	†	†	†	†	†
Cleveland	-2.7	†	†	†	†	†
DC	-3.9	†	†	†	†	†
Houston	0.7	†	†	†	†	†
Los Angeles	0.6	†	†	†	†	†
New York City	3.9	†	†	†	†	†
San Diego	2.5	-7.4	3.2	5.4	5.7	5.5
National Public	2.0*	1.7*	2.5*	2.4*	2.2*	1.5*
Large City	1.9	0.7	3.1*	2.6*	1.8	1.1
Changes 2003 to 2007						
Atlanta	8.8	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	14.7*	†	†	†	†	†
Charlotte	3.5	3.5	4.1	3.4	3.4	3.3
Chicago	9.5*	†	†	†	†	†
Cleveland	-2.6	†	†	†	†	†
DC	1.2	†	†	†	†	†
Houston	9.8*	†	†	†	†	†
Los Angeles	6.8	†	†	†	†	†
New York City	5.3*	2.8	7.1*	7.1*	6.5	2.9
San Diego	9.5*	4.2	10.2*	11.5*	10.6*	10.7*
National Public	5.0*	4.5*	5.9*	5.6*	5.0*	4.3*
Large City	6.4*	3.2	8.2*	8.0*	7.0*	5.5*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.19 Changes in the average scale score of grade 8 White public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	10.9*	†	†	†	†	†
Charlotte	3.1	-0.8	2.2	4.9	4.9	4.5
Chicago	8.0	†	†	†	†	†
Cleveland	-3.3	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	0.8	†	†	†	†	†
Los Angeles	2.3	†	†	†	†	†
New York City	-2.1	†	†	†	†	†
San Diego	6.9*	6.2	7.2	5.7	6.5	8.8*
National Public	0.7*	-0.3	0.4	0.5	1.1*	1.9*
Large City	2.7*	1.9	1.7	2.3	3.2*	4.3*
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	3.2	4.0	4.0	3.8	3.0	1.2
Boston	7.3	†	†	†	†	†
Charlotte	4.0	4.4	3.0	2.7	4.4	5.6
Chicago	2.9	†	†	†	†	†
Cleveland	-0.2	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	14.2*	†	†	†	†	†
Los Angeles	5.3	†	†	†	†	†
New York City	2.8	†	†	†	†	†
San Diego	0.9	1.1	0.4	0.4	1.5	1.1
National Public	2.5*	1.9*	2.4*	2.7*	2.7*	2.6*
Large City	3.6*	2.6	4.1*	3.9*	3.9*	3.6
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	†	†	†	†	†	†
Boston	18.1*	†	†	†	†	†
Charlotte	7.1*	3.6	5.1	7.6*	9.3*	10.1*
Chicago	10.9	†	†	†	†	†
Cleveland	-3.5	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	14.9*	†	†	†	†	†
Los Angeles	7.6	†	†	†	†	†
New York City	0.7	†	†	†	†	†
San Diego	7.8*	7.3	7.6	6.1	7.9*	9.9*
National Public	3.2*	1.7*	2.8*	3.2*	3.8*	4.5*
Large City	6.3*	4.4*	5.8*	6.2*	7.0*	7.9*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.20 Changes in the average scale score of grade 4 Hispanic public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-0.9	0.7	-0.6	-0.7	-1.1	-2.9
Charlotte	3.3	†	†	†	†	†
Chicago	5.2	7.1	6.9	4.5	3.7	3.8
Cleveland	1.1	†	†	†	†	†
DC	3.7	†	†	†	†	†
Houston	2.6	6.5	4.0	2.8	0.9	-1.3
Los Angeles	1.5	1.8	-0.2	0.0	1.4	4.3
New York City	1.6	1.8	3.4	3.2	0.9	-1.0
San Diego	1.4	2.6	2.9	2.0	1.0	-1.5
National Public	1.7*	2.7*	2.2*	1.7*	1.3	0.5
Large City	1.1	1.4	1.1	1.3	1.1	0.6
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	-2.2	-9.3*	-1.2	0.1	-0.3	-0.5
Boston	2.8	-5.6	3.1	4.4	5.5	6.7
Charlotte	-1.9	†	†	†	†	†
Chicago	-0.3	-6.3	-0.4	2.7	1.8	0.7
Cleveland	-18.9	†	†	†	†	†
DC	3.5	†	†	†	†	†
Houston	-2.1	-9.1*	-3.0	0.1	1.6	-0.2
Los Angeles	0.5	-4.2	3.0	3.5	1.9	-1.8
New York City	-3.8	-11.2*	-6.2	-3.8	-0.1	2.5
San Diego	-0.9	-11.5	-0.1	1.9	2.6	2.7
National Public	1.9*	-3.3*	3.5*	3.9	3.1*	2.1*
Large City	0.7	-5.4*	1.8	3.0*	2.6*	1.3
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	1.9	-4.9	2.5	3.6	4.4	3.8
Charlotte	1.4	†	†	†	†	†
Chicago	4.9	0.8	6.5	7.2	5.5	4.5
Cleveland	-17.8	†	†	†	†	†
DC	7.2	†	†	†	†	†
Houston	0.5	-2.6	1.1	2.9	2.4	-1.5
Los Angeles	1.9	-2.5	2.8	3.5	3.3	2.5
New York City	-2.1	-9.5*	-2.9	-0.6	0.9	1.5
San Diego	0.5	-8.9	2.8	4.0	3.6	1.2
National Public	3.5*	-0.6	5.7*	5.6*	4.4*	2.6*
Large City	1.8	-4.0	2.9*	4.3*	3.7*	1.9

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.21 Changes in the average scale score of grade 8 Hispanic public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	5.6	6.6	7.2	5.7	6.0	2.7
Charlotte	2.5	†	†	†	†	†
Chicago	2.9	4.3	3.0	2.8	2.6	1.7
Cleveland	12.2	†	†	†	†	†
DC	7.1	†	†	†	†	†
Houston	3.1	1.0	3.9	4.3	4.1	2.3
Los Angeles	6.8*	10.0*	6.8*	5.9*	5.5*	5.9*
New York City	0.3	4.2	2.9	0.5	-1.5	-4.7
San Diego	-0.1	-5.2	-0.1	0.8	0.8	3.3
National Public	1.4	4.1*	1.9	0.8	0.4	-0.2
Large City	3.8*	6.2*	5.1*	3.1	2.7	1.8
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	1.8	-4.5	2.8	4.7	4.1	1.8
Boston	-5.9	-6.1	-5.4	-3.7	-5.8	-8.5
Charlotte	-0.3	†	†	†	†	†
Chicago	2.3	-0.9	4.1	3.5	2.6	2.3
Cleveland	-10.4	†	†	†	†	†
DC	-0.9	†	†	†	†	†
Houston	0.8	-1.1	1.5	1.5	1.2	1.1
Los Angeles	0.7	-0.3	2.2	2.3	0.8	-1.4
New York City	-5.4	-7.8	-7.9*	-5.6	-4.3	-1.6
San Diego	-2.6	-6.1	-1.6	-1.8	-1.7	-1.8
National Public	0.4	-2.3	1.1	1.8*	1.1	0.4
Large City	-1.4	-3.3	-0.7	-0.1	-1.1	-1.7
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-0.3	0.5	1.9	2.1	0.1	-5.8
Charlotte	2.2	†	†	†	†	†
Chicago	5.2	3.5	7.1*	6.3*	5.2	4.1
Cleveland	1.8	†	†	†	†	†
DC	6.2	†	†	†	†	†
Houston	4.0	-0.1	5.4	5.8*	5.3*	3.4
Los Angeles	7.5*	9.7*	9.0*	8.2*	6.3*	4.5
New York City	-5.2	-3.6	-5.0	-5.1	-5.7	-6.4
San Diego	-2.7	-11.4*	-1.6	-0.9	-0.9	1.5
National Public	1.8*	1.8	3.0*	2.6*	1.4	0.3
Large City	2.4	2.9	4.3*	3.0	1.6	0.1

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.22 Changes in the average scale score of grade 4 Hispanic public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	9.3*	6.7	10.6*	11.1*	9.6*	8.7*
Charlotte	1.1	†	†	†	†	†
Chicago	1.1	-1.4	-0.1	1.0	2.3	3.7
Cleveland	7.0	†	†	†	†	†
DC	8.5*	†	†	†	†	†
Houston	5.2*	2.0	5.0*	6.1*	6.3*	6.6*
Los Angeles	4.4*	-0.8	2.4	4.7*	7.2*	8.4*
New York City	7.2*	7.2*	8.1*	8.2*	7.3*	5.2*
San Diego	5.3*	0.9	4.5*	6.0*	7.4*	7.9*
National Public	3.5*	2.0*	3.7*	4.4*	4.2*	3.5*
Large City	3.7*	1.1	3.3*	4.4*	4.9*	4.7*
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	0.1	-0.3	0.5	0.6	0.6	-0.9
Boston	5.0*	-1.2	4.8	5.5*	7.2*	8.6*
Charlotte	-1.5	†	†	†	†	†
Chicago	1.5	-4.4	2.5	3.6	3.6	2.2
Cleveland	-14.8	†	†	†	†	†
DC	1.9	†	†	†	†	†
Houston	2.6	2.5	4.2*	3.4*	2.6	0.5
Los Angeles	1.2	-1.8	1.8	2.2	2.1	1.6
New York City	4.5*	0.8	3.6	4.8*	6.0*	7.1*
San Diego	0.2	-10.6*	-1.2	3.1	4.3*	5.5
National Public	1.7*	-1.2	1.9*	2.5*	2.7*	2.6*
Large City	1.7*	-2.2	1.9	2.8*	3.0*	3.2*
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	14.3*	5.5	15.4*	16.6*	16.7*	17.3*
Charlotte	-0.4	†	†	†	†	†
Chicago	2.6	-5.8	2.4	4.6	5.9*	6.0*
Cleveland	-7.8	†	†	†	†	†
DC	10.4*	†	†	†	†	†
Houston	7.9*	4.4	9.2*	9.5*	8.9*	7.2*
Los Angeles	5.6*	-2.6	4.2*	6.9*	9.4*	10.0*
New York City	11.7*	8.1*	11.8*	13.0*	13.3*	12.3*
San Diego	5.5*	-9.8*	3.2	9.1*	11.7*	13.4*
National Public	5.2*	0.8	5.6*	6.9*	6.8*	6.1*
Large City	5.4*	-1.2	5.2*	7.2*	7.9*	7.9*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.23 Changes in the average scale score of grade 8 Hispanic public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	4.0	-5.7	2.2	7.1	9.3*	7.1
Charlotte	-0.3	†	†	†	†	†
Chicago	4.6	6.8	4.4	3.5	3.9	4.5
Cleveland	-0.6	†	†	†	†	†
DC	5.5	†	†	†	†	†
Houston	5.3*	2.3	3.7	5.8*	7.2*	7.3*
Los Angeles	4.9*	3.2	5.2*	5.0*	4.9*	6.3*
New York City	-0.6	2.0	0.8	-0.1	-1.9	-3.6
San Diego	7.9*	-1.5	8.2*	12.0*	11.1*	9.6*
National Public	2.7*	1.3	2.9*	3.3*	3.6*	2.5*
Large City	2.4*	1.6	2.5	2.8*	3.4*	1.9
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	6.0*	6.5	8.2*	6.9*	4.0	4.2
Boston	11.8*	16.4	13.4*	9.3*	7.8*	12.0*
Charlotte	2.6	†	†	†	†	†
Chicago	0.8	-1.6	1.4	1.3	1.4	1.4
Cleveland	1.1	†	†	†	†	†
DC	-2.5	†	†	†	†	†
Houston	5.0*	5.6	5.1*	4.8*	4.9*	4.7*
Los Angeles	7.8*	7.1*	7.7*	7.9*	7.8*	8.7*
New York City	4.1	2.8	4.4	4.4	4.2	4.4
San Diego	2.1	3.7	2.0	-0.5	0.6	4.5
National Public	3.5*	3.3*	3.7*	3.4*	3.4*	3.4*
Large City	3.2*	2.3	3.7*	3.3*	3.2*	3.2
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	15.8*	10.7	15.6*	16.3*	17.1*	19.1*
Charlotte	2.4	†	†	†	†	†
Chicago	5.4	5.2	5.8	4.8	5.3	5.9
Cleveland	0.6	†	†	†	†	†
DC	2.9	†	†	†	†	†
Houston	10.3	7.8	8.8	10.6	12.1	12.0
Los Angeles	12.8	10.3	12.9	12.9	12.7	15.0
New York City	3.5	4.9	5.2	4.4	2.3	0.8
San Diego	10.0*	2.2	10.1*	11.6*	11.8*	14.1*
National Public	6.2*	4.6*	6.7*	6.7*	6.9*	5.9*
Large City	5.6*	3.9*	6.2*	6.1*	6.6*	5.1*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.24 Changes in the average scale score of grade 4 Asian public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-1.5	†	†	†	†	†
Charlotte	†	†	†	†	†	†
Chicago	†	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	7.0	†	†	†	†	†
Los Angeles	7.0	†	†	†	†	†
New York City	6.1	†	†	†	†	†
San Diego	-1.5	†	†	†	†	†
National Public	2.8	2.7	3.2	3.4	3.0	1.8
Large City	-0.2	1.6	0.5	-0.1	0.2	-3.1
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	†	†	†	†	†	†
Boston	3.8	†	†	†	†	†
Charlotte	†	†	†	†	†	†
Chicago	†	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	-6.2	†	†	†	†	†
Los Angeles	-4.2	†	†	†	†	†
New York City	-2.7	†	†	†	†	†
San Diego	1.7	†	†	†	†	†
National Public	3.7*	2.6*	5.4*	3.9*	3.6*	2.9
Large City	4.3	-1.4	5.4	5.3	5.1	7.1
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	2.3	†	†	†	†	†
Charlotte	15.7*	†	†	†	†	†
Chicago	†	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	0.9	†	†	†	†	†
Los Angeles	2.9	†	†	†	†	†
New York City	3.4	†	†	†	7.0	†
San Diego	0.2	†	†	†	†	†
National Public	6.5*	5.3*	8.6*	7.3*	6.6*	4.6*
Large City	4.1	0.3	5.9	5.2	5.3	4.1

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.25 Changes in the average scale score of grade 8 Asian public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	7.0	†	†	†	†	†
Charlotte	†	†	†	†	†	†
Chicago	9.5	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	†	†	†	†	†	†
Los Angeles	6.2	†	†	†	†	†
New York City	8.3	†	†	†	†	†
San Diego	3.0	†	†	†	†	†
National Public	1.1	0.1	1.7	1.4	0.7	1.5
Large City	6.0*	7.8*	7.0*	5.4	4.6	5.1
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	†	†	†	†	†	†
Boston	-5.1	†	†	†	†	†
Charlotte	†	†	†	†	†	†
Chicago	†	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	5.1	†	†	†	†	†
Los Angeles	3.3	†	†	†	†	†
New York City	-4.1	†	†	†	†	†
San Diego	1.0	†	†	†	†	†
National Public	-0.3	-2.4	0.6	0.6	0.7	-0.8
Large City	-3.7	-11.6	-2.7	-1.3	-0.8	-1.9
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	1.9	†	†	†	†	†
Charlotte	†	†	†	†	†	†
Chicago	†	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	†	†	†	†	†	†
Los Angeles	9.5*	†	†	†	†	†
New York City	4.2	†	†	†	†	†
San Diego	3.9	†	†	†	†	†
National Public	0.8	-2.3	2.2	2.0	1.4	0.7
Large City	2.3	-3.8	4.2	4.0	3.8	3.2

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.26 Changes in the average scale score of grade 4 Asian public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	14.5*	†	†	†	†	†
Charlotte	4.4	†	†	†	†	†
Chicago	†	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	†	†	†	†	†	†
Los Angeles	5.1	†	†	†	†	†
New York City	8.5*	†	†	†	†	†
San Diego	7.4*	†	†	†	†	†
National Public	5.3*	5.1*	5.1*	5.1*	5.2*	6.2*
Large City	2.2	2.1	2.3	1.8	1.3	3.2
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	†	†	†	†	†	†
Boston	-3.2	†	†	†	†	†
Charlotte	0.5	†	†	†	†	†
Chicago	†	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	†	†	†	†	†	†
Los Angeles	1.6	†	†	†	†	†
New York City	3.6	†	†	†	†	†
San Diego	2.7	†	†	†	†	†
National Public	2.6*	1.2	3.7*	3.4*	3.2*	1.4
Large City	3.1	-2.2	4.5	5.4	5.0	3.0
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	11.3*	†	†	†	†	†
Charlotte	4.9	†	†	†	†	†
Chicago	9.3	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	9.7	†	†	†	†	†
Los Angeles	6.7	†	†	†	†	†
New York City	12.1*	†	13.0*	14.1*	12.7*	10.6*
San Diego	10.0*	†	8.8*	†	†	11.1*
National Public	7.9*	6.3*	8.9*	8.5*	8.4*	7.5*
Large City	5.3	-0.1	6.8*	7.2	6.2	6.2

*Statistically different from large cities at $p < .05$; ** Statistically different from national public at $p < .05$; *** Statistically different from 2003 at $p < .05$; ‡ Reporting standards not met

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

Table B.27 Changes in the average scale score of grade 8 Asian public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	4.7	†	†	†	†	†
Charlotte	5.1	†	†	†	†	†
Chicago	4.4	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	†	†	†	†	†	†
Los Angeles	16.4*	†	†	†	†	†
New York City	11.4	†	†	†	†	†
San Diego	3.7	†	†	†	†	†
National Public	5.3*	4.1*	5.3*	4.8*	5.1*	7.3*
Large City	7.4*	1.6*	6.3*	8.7*	8.1*	12.3*
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	†	†	†	†	†	†
Boston	0.8	†	†	†	†	†
Charlotte	7.0	†	†	†	†	†
Chicago	†	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	9.9	†	†	†	†	†
Los Angeles	-1.4	†	†	†	†	†
New York City	3.8	†	†	†	†	†
San Diego	5.6	†	†	†	†	†
National Public	2.0	1.2	2.2	2.8	3.0	1.0
Large City	2.2	2.1	3.3	3.1	2.6	0.1
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	5.5	†	†	†	†	†
Charlotte	12.1	†	†	†	†	†
Chicago	†	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	†	†	†	†	†	†
Houston	†	†	†	†	†	†
Los Angeles	15.0*	†	†	†	†	†
New York City	15.2*	†	†	†	†	†
San Diego	9.3*	†	†	†	†	†
National Public	7.3*	5.2*	7.4*	7.5*	8.1*	8.3*
Large City	9.6*	3.7	9.6*	11.8*	10.8*	12.4*

*Statistically different from large cities at $p < .05$; ** Statistically different from national public at $p < .05$; *** Statistically different from 2003 at $p < .05$; ‡ Reporting standards not met

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

Table B.28 Changes in the average scale score of grade 4 National School Lunch Program (NSLP)-eligible public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	1.0	1.8	1.6	1.3	-0.2	0.3
Austin	—	†	†	†	†	†
Boston	0.2	0.4	-0.1	0.4	0.5	0.0
Charlotte	4.8	5.1	5.3	4.7	4.0	4.9
Chicago	-0.2	2.3	0.5	-0.6	-0.8	-2.4
Cleveland	3.2	6.7	3.1	2.2	2.6	1.4
DC	0.3	1.7	0.1	-0.4	-0.7	0.5
Houston	1.4	3.5	1.9	1.4	0.4	-0.1
Los Angeles	1.3	1.8	-0.4	-0.5	1.2	4.2
New York City	3.1	3.5	4.1	3.4	2.3	2.4
San Diego	1.9	0.6	1.8	2.1	2.3	2.9
National Public	1.1*	2.4*	1.3*	1.0*	0.6	0.1
Large City	0.8	1.9	0.9	0.7	0.4	0.2
Changes 2005 to 2007						
Atlanta	4.8*	3.6	6.7*	6.0*	5.5*	2.2
Austin	-1.6	-6.9	-0.2	0.7	-0.3	-1.2
Boston	2.2	-2.7	2.7	3.2	3.3	4.6
Charlotte	-0.8	-7.3	1.3	0.8	0.7	0.7
Chicago	3.5	-0.5	2.4	4.6	5.3	5.8*
Cleveland	-6.1	-22.1	-3.8	-1.6	-1.2	-1.8
DC	0.7	-1.7	0.3	1.6	2.3	1.0
Houston	-0.9	-7.5*	-1.2	1.4	2.2	0.4
Los Angeles	0.5	-4.1	3.1	3.7	1.8	-1.9
New York City	-1.4	-6.7*	-2.1	0.1	0.8	0.8
San Diego	-1.0	-8.5	0.8	1.9	1.7	-0.7
National Public	1.8*	-1.4	3.5*	3.4*	2.5*	1.1*
Large City	1.3	-3.8*	2.6*	3.2*	2.9*	1.5
Changes 2003 to 2007						
Atlanta	5.7*	5.3	8.3*	7.3*	5.3*	2.5
Austin	—	†	†	†	†	†
Boston	2.5	-2.3	2.6	3.6	3.8	4.5
Charlotte	4.0	-2.2	6.6*	5.4*	4.8	5.6
Chicago	3.3	1.8	2.9	4.0	4.5	3.4
Cleveland	-2.9	-15.4	-0.7	0.6	1.4	-0.4
DC	1.0	0.0	0.4	1.2	1.6	1.5
Houston	0.5	-3.9	0.7	2.8	2.6	0.3
Los Angeles	1.8	-2.3	2.7	3.2	3.0	2.4
New York City	1.7	-3.2	2.0	3.5	3.1	3.2
San Diego	1.0	-7.9	2.6	4.0	4.0	2.2
National Public	2.9*	0.9	4.8*	4.4*	3.1*	1.2*
Large City	2.1*	-1.8	3.4*	3.9*	3.4*	1.7

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.29 Changes in the average scale score of grade 8 NSLP-eligible public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	-0.8	-1.3	-2.1	-1.5	-0.6	1.5
Austin	—	†	†	†	†	†
Boston	0.7	-0.2	0.9	0.4	1.1	1.5
Charlotte	-1.5	-7.2	-1.7	0.9	-0.2	0.8
Chicago	-0.3	-5.1	-0.3	1.1	1.7	1.0
Cleveland	-1.4	-6.9	-1.3	-0.4	0.3	1.2
DC	2.6	2.6	2.4	1.7	1.9	4.4
Houston	2.6	-1.4	2.9	4.3*	4.6*	2.7
Los Angeles	5.8*	9.2*	5.5	5.1	5.0*	4.1
New York City	2.0	4.5	1.7	1.2	1.3	1.4
San Diego	-0.5	-6.0	-0.6	0.6	1.1	2.4
National Public	0.5	1.0	0.5	0.3	0.3	0.2
Large City	1.9*	1.5	2.1	2.0*	2.1*	2.0
Changes 2005 to 2007						
Atlanta	2.7	-1.1	4.7	5.1*	3.8	1.1
Austin	2.0	-0.9	3.8	3.9	2.8	0.3
Boston	1.5	0.3	2.8	3.2	1.2	-0.2
Charlotte	1.0	0.6	0.5	0.3	1.8	2.0
Chicago	-0.6	-3.2	-0.8	0.4	-0.2	0.8
Cleveland	2.2	-2.2	4.1	4.4	3.6	1.2
DC	-2.9	-3.9	-2.7	-2.0	-2.0	-4.1
Houston	1.8	0.1	3.5	2.6	1.5	1.1
Los Angeles	1.8	-0.9	3.3	3.9	2.4	0.3
New York City	-3.4	-6.4	-3.7	-2.4	-2.2	-2.4
San Diego	-4.0	-6.3	-3.1	-3.5	-3.3	-3.7
National Public	0.3	-1.3	1.3*	1.3*	0.6	-0.4
Large City	-1.2	-3.7*	-0.5	-0.2	-0.7	-1.1
Changes 2003 to 2007						
Atlanta	1.9	-2.4	2.6	3.7	3.2	2.6
Austin	—	†	†	†	†	†
Boston	2.2	0.1	3.7	3.6	2.3	1.3
Charlotte	-0.4	-6.6	-1.3	1.2	1.6	2.8
Chicago	-0.9	-8.4*	-1.0	1.5	1.6	1.8
Cleveland	0.8	-9.1	2.8	4.1	3.9	2.4
DC	-0.3	-1.3	-0.3	-0.2	-0.2	0.2
Houston	4.4*	-1.4	6.4*	6.9*	6.1*	3.8*
Los Angeles	7.6*	8.3*	8.8*	9.0*	7.5*	4.5
New York City	-1.4	-1.9	-2.0	-1.1	-0.9	-0.9
San Diego	-4.5	-12.4*	-3.8	-2.9	-2.2	-1.3
National Public	0.8	-0.3	1.8*	1.6*	0.9	-0.2
Large City	0.7	-2.2	1.7	1.8	1.4	0.9

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.30 Changes in the average scale score of grade 4 NSLP-eligible public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	2.9*	3.3	3.2	2.9	2.9	2.4
Austin	—	†	†	†	†	†
Boston	8.5*	5.7*	8.4*	9.5*	9.9*	9.1*
Charlotte	0.5	-0.6	0.9	1.4	0.9	-0.3
Chicago	0.6	-1.8	-0.8	0.4	1.9	3.2
Cleveland	5.5*	4.8	5.9	6.6*	5.9*	4.1
DC	5.9*	6.0*	6.0*	5.7*	5.8*	5.7*
Houston	4.0*	1.4	4.2*	5.0*	5.0*	4.3*
Los Angeles	3.6*	-0.8	1.8	3.8*	6.3*	7.0*
New York City	4.9*	4.4*	5.4*	5.3*	4.9*	4.4
San Diego	6.1*	-0.4	4.6*	6.7*	8.6*	11.0*
National Public	3.6*	2.4*	3.6*	4.0*	4.0*	3.8*
Large City	3.5*	1.5	3.4*	4.1*	4.2*	4.0*
Changes 2005 to 2007						
Atlanta	2.5	-0.2	2.0	2.6	2.9	5.2*
Austin	-0.3	-0.7	-0.4	-0.2	0.1	-0.4
Boston	2.8	-1.2	3.0	3.5*	4.1*	4.8*
Charlotte	0.7	-3.7	0.7	1.9	2.4	2.3
Chicago	3.6*	0.5	4.5*	4.8*	4.7*	3.7
Cleveland	-8.7*	-16.8*	-9.2*	-7.4*	-6.1*	-4.0
DC	1.5	-3.6	-0.1	2.3	3.5*	5.1*
Houston	3.2*	1.2	4.1*	4.0*	3.8*	3.2
Los Angeles	2.0	-0.8	2.4	2.9	2.7	2.9
New York City	6.3*	3.7	5.8*	6.5*	7.4*	7.9*
San Diego	-1.0	-9.4*	-1.9	1.2	2.3	2.6
National Public	1.8*	-0.3	2.4*	2.5*	2.3*	2.1*
Large City	2.2*	-1.3	2.2*	2.8*	3.2*	3.9*
Changes 2003 to 2007						
Atlanta	5.4*	3.1	5.2*	5.5*	5.8*	7.5*
Austin	—	†	†	†	†	†
Boston	11.3*	4.5	11.4*	13.0*	14.0*	13.9*
Charlotte	1.2	-4.3	1.5	3.3	3.2	2.0
Chicago	4.2*	-1.4	3.7	5.2*	6.6	6.9*
Cleveland	-3.2	-12.0*	-3.3	-0.8	-0.2	0.1
DC	7.3*	2.4	5.9*	8.0*	9.4*	10.9*
Houston	7.2*	2.7	8.2*	9.0*	8.8*	7.4*
Los Angeles	5.6*	-1.6	4.1*	6.7*	8.9*	9.8*
New York City	11.1*	8.1*	11.2*	11.9*	12.3*	12.3*
San Diego	5.0*	-9.8*	2.7	7.8*	10.9*	13.6*
National Public	5.4*	2.1*	6.0*	6.5*	6.3*	6.0*
Large City	5.6*	0.2	5.5*	6.9*	7.4*	7.9*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.31 Changes in the average scale score of grade 8 NSLP-eligible public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	1.8	-0.5	0.7	1.4	2.6	4.9*
Austin	—	†	†	†	†	†
Boston	4.8*	-4.2	4.0	7.3*	8.2*	8.5*
Charlotte	4.2	1.2	5.3	5.7	5.4*	3.4
Chicago	2.5	3.8	3.1	2.0	1.7	2.1
Cleveland	-3.5	-9.0*	-4.0*	-2.3	-2.2	0.0
DC	6.1*	7.0*	6.9*	5.6*	5.1*	6.1*
Houston	4.2*	-0.1	2.5	4.6*	6.4*	7.4*
Los Angeles	4.2*	3.3	5.0	4.0	4.0	4.6*
New York City	3.9	4.0	2.3	2.7	3.5	7.0
San Diego	4.2	-3.8	5.4	7.9*	6.8*	4.9
National Public	2.8*	2.2*	3.0*	2.9*	2.8*	3.0*
Large City	3.0*	2.7*	2.5*	2.5*	3.1*	4.4*
Changes 2005 to 2007						
Atlanta	9.9	11.0*	11.6*	9.4*	7.6*	9.9*
Austin	7.8*	7.2	10.4*	9.7*	6.8*	4.9
Boston	8.0*	13.0*	9.1*	6.9*	6.1*	4.8
Charlotte	4.9*	9.5*	5.5*	3.0	3.6	3.0
Chicago	1.5	-2.5	1.2	2.7	2.6	3.5
Cleveland	2.9	-3.1	4.2	5.2*	6.0*	2.1
DC	-1.0	-5.5	-1.6	0.1	1.1	1.0
Houston	5.2*	3.8	5.2*	5.6*	5.6*	5.6*
Los Angeles	9.0*	8.6*	8.7*	8.7*	8.7*	10.3*
New York City	2.9	3.5	2.6	2.7	2.7	3.2
San Diego	2.5	3.8	1.7	0.6	1.3	4.9
National Public	3.1*	2.8*	3.4*	3.2*	3.2*	3.1*
Large City	4.5*	3.4*	4.7*	4.6*	5.0*	4.9*
Changes 2003 to 2007						
Atlanta	11.7*	10.5*	12.3*	10.8*	10.1*	14.8*
Austin	—	†	†	†	†	†
Boston	12.7*	8.8	13.1*	14.1	14.3*	13.2*
Charlotte	9.1*	10.7*	10.8*	8.7*	8.9*	6.4*
Chicago	4.0	1.2	4.3	4.7*	4.3	5.6
Cleveland	-0.6	-12.1	0.2	2.9	3.7	2.0
DC	5.2*	1.6	5.3*	5.7*	6.1*	7.1*
Houston	9.3*	3.7	7.7*	10.3*	12.0*	13.0*
Los Angeles	13.2*	12.0*	13.7*	12.7*	12.7*	14.9*
New York City	6.8*	7.5*	4.9*	5.3*	6.2*	10.2*
San Diego	6.7	0.1	7.1*	8.5*	8.1*	9.8
National Public	5.9*	5.1*	6.4*	6.0*	6.1*	6.1*
Large City	7.5*	6.0*	7.1*	7.2*	8.1*	9.2*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.32 Changes in the average scale score of grade 4 limited English proficient (LEP) public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-3.2	†	†	†	†	†
Charlotte	0.3	†	†	†	†	†
Chicago	-1.1	4.4	2.7	1.5	-3.1	-10.8*
Cleveland	†	†	†	†	†	†
DC	1.6	†	†	†	†	†
Houston	6.9*	10.2*	7.7*	6.8*	5.1	4.8
Los Angeles	-0.3	2.2	-0.4	-1.3	-1.8	-0.4
New York City	-1.8	†	†	†	†	†
San Diego	1.7	1.8	3.2	2.4	2.1	-0.9
National Public	0.2	2.1	1.1	0.2	-0.6	-1.9
Large City	-0.3	1.6	0.2	-0.2	-1.1	-2.0
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	1.4	-10.7	-1.0	4.2	7.0	7.3
Boston	8.0	†	†	†	†	†
Charlotte	-1.2	†	†	†	†	†
Chicago	4.6	-8.1	1.2	4.6	10.2*	14.9*
Cleveland	†	†	†	†	†	†
DC	10.0	†	†	†	†	†
Houston	-5.6*	-12.3*	-6.8*	-4.5	-1.8	-2.4
Los Angeles	-4.8*	-10.7*	-2.8	-2.4	-2.6	-5.3
New York City	-4.0	†	†	†	†	†
San Diego	1.2	-10.3	1.1	4.2	5.7	5.1
National Public	-0.5	-9.5*	-0.3	2.0*	2.7*	2.6*
Large City	-1.6	-11.6*	-1.4	0.8	2.2	2.3
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	4.8	†	†	†	†	†
Charlotte	-0.9	†	†	†	†	†
Chicago	3.5	-3.7	3.9	6.1	7.1	4.1
Cleveland	-11.2	†	†	†	†	†
DC	11.6	†	†	†	†	†
Houston	1.4	-2.1	0.9	2.3	3.3	2.4
Los Angeles	-5.1*	-8.4*	-3.2	-3.7	-4.4	-5.7*
New York City	-5.8	†	†	†	†	†
San Diego	2.9	-8.5	4.2	6.6*	7.8*	4.3
National Public	-0.3	-7.5*	0.9	2.2*	2.1	0.7
Large City	-1.9	-10.0*	-1.2	0.5	1.1	0.3

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.33 Changes in the average scale score of grade 8 LEP public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-3.4	†	†	†	†	†
Charlotte	6.5	†	†	†	†	†
Chicago	-1.9	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	3.3	†	†	†	†	†
Houston	3.0	†	†	†	†	†
Los Angeles	8.1*	11.2*	9.5*	8.0*	6.9*	5.1
New York City	3.6	†	†	†	†	†
San Diego	-5.8	†	†	†	†	†
National Public	0.6	2.4	1.9	1.0	-0.4	-1.7
Large City	5.0*	5.9*	5.9*	5.8*	4.8*	2.8
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	-5.3	†	†	†	†	†
Boston	0.1	†	†	†	†	†
Charlotte	-11.7	†	†	†	†	†
Chicago	-4.6	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	-14.7	†	†	†	†	†
Houston	-9.6*	†	†	†	†	†
Los Angeles	-2.3	-4.8	-2.7	-1.7	-1.1	-1.2
New York City	-4.6	†	†	†	†	†
San Diego	-5.7	†	†	†	†	†
National Public	-1.7	-6.2*	-2.6	-1.3	0.4	1.1
Large City	-6.2*	-10.7*	-6.7*	-4.9*	-4.5*	-4.2
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-3.4	†	†	†	†	†
Charlotte	-5.2	†	†	†	†	†
Chicago	-6.5	†	†	†	†	†
Cleveland	-0.8	†	†	†	†	†
DC	-11.4	†	†	†	†	†
Houston	-6.6	†	†	†	†	†
Los Angeles	5.9*	6.3	6.8	6.4*	5.8	4.0
New York City	-0.9	†	†	†	†	†
San Diego	-11.4*	†	†	†	†	†
National Public	-1.1	-3.8	-0.8	-0.2	0.0	-0.6
Large City	-1.1	-4.7	-0.8	1.0	0.3	-1.4

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.34 Changes in the average scale score of grade 4 LEP public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	8.7	†	†	†	†	†
Charlotte	2.1	†	†	†	†	†
Chicago	-3.9	-4.2	-3.5	-3.8	-3.9	-3.9
Cleveland	†	†	†	†	†	†
DC	4.3	†	†	†	†	†
Houston	6.0*	3.4	6.5*	6.9*	6.9*	6.4*
Los Angeles	2.4	-1.9	1.1	2.7	4.4*	5.4*
New York City	6.3	†	†	†	†	†
San Diego	4.7*	0.1	4.2	5.3*	5.9*	7.9*
National Public	2.5*	1.1	2.3*	2.9*	3.2*	3.1*
Large City	2.7*	0.3	1.9	3.1*	3.9*	4.6*
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	3.0	-0.3	3.3	4.1	4.2	3.9
Boston	8.4*	†	†	†	†	†
Charlotte	-1.2	†	†	†	†	†
Chicago	4.6	-7.2	3.4	7.9*	9.4*	9.3*
Cleveland	†	†	†	†	†	†
DC	0.6	†	†	†	†	†
Houston	2.7	1.6	3.7	3.1	2.8	2.1
Los Angeles	-1.5	-5.2*	-1.1	-0.6	-0.4	-0.1
New York City	5.1	†	†	†	†	†
San Diego	0.6	-9.8*	-1.2	3.2	5.2*	5.6
National Public	0.9	-3.8*	0.8	2.0*	2.6*	2.6*
Large City	0.6	-5.3*	0.5	1.9	2.7*	3.4*
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	17.1*	†	†	†	†	†
Charlotte	0.9	†	†	†	†	†
Chicago	0.7	-11.4*	-0.1	4.2	5.5	5.4
Cleveland	-3.3	†	†	†	†	†
DC	4.9	†	†	†	†	†
Houston	8.7*	5.0	10.2*	10.0*	9.7*	8.5*
Los Angeles	0.9	-7.0*	0.1	2.2	4.0*	5.3*
New York City	11.3*	†	†	†	†	†
San Diego	5.3*	-9.7*	3.0	8.6*	11.1*	13.6*
National Public	3.4*	-2.7*	3.1*	5.0*	5.9*	5.8*
Large City	3.4*	-5.0*	2.4	5.0*	6.6*	8.0*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.35 Changes in the average scale score of grade 8 LEP public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-7.2	†	†	†	†	†
Charlotte	-6.0	†	†	†	†	†
Chicago	1.8	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	2.2	†	†	†	†	†
Houston	5.2	†	†	†	†	†
Los Angeles	1.5	2.5	2.4	3.1	1.9	-2.4
New York City	-6.7	†	†	†	†	†
San Diego	-1.7	†	†	†	†	†
National Public	1.3	-0.8	0.8	1.5	1.8	2.9
Large City	-0.6	-0.8	0.0	-0.6	-0.7	-0.6
Changes 2005 to 2007						
Atlanta	†	†	†	†	†	†
Austin	5.0	†	†	†	†	†
Boston	15.5	†	†	†	†	†
Charlotte	1.0	†	†	†	†	†
Chicago	4.7	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	-9.5	†	†	†	†	†
Houston	-2.8	†	†	†	†	†
Los Angeles	5.3*	3.6	4.9	4.9	5.5	7.5*
New York City	5.0	†	†	†	†	†
San Diego	1.9	†	†	†	†	†
National Public	1.4	-0.2	1.8	1.9	2.0	1.7
Large City	0.6	-2.0	1.3	1.5	1.1	1.2
Changes 2003 to 2007						
Atlanta	†	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	8.3	†	†	†	†	†
Charlotte	-5.1	†	†	†	†	†
Chicago	6.5	†	†	†	†	†
Cleveland	†	†	†	†	†	†
DC	-7.3	†	†	†	†	†
Houston	2.4	†	†	†	†	†
Los Angeles	6.8*	6.1	7.3*	8.0*	7.4*	5.1
New York City	-1.7	†	†	†	†	†
San Diego	0.2	†	†	†	†	†
National Public	2.7*	-1.0	2.6	3.4*	3.8*	4.6*
Large City	0.1	-2.8	1.3	0.9	0.4	0.5

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.36 Changes in the average scale score of grade 4 Individualized Education Program (IEP) public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	-9.5	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-1.7	1.3	0.7	†	-3.2	†
Charlotte	0.5	†	†	†	†	†
Chicago	8.0	3.3	†	†	†	†
Cleveland	15.6	†	†	†	†	†
DC	3.9	3.3	3.3	4.8	4.1	4.1
Houston	0.2	†	†	†	†	†
Los Angeles	-1.6	†	†	†	†	†
New York City	1.2	†	†	†	†	†
San Diego	-0.7	†	†	†	†	†
National Public	3.0	3.6	3.2	2.5	2.8	2.8
Large City	3.7	4.8	4.4	3.2	2.8	3.6
Changes 2005 to 2007						
Atlanta	5.1	†	†	†	†	†
Austin	-2.5	†	†	†	†	†
Boston	1.0	†	†	†	3.4	9.1
Charlotte	-9.2	†	†	†	†	†
Chicago	-2.6	†	†	†	†	†
Cleveland	-29.5	†	†	†	†	†
DC	-2.7	-1.9	-0.8	-1.4	-3.1	-6.1
Houston	-17.2*	†	†	†	†	†
Los Angeles	0.0	†	†	†	†	†
New York City	-3.5	†	†	†	†	†
San Diego	-12.6*	†	†	†	†	†
National Public	-1.4*	-9.3*	-2.8*	0.2	1.9*	3.2*
Large City	-5.1*	-16.3*	-7.3*	-3.6	-0.8	2.5
Changes 2003 to 2007						
Atlanta	-4.4	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-0.7	†	†	†	0.3	†
Charlotte	-8.8	†	†	†	†	†
Chicago	5.4	†	†	†	†	†
Cleveland	-13.9	†	†	†	†	†
DC	1.3	1.5	2.4	3.4	1.0	-2.0
Houston	-17.0*	-28.4*	-21.1*	-17.9*	†	-3.8
Los Angeles	-1.6	†	†	†	†	†
New York City	-2.3	-9.2	-3.8	-2.0	0.7	2.8
San Diego	-13.3*	†	†	†	†	†
National Public	1.6*	-5.7*	0.4	2.7*	4.7*	6.0*
Large City	-1.4	-11.5*	-2.9	-0.4	2.0	6.1*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.37 Changes in the average scale score of grade 8 IEP public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	-2.5	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-1.9	†	†	†	†	†
Charlotte	-13.7*	†	†	†	†	†
Chicago	-6.3	-9.9	-9.7*	-5.6	-3.7	-2.7
Cleveland	-9.9	†	†	†	†	†
DC	0.9	0.8	2.4	1.9	1.6	-2.2
Houston	-11.4*	†	†	†	†	†
Los Angeles	7.0	†	†	†	†	†
New York City	1.4	†	†	†	†	†
San Diego	1.5	†	†	†	†	†
National Public	-0.3	-0.7	-1.0	-0.8	0.0	1.0
Large City	-0.2	-1.5	0.0	0.2	1.0	-0.5
Changes 2005 to 2007						
Atlanta	1.8	†	†	†	†	†
Austin	6.8	†	†	†	†	†
Boston	4.1	†	†	†	†	†
Charlotte	10.8	†	†	†	†	†
Chicago	-0.6	-6.9	-0.4	0.1	0.2	3.8
Cleveland	-5.9	†	†	†	†	†
DC	-0.5	-4.3	-3.2	-1.5	0.8	5.9
Houston	-2.5	†	†	†	†	†
Los Angeles	-4.5	†	†	†	†	†
New York City	3.2	†	†	†	†	†
San Diego	-0.8	†	†	†	†	†
National Public	-1.6*	-6.1*	-2.3*	-1.0	0.1	1.2
Large City	-1.4	-6.7*	-2.2	-1.0	0.0	2.8
Changes 2003 to 2007						
Atlanta	-0.7	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	2.2	†	†	†	†	†
Charlotte	-2.9	†	†	†	†	†
Chicago	-6.9	-16.7*	-10.1*	-5.4	-3.6	1.1
Cleveland	-15.8*	†	†	†	†	†
DC	0.4	-3.5	-0.9	0.5	2.4	3.7
Houston	-13.9*	-19.6*	-16.5*	†	†	†
Los Angeles	2.6	†	†	†	†	†
New York City	4.6	†	†	†	†	†
San Diego	0.8	†	†	†	†	†
National Public	-1.9*	-6.8*	-3.2*	-1.8*	0.1	2.2*
Large City	-1.6	-8.2*	-2.2	-0.9	1.0	2.3

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.38 Changes in the average scale score of grade 4 IEP public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	-2.5	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	8.0*	†	†	†	†	†
Charlotte	1.9	†	†	†	†	†
Chicago	-0.1	†	†	†	†	†
Cleveland	6.3	†	†	†	†	†
DC	9.1*	6.8	7.5*	9.2*	9.7*	12.4*
Houston	-4.2	†	†	†	†	†
Los Angeles	-0.3	†	†	†	†	†
New York City	4.7	†	†	†	†	†
San Diego	1.2	†	†	†	†	†
National Public	3.6*	1.9*	3.4*	4.1*	4.4*	4.5*
Large City	4.4*	1.5	3.2	4.7*	5.5*	7.3*
Changes 2005 to 2007						
Atlanta	6.1	†	†	†	†	†
Austin	1.8	†	†	†	†	†
Boston	0.9	†	†	†	†	†
Charlotte	-9.0	†	†	†	†	†
Chicago	-1.3	†	†	†	†	†
Cleveland	-19.5*	†	†	†	†	†
DC	-0.1	†	†	†	†	†
Houston	-4.9	†	†	†	†	†
Los Angeles	-0.2	†	†	†	†	†
New York City	5.9*	†	†	†	†	†
San Diego	-14.7*	†	†	†	†	†
National Public	0.8	-5.6*	0.1	2.3*	3.5*	3.5*
Large City	-1.9	-8.6*	-3.9	-1.1	1.6	2.3
Changes 2003 to 2007						
Atlanta	3.6	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	8.9*	†	9.1	11.6*	13.3*	†
Charlotte	-7.2	†	†	†	†	†
Chicago	-1.4	-10.2*	-3.9	-2.0	0.3	8.9
Cleveland	-13.2*	†	†	†	†	†
DC	9.0*	†	†	†	†	†
Houston	-9.1*	†	†	†	†	†
Los Angeles	-0.5	†	†	†	†	†
New York City	10.5*	†	†	†	†	14.3*
San Diego	-13.5*	†	†	†	†	†
National Public	4.4*	-3.7*	3.5*	6.4*	8.0*	7.9*
Large City	2.5	-7.1*	-0.6	3.6*	7.1*	9.6*

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.39 Changes in the average scale score of grade 8 IEP public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by TUDA district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	-6.1	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	-0.7	†	†	†	†	†
Charlotte	-12.3*	†	†	†	†	†
Chicago	4.6	0.0	3.4	6.0	6.8	6.6
Cleveland	-4.7	†	†	†	†	†
DC	5.7	6.0	4.4	5.2	6.5	6.2
Houston	-8.7*	†	†	†	†	†
Los Angeles	-3.5	†	†	†	†	†
New York City	5.7	†	†	†	†	†
San Diego	0.1	†	†	†	†	†
National Public	0.5	-1.2	-0.2	0.6	1.5	2.1
Large City	-0.2	0.1	0.2	0.0	-0.5	-0.8
Changes 2005 to 2007						
Atlanta	13.9*	†	†	†	†	†
Austin	8.3	†	†	†	†	†
Boston	14.2*	†	†	†	†	†
Charlotte	15.2*	†	†	†	†	†
Chicago	1.0	-2.4	-1.2	-0.6	2.6	†
Cleveland	-8.1	†	†	†	†	†
DC	-1.4	-6.4	-2.5	-0.6	1.1	1.6
Houston	3.8	†	†	†	†	†
Los Angeles	6.0	†	†	†	†	†
New York City	4.8	†	†	†	†	†
San Diego	-2.2	†	†	†	†	†
National Public	0.3	-2.9	0.3	1.1	1.1	2.1
Large City	1.2	-3.2	1.2	2.4	2.7	2.9
Changes 2003 to 2007						
Atlanta	7.8	†	†	†	†	†
Austin	—	†	†	†	†	†
Boston	13.5*	†	†	†	†	†
Charlotte	2.8	†	†	†	†	†
Chicago	5.6	-2.3	2.3	5.4	9.4	†
Cleveland	-12.8	†	†	†	†	†
DC	4.3	-0.5	1.9	4.6	7.6	7.8
Houston	-4.9	†	†	†	†	†
Los Angeles	2.6	†	†	†	†	†
New York City	10.5*	†	†	†	†	†
San Diego	-2.1	†	†	†	†	†
National Public	0.9	-4.1*	0.1	1.7	2.6*	4.2*
Large City	1.0	-3.1	1.4	2.5	2.2	2.2

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met

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

Table B.40 Percentile of grade 4 NAEP reading subscale adjusted averages for TUDA districts corresponding to the subscale score distribution of the national public school sample, 2007

	Composite	Literary	Information
Atlanta	32.17	33.16	32.04
Austin	36.08	37.53	35.48
Boston	41.17	42.81	40.25
Charlotte	38.29	38.26	39.33
Chicago	28.92	30.44	28.33
Cleveland	24.56	26.02	24.12
DC	24.01	25.49	23.55
Houston	34.19	34.74	34.51
Los Angeles	27.62	28.79	27.56
New York City	35.82	37.53	34.91
San Diego	34.55	36.13	33.72

Note: In order to reveal the strengths of each TUDA across different subscales, we compute the percentile to which each adjusted TUDA subscale mean corresponds on the subscale score distribution of the national public school sample. Note that the NAEP subscales are not all reported on the same metric; hence the subscale means are not directly comparable. Instead, our analyses allow indirect, normative comparisons between subscales (within a district) by looking at the percentile to which a given district's adjusted subscale mean corresponds in the national public school samples.

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

Table B.41 Percentile of grade 8 NAEP reading subscale adjusted averages for TUDA districts corresponding to the subscale score distribution of the national public school sample, 2007

	Composite	Literary	Information	Task
Atlanta	31.81	32.47	31.97	35.81
Austin	35.96	38.56	34.91	37.86
Boston	37.96	37.64	40.32	38.54
Charlotte	35.61	39.20	34.03	36.54
Chicago	36.69	39.56	36.87	35.67
Cleveland	34.42	37.91	34.49	32.87
DC	27.59	31.08	27.16	28.19
Houston	36.22	38.40	35.88	37.47
Los Angeles	29.20	31.51	30.78	28.04
New York City	30.36	30.31	33.55	30.70
San Diego	29.62	30.80	31.16	30.49

Note: In order to reveal the strengths of each TUDA across different subscales, we compute the percentile to which each adjusted TUDA subscale mean corresponds on the subscale score distribution of the national public school sample. Note that the NAEP subscales are not all reported on the same metric; hence the subscale means are not directly comparable. Instead, our analyses allow indirect, normative comparisons between subscales (within a district) by looking at the percentile to which a given district's adjusted subscale mean corresponds in the national public school samples.

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

Table B.42 Percentile of grade 4 NAEP mathematics subscale adjusted averages for TUDA districts corresponding to the subscale score distribution of the national public school sample, 2007

	Composite	Numbers	Algebra	Geometry	Data	Measurement
Atlanta	31.09	33.89	31.76	35.23	31.23	26.53
Austin	43.72	44.82	45.33	44.75	38.13	43.76
Boston	44.92	48.27	42.40	45.46	39.57	43.53
Charlotte	44.08	44.62	49.42	51.08	41.20	38.80
Chicago	26.04	26.45	27.83	26.59	29.87	25.89
Cleveland	21.47	22.41	22.60	23.68	23.87	21.29
DC	21.27	23.72	22.11	24.12	20.06	19.75
Houston	45.35	44.84	48.04	44.09	43.05	47.29
Los Angeles	27.69	31.55	30.72	28.03	26.45	23.08
New York City	39.69	42.08	40.07	38.17	35.93	38.77
San Diego	34.26	34.24	36.96	41.94	31.25	35.52

Note: In order to reveal the strengths of each TUDA across different subscales, we compute the percentile to which each adjusted TUDA subscale mean corresponds on the subscale score distribution of the national public school sample. Note that the NAEP subscales are not all reported on the same metric; hence the subscale means are not directly comparable. Instead, our analyses allow indirect, normative comparisons between subscales (within a district) by looking at the percentile to which a given district's adjusted subscale mean corresponds in the national public school samples.

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

Table B.43 Percentile of grade 8 NAEP mathematics subscale adjusted averges for TUDA districts corresponding to the subscale score distribution of the national public school sample, 2007

	Composite	Numbers	Algebra	Geometry	Data	Measurement
Atlanta	31.63	31.28	32.96	30.71	33.46	31.96
Austin	45.58	44.57	41.30	50.08	44.63	49.52
Boston	42.92	41.39	44.21	44.14	41.90	42.51
Charlotte	44.73	39.30	48.24	48.08	43.11	42.96
Chicago	31.51	32.32	31.23	33.65	31.89	31.51
Cleveland	31.34	27.84	33.82	35.01	30.60	31.51
DC	23.42	24.94	26.77	21.30	26.62	21.47
Houston	41.74	42.02	39.22	46.55	39.31	43.50
Los Angeles	27.82	28.32	31.43	31.66	25.15	24.34
New York City	35.03	32.34	38.44	36.26	32.81	35.62
San Diego	32.48	31.04	38.92	31.62	28.19	32.27

Note: In order to reveal the strengths of each TUDA across different subscales, we compute the percentile to which each adjusted TUDA subscale mean corresponds on the subscale score distribution of the national public school sample. Note that the NAEP subscales are all not reported on the same metric; hence the subscale means are not directly comparable. Instead, our analyses allow indirect, normative comparisons between subscales (within a district) by looking at the percentile to which a given district's adjusted subscale mean corresponds in the national public school samples.

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

Table B.44 Percentile of grade 4 NAEP science subscale adjusted averages for TUDA districts corresponding to the subscale score distribution of the national public school sample, 2005

	Composite	Physical Science	Earth Science	Life Science
Atlanta	30.78	30.67	33.92	29.63
Austin	38.60	37.26	40.32	39.86
Boston	30.72	30.86	30.75	32.60
Charlotte	30.56	28.58	32.00	33.19
Chicago	24.31	25.30	23.97	26.25
Cleveland	28.49	27.73	28.40	31.25
Houston	36.41	34.93	38.44	37.57
Los Angeles	26.42	26.88	26.23	28.27
New York City	26.78	27.20	27.54	27.62
San Diego	27.15	26.84	25.43	31.55

Note: In order to reveal the strengths of each TUDA across different subscales, we compute the percentile to which each adjusted TUDA subscale mean corresponds on the subscale score distribution of the national public school sample. Note that the NAEP subscales are not all reported on the same metric; hence the subscale means are not directly comparable. Instead, our analyses allow indirect, normative comparisons between subscales (within a district) by looking at the percentile to which a given district's adjusted subscale mean corresponds in the national public school samples.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Science Assessments

Table B.45 Percentile of grade 8 NAEP science subscale adjusted averages for TUDA districts corresponding to the subscale score distribution of the national public school sample, 2005

	Composite	Physical Science	Earth Science	Life Science
Atlanta	25.18	23.98	28.76	25.47
Austin	34.45	35.40	34.45	34.96
Boston	32.65	31.19	34.52	33.70
Charlotte	31.68	30.15	36.23	30.92
Chicago	26.79	26.89	27.59	27.74
Cleveland	29.06	27.52	29.81	31.44
Houston	31.14	31.88	32.38	30.97
Los Angeles	27.73	29.27	29.08	27.25
New York City	25.62	27.50	27.40	24.72
San Diego	28.68	28.11	29.58	30.16

Note: In order to reveal the strengths of each TUDA across different subscales, we compute the percentile to which each adjusted TUDA subscale mean corresponds on the subscale score distribution of the national public school sample. Note that the NAEP subscales are not all reported on the same metric; hence the subscale means are not directly comparable. Instead, our analyses allow indirect, normative comparisons between subscales (within a district) by looking at the percentile to which a given district's adjusted subscale mean corresponds in the national public school samples.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Science Assessments

Table B.46 District funding per pupil and percentage of total expenditures devoted to instruction, 2003-2009

		2002-03			2006-07			2008-09	
	Instructional APPE	Total APPE	Percent of Total	Instructional APPE	Total APPE	Percent of Total	Instructional APPE	Total APPE	Percent of Total
Atlanta	$6,442	$11,435	56.3%	$6,939	$12,745	54.4%	$6,684	$13,516	49.5%
Austin	4,420	7,580	58.3	4,691	8,182	57.3	5,156	9,035	57.1
Baltimore City	6,036	9,639	62.6	7,274	12,440	58.5	8,355	14,201	58.8
Boston	7,837	13,730	57.1	11,129	19,435	57.3	11,737	20,324	57.8
Charlotte	4,441	7,188	61.8	4,991	8,081	61.8	5,045	8,115	62.2
Chicago	4,937	7,967	62.0	5,774	9,666	59.7	6,207	10,392	59.7
Cleveland	5,782	10,199	56.7	6,812	11,383	59.8	7,416	12,393	59.8
Detroit	5,089	9,063	56.2	6,503	11,896	54.7	6,522	12,016	54.3
District of Columbia	6,976	13,328	52.3	6,226	14,324	43.5	6,542	14,594	44.8
Fresno	4,651	7,769	59.9	5,237	8,995	58.2	5,990	10,053	59.6
Houston	4,277	7,236	59.1	4,732	7,994	59.2	5,048	8,604	58.7
Jefferson County	4,218	7,663	55.0	5,206	9,698	53.7	5,350	9,966	53.7
Los Angeles	4,892	8,447	57.9	6,256	10,364	60.4	6,666	11,357	58.7
Miami-Dade County	4,246	6,956	61.0	5,694	9,371	60.8	6,057	9,933	61.0
Milwaukee	6,156	10,352	59.5	6,990	11,725	59.6	7,242	12,705	57.0
New York City	8,960	11,920	75.2	12,494	16,443	76.0	--	17,923	--
Philadelphia	4,333	7,554	57.4	4,716	8,985	52.5	5,051	9,399	53.7
San Diego	4,973	8,482	58.6	5,441	9,682	56.2	5,767	10,305	56.0

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

Table B.47 Percentage of district staffing levels that are teachers and student/teacher ratios, 2003-2009

		2002-03			2006-07		2008-09	
	% of Staff that are Teachers	Pupil/Teacher Ratio		% of Staff that are Teachers	Pupil/Teacher Ratio		% of Staff that are Teachers	Pupil/Teacher Ratio
Atlanta	52.2%	14.2		53.5%	13.7		54.0%	13.0
Austin	50.1	14.6		52.7	14.4		52.0	14.2
Baltimore City	57.5	14.7		51.4	14.3		50.7	14.1
Boston	46.5	13.6		60.8	13.2		56.3	12.8
Charlotte	51.2	15.1		53.2	13.7		50.5	14.5
Chicago	85.0	17.7		77.9	21.8		84.4	19.6
Cleveland	50.4	10.7		43.3	15.8		44.8	13.9
Detroit	31.5	30.6		43.8	16.5		43.0	16.4
District of Columbia	43.3	13.5		--	--		41.8	12.5
Fresno	52.7	20.6		56.8	19.9		53.6	19.5
Houston	44.4	17.1		49.7	16.8		49.0	16.7
Jefferson County	39.7	17.9		46.1	15.5		43.4	16.1
Los Angeles	47.7	21.0		48.6	20.6		47.1	19.6
Miami-Dade County	51.1	20.0		54.6	17.1		57.5	15.4
Milwaukee	45.4	15.0		50.7	17.6		47.5	16.6
New York City	51.1	16.4		84.1	14.1		--	--
Philadelphia	41.3	19.5		78.7	18.0		49.2	15.6
San Diego	51.5	18.8		53.6	18.4		51.6	19.3

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

APPENDIX C

NAEP ANALYSIS METHODOLOGY

C.1 Background of the Full Population Estimates

Since the late 1990s, the rates at which sampled students with disabilities and English language learners participate in NAEP have fluctuated across time and across jurisdictions. Reporting of trends requires consistency in practices across years, and the lack of consistency in the inclusion of students with disabilities has called the validity of NAEP trends into question (Forgione, 1999; McLaughlin, 2000, 2001, 2003).

In early 2001, to support an internal evaluation of the impact of changing exclusion rates on reports of statistically significant gains across states, the National Center for Education Statistics (NCES) sponsored research on imputation procedures of NAEP scores for the excluded students and provided adjusted or full population estimates (FPEs) for the 1996 to 2000 NAEP mathematics gains. The same method was subsequently used to produce FPEs for grades 4 and 8 in reading, writing, mathematics, and science for each year these assessments were administered since 2000 (McLaughlin, 2005).

In 2004, the FPE methodology was tested for sensitivity to violation of assumptions (Wise, Hoffman, & Becker, 2004). Overall, under the assumptions of the model, the FPEs were unbiased. Violations of these assumptions led to slightly biased estimates which, at the jurisdiction level, were considered negligible.

The basis of the methodology used to produce the FPEs that were used in the analyses described in this report is one of many statistical scenarios. More recently, for example, Braun, Zhang, and Vezzu (2006) introduced an alternative approach to address the exclusion problem. Their approach is also an imputation procedure based on the same basic assumptions used in McLaughlin (2005). When both approaches were compared, their performances were found to be equivalent (Wise, Le, Hoffman, & Becker, 2006).

In 2009, the National Institute of Statistical Sciences and the NAEP-Education Statistics Services Institute (ESSI) task force on FPEs found that methods used to calculate FPEs were sufficiently sound that there was no identified need for drastic modifications, although it also noted that NCES should support studies to extend and further validate the methodology for imputing plausible values.

The task force recommended that NCES publish the adjusted estimates, which were thought to be "more consistent with the goal of provid[ing] high-quality indicators of performance for well-defined populations of students enrolled." Further, the task force recommended that "NCES set as its goal to report expanded population estimates as the primary (or only) measure of NAEP performance."[1,2]

C.2 Adjusted Analyses

Variables Used in Regression Analyses to Calculate "Adjusted" Scores

- **Race/ethnicity**

  In the NAEP files, student race/ethnicity information is obtained from school records and classified under six categories: White, African American, Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, and unclassifiable. When school-reported information was missing, student-reported data from the student background questionnaire were used to establish race/ethnicity. We categorized as unclassifiable the students whose race/ethnicity based on school

---

[1] Available at http://niss.org/sites/default/files/pdfs/technicalreports/tr172.pdf

[2] Additional information on the inclusion of special needs students on NAEP is available at http://nces.ed.gov/nationsreportcard/about/inclusion.asp #research

records were unclassifiable or missing and (1) who self-reported their race as multicultural but not Hispanic or (2) who did not self-report race information.

- **Students with disabilities or special education (SD) status**
  Student has an Individualized Educational Program (IEP), for reasons other than being gifted or talented; or a student has a Section 504 Plan.

- **English language learner (ELL) status**
  Student is currently classified as an English language learner and is receiving services**.**

- **National School Lunch Program (NSLP) eligibility**
  Eligibility for the National School Lunch Program is determined by a student's family income in relation to the federally established poverty level. Based on available school records, students were classified as either currently eligible for the National School Lunch Program or currently not eligible. If the school record indicated the information was not available, the student was classified as not eligible.

- **Parental education**
  This variable shows the highest level of education attained by either parent: did not complete high school, graduated from high school, some education after high school, graduated from college. This indicator is available only for grade 8 students.

- **Literacy materials**
  The presence of literacy materials in the home is associated with both socioeconomic status and student achievement. The measure reported here is based on questions in both grade 4 and grade 8 student background questionnaires that ask about the availability of computers, newspapers, magazines, and more than 25 books in the home. A summary score has been created to indicate how many of these four types of literacy materials are present.[3]

Information on race/ethnicity and NSLP, ELL, and SD status come from the school and are available for all students. However, data on background characteristics for students who do not participate in NAEP are not available: excluded students do not fill out the background questionnaire. Therefore, data on *literacy materials* and *parent education* are available only for the included population. The calculation of adjusted scores controlling for background characteristics was conducted on the reported sample only.

Estimating adjusted average or mean scores

The method used in calculating the adjusted district averages is discussed below.

Let y_{ijv} be plausible value v of student j in district i, and

X_{ijk} be the demographic characteristic k of student j in district i.

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

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

where μ is the overall average,

α_i is the district i effect, and

[3] This summary score has been used for reporting NAEP background variables for a number of years and has been shown to be associated with students' achievement scores (e.g., NAEP 1996 Mathematics Cross-State Data Compendium).

β_k is the effect of the demographic characteristic *k* of student *j* in district *i*.

Letting the subscript • indicate average, then the average scale score in district *i* is expressed as

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

Subtracting [2] from [1] we can estimate the regression in [3]

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

and obtain estimates of β_k directly, without any contamination from the α_i because α_i has been subtracted out before the regression.

With the estimates $\hat{\beta}_k$, we compute the average effect of the demographic characteristics of student *j* in district *i*:

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

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

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

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

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

C.3 2007 NAEP Results for Selected Groups of Students

Table C.1 Average scale scores of grade 4 public school students in the NAEP reading assessment overall and by selected characteristics, based on the full population estimates, by district, 2007

Districts	Overall		African American Students		Hispanic Students		Eligible for National School Lunch Program		Students with Disabilities		English Language Learners	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Atlanta	204	1.4	196	1.4	196	5.6	194	1.3	168	5.0	—	†
Austin	207	1.9	194	4.3	194	2.0	192	1.9	168	4.8	180	2.1
Boston	206	2.7	202	3.3	199	2.8	203	2.6	176	7.9	191	3.7
Charlotte	220	1.5	204	2.1	201	2.8	202	1.6	179	4.1	190	3.4
Chicago	197	1.5	192	2.1	195	2.1	194	1.5	162	3.7	175	2.3
Cleveland	185	4.4	181	3.6	167	15.6	185	4.4	128	13.9	150	24.5
District of Columbia	190	1.4	185	1.3	192	5.9	180	1.6	144	5.3	182	6.4
Houston	198	1.2	199	2.0	192	1.4	193	1.2	155	3.5	179	1.5
Los Angeles	194	1.3	194	5.2	189	1.6	189	1.6	159	4.2	176	1.8
New York City	211	1.2	205	1.7	201	2.0	207	1.6	178	3.6	180	3.3
San Diego	207	2.0	196	3.5	192	1.9	195	1.8	163	4.4	185	2.0

— Too few cases for a reliable estimate.
† Not applicable.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessments: Full Population Estimates.

Table C.2 Average scale scores of grade 8 public school students in the NAEP reading assessment overall and by selected characteristics, based on the full population estimates, by district, 2007

Districts	Overall		African American Students		Hispanic Students		Eligible for National School Lunch Program		Students with Disabilities		English Language Learners	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Atlanta	241	1.4	238	1.4	—	†	235	1.8	201	5.8	—	†
Austin	252	1.6	234	3.2	239	1.8	235	1.7	216	3.3	204	2.6
Boston	249	2.3	246	2.1	235	3.9	245	2.7	216	5.3	205	13.7
Charlotte	257	1.6	245	1.7	243	5.8	241	2.5	222	4.6	222	6.7
Chicago	246	1.6	237	2.0	250	2.1	243	1.7	207	2.9	206	5.1
Cleveland	236	1.8	233	1.7	227	7.6	236	1.8	188	6.7	202	13.2
District of Columbia	234	1.6	231	1.5	241	4.4	228	1.9	196	7.8	213	8.4
Houston	246	1.3	242	1.8	240	1.4	240	1.3	199	3.3	198	3.2
Los Angeles	238	1.1	227	4.8	234	1.2	235	1.2	195	3.2	209	1.6
New York City	248	2.0	240	2.7	239	3.1	244	2.0	214	2.5	210	8.9
San Diego	248	1.2	238	2.9	233	2.1	234	2.3	209	3.7	207	2.3

— Too few cases for a reliable estimate.
† Not applicable.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessments: Full Population Estimates.

Table C.3 Average scale scores of grade 4 public school students in the NAEP mathematics assessment, overall and by selected characteristics, based on the full population estimates, by district, 2007

Districts	Overall		African American Students		Hispanic Students		Eligible for National School Lunch Program		Students with Disabilities		English Language Learners	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Atlanta	223	1.0	216	1.1	222	4.3	215	1.1	202	4.5	—	†
Austin	238	1.2	223	1.4	230	1.1	227	1.0	216	3.1	224	1.3
Boston	231	1.3	224	1.5	228	1.8	228	1.4	208	2.9	225	2.6
Charlotte	242	1.0	229	1.1	230	2.8	229	1.1	213	3.1	225	3.6
Chicago	218	1.1	212	1.5	217	1.6	214	1.0	192	2.8	203	2.2
Cleveland	210	1.7	205	1.7	206	6.1	210	1.7	177	4.5	195	7.0
District of Columbia	212	0.9	207	0.9	213	2.2	205	0.9	183	2.5	203	2.6
Houston	232	1.2	222	1.9	232	1.3	229	1.2	202	3.3	228	1.6
Los Angeles	220	0.9	216	2.1	216	1.0	216	1.1	195	3.6	208	1.0
New York City	235	1.3	227	1.6	230	1.2	233	1.4	213	2.1	216	2.0
San Diego	232	1.5	220	3.6	221	1.5	222	1.5	193	3.5	216	1.5

SE=standard error
— Too few cases for a reliable estimate.
† Not applicable.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments: Full Population Estimates.

Table C.4 Average scale scores of grade 8 public school students in the NAEP mathematics assessment, overall and by selected characteristics, based on the full population estimates, by district, 2007

Districts	Overall		African American Students		Hispanic Students		Eligible for National School Lunch Program		Students with Disabilities		English Language Learners	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Atlanta	255	1.7	252	1.6	—	†	250	1.7	217	5.6	—	†
Austin	280	1.2	262	2.4	268	1.7	264	1.5	243	3.1	242	3.2
Boston	272	1.6	259	1.9	266	2.6	266	1.7	239	5.0	236	5.7
Charlotte	281	1.1	266	1.4	261	3.1	263	1.3	252	3.3	250	4.4
Chicago	258	1.9	246	2.5	262	1.8	254	1.8	223	4.0	236	4.7
Cleveland	248	1.4	246	1.8	247	3.8	248	1.4	205	8.2	—	†
District of Columbia	243	1.4	241	1.4	246	3.8	237	1.8	204	4.8	221	6.4
Houston	270	1.4	260	1.8	267	1.2	264	1.4	228	5.0	237	3.1
Los Angeles	256	1.2	243	2.6	252	1.4	253	1.3	216	3.6	229	2.1
New York City	269	1.8	258	2.2	262	2.5	266	1.6	234	3.4	235	3.1
San Diego	270	1.5	254	2.9	257	2.1	257	2.6	224	4.4	234	2.5

— Too few cases for a reliable estimate.
† Not applicable.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments: Full Population Estimates.

Table C.5 Average scale scores of grade 4 public school students in the NAEP reading assessment, overall and by selected characteristics, by district, 2007

Districts	Overall		African American Students		Hispanic Students		Eligible for National School Lunch Program		Students with Disabilities		English Language Learners	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Atlanta	207	1.5	200	1.4	—	†	198	1.4	191	5.7	—	†
Austin	218	2.0	201	4.1	206	2.4	203	2.2	190	5.4	194	3.4
Boston	210	1.9	204	2.4	204	1.9	207	1.7	183	2.8	197	3.0
Charlotte	222	1.5	206	2.1	207	2.2	205	1.6	187	5.1	196	2.6
Chicago	201	1.5	193	2.2	201	2.2	197	1.5	172	4.6	182	2.3
Cleveland	198	1.7	192	1.8	200	3.9	198	1.7	—	†	—	†
District of Columbia	197	0.9	192	1.0	206	3.6	188	1.1	162	4.7	198	4.1
Houston	206	1.2	205	1.5	200	1.6	201	1.2	174	3.8	186	2.1
Los Angeles	196	1.3	196	4.9	190	1.6	191	1.7	166	4.2	177	1.8
New York City	213	1.1	206	1.7	203	1.6	209	1.4	181	3.5	181	2.7
San Diego	210	1.8	199	2.8	196	1.9	198	1.6	171	4.8	189	1.9

— Too few cases for a reliable estimate.
† Not applicable.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessments.

Table C.6 Average scale scores of grade 8 public school students in the NAEP reading assessment, overall and by selected characteristics, by district, 2007

Districts	Overall		African American Students		Hispanic Students		Eligible for National School Lunch Program		Students with Disabilities		English Language Learners	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Atlanta	245	1.4	242	1.3	—	†	240	1.6	—	†	—	†
Austin	257	2.0	238	2.9	244	2.2	240	2.0	228	3.5	210	3.5
Boston	254	1.6	250	2.2	241	3.2	249	2.0	223	2.1	210	5.3
Charlotte	260	1.2	246	1.5	251	4.3	245	1.8	228	3.9	228	4.5
Chicago	250	1.5	240	1.8	255	1.5	247	1.5	213	3.2	217	4.5
Cleveland	246	1.5	243	1.7	249	3.7	246	1.5	210	4.6	—	†
District of Columbia	241	0.7	238	0.9	249	3.2	234	1.0	210	4.2	—	†
Houston	252	1.4	249	1.6	246	1.4	247	1.3	217	3.7	209	2.8
Los Angeles	240	1.0	229	4.7	236	1.1	237	1.0	200	3.3	212	1.4
New York City	249	1.9	240	2.7	241	2.8	246	1.8	216	2.9	209	3.9
San Diego	250	1.2	240	3.2	235	2.0	236	2.1	214	4.1	209	2.5

— Too few cases for a reliable estimate.
† Not applicable.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessments.

Table C.7 Average scale scores of grade 4 public school students in the NAEP mathematics assessment, overall and by selected characteristics, by district, 2007

Districts	Overall		African American Students		Hispanic Students		Eligible for National School Lunch Program		Students with Disabilities		English Language Learners	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Atlanta	224	0.9	217	1.1	223	4.4	216	1.0	207	4.5	—	†
Austin	241	1.2	226	1.2	233	1.1	229	0.9	226	2.3	226	1.3
Boston	233	1.1	226	1.4	230	1.6	231	1.2	214	1.9	228	2.0
Charlotte	244	1.1	230	1.3	234	2.6	231	1.3	222	3.4	230	3.4
Chicago	220	1.0	213	1.6	219	1.5	216	1.0	196	2.8	207	2.3
Cleveland	215	1.6	210	1.5	215	5.3	215	1.6	—	†	205	5.8
District of Columbia	214	0.8	209	0.8	220	2.4	207	0.9	188	2.4	209	2.8
Houston	234	1.1	225	1.7	234	1.2	231	1.1	214	3.0	229	1.5
Los Angeles	221	0.9	216	2.3	217	1.0	217	1.1	196	3.2	208	1.0
New York City	236	1.3	227	1.6	230	1.3	234	1.4	213	2.2	216	2.0
San Diego	234	1.4	222	3.4	223	1.5	224	1.6	201	3.9	217	1.7

— Too few cases for a reliable estimate.
† Not applicable.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments.

Table C.8 Average scale scores of grade 8 public school students in the NAEP mathematics assessment, overall and by selected characteristics, by district, 2007

Districts	Overall		African American Students		Hispanic Students		Eligible for National School Lunch Program		Students with Disabilities		English Language Learners	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Atlanta	256	1.5	253	1.4	—	†	251	1.5	—	†	—	†
Austin	283	1.1	265	2.2	271	1.4	267	1.3	252	2.7	245	2.6
Boston	276	1.0	263	1.7	270	2.0	271	1.3	247	3.3	242	4.6
Charlotte	283	1.2	267	1.4	264	3.1	265	1.3	256	3.6	252	3.6
Chicago	260	1.9	248	2.5	265	1.9	257	1.7	228	3.8	240	4.2
Cleveland	257	1.7	253	1.8	258	4.1	257	1.7	222	4.3	—	†
District of Columbia	248	0.9	245	0.9	251	3.0	243	1.2	211	3.1	226	4.3
Houston	273	1.2	265	1.5	270	1.1	268	1.1	240	3.7	241	2.3
Los Angeles	257	1.1	245	2.4	253	1.2	254	1.1	220	2.8	230	1.9
New York City	270	1.8	258	2.2	262	2.5	267	1.6	235	3.1	235	3.3
San Diego	272	1.4	258	3.2	259	2.1	260	2.6	234	4.1	237	2.4

— Too few cases for a reliable estimate.
† Not applicable.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments.

Table C.9 Changes in the average scale score of grade 4 public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	3.4	3.0	3.3	3.5	3.7	3.4
Austin	—	†	†	†	†	†
Boston	0.6	0.2	0.1	0.9	1.0	0.8
Charlotte	1.9	1.5	1.2	1.2	2.0	3.9
Chicago	0.7	2.3	0.9	0.2	0.3	-0.3
Cleveland	3.2	6.7	3.0	2.2	2.5	1.4
DC	1.4	2.1	1.7	1.2	1.6	0.2
Houston	4.2	4.6	4.0	3.4	3.5	5.5
Los Angeles	2.3	2.1	0.3	0.4	2.7	6.2*
New York City	2.5	3.7	4.2	2.9	1.2	0.5
San Diego	-0.3	-0.5	0.5	-0.4	-0.6	-0.8
National Public	0.6	1.9*	0.9*	0.5	0.1	-0.3
Large City	1.3	2.8*	1.8	1.4	0.8	-0.5
Changes 2005 to 2007						
Atlanta	4.5*	4.7	6.9*	5.7*	3.9	1.1
Austin	-0.3	-6.2	-0.1	0.8	1.8	2.3
Boston	2.9	-2.1	3.2	3.7	4.3	5.6
Charlotte	0.5	-3.4	1.4	1.7	1.9	1.0
Chicago	2.3	-0.8	1.9	3.5	3.6	3.3
Cleveland	-6.1	-22.1	-3.7	-1.6	-1.2	-1.8
DC	3.1	0.6	2.6	3.8*	3.8*	4.6*
Houston	-4.3*	-9.4*	-3.6	-1.4	-2.1	-5.2
Los Angeles	0.5	-2.5	3.6	3.9	1.3	-3.9
New York City	0.1	-5.9*	-1.4	1.3	2.6	4.1
San Diego	2.1	-4.9	3.2	4.2	4.4	3.6
National Public	2.1*	0.7	3.6*	2.8*	1.9*	1.4*
Large City	2.1*	-2.3	3.3*	3.6*	3.2*	2.7*
Changes 2003 to 2007						
Atlanta	7.9*	7.7*	10.3*	9.2*	7.6*	4.5
Austin	—	†	†	†	†	†
Boston	3.5	-1.9	3.2	4.6	5.3	6.3
Charlotte	2.5	-1.9	2.6	2.9	3.9	4.9
Chicago	3.0	1.5	2.8	3.7	3.9	3.0
Cleveland	-2.9	-15.4	-0.7	0.6	1.4	-0.4
DC	4.4*	2.7	4.3*	5.0*	5.5*	4.8*
Houston	-0.1	-4.8	0.4	2.0	1.4	0.3
Los Angeles	2.8	-0.3	3.9	4.3	3.9*	2.3
New York City	2.6	-2.2	2.9	4.2*	3.8	4.6
San Diego	1.8	-5.4	3.7	3.8	3.9	2.9
National Public	2.7*	2.5*	4.5*	3.3*	2.0*	1.1*
Large City	3.4*	0.5	5.1*	5.0*	4.0*	2.2

— Austin did not participate in the 2003 NAEP assessment.

† Not applicable.

* Difference is statistically significant at $p < .05$.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005 and 2007 Reading Assessments: Full Population Estimates.

Table C.10 Changes in the average scale score of grade 8 public school students in the NAEP reading assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	0.3	-0.2	-0.7	-0.8	-0.5	3.9
Austin	—	†	†	†	†	†
Boston	1.4	0.3	1.5	1.9	2.1	1.4
Charlotte	-2.5	-8.2*	-2.5	-2.1	-1.2	1.7
Chicago	0.9	-4.3	0.2	2.1	3.0	3.6
Cleveland	-1.4	-6.9	-1.3	-0.4	0.3	1.2
DC	0.0	0.4	-0.7	-1.4	-0.5	1.8
Houston	3.0*	-1.1	3.1	4.4*	4.5*	4.1*
Los Angeles	4.9*	8.9*	5.1*	3.8	3.5	3.4
New York City	-0.1	3.3	0.3	-0.6	-1.3	-2.2
San Diego	0.8	-2.0	0.1	1.0	2.1	2.8
National Public	-1.1*	-1.1	-1.4*	-1.3*	-1.1*	-0.6
Large City	1.9*	1.6	1.9	1.9	1.9*	2.2*
Changes 2005 to 2007						
Atlanta	2.9	0.3	4.7	4.5*	3.8	1.4
Austin	1.7	-2.0	3.2	3.3	2.8	1.0
Boston	0.2	-0.5	1.0	0.6	-0.9	0.6
Charlotte	-0.8	-0.5	-0.9	-0.2	-0.3	-1.9
Chicago	-0.7	-2.8	-0.1	0.3	-0.5	-0.5
Cleveland	2.2	-2.2	4.1	4.4	3.6	1.2
DC	-0.9	-2.9	-0.9	-0.2	0.1	-0.6
Houston	1.6	-0.3	3.1	2.2	1.4	1.5
Los Angeles	0.9	-0.9	2.4	2.6	0.8	-0.2
New York City	-1.8	-5.7	-2.4	-1.1	-0.5	0.6
San Diego	-1.5	-4.6	-1.2	-0.4	0.0	-1.2
National Public	0.2	-0.6	1.4*	0.8*	0.3	-0.6*
Large City	-1.0	-3.1	-0.2	-0.3	-0.6	-0.6
Changes 2003 to 2007						
Atlanta	3.3	0.0	4.1	3.7	3.4	5.3
Austin	—	†	†	†	†	†
Boston	1.6	-0.2	2.5	2.5	1.3	2.0
Charlotte	-3.2	-8.7	-3.4	-2.3	-1.6	-0.2
Chicago	0.2	-7.1*	0.1	2.4	2.5	3.1
Cleveland	0.8	-9.1	2.8	4.1	3.9	2.4
DC	-1.0	-2.4	-1.6	-1.6	-0.4	1.2
Houston	4.6*	-1.4	6.1*	6.6*	5.9*	5.6
Los Angeles	5.9*	7.9*	7.5*	6.4*	4.3*	3.2
New York City	-1.9	-2.4	-2.2	-1.7	-1.8	-1.7
San Diego	-0.7	-6.6*	-1.0	0.6	2.1	1.6
National Public	-0.8*	-1.7*	0.0	-0.4	-0.8*	-1.2*
Large City	0.9	-1.5	1.6	1.6	1.3	1.6

— Austin did not participate in the 2003 NAEP assessment.
† Not applicable.
* Difference is statistically significant at $p < .05$.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005 and 2007 Reading Assessments: Full Population Estimates.

Table C.11 Changes in the average scale score of grade 4 public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	4.8*	4.6*	4.5*	4.9*	5.3*	4.8
Austin	—	†	†	†	†	†
Boston	9.0*	5.9*	8.7*	9.9*	10.5*	10.1*
Charlotte	2.7	0.2	2.6	2.5	3.5	4.9*
Chicago	1.4	-1.8	-0.5	1.1	3.1	5.3
Cleveland	5.3*	4.7	5.7*	6.5*	5.8*	4.0
DC	5.7*	6.1*	5.8*	5.7*	5.9*	4.9*
Houston	5.8*	2.4	5.5*	6.5*	7.1*	7.7*
Los Angeles	4.1*	-0.6	2.1	4.7*	6.6*	7.4*
New York City	4.5*	4.3*	5.3*	5.2*	4.5*	3.3
San Diego	5.5*	0.8	5.2*	7.0*	7.5*	7.2*
National Public	3.1*	2.6*	3.6*	3.5*	3.1*	3.0*
Large City	3.7*	2.0*	3.8*	4.5*	4.2*	3.9*
Changes 2005 to 2007						
Atlanta	2.6*	0.3	2.7	3.1*	3.9*	3.0
Austin	0.5	-0.1	-0.2	0.5	1.6	0.8
Boston	3.1	-0.6	3.7*	4.0*	4.4*	4.0
Charlotte	-1.2	-2.8	-0.6	0.1	-0.9	-2.0
Chicago	3.6	0.9	4.6	4.6*	4.5*	3.4
Cleveland	-8.6*	-16.6*	-9.0*	-7.3*	-6.0*	-4.0
DC	2.4*	-3.1	0.8	3.1*	4.0*	7.3*
Houston	1.3	0.4	2.8	2.1	1.4	-0.3
Los Angeles	1.4	-0.2	2.7	2.6	1.8	0.0
New York City	5.9*	3.8	5.8*	6.3*	6.7*	6.6*
San Diego	1.5	-6.8*	1.3	3.5	4.9*	4.3
National Public	1.9*	0.6	2.4*	2.4*	2.4*	1.8*
Large City	2.0*	-0.8	2.3*	2.9*	3.2*	2.5*
Changes 2003 to 2007						
Atlanta	7.4*	5.0*	7.2*	8.1*	9.1*	7.8*
Austin	—	†	†	†	†	†
Boston	12.1*	5.2	12.4*	14.0*	15.0*	14.0*
Charlotte	1.5	-2.6	2.0	2.5	2.6	2.9
Chicago	5.1*	-0.8	4.1*	5.7*	7.6*	8.7*
Cleveland	-3.2	-12.0*	-3.3	-0.8	-0.2	0.0
DC	8.1*	3.0	6.6*	8.8*	9.8*	12.2*
Houston	7.1*	2.8	8.3*	8.6*	8.4*	7.4*
Los Angeles	5.4*	-0.8	4.8*	7.3*	8.5*	7.4*
New York City	10.4*	8.1*	11.1*	11.5*	11.3*	10.0*
San Diego	7.0*	-6.0*	6.5*	10.5*	12.4*	11.5*
National Public	5.1*	3.2*	6.0*	5.9*	5.5*	4.8*
Large City	5.7*	1.2	6.2*	7.3*	7.4*	6.4*

— Austin did not participate in the 2003 NAEP assessment.

† Not applicable.

* Difference is statistically significant at $p < .05$.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005 and 2007 Mathematics Assessments: Full Population Estimates.

Table C.12 Changes in the average scale score of grade 8 public school students in the NAEP mathematics assessment, overall and at selected ranges of the achievement scale distribution, based on the full population estimates, by district, large city, and national public: 2003, 2005, and 2007

District	Overall	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Changes 2003 to 2005						
Atlanta	0.9	-0.4	0.7	0.8	1.3	2.2
Austin	—	†	†	†	†	†
Boston	5.7*	-1.2	6.0*	8.2*	7.8*	7.9*
Charlotte	1.3	0.2	1.5	1.1	1.2	2.4
Chicago	4.2*	4.3	4.0	3.5	3.8	5.1
Cleveland	-3.5	-9.0*	-4.0	-2.3	-2.2	0.0
DC	2.7*	4.6	3.1	1.5	1.2	3.1
Houston	3.5*	-1.3	2.6	4.7*	5.9*	5.6*
Los Angeles	5.1*	2.3	5.0	5.0*	5.5*	7.4*
New York City	1.4	3.0	1.1	0.9	0.2	1.9
San Diego	4.7*	-0.5	6.1*	6.8*	5.0*	6.4*
National Public	1.2*	0.8	1.2*	1.1*	1.1*	1.9*
Large City	3.1*	1.9	2.4*	3.0*	3.6*	4.5*
Changes 2005 to 2007						
Atlanta	9.9*	11.0*	11.4*	9.3*	8.7*	9.2*
Austin	4.0*	5.9	6.4*	3.7	2.9	1.2
Boston	7.4*	11.9*	8.2*	6.7*	6.4*	3.4
Charlotte	2.9	6.0*	2.2	1.6	1.1	3.4
Chicago	1.1	-2.7	1.4	2.5	2.1	2.5
Cleveland	2.9	-3.1	4.2	5.2*	6.0*	2.1
DC	0.4	-3.3	0.4	1.6	2.4	0.8
Houston	5.6*	3.9	5.3*	5.7*	5.7*	7.4*
Los Angeles	6.9*	8.0*	6.8*	6.6*	6.6*	6.3*
New York City	3.3	3.7	3.1	3.2	3.0	3.7
San Diego	2.1	2.3	1.2	1.0	2.5	3.5
National Public	2.3*	2.2*	2.3*	2.3*	2.5*	2.3*
Large City	3.3*	2.7*	3.5*	3.5*	3.6*	3.0*
Changes 2003 to 2007						
Atlanta	10.8*	10.6*	12.1*	10.1*	10.0*	11.4*
Austin	—	†	†	†	†	†
Boston	13.1*	10.7*	14.2*	14.9*	14.3*	11.3*
Charlotte	4.2*	6.2*	3.7	2.7	2.4	5.8*
Chicago	5.3*	1.6	5.4	6.0*	6.0*	7.6*
Cleveland	-0.6	-12.1	0.2	2.9	3.7	2.0
DC	3.0	1.3	3.5	3.1	3.5*	3.8
Houston	9.1*	2.5	7.9*	10.4*	11.6*	13.1*
Los Angeles	11.9*	10.3*	11.8*	11.7*	12.1*	13.8*
New York City	4.8	6.7*	4.2	4.0	3.2	5.6
San Diego	6.8*	1.8	7.3*	7.8*	7.4*	9.8*
National Public	3.6*	3.1*	3.5*	3.4*	3.6*	4.2*
Large City	6.4*	4.7*	5.9*	6.5*	7.2*	7.5*

— Austin did not participate in the 2003 NAEP assessment.
† Not applicable.
* Difference is statistically significant at $p < .05$.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005 and 2007 Mathematics Assessments: Full Population Estimates.

APPENDIX D

ALIGNMENT ANALYSIS METHODOLOGY

Appendix D – Alignment Analysis Methodology

D.1. Sources for Alignment Analyses – State and District Standards

Atlanta

Georgia Performance Standards:

Reading:
https://www.georgiastandards.org/standards/GPS%20Support%20Docs/gps_summary_ela.pdf

Mathematics:
https://www.georgiastandards.org/Standards/Pages/BrowseStandards/BrowseGPS.aspx

Science:
https://www.georgiastandards.org/Standards/Pages/BrowseStandards/ScienceStandards.aspx

Boston

Reading:
www.bostonpublicschools.org/node/353
http://www.doe.mass.edu/acls/frameworks/ELA.pdf

Mathematics:
http://bostonpublicschools.org/node/353

Science:
Massachusetts Science and Technology/Engineering Curriculum Framework (May 2001):
http://www.doe.mass.edu/frameworks/scitech/2001/0501.pdf
Grade 3 (implemented academic year 2003–2004): "Water" http://www.fossweb.com/modules3-6/Water/index.html
Grade 3 (implemented academic year 2004–2005): "Physics of Sound"
http://www.fossweb.com/modules3-6/PhysicsofSound/index.html
Grade 4 (implemented academic year 2004–2005): "Magnetism/Electricity"
http://www.fossweb.com/modules3-6/MagnetismandElectricity/index.html
Grade 5 (implemented academic year 2003–2004): "Levers and Pulleys"
http://www.fossweb.com/modules3-6/LeversandPulleys/index.html
Grade 7 (implemented academic year 2003–2004): "Diversity of Life"
http://www.fossweb.com/modulesMS/DiversityofLife/index.html
Grade 7 (implemented academic year 2004–2005): "Earth History"
http://www.fossweb.com/modulesMS/EarthHistory/index.html
Grade 8 (implemented academic year 2003–2004): "Planetary Science"
http://www.fossweb.com/modulesMS/PlanetaryScience/index.html
Grade 8 (implemented academic year 2004–2005): "Populations and Ecosystems"
http://www.fossweb.com/modulesMS/PopulationsandEcosystems/index.html

Charlotte

Reading:
http://www.dpi.state.nc.us/curriculum/languagearts/scos/2004/19grade3
http://www.dpi.state.nc.us/curriculum/languagearts/scos/2004/19grade4
http://www.dpi.state.nc.us/curriculum/languagearts/scos/2004/19grade5
http://www.dpi.state.nc.us/curriculum/languagearts/scos/2004/19grade7
http://www.dpi.state.nc.us/curriculum/languagearts/scos/2004/19grade8

Mathematics:
http://www.cms.k12.nc.us/cmsdepartments/ci/mathandscience/Pages/K-8Mathematics.aspx

Science:
North Carolina Standard Course of Study:
http://www.ncpublicschools.org/curriculum/science/scos/2004/

Cleveland

Reading:
http://www.cmsdnet.net/en/Academics/ScopeAndSequence.aspx
http://education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&TopicRelationID=1699&ContentID=489&Content=67593
http://www.genevaschools.org/standards/

Mathematics:
http://www.cmsdnet.net/en/Academics/ScopeAndSequence.aspx

Science:
Ohio's Academic Content Standards for Science:
http://www.ode.state.oh.us/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&TopicRelationID=1705&ContentID=834&Content=72481

D. 2. Column Labels on Standardized Alignment Charts

Grade 4

Column A:	NAEP content for grade 4 (2003–2007)
Column B:	Code representing the depth of cognitive demand implied for the NAEP content
Column C:	Aligned state content standards for grades 3 and 4 in effect in 2006–2007 (2004–2005 for science)
Column D:	Aligned state content standards for grade 5 in effect in 2006–2007 (2004–2005 for science)
Column E:	Cognitive demand code for the grades 3 and 4 state standard
Column F:	NAEP-to-state grade-level match code
Column G:	NAEP-to-state content match code
Column H:	Aligned district content standards for grades 3 and 4 in effect in 2006–2007 (2004–2005 for science)
Column I:	Aligned district content standards for grade 5 in effect in 2006–2007 (2004–2005 for science)
Column J:	Cognitive demand code for the district standard
Column K:	NAEP-to-district grade-level match
Column L:	NAEP-to-district content match
Column M:	Comments/clarifications

Grade 8

Column A:	NAEP content for grade 8 (2003–2007)
Column B:	Cognitive demand code for the NAEP content
Column C:	Aligned state content standards for grades 7 and 8 in effect in 2006–2007 (2004–2005 for science)
Column D:	Cognitive demand code for the grades 7 and 8 state standard
Column E:	NAEP-to-state grade-level match code
Column F:	NAEP-to-state content match code
Column G:	Aligned district content standards for grades 7 and 8 in effect in 2006–2007 (2004–2005 for science)
Column H:	Cognitive demand code for the district standard
Column I:	NAEP-to-district grade-level match
Column J:	NAEP-to-district content match
Column K:	Comments/clarifications

D. 3. Process for Aligning NAEP and State and District Standards

One of the two coders examined the state and district content standards and determined alignment between the state content standard and the NAEP content objectives in column A. By standards, we mean the expressions of content by grade level, including benchmarks, indicators, objectives, and/or expectations, depending on the nomenclature of the state and/or district standards. Matches were determined on the basis of the decision rules described below in appendix D.4.

Instructions to coders:

For grade 4, read the state standards for grade 3 to determine whether the content can be matched to the NAEP content objectives. Then read grade 4 and then grade 5. Repeat this process of seeking matches for the district standards for grades 3, 4, and then 5. For grade 8, repeat this process with the state and district standards, first for grade 7 and then for grade 8.

- o As elaborated in appendix D.4, look first at the specific mathematics and science content, then look at the verbs that describe what students should know and be able to do. Finally, to help in determining matches, look for key technical concepts and terms and for examples that elaborate on the intent of the standards.
- o For reading, the first determination was the "context" for reading referred to in the state and district standards. In some cases, a standard statement referred specifically to "literary" or "informational" text; in other cases, the standard referred to reading in either genre. When a standard referred to both genres, the standard was coded twice.
  After determining the "context," the aspect of reading was determined.
- o The coder should annotate the state standards document to provide evidence of his or her decisions and thought processes about what has been included and what has been omitted from the chart. This annotation entails a notation of what NAEP content objective the content is matched to (where there are questions or nuanced judgments) and which state or district standards are not matched.

- Where matches exist at grade 4, enter grades 3 and 4 state standards in the cell in column C of the grade 4 matrix *even if the standard extends beyond the NAEP content statement* (e.g., a standard might refer to understanding aspects of literary and informational text; use brackets to indicate aspects that do not apply).

- Enter matching grade 5 standards in column D *when the topic/content has not already been addressed at grade 3 or 4 (or when it is further elaborated at grade 5).*

- Repeat this matching process with the district standards to populate columns H and I for grade 4. Annotate the district standards document the same way as the state standards document.

- For grade 8, repeat this matching process with the state standards and district standards, entering matching grades 7 and 8 state content in column C, matching grades 7 and 8 district content in column G, and annotating the state and district standards documents as for grade 4. Note that there is no equivalent to the grade 5 column on the grade 8 alignment chart.

D. 4. Content-Matching Decision Rules

Summary of Mathematics Content-Matching Decision Rules

1. In reviewing state and district standards, indicators, or expectations, a match with a particular NAEP content objective is determined by examining the content, the verbs, and the elaboration (if any) in the text of the state or district content standards. Here are some possibilities and related actions:

    a. The mathematics content matches exactly or closely (e.g., add whole numbers or length).
    b. A state or district objective is broader than a single NAEP content objective and will be matched to more than one content objective.
    c. Two or more state or district standards (at the same or at different grades) will be matched to a single NAEP content objective.
    d. The content of the state or district standard appears to match more than one NAEP content objective. In this case, look to the verbs to determine where the most appropriate match occurs (e.g., add, describe, or measure) or whether, even after considering the verbs, the same content is appropriately matched to more than one NAEP content objective.

    For further support in making matching decisions, turn to the elaborating language (e.g., including prisms and cylinders or using concrete models like riding an escalator).

2. Content is entered for grade 5 only (a) when it matches an NAEP content objective AND grades 3 and 4 are missing or (b) when grade 5 elaborates and expands on grades 3 and 4 in ways that represent a closer or clearer match.

3. In making these matching decisions, **err on the side of being inclusive.** It is much easier to decide later in the process that a match does not really hold than it is to reconstruct a match that was not made.

Summary of Science Content-Matching Decision Rules

1. State and district standards that focus only on science process and/or skills will not be matched to the NAEP content objectives.

    a. This exclusion reduces the number of inferences made during the matching process because the excluded standards are frequently devoid of specific content. This absence of content requires that either the standard be assumed appropriate for alignment to all NAEP content objectives or the coders have to decide which content is most appropriate to the stated skill, a step that would significantly increase the subjectivity and decrease the reliability of the alignment task.

    b. Consider the following example from *Texas Essential Knowledge and Skills*: "Science is a way of learning about the natural world. Students should know how science has built a vast body of changing and increasing knowledge described by physical, mathematical, and conceptual models, and also should know that science may not answer all questions." Because this standard does not refer to specific content, it would be difficult for a coder to identify the NAEP content objective(s) to which it is most closely aligned. This difficulty would most likely result in matches that would be difficult to replicate. For this reason, if Texas were included in the study, this standard would not be matched to the NAEP content objectives.

2. In reviewing state and district standards, indicators, or expectations, a match with a particular NAEP content objective is determined by examining the content, the verbs, and the elaboration (if any) in the text of the state and/or district content standards. Here are some options:

   a. The science content matches exactly or closely.
   b. The content of the NAEP content objective is foundational to the content encompassed by the state and/or district standard.
   c. A state or district standard is broader than a single NAEP content objective and will match to more than one NAEP objective.
   d. Two or more state or district standards (at the same or at different grades) match to a single NAEP content objective.

   The supporting curricular and instructional materials may be referenced to help determine the content boundaries of a state or district standard.

3. Content is entered for grade 5 only when (a) it matches an NAEP content objective AND there is no applicable grade 3 or grade 4 state and/or district standard OR (b) when grade 5 elaborates and expands on grades 3 and 4 in ways that represent a closer or clearer match.

4. In making these matching decisions, **err on the side of being inclusive**. It is much easier to decide later in the process that a match does not really hold than it is to reconstruct a match that was not made.

5. When available, published research on how students learn is used to support identified implicit matches. Resources for such research include *Benchmarks for Science Literacy* (AAAS, 1993); *Atlas of Science Literacy*, Vol. 1 (AAAS, 2001); and *Atlas of Science Literacy*, Vol. 2 (AAAS, 2007).

Summary of Reading Content-Matching Decision Rules

1. When state or district standards seem to refer to multiple NAEP contexts for reading, they should be considered for matching in all relevant contexts.

2. When state and district standards include multiple elements of text, they should be considered for matching to all relevant NAEP tasks, in all contexts for reading.

   a. Bracket the elements of text that do not directly relate to the NAEP task, e.g., "Identify and analyze the elements of plot [character, and setting] in the stories they read [and write].
   b. Depending on the wording of the standard, *plot* may be coded under "forming a general understanding" as *major events* or under "developing an interpretation" as *interpreting major events* or *extending initial understanding to interpret and develop deeper understanding of problem/conflict/resolution*.

3. State and district standards often include elements of text for each context for reading (e.g., *plot and theme* for literary texts and *events and main ideas* for informational texts) across multiple aspects of reading (forming a general understanding, developing an interpretation, making reading/text connections, or examining content and structure).

   Each example should be considered individually for matching. For example, the standards presented above – "Identify and analyze the elements of plot [character, and setting] in the stories they read [and write]" – would be considered in both forming a general understanding because of the verb "identify" and developing an interpretation because of the verb "analyze."

4. *Cause and effect* can have different meanings in literary and informational texts.

   a. In literary texts, the term refers to the problem/conflict/resolution elements of text and as such is most often coded as forming a general understanding or developing an interpretation.
   b. In informational texts, the term refers to events, main ideas, and supporting details and here, too, is most often coded as forming a general understanding or developing an interpretation.

5. *Point of view* may be codable for literary or informational texts.

   a. In informational texts, key words and phrases to suggest *point of view* are perspective, ideas author wants to convey, author's use of information, or quality of ideas presented.
   b. In literary texts, the term may apply to the author's point of view or to the point of view of different characters as depicted by the author.

6. The author's *point of view* is different from his or her *purpose.*

   a. Standards referring to literary text often use the term *author's intention.*
   b. Standards referring to informational or procedural text often use the term *central purpose*, at times without reference specifically to an author.

7. *Structure and organization* can refer to multiple aspects of texts within the three contexts for reading; the terms may refer to the structure and organization as a whole or aspects of the structure as determined by the author.

a. Traditional story structure/story grammar (beginning, middle, end); stanzas, verse, etc. in poetry, in literary texts
b. Developmental structures of argumentative, persuasive, chronological texts, in informational texts
c. Directly stated or implied steps in a procedural document
d. Adjunct aids, graphics, charts, tables, etc. in informational or procedural texts or in documents

8. If standards seem ambiguous, interpret them literally—at the lowest possible level of implementation—and annotate the decision-making process.

9. Standards referring to making predictions or asking questions or other classroom-based activities should not be considered unless they also contain a testable behavior. Standards referring to reading and *writing* should be considered only on the basis of the reading behaviors they imply because classroom-based writing is not similar to writing in response to an open-ended test question.

10. Standards referring to vocabulary should be coded as follows:

a. *Forming a general understanding* if readers must apply fundamental word knowledge
b. *Developing an interpretation* if readers must use a strategy such as analysis of context clues
c. *Making reader/text connections* if the standard implies that the reader must draw upon her background knowledge, for example, to find a connotative meaning of a word
d. *Examining content and structure* if the standard implies determining why an author made specific vocabulary choices

The following table provided additional support in understanding the interactions of the contexts for reading and the aspects for reading assessed by NAEP.

Reading for Literary Experience	**Reading for Information**	**Reading to Perform a Task**
Forming a General Understanding	Forming a General Understanding	Forming a General Understanding
Considering text as a whole and providing a global understanding	Considering text as a whole and providing a global understanding	Considering text as a whole and providing a global understanding
Theme	Central purpose	Central purpose
Major characters	Major ideas	Key information
Major events	Supporting ideas	Key organizing features
Problem/conflict/resolution	Adjunct aids	Key graphics
Vocabulary	Vocabulary	Interrelationship of key graphics
Literary devices		Vocabulary
Developing an Interpretation	Developing an Interpretation	Developing an Interpretation
Linking across parts of texts	Linking across parts of texts	Linking across parts of texts
Making inferences	Drawing inferences about relationships of two pieces of text	Drawing inferences about relationships of two pieces of text
Providing evidence for actions	Providing evidence to determine reasons for an	Providing evidence to determine reasons for an

	action	action
Theme	Central purpose	Central purpose
Major characters	Major ideas	Major ideas
Major events	Supporting ideas	Supporting ideas
Problem/conflict/resolution	Adjunct aids	Adjunct aids
Vocabulary	Vocabulary	Vocabulary
Literary devices		
Making Reader/Text Connections	Making Reader/Text Connections	Making Reader/ Text Connections
Comparing information in text with knowledge and experience (refers primarily to own self)	Comparing information in text with knowledge and experience (refers primarily to own self)	Comparing information in text with knowledge and experience (refers primarily to own self)
Applying ideas in text to real world (refers to self and others)	Applying ideas in text to real world (refers to self and others)	Applying ideas in text to real world (refers to self and others)
Theme	Central purpose	Central purpose
Major characters	Major ideas	Key information
Major events	Supporting ideas	Key organizing features
Problem/conflict/resolution	Adjunct aids	Key graphics
Vocabulary	Vocabulary	Vocabulary
Literary devices		
Examining Content and Structure	Examining Content and Structure	Examining Content and Structure
Critically evaluating	Critically evaluating	Critically evaluating
Comparing and contrasting	Comparing and contrasting	Comparing and contrasting
Literary devices: understanding the effects of features such as irony, humor, organization/structure	Understanding the effects of organization	Central purpose
Theme	Central purpose	Key information
Major characters	Major ideas	Key organizing features
Major events	Supporting ideas	Key graphics
Problem/conflict/resolution	Adjunct aids	Interrelationship of key graphics
Vocabulary	Vocabulary	Vocabulary

Resources for Use in Understanding NAEP Reading

National Assessment Governing Board. (1990). *Assessment and exercise specifications for National Assessment of Educational Progress in Reading 1992 - 1998*. Washington, DC: National Assessment Governing Board & The Council of the Chief State School Officers.

National Assessment Governing Board. (2003). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board.

D. 5. Content-Matching Exercises

Mathematics Content-Matching Exercises

A. Where does it go? Why?

	State or district standard (grades 3–5)	Where does it go in NAEP standards? Why?
1	Use place value to read, write, compare and order decimals, involving tenths and hundredths, including money, using concrete models (such as play money [dollars, dimes, and pennies] to model, record, read, and compare decimals numbers).	*Grade 4, 1(J) Order or compare whole numbers, decimals, or fractions.* *This state/district standard seems to fit best in the "number sense" content. It also mentions comparing and ordering decimals, which is why I selected (J). It could also fit with (E) or (B).*
2	Identify models of parallel and perpendicular lines on two-dimensional shapes (such as opposite sides of a rectangle are parallel, or consecutive sides of a square are perpendicular).	*Geometry 4.A – Describing perpendicularity/parallelism implies also being able to identify it*
3	Use scientific notation to represent very large and very small numbers.	*None – not included in grade 4 standards*
4	Develop fluency with addition and subtraction of non-negative rational numbers with like denominators, including decimal fractions through hundredths.	*Number operations: add and subtract fractions with like denominators, or decimals through hundredths*
5	The student is expected to measure to solve problems involving length, including perimeter, time, temperature, and area.	*Measurement 2b. Solve problems involving conversions with the same system and 2d. Determine appropriate size of unit….*

B. Is it a match? Why or why not?

	NAEP content objective	State or district standard	Is it a match? Why/why not?
1	Identify or describe (informally) real-world objects using simple plane figures (e.g., triangles, rectangles, squares, and circles) and simple solid figures (e.g., cubes, spheres, and cylinders).	The student identifies and describes lines, shapes, and solids using formal geometric language.	*Yes – Both involve identifying and describing geometric figures. Although the state standard uses the word "formal," it is implied that students are also able to provide informal descriptions.*
2	Recognize which attributes (such as shape and area) change or don't change when plane figures are cut up	Demonstrate translations, reflections, and rotations using concrete models (such as riding an	*No – Although both involve rearranging figures, the state/district standard does*

	NAEP content objective	State or district standard	Is it a match? Why/why not?
	or rearranged.	escalator [translation], flipping a pancake [reflection], or turning a cartwheel [rotation]).	*not require a student to recognize which attributes change or don't change.*
3	Apply basic properties of operations.	Justify why an answer is reasonable and explain the solution process.	*No – This standard would fit better with NAEP 5(f) – "Explain or justify a mathematical concept or relationship."*
4	Solve problems involving perimeter of plane figures.	Develop strategies to determine the area of rectangles and the perimeter of plane figures.	*Yes – Perimeter matches perfectly, and we would look to match to area elsewhere.*
5	For a given set of data, complete a graph (limits of time make it difficult to construct graphs completely).	Generate a table of paired numbers based on a real-life situation such as insects and legs.	*No – One is about graphs and the other is about tables.*

C. Grade 5 matches

	NAEP specification	Grade 3 or 4 standard	Grade 5 standard	Enter or ignore? Why/why not?
1	Construct geometric figures with vertices at points on a coordinate grid.	None	The student is expected to locate and name points on a coordinate grid using ordered pairs of whole numbers.	*Although the grade 5 standard does not mention constructing geometric figures, I would err on the side of being inclusive here and enter it, since they both require students to locate and name points on a coordinate grid.*
2	Describe attributes of two- and three-dimensional shapes.	Describe shapes and solids in terms of vertices, edges, and faces.	Identify vertices, edges and faces of solid figures.	*Ignore – The grade 4 standards require a higher cognitive demand than the grade 5 standards. Nothing is added in grade 5.*
3	Use letters and symbols to represent an unknown quantity in a simple mathematical expression.	Explore the concept of a variable when finding missing addends in real-life situations.	Select from and use diagrams and number sentences to represent real-life situations.	*Enter –The grade 5 standard expands on the grade 4 standard with reference to "number sentences" and comes closer to matching NAEP.*
4	Explore properties of paths between points.	Use a rectangular coordinate system to solve problems.	Classify plane figures according to types of symmetry (line, rotational).	*Ignore – Grade 5 is not related to either NAEP or to grade 3 or 4 content.*

5	Recognize, describe, or extend numerical or geometric patterns.	Make predictions and solve problems using whole numbers and geometric patterns.	Make predictions and solve problems using whole numbers and geometric patterns (such as what is the 7th number of the pattern 12, 15, 18, 21).	*Ignore – Grade 5 adds nothing to grades 3 and 4.*

Science Content-Matching Exercises

A. Where does it go? Why?

	State/District Standard (grades 3–5)	**Where does it go in NAEP standards? Why? (grade 4)**
1	4b.11 (A) The student is expected to test properties of soils including texture, capacity to retain water, and ability to support life.	*ES.A.4a Students know some facts about the composition of soil; for example, students can separate soil samples into component parts.* *The state standard requires students to be able to test various properties of soil, each of which (properties) depends on the composition of the soil.*
2	3b.11 (D) The student is expected to describe the characteristics of the Sun.	*ES.D.1a Students can explain how the Earth differs from the Sun and the Moon.* *The state standard requires student to be able to describe the characteristics of the sun. NAEP requires students to explain how the Sun, Earth, and Moon differ. This requires students to know the characteristics of each body. Thus, the content of the state standard is encompassed within the content of this NAEP content objective.*
3	4a.2 Students identify the physical properties of matter and observe the addition or reduction of heat as an example of what can cause changes in states of matter.	*PS.A.1a Students can classify/identify common objects and substances by physical characteristics such as state of matter, texture, color, size, shape, hardness, and opacity.* *The state standard requires students to identify the physical properties of matter, whereas the NAEP content objective requires students to use physical properties to classify and/or identify common objects.*
4	4a.2 As students learn science skills they identify the role of the Sun as our major source of energy.	*PS.3a.(2) Students can explain/trace how coal, gasoline, and wood originally got their stored energy from the sun.* *The state standard requires students to know that the Sun is the major source of Earth's*

		energy. The NAEP content objective requires the student to know that the Sun is the source of the energy in coal, gasoline, and wood.
5	3b.9 The student knows that many likenesses between offspring and parents are inherited from the parents.	*LS.A.2c Students can describe/identify similarities and differences between multiple offspring of the same parents and between parents and offspring.* *The state standard requires students to know that many similarities between parent and offspring are inherited. The NAEP content objective requires students to be able to identify some of these similarities.*

B. Is it a match? Why or why not?

	NAEP content objective	**State or district standard**	**Is it a match? Why/why not?**
1	ES.A.1a Students can classify substances as soil, sand, or rock.	3a.2 Students identify the importance of components of the natural world including rocks, soils, water, and atmospheric gases.	*No – State/district requires students only to know that the provided components are important. It does not require that students be able to distinguish among or between them or to be able to classify "unknown" objects by type.*
2	ES.A.3a.(2) Students can explain that molten rock comes out of volcanoes, hardens and becomes part of the landscape.	SCI.4.10.A.05 Define and examine lava.	*Yes – This is at least a partial match because defining and examining lava requires students to understand what lava is and where it comes from.*
3	ES.C.2c Students can describe weather changes, list ways of measuring them, and can offer simple explanations for how the weather changes.	SCI.5.06.A.01 Identify and describe examples of change such as daily or weekly changes in weather.	*Yes – This is at least a partial match. The state/district standards require students to be able to describe examples of changes in weather but not to explain how/why those changes occur.*
4	PS.A.3b.(2) Students can distinguish between conductors/nonconductors. When presented with a variety of materials (conductors and nonconductors), a D cell battery, a battery holder, three wires, and a light bulb in a socket, students can construct a testing circuit and sort the objects into two categories.	SCI.4.05.A Identify and describe the roles of some organisms in living systems such as plants in a school-yard, and parts in non-living systems such as a light bulb in a circuit.	*Yes – This represents a partial match. The state/district standard requires students to be able to describe the roles of the components in a circuit. It is not a complete match, however, because students are not required to know*

			which components are or are not conductors.
5	LS.A.3b.(1) Students can offer simple explanations for why things look the way they do, e.g., fish are streamlined, lions have big teeth, etc.	3b.9 (A) The student is expected to observe and identify characteristics among species that allow each to survive and reproduce.	*Yes – This is at least a partial match because the state/district standard requires students to identify characteristics of organisms (e.g., shape, size of teeth/eyes) that are advantageous from a survival and reproductive perspective. The survival and reproductive aspect of the local standard helps explain why organisms look as they do.*

Reading Content-Matching Exercises

A. Where does it go? Why?

	State/District Standard (grades 3–5)	**Where does it go in NAEP standards? Why? (grade 4)**
1	Determining "context for reading" or text type" There are three contexts for reading, which denote the genre or text type included on NAEP: reading for a literary experience (narratives, poetry, some essays), reading for information (expository texts such as persuasion, argumentation, informational, biographical sketches, etc.), and reading to perform a task (procedural text and documents). Key words within the state or district standards point to the "elements" of text about which NAEP question are developed	*Reading for a literary experience: theme, major characters, major events, problem/conflict/resolution, vocabulary, literary devices.* *Reading for information: central purpose, major ideas, supporting ideas, adjunct aids, vocabulary* *Reading to perform a task: central purpose, key information, key organizing features, key graphics, interrelationship of key graphics, vocabulary*
2	11.2 Identify themes as lessons in folktales, fables, and Greek myths for children	*The text types position this in "reading for a literary experiences;" the word "identify" suggests basic, general, or superficial comprehension or "forming a general understanding;" and the word "themes" places the standard in A2.* *A2. Constructing an initial theme or message.*
3	9.3 Identify similarities and differences between the characters or events in a literary work and the actual experiences in an author's life	*This standard assumes that the reader has gained prior knowledge about the life of the author either through reading or from a teacher; and the reader is then able to compare that knowledge to what is presented in a literary work.*

		The behaviors required here would involve comparing/contrasting and also bringing forth one's own knowledge, so the standard would be classified as "making reader/text connections." *C2. Understanding aspects of text by applying ideas in text to the real world (in this case, to prior knowledge)* *C5. Understanding major events by comparing them to own experiences and knowledge (in this case knowledge of the life of the author)*
4	12.2. Identify and analyze the elements of plot, character, and setting in the stories they read [and write].	*The standard refers to literary text, and the verbs "identify" and "analyze" place it in two aspects of reading: " forming a general understanding" and "developing an interpretation."* *A.1. Considering text as a whole to develop an initial impression or understanding* *A.3. Identifying and constructing an initial understanding of main characters.* *A.4. Constructing an initial understanding of problem/conflict/resolution [as proxy for "plot"]* *A5. Being able to identify and state or retell major events.* *B1. Extending initial understanding to develop more complete comprehension of text* *B3. Interpreting major characters, ways in which they change, and reasons for their actions* *B.4. Interpreting major events* *B.5. Extending initial understanding to interpret and develop deeper understanding of problem/conflict/resolution*
5	8.14 Make judgments about setting, characters, and events and support them with evidence from the text	*The standard refers to literary text, and the verbs "make judgment" and "support . . . with evidence" places the standard in developing an interpretation. The reader has to move beyond her initial understanding to deepen comprehension. The standards refer to multiple "elements" of literary text and would be matched to the NAEP tasks.*
6	10.2. Distinguish among forms of literature such as poetry, prose, fiction, nonfiction, and drama and apply knowledge as a strategy for reading [and writing]	*This standard refers to literary and informational text and covers several reading behaviors, specifically distinguishing among genres, learning about their characteristics, and then using that information to deepen comprehension through critical reading of new text. The standard illustrates the complexity of the "examining content and structure" reading context.* *D.1. Critically evaluating and assessing [literary] text in light of own understanding of quality writing or established criteria* *H3. Considering and evaluating the effects of organization on text*
7	13.6 Identify and use knowledge of common textual features (paragraphs, topic sentences, concluding sentences, glossaries) 13.8 Identify and use knowledge of common organizational structures (chronological order)	*References to structural features of text are most common for informational and procedural text. (Procedural text is not included in Grade 4 NAEP Reading, however.) These standards suggest that readers know to look for organizational features in text and use them to form a general understanding.* *E5 Recognizing key organizing features.*

B. Is it a match? Why or why not?

	NAEP content objective	State or district standard	Is it a match? Why/why not?
1	A.3. Identifying and constructing an initial understanding of major characters A.4. Constructing an initial understanding of problem/conflict/resolution B.1. Interpreting major events B.2. Extending initial understanding to interpret and develop deeper understanding of problem/conflict/resolution B.3. Interpreting major characters, ways in which they change, and reasons for their actions	12.4. Locate and analyze elements of plot and characterization and then use an understanding of these elements to determine how qualities of the central characters influence the resolution of the conflict	*Yes. This standard illustrates how states often combine numerous aspects of text and several reading behaviors into one standard. The word "locate" suggests "reading for an initial understanding," while "analyze" suggests reading more deeply, in "developing an interpretation."* *The standard also touches on various elements of text, each of which is covered by a separate NAEP content objective.*
2	F.6. Interpreting adjunct aids as a means of deepening understanding of text	13.19. Identify and use knowledge of common graphic features (charts, maps, diagrams) 13.b. Identify and use knowledge of common graphic features to analyze nonfiction texts	*Yes. The term "adjunct aids" is defined broadly to include graphic elements inserted into texts to convey information and to supplement what is included in text.* *Both standards require readers not just to "identify" but also to "use" graphic elements, hence the match at the "developing an interpretation" level of NAEP.*
3	H3. Considering and evaluating the effects of organization on text	R1.I.d. Identifies and uses knowledge of common organizational structures (e.g., chronological order, cause and effect)	*Yes – partial match. This is a partial match because the standard does not ask for an evaluation of the effectiveness of the organizational structure.*
4	D.2. Considering and evaluating author's craft or choices	13c Identify and analyze the characteristics of various genres as forms chosen by an author to accomplish a task.	*Yes – partial match. This is a partial match because the standard is less inclusive than the NAEP objective.*
5	A.3. Identifying and constructing an initial understanding of major characters A.4. Constructing an initial understanding of problem/conflict/resolution A5. Being able to identify and state or	17.a. Identify and analyze elements of plot and character presented through dialogue in scripts that are read, [viewed, listened [to], or performed.]	*No. Although NAEP assesses elements of plot and character as part of its assessment of understanding of literature, it does not include dramatic literature as cued by the word*

	retell major events B.3. Interpreting major characters, ways in which they change, and reasons for their actions B.4. Interpreting major events B.5. Extending initial understanding to interpret and develop deeper understanding of problem/conflict/resolution		*"script." This standard would apply to dialogue in narrative prose without the word "script."* *It would be a partial match with grades 5–6 standard 8.20: Identify and analyze the author's use of* ***dialogue*** *and description.*

D. 6. Content-Matching Verification Process

Using the annotated state and district standards, the other trained coder verified the primary coder's alignment decisions independently to ensure that state and district content standards at the relevant grade levels had been correctly matched to the NAEP content. The verifier's tasks were set out as follows:

- The verifier will examine each matching decision made, using the annotated state and district standards to verify that all matches are appropriate, all matching state and district content has been matched, and non-matched content does not, in fact, match.
- Where matches seem questionable, the verifier will first consult the state and district standards and, if available, accompanying material (e.g., content examples, proposed instructional activities) to determine the appropriateness of the match.
- Where matches seem questionable or missing, the verifier will note each case and conduct a review with the primary coder to try to reach agreement and produce verified drafts of the charts.
- Where agreement does not emerge, the cells will be noted for the content lead.
- Finally, the content lead will resolve any issues that emerged from the verification process and will review the complete set of matching decisions to create final drafts of the charts.

D. 7. Process for Coding Content Matches and Cognitive Demand

This part of the alignment process entailed the assignment of seven separate codes to the content objectives and standards that had been entered into columns A, C, D, H, and I for grade 4 and columns A, C, and G for grade 8 of the alignment charts.

Assigning NAEP-to-State/District Grade-Level Match Codes

Purpose: These codes indicate the extent of overlap between the NAEP content listed in column A (for grades 4 and 8) and the instructional content referenced by the state and district standard statements (at grades 3, 4, and 5 for NAEP grade 4 and at grades 7 and 8 for NAEP grade 8). Assigning this code required a straightforward review of the column entries as described below.

Codes: For each content area, two coders and the content lead independently assigned one of the following codes. [4]

- Code **N** (for **n**ot): NAEP content is not covered anywhere in the state or district standards.
- Code **M** (for **m**atch): NAEP content is covered (even partially) at grades 3 or 4 for the grade 4 alignment and at grades 7 or 8 for the grade 8 alignment.
- Code **L** (for **l**ater): NAEP content is covered (even partially) at grade 5 for the grade 4 alignment. (NOTE: Code **L** is applicable only to the grade 4 alignment charts because the NAEP content objectives and state/district standards were not reviewed beyond grade 8.)

Placement: The codes were entered in the following columns:[5]

- Column F (at grade 4) for the NAEP/state grade-level match
- Column K (at grade 4) for the NAEP/district grade-level match

[4] For reading, one coder reviewed independently, a second coder reviewed that work, and the content lead also independently coded the standards. Then the team reconciled differences. Much of the reconciliation process led to the coding rules presented earlier.

[5] These columns differ for reading because the coding matrices differed.

- Column E (at grade 8) for the NAEP/state grade-level match
- Column I (at grade 8) for the NAEP/district grade-level match

Decision Process:

Step 1: For grade 4, examine the NAEP content objectives in column A and the state standards in columns C and D. For grade 8, examine the NAEP content objectives in column A and the state standards in column C. For every row:

- If the state column cell is empty, then code N for not matched.
- If the state column cell for grades 3 and 4 or for grades 7 and 8 contains an entry (**regardless of the degree of content match**), then code M for matched.
- If the state column cell for grades 3 and 4 is empty, but the cell for grade 5 contains an entry, then code L for matched later.

Step 2: For grade 4, examine the NAEP content objectives in column A and the district standards in columns H and I. For grade 8, examine the NAEP content objectives in column A and the district standards in column G. For every row:

- If the district column cell is empty, then code N for not matched.
- If the district column cell for grades 3 and 4 or for grades 7 and 8 contains an entry (**regardless of the degree of content match**), then code M for matched.
- If the district column cell for grades 3 and 4 is empty, but the cell for grade 5 contains an entry, then code L for matched later.

Assigning NAEP-to-State/District Content-Match Codes

Purpose: These codes indicate the extent of content overlap or alignment between the NAEP content listed in column A and the instructional content described by the state and district standards/grade-level standards.

Codes: For each content area, two coders and the content lead independently assigned one of the following codes:

- Code **N** (for **n**o): There is no explicit or implicit match or reference to the NAEP content in the state or district standards at any grade level.
- Code **P** (for **p**artial): There is some—even minimal—explicit or implicit match or reference to the NAEP content in the state or district standards at any grade level.
- Code **C** (for **c**omplete): There is a complete match of the NAEP content and the state or district standards at all grades; that is, a reasonable person, with reasonably strong content knowledge, would say that the entries refer to essentially the same content and skills.

Placement: The codes were entered in the following columns:

- Column G (at grade 4) for the NAEP/state grade-level match
- Column L (at grade 4) for the NAEP/district grade-level match
- Column F (at grade 8) for the NAEP/state grade-level match
- Column J (at grade 8) for the NAEP/district grade-level match

Decision Process:

Step 1: For grade 4, examine the NAEP content objectives in column A and the state standards in columns C and D. For grade 8, examine the NAEP content objectives in column A and the state standards in column C. For every row:

- If the state column cell is empty, then code N for not matched.
- If the state column cells for grades 3 and 4 or for grades 7 and 8 contain an entry, then code P for partial or C for complete on the basis of the following decision rule:

  If the NAEP and the state content address the same mathematical expectations such that a reasonable person with reasonably strong content knowledge would say that the entries refer to essentially the same content and skills, then code C; otherwise, code P.

Step 2: For grade 4, examine the NAEP content objectives in column A and the district standards in columns H and I. For grade 8, examine the NAEP content objectives in column A and the district standards in column G. For every row:

- If the district column cell is empty, then code N for not matched.
- If the district column cells for grades 3 and 4 or for grades 7 and 8 contain an entry, then code P for partial or C for complete on the basis of the following decision rule:

  If the NAEP and the district content address the same mathematical expectations such that a reasonable person with reasonably strong content knowledge would say that the entries refer to essentially the same content and skills, then code C; otherwise, code P.

Assigning Cognitive Demand Codes

Purpose: These codes refer to the cognitive demands of the NAEP content objectives and the strategies or skills implicitly or explicitly referenced in the state and district standards. Working independently, the content lead and two coders applied the cognitive demand coding guidelines to assign a cognitive demand code.

Codes: For each content area, two coders and the content lead independently assigned one of the following codes.

- Code **L** (for **l**ow): The task the student must perform is relatively simple for a student with grade-level skills and appropriate background knowledge and experiences. The task places low cognitive demand on the student. Examples include recalling factual information, locating a piece of text that explicitly states the answer to a question, or performing a simple task.
- Code **M** (for **m**edium): The task requires some cognitive engagement and mental processing beyond recalling or reproducing information. The task may involve making some decisions about solving a problem or drawing inferences by looking across several sections of text.
- Code **H** (for **h**igh): The task requires a greater depth of cognitive processing, including planning, using evidence, and applying demanding cognitive reasoning. The task may involve justifying a response, explaining the procedures followed, substantiating one's thinking, and thinking abstractly.

See appendix D.9 for more detailed descriptions of Webb's related Depth of Knowledge Levels 1, 2, and 3 for mathematics and science, and see appendix D.10 for Karen Wixson's discussion of this topic for reading.

Placement: The codes were entered in the following columns:

- For each NAEP content objective, enter code in column B.
- For each state standard for grades 4 and 8:
  - Enter grade 4 code in column E.
  - Enter grade 8 code in column D (at grade 8).
- For each district standard for grades 4 and 8:
  - Enter grade 4 in column J.
  - Enter grade 8 in column H.

Decision Process:

Step 1: Analysis of Verbs: Coding for cognitive demand begins with an analysis of the operative verbs in the NAEP content objectives and in the state and district standards. The coding is adjusted as necessary, based on the context and content of the standard. The following table provides examples of verbs that **may** connote cognitive demand at the low, medium, and high levels. The lists are not exhaustive and should be considered to be only the first reference point for coding. When a content objective or standard contains more than one verb, the overall intent and content must be examined before automatically assigning the level to the "higher" verb.[6] Often the Depth of Knowledge descriptions in appendices D.9 and D.10 helped clarify the correct coding.

Verbs expressing what students need to know and be able to do		
Low Cognitive Demand	Medium Cognitive Demand	High Cognitive Demand
Identify	Explain	Apply
Recognize	Compare	Verify
Find	Represent	Justify
Calculate	Summarize	Analyze
Express	Describe	Evaluate
Extend	Interpret/infer	Develop
Write	Determine	Model
Order	Construct	Prove
Add	Use	Generalize
Subtract	Display	Create
Multiply	Measure	Plan
Divide	Estimate	Support/justify thinking
Rename	Graph	Extend knowledge
Recall	Connect	Recognize and discuss interactions
Measure	Make	
Understand at a basic level	Translate	Conduct investigations

[6] For reading, the code assigned reflects the verb denoting the greatest level of reading comprehension. Thus, if a standard included the word "analyze" or "evaluate," it was coded as having a high level of demand. Considerations of cognitive demand for reading have to be tempered by an awareness of the kinds of items on a state test that measure the standard, in that reading a long passage and answering constructed-response items about it will arguably require a higher level of cognitive demand than responding to multiple-choice items about a short passage.

Investigate Explore Added for reading: Locate	Draw Select Classify Assemble Distinguish Solve Organize Observe and predict Examine and Predict Relate Differentiate Added for reading: Examine	Added for reading: Reorganize

Step 2: Confirmation of Initial Determination: Even the most careful analysis of the verbs of NAEP content or state and district standards may not be the definitive step in determining cognitive demands, so it is important to confirm the initial coding decision. This should be done by looking carefully beyond the verb to the specifics of the mathematics, reading, or science content to get a fuller sense of what the statement represents, that is, the totality of what students should know and be able to do. It is possible that the NAEP content statement or the standards themselves will provide this fuller sense of expected student behaviors; however, in some cases it may be necessary to look at accompanying documentation such as content examples or suggested instructional activities.

D. 8. Content Coding Exercises

Mathematics Content Coding Exercises

A. Grade-Level Match Codes

NAEP Content	State Grades 3 and 4 Content	State Grade 5 Content	Code: N, M, L	Justification
Describe attributes of two- and three-dimensional shapes	Describe shapes in terms of vertices, edges, and faces		M	*Match is with grades 3 and 4 content.*
Identify or draw angles and other geometric figures		Identify acute, right, obtuse, and straight angles	L	*Match is with grade 5 content.*
Determine situations in which a highly accurate measurement is important			N	*No match*

B. Content Match Codes

NAEP Content	State Grades 3 and 4 Content	State Grade 5 Content	Code: N, P, C	Justification
Describe attributes of two- and three-dimensional shapes	Describe shapes in terms of vertices, edges, and faces		P	*State content addresses only one aspect of attributes*
Identify or draw angles and other geometric figures		Identify acute, right, obtuse, and straight angles	P	*State content addresses only angles and does not involve "draw"*
Determine situations in which a highly accurate measurement is important			N	*No matching content*
Compare two sets of related data	Describe relationships between two sets of data such as ordered pairs in a table		C	*Content is equivalent*
Select or use appropriate types of unit for the attribute being measured, such as length, time, or temperature	Use a thermometer to measure temperature		P	*State content addresses only temperature and does not address selection or use of units*
Solve problems involving areas of squares and rectangles	Use concrete models of square units to determine the area of shapes		P	*State content addresses only a part of the NAEP content*

Solve problems involving perimeter of plane figures		Use linear measure to find the perimeter of a shape	C	*Content is equivalent*

C. Cognitive Demand Codes

Content objective	Code (L, M, H)	Justification
Describe attributes of two- and three dimensional shapes	M	*Addresses conceptual understanding of these shapes and goes beyond just recall*
Identify or draw angles and other geometric figures	L	*Requires merely a recall of terms*
Determine situations in which a highly accurate measurement is important	M	*Requires some analysis and comparison*
Identify congruent shapes	L	*Merely identifying shapes that share a characteristic*
Verify a conclusion using algebraic properties	H	*Verification requires reasoning and strategic thinking*
Compare two sets of related data	M	*Requires comparison*
List all possible outcomes of a probability experiment	M	*Likely to require the engagement of some mental processing beyond a habitual response*
Select or use appropriate types of unit for the attribute being measured, such as length, time, or temperature	M	*If only selection, this is likely to be L, but "using" these units pushes this to M*
Solve problems involving areas of squares and rectangles	M	*Requires problem solving*
Solve problems involving perimeter of plane figures	M	*Requires problem solving*
Use place value to read, write, compare and order whole numbers through millions	M	*Comparing and ordering are medium depth of knowledge*
Justify why an answer is reasonable, and explain the solution process	H	*Verification requires reasoning and strategic thinking*

Science Content Coding Exercises

A. Grade-Level Match Codes

NAEP Content	State Grades 3 and 4 Content	State Grade 5 Content	Code (N, M, L)	Justification
Students can classify substances as soil, sand, or rock.	Compare distinct properties of rocks (e.g., color, layering, texture)		M	*Match is with grades 3 and 4 content*
Students can describe basic requirements for living things, e.g., plants and animals need food for energy and growth.		Summarize that organisms can survive only in ecosystems in which their needs can be met (e.g., food, water, shelter, air, carrying capacity, waste disposal). The world has different ecosystems, and distinct ecosystems support the lives of different types of organisms.	L	*Match is with grade 5 content*
Students know that water exists not only on the Earth's surface but beneath the Earth's surface as well; for example, students can identify caves and springs as evidence of underground water.			N	*No match*

B. Content Match Codes

NAEP Content	State Grades 3 and 4 Content	State Grade 5 Content	Code (N, P, C)	Justification
Students can identify/describe where plants and animals get their energy.	• Analyze plant and animal structures and functions needed for survival and describe the flow of energy through a system that all organisms use to survive. • Relate animal structures to their		C	*NAEP content is completely covered by collection of state standards*

	specific survival functions (e.g., obtaining food, escaping or hiding from enemies). • Relate plant structures to their specific functions (e.g., growth, survival, reproduction).			
Students can identify useful properties of common materials; for example, solubility—students can give evidence (taste, color, smell) that, even though solids seem to disappear in water, they are still there (that is, they have dissolved).	• Identify and describe the physical properties of matter in its various states. • Describe objects by the properties of the materials from which they are made so that these properties can be used to separate or sort a group of objects (e.g., paper, glass, plastic, metal).		P	*State content does not address usefulness of properties*
Weighing—Given one or more objects and an appropriate balance, students can correctly measure and record the weight of the object(s).			N	*No matching content*
Students can use metric devices to measure linear dimensions of objects, weight, volume, temperature.		Define temperature as the measure of thermal energy and describe the way it is measured.	P	*State content standard does not address devices used to measure linear dimensions of objects, volume, or weight*
Select or use appropriate types of units for the attribute being measured such as length, time, or temperature	Use a thermometer to measure temperature		P	*State content only addresses temperature and does not address selection or use of units*
Given daily weather data, students can make weather charts.	• Analyze weather and changes that occur over a period of time.		C	*NAEP content is completely covered by collection of state standard*

	• Record local weather information on a calendar or map and describe changes over a period of time (e.g., barometric pressure, temperature, precipitation symbols, cloud conditions).			

C. Cognitive Demand Codes

Content objective	**Code (L, M, H)**	**Justification**
Describe how wind and ice shape and reshape Earth's land surface by eroding rock and soil in some areas and depositing them in other areas, producing characteristic landforms (e.g., dunes, deltas, glacial moraines).	M	*Requires student to describe a process*
Define temperature as the measure of thermal energy and describe the way it is measured.	L	*Merely recall of terms*
Identify and describe the physical properties of matter in its various states.	M	*Requires student to describe properties*
Classify animals according to their characteristics (e.g., body coverings, body structure).	M	*Requires student to classify/organize organisms into groups*
Analyze plant and animal structures and functions needed for survival and describe the flow of energy through a system that all organisms use to survive.	H	*Requires student to analyze the relationship between structure and function*

Reading Content Coding Exercises

A. Grade-Level Match Codes

NAEP Content	State Grades 3 and 4 Content	State Grade 5 Content	Code (N, M, L)	Justification
Deepening understanding by attention to literary devices	Identify and show the relevance of foreshadowing clues	Identify and analyze sensory details and figurative language	M	*Match is with grade 3 - 4 content*
Providing evidence to determine reasons for an action	Distinguish cause from effect		M	*Match is with grade 3-4 content*
Recognizing and understanding major ideas the author wants to convey	Summarize main ideas [and supporting details]	Identify and analyze main ideas [supporting ideas, and supporting details]	M	*Match is with grade 3–4 content*
Considering and evaluating the effects of organization on text	Identify and use knowledge of common organizational structures (e.g., chronological order, cause and effect)		M	*Match is with grade 3–4 content*
Comparing information in text with own knowledge and experience and applying information in text to real world	Use comprehension strategies such as prior knowledge, predicting, visualizing, questioning and summarizing to understand text	Relate a literary work to its setting, identify and analyze characteristics of various genres (poetry, fiction, nonfiction, short story, and drama) as forms with distinct characteristics and purposes	M	*Match is with grade 3–4 content [Grade 5 assumes student brings own knowledge of genres to comprehension]*
Considering and evaluating author's craft or choices		Identify and analyze author's use of dialogue and description	L	*Match is with grade 5 content*
Interpreting information in text by determining how information supports main idea		Identify and analyze main ideas, supporting ideas, and details	L	*Match is with grade 5 content*
Deepening understanding of text by attention to vocabulary		Determining the meaning of unfamiliar words through context	L	*Match is with grade 5 content*

		clues, definition, and structural analysis (using knowledge of Greek and Latin roots, suffixes and prefixes)		
Critically analyzing text to be able to reorganize ideas presented in text			***N***	*No matches*
Evaluating the vocabulary choices the author has made			***N***	*No matches*
Understanding author's purpose for writing (e.g., to convince, argue, etc.) and opinions about the topic			***N***	*No matches*

B. Content Match Codes

NAEP Content	**State Grades 3 and 4 Content**	**State Grade 5 Content**	**Code (N, P, C)**	**Justification**
Considering the text as a whole to form a general understanding of the concepts or information the author is presenting	• Use comprehension strategies such as prior knowledge, predicting, visualizing, questioning and summarizing to understand text • Use comprehension strategies to access text: accessing prior knowledge, predicting, questioning, visualization, summarizing and structural analysis		C	*NAEP content is completely covered by these two standards*
Recognizing key organizing features	• Identify and use knowledge of common textual features, graphic features, and organization structures in order to gain meaning from a variety of informational	• [Identify and analyze sensory language in literary text] and recognize organizational structures and text features in informational text	C	*NAEP content is completely covered; grade 5 standard covers two kinds of text*

	materials			
Extending initial understanding to interpret and develop deeper understanding of problem/conflict/resolution	• Identifies and analyzes the elements of plot, character, and setting in stories read, [written, viewed, or performed]	• Identifies and analyzes the elements of [setting], characterization, and conflict in plot • Makes judgment and inferences about setting, characters, and events and supports them with elaborating and convincing evidence from the text	C	*NAEP content is completely covered*
Understanding literary devices and the effects of author's choice of literary devices on text features such as irony/humor or understanding organization/structure		Identifies and analyzes author's use of dialogue and description	P	*Grade 5 standard partially addresses the NAEP content, in that dialogue and description do not fully represent the range of literary devices students at this age can understand*
Understanding problem/conflict/ resolution by comparing text to own experiences			N	*NAEP content is not covered*
Critically evaluating the quality of ideas and arguments presented in text			N	*NAEP content is not covered*

C. Cognitive Demand Codes

Content Objective	**Code (L, M, H)**	**Justification**
Critically analyzing author's use of information presented in the text	H	*Requires reader to make a judgment about text and ideas based on established criteria for accuracy or validity and on own opinions*
Critically analyzing text to be able to reorganize ideas presented in text	H	*Requires reader to step back from text and to analyze ways in which ideas might have been*

		presented differently, for example, to present an argument more persuasively or to make a cause-and-effect relationship easier to understand
Understanding major characters by considering them through own experiences and knowledge	M	*Requires reader to compare characters to behavioral notes from another time period or culture, to current norms, or to own experiences*
Determining the importance of major ideas to the topic of the text	M	*Requires reader to look across a text to determine the relative importance of major ideas and their relationship, for example, by sorting fact from opinion*
Recognizing and understanding major ideas the author wants to convey	L	*Requires reader to locate explicitly stated, main ideas in a text and understand them at a superficial/literal level; such understanding should be the foundation for higher levels of comprehension*
Considering the text or document as a whole to determine the central purpose [for procedural texts]	L	*Requires reader to place self in position of the user of a document and to determine at a general or superficial level the information the text provides (e.g., schedules for bus routes) or the sequential procedures to be followed to perform a task (e.g., the steps to complete an order form)*

D. 9 Norman Webb's Descriptors for Depth of Knowledge – Mathematics and Science

Mathematics

Level 1 (Recall) includes the recall of information such as a fact, a definition, a term, or a simple procedure, as well as performing a simple algorithm or applying a formula. In mathematics, a one-step, well-defined, and straight algorithmic procedure should be included at this lowest level. Other key words that signify a Level 1 include "identify," "recall," "recognize," "use," and "measure." Verbs such as "describe" and "explain" could be classified at different levels, depending on what is to be described and explained.

Level 2 (Skill/Concept) includes the engagement of some mental processing beyond a habitual response. A Level 2 assessment item requires students to make some decisions as to how to approach the problem or activity, whereas Level 1 requires students to demonstrate a rote response, perform a well-known algorithm, follow a set procedure (like a recipe), or perform a clearly defined series of steps. Keywords that generally distinguish a Level 2 item include "classify," "organize," "estimate," "make observations," "collect and display data," and "compare data." These actions imply more than one step. For example, to compare data requires first identifying characteristics of the objects or the phenomenon and then grouping or ordering the objects. Some action verbs, such as "explain," "describe," or "interpret," could be classified at different levels depending on the object of the action. For example, if an item required students to explain how light affects mass by indicating there is a relationship between light and heat, this is considered a Level 2. Interpreting information from a simple graph, requiring reading information from the graph, is also a Level 2. Interpreting information from a complex graph that requires some decisions on what features of the graph need to be considered and how information from the graph can be aggregated is a Level 3. Caution is warranted in interpreting Level 2 solely as skills, because some reviewers will interpret skills very narrowly, as primarily numerical skills, and such interpretation excludes from this level other skills, such as visualization skills and probability skills, which may be more complex simply because they are less common. Other Level 2 activities include explaining the purpose and use of experimental procedures; carrying out experimental procedures; making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts.

Level 3 (Strategic Thinking) requires a higher level of thinking than the previous two levels. It requires reasoning, planning, and using evidence. In most instances, requiring students to explain their thinking is a Level 3. Activities that require students to make conjectures are also at this level. The cognitive demands at Level 3 are complex and abstract. The complexity does not result because there are multiple answers—a possibility for both Levels 1 and 2—but because the task requires more demanding reasoning. An activity, however, that has more than one possible answer and requires students to justify the response they give would most likely be a Level 3. Other Level 3 activities include drawing conclusions from observations, citing evidence and developing a logical argument for concepts, explaining phenomena in terms of concepts, and using concepts to solve problems.

Science

This alignment analysis used four levels of depth of knowledge (DOK). Because the highest (fourth) DOK level is rare or even absent in most standardized assessments, reviewers will in fact be making distinctions among DOK levels 1, 2, and 3. Please note that in science, "knowledge" can refer both to content knowledge and knowledge of science processes. This meaning of knowledge is consistent with the National Science Education Standards (NSES), whose first content standard is "Science as Inquiry."

Level 1. Recall and Reproduction
Level 1 is the recall of information such as a fact, a definition, a term, or a simple procedure, as well as performing a **simple** science process or procedure. Level 1 only requires students to demonstrate a rote response, use a well-known formula, follow a set procedure (like a recipe), or perform a clearly defined series of steps. A "simple" procedure is well defined and typically involves only **one step**. Verbs such as "identify," "recall," "recognize," "use," "calculate," and "measure" generally represent cognitive work at the recall and reproduction level. Simple word problems that can be directly translated into and solved by a formula are considered Level 1. Verbs such as "describe" and "explain" could be classified at different DOK levels, depending on the complexity of what is to be described and explained.

A student answering a Level 1 item either knows the answer or does not, that is, the answer does not need to be "figured out" or "solved." In other words, if the knowledge necessary to answer an item does not need to be further acted upon in order to reach the answer, the item is at Level 1. If the knowledge necessary to answer the item does not automatically provide the answer but needs to be acted upon to reach the answer, the item is at least at Level 2. Some examples that represent but do not constitute all of, Level 1 performance are:

- Recall or recognize a fact, term, or property.
- Represent in words or diagrams a scientific concept or relationship.
- Provide or recognize a standard scientific representation for simple phenomenon.
- Perform a routine procedure such as measuring length.

Level 2. Skills and Concepts
Level 2 includes the engagement of some mental processing beyond recalling or reproducing a response. The content knowledge or process involved is **more complex** than in Level 1. Items require students to make some decisions as to how to approach the question or problem. Keywords that generally distinguish a Level 2 item include "classify," "organize," "estimate," "make observations," "collect and display data," and "compare data." These actions imply **more than one step**. For example, to compare data requires first identifying characteristics of the objects or phenomenon and then grouping or ordering the objects. Level 2 activities include making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts. Some action verbs, such as "explain," "describe," or "interpret," could be classified at different DOK levels, depending on the complexity of the action. For example, interpreting information from a simple graph, an activity that requires reading information from the graph, is a Level 2. An item that requires interpretation from a complex graph, such as making decisions regarding features of the graph that need to be considered and how information from the graph can be aggregated, is at Level 3. Some examples that represent, but do not constitute all of, Level 2 performance, are:

- Specify and explain the relationship between facts, terms, properties, or variables.
- Describe and explain examples and non-examples of science concepts.
- Select a procedure according to specified criteria and perform it.
- Formulate a routine problem given data and conditions.
- Organize, represent, and interpret data.

Level 3. Strategic Thinking
Level 3 requires a higher level of thinking than the previous two levels. It requires reasoning, planning, and using evidence. The cognitive demands at Level 3 are **complex and abstract**. The complexity does not result simply because that there could be multiple answers (a possibility for both Levels 1 and 2), but because the multistep task requires **more demanding reasoning**. In most instances, requiring students to explain their thinking is at Level 3; requiring a very simple explanation or a word or two should be at

Level 2. An activity that has more than one possible answer and requires students to justify the response they give would most likely be a Level 3. Experimental designs in Level 3 typically involve more than one dependent variable. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve non-routine problems. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Identify research questions and design investigations for a scientific problem.
- Solve non-routine problems.
- Develop a scientific model for a complex situation.
- Form conclusions from experimental data.

D.10. Reading Descriptors for Depth of Knowledge

(based on Karen Wixson Descriptors, 1999, NAEP Achievement Levels for Grades 4 and 8, 2002)

Level 1

Level 1 requires students to receive or recite facts or to use simple skills or abilities to make sense of text. Level 1 tasks require students to comprehend text at an overall level, including making relatively obvious connections across parts of texts. Standards and items at this level require only a literal or superficial understanding of text and often ask students simply to locate information in text. Students' thinking remains at the text level, without analysis or connections beyond the printed word. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Support ideas by reference to details in the text.
- Use a dictionary to find the meaning of words.
- Identify figurative language in a reading passage.

Level 2

Level 2 requires students to engage in some mental processing beyond recalling or locating and reproducing what is presented in the text; it requires both comprehension and subsequent processing of the text as a whole or making some connections across portions of text or to one's own knowledge and experience. Inferences require thinking beyond the sentence level. Students can identify some important concepts, such as theme, but not in a deep or nuanced way. Standards and items at this level may include words such as "summarize," "interpret," "infer," "classify," "organize," "collect," "display," "compare," and "determine" whether fact or opinion. Literal main ideas are stressed. A Level 2 assessment item may require students to apply some of the skills and concepts that are covered in Level 1 but to think about the text in a somewhat more sophisticated way. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Use context cues to identify the meaning of unfamiliar words.
- Predict a logical outcome based on information in a reading selection.
- Identify and summarize the major events in a narrative.
- Determine whether something is fact or opinion.

Level 3

Deep knowledge becomes more of a focus at Level 3. Using their understanding of the text as a base, students are encouraged to go beyond the actual words and ideas an author has explicitly stated. Students may be encouraged to explain, generalize, or connect ideas that appear in multiple parts of a text and even to draw upon their knowledge of other texts. Standards and items at Level 3 involve reasoning and planning. Students must be able to support their thinking. Items may involve abstract theme identification, inference across an entire passage, or students' application of prior knowledge. Items may also involve more superficial connections between texts. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Determine the author's purpose and describe how it affects the interpretation of a reading selection.
- Summarize information from multiple sources to address a specific topic.
- Analyze and describe the characteristics of various types of literature.

A final reference point for reading is the comparison of NAEP and state reading tests. State reading tests differ widely in passage length and complexity (Linn et al., 1999) and often have less rigorous selection criteria than NAEP. As part of our understanding of cognitive demand, we will compare the NAEP specifications for passage selection (length and type) and the balance of multiple-choice and open-ended items and the specifications for the grades 4 and 8 reading tests taken by students in our target school districts. Analysis of publicly released NAEP items and state reading test passages included on state Websites will provide an estimate of the reading challenges that students encounter on NAEP reading and on their state reading tests. For example, NAEP passages for grade 4 can be as long as 1,000 words and an item set may include 10 or more items. If the average number of words in passages on a state test for grade 3 is only 400 words and if passages are accompanied by no more than five items, we will assume that the cognitive demands students encounter on NAEP and the state assessment differ considerably. This assumed difference will be factored into determining the cognitive demand expected at the state level.

APPENDIX E

CASE STUDY METHODOLOGY PROTOCOL

Instructional Case Study Methodology

The purpose of the site visits was to gain a more nuanced understanding of the achievement patterns on NAEP performance from 2003 to 2007. Fur selected TUDA districts each received a three-day visit from an expert team composed of the Council of Great City Schools' director of research and director of academic achievement and two to three other team members with instructional leadership experience in urban school districts. The Council's executive director participated in three of the four visits.

Prior to the visit, each district received a detailed letter outlining a schedule of interviews and an extensive list of materials to be gathered for team review. A phone conversation was held to clarify the precise list of interviewees, as well as documents for the instructional program, professional development, and strategic plans for the study period. District staff members were encouraged to explore their archives to find many of the documents. Two days of interview sessions ranging from 30 minutes to an hour were scheduled for a variety of past and present district leaders, central office staff, school site principals and teachers, coaches, and community members. The team examined the district's broad instructional strategies; materials; core reading, science and math programs; assessment programs; and professional development efforts to improve student achievement for that time period. It also reviewed district priorities and analyzed how the strategies and programs of the school system reflected those priorities.

The team used a protocol that contained 10 categories from the *Foundations for Success* research[1] that compared characteristics of urban districts that were making faster student achievement gains with similar districts making slower achievement gains. Since 2003, the Council of Great City Schools has used this research to guide interviews and document reviews for its strategic support teams. Member districts invite these teams of practitioners and council staff to address specific district concerns. The team begins interviews with opening questions, forming the basis for a general conversation around each of the 10 areas. An expert advisory panel reviewed the protocol and suggested modifications in keeping with the goals of this research project.

On site, after a brief introduction, the team framed questions around the following categories: political preconditions (context), goals, accountability, curriculum and instruction; professional development; teacher quality and principal capacity; quality of implementation, assessments and data; addressing low-performing students and schools, early childhood and elementary programs, and grades six through eight. This appendix contains a complete copy of the protocol.

The bulleted statements listed under each category are indicators that the Council has observed in other districts that have made faster gains on state tests. These indicators are not to be considered questions in themselves; rather, they prompt more generic questions from which one might infer the degree to which these indicators may be present in the district. The team also recognized that members needed to stay open to other indicators that arose in discussions. The indicators on the protocol were not exclusive.

Based on the participant's response to an opening or general question, the team would raise more specific questions, such as:

---

[1] Source: J. Snipes, F. Dolittle, and C. Herlihy. *Foundations for Success: Case Studies of How Urban School Systems Improve Student Achievement,* MDRC for the Council of the Great City Schools, 2002.

- Why was this developed?
- How was it developed?
- Describe the implementation process.
- How many schools/teachers/students were involved in the implementation?
- How was the level of implementation measured?
- How was progress monitored?
- How was success measured?
- Were there any modifications based on data?
- Is it still in place? To what degree? If it is no longer used, how was the decision made to stop?

Following two days of interviews and of reviewing documents, the team met for one day to review findings and discuss themes that emerged from interviews and document review. These findings were summarized in case studies for each district and incorporated into this final report.

Site Visit Protocol

Brief Introduction

Introduce the purpose, time period of the study, and the members of the team. Ask interviewees to introduce themselves and describe their role with the district.

Unless otherwise noted, questions are asked of board members, superintendent, central-office staff, community leaders involved in the schools, parents, teachers, principals, union president, and all groups interviewed.

Political Preconditions

Opening Questions

1. What particular context should we understand about the district during the period of this study?

2. What was the culture like during the period of study? (Board/superintendent relationships, clear vision, board members' relationships with each other, community support, etc.)

3. What factors led to the district's decision to participate in the TUDA project taking the NAEP test?

4. How were NAEP results reported to and in the press? (Board, superintendent, central office, teacher union, parents) How does that compare to how other measures of student achievement are treated?

5. What has been the reaction to district performance on NAEP?

Listen for:

- Specificity and clarity about what they wanted to do.
    - Vision, theory of action grounded in achievement
    - Sense of urgency
    - Level of focus on academics, school safety
    - Level of community support/ community concerns
    - Board updates on academic reforms, progress, and issues

- Coherence and alignment of what we hear to their vision and theory of action, unity of board and superintendent
    - Board and superintendent roles distinct but aligned
    - Systems in place for dealing with issues or changes in demographics
    - Financial systems in place sufficient to support academics and reforms

- Use of persuasion by superintendent and board

- Use of power and policy by superintendent and board

- Stability (cohesive board, superintendent longevity, nurtured vision over time)

- Clear theory of action for the reform and improvement of the school district.
- Vision for district improvement and reform was grounded in instruction and achievement.
- Vision was sustained and nurtured over an extended period.
- District leadership and administration had a sense of urgency for improving student achievement.
- School board was generally cohesive in pursuit of the vision and was able to sustain its direction.
- School board hired a superintendent who shared the same vision and worked in collaboration with the administration to see that it progressed.
- School board and the administration had mutually supportive but distinct roles in the district's improvement.
- District school board was focused on policy and did not micromanage administrative issues.
- School board meetings contained regular updates or status reports on district academic reforms and progress.
- District had operating and financial systems that were effective enough to support the instructional program.
- District was able to identify sufficient resources to seed improvement efforts and pursued external funding that was tied to reform goals.
- District did not experience unusual turnover rates in its leadership.
- District leadership was familiar with the reforms and strategies in other cities and what worked and did not work.
- District had a strategic plan for improving academic performance, based on a careful review of student and district needs.
- The community generally had a clear understanding of the district's strategic plan and vision for reform and supported it.

Goals

Opening Questions:

1. Please describe any district goals and objectives that went beyond state standards during the period of this study. (Or, what are your district goals? How have they changed since 2003-07?)

2. How did individual schools set their goals?

Listen for:

- Connection of vision and concrete, measureable, time-specific goals
- Link between school and district academic goals
- Clear goals for subpopulations, cultural diversity, focus on student learning
- Same understanding of goals throughout the organization
- School improvement plans explicitly state the school and district goals
- Goals stretch beyond NCLB requirements

- [ ] The district had translated its broad vision into a set of academic attainment goals that were measurable, concrete, and time specific.
- [ ] The district's academic goals existed at both the overall system level and the school level
- [ ] Individual improvement goals at the school level "rolled up" to a set of districtwide improvements.
- [ ] District and school-by-school goals included subgroup objectives that were clear and measurable.
- [ ] School district staff were familiar with and in accord with district and school goals.
- [ ] District and school goals were contained explicitly in school improvement plans.
- [ ] District and school goals contained "stretch" goals beyond those existing solely in state reading, math, and science test scores.
- [ ] District and school goals did not reflect "safe harbor" or other minimal objectives designed solely to keep schools out of sanctions.

Accountability

Opening Questions:

1. In what ways were you accountable for student achievement? (central office and school staff)

2. What was the impact of NCLB in your district? Reconstitution of schools, principal reassignment, etc.

3. What was the district's accountability system? What measures did it use? What, if any, relationship did it have to NAEP? (superintendent, principals, teachers, senior staff)

4. Was there any form of rewards or sanctions for administrators, teachers, or students? If these existed, what percentage received them?

Listen for:

- Who was held accountable for student achievement and attaining district goals
- What measures were used, and to what extent were those measures reflected in the evaluation process
- What rewards and sanctions existed, and to what extent were they used
- How central-office leadership was viewed. How principals and teachers were viewed.
- How progress was reported to the public

☐ District had a mechanism to hold its staff accountable for making progress on the goals and objectives.

☐ The school board's reporting to the community reflected progress on goals.

☐ The superintendent was held explicitly responsible for progress on the goals and was evaluated accordingly.

☐ Senior central-office staff were also held responsible, either through performance contracts or personnel evaluations, for district progress toward the academic goals.

☐ Principals were evaluated to a meaningful degree although not solely, on the progress their schools had made on the academic goals.

☐ Superintendent could remove principals for lack of performance.

☐ Teachers, either individually or as a group, were evaluated to a meaningful degree although not solely, on the progress their schools and classrooms had made on the academic goals.

☐ Schools viewed central office as leading and supporting reforms rather than focusing on compliance.

Curriculum and Instruction

Opening Questions:

1. How did your district's curriculum documents inform teachers about district expectations for student learning? (central office)

2. How did you know what students were expected to learn and the level of mastery they were to achieve in reading, math, and science? (principals, teachers)

3. What was the design for curriculum documents? How were connections made to the state curriculum expectations?

4. How did teachers know what was nonnegotiable and what they were free to modify?

5. Describe the reading, math, and science programs and textbooks in place in the district at that time.

6. How were textbooks selected?

7. To what extent was NAEP considered in the writing of the curriculum? To what degree, if any, did the district's curriculum/instructional practices reflect NAEP? (central office)

8. Where do you think your curriculum was closely matched to NAEP? (central office)

9. What process was in place to establish and monitor the level of rigor for instruction? (central office, administrative offices, principals) How was the fidelity to a program monitored or measured?

10. How did your curriculum link to special populations—including English language learners, special education, gifted and talented?

11. How did you know that children were learning what they are supposed to know at each grade level and course?

12. What kinds of interventions were in place for students who were performing below grade level?

Listen for:

- Alignment of written, taught, and tested curricula
- Presence, quality, and use of pacing guides or other curriculum documents
- Support for the classroom use of curriculum and textbook/support materials
- Coherence and clarity of the curriculum as a guide to classroom teaching and learning

☐ State academic standards in core subjects were clear and specific and could guide the development of curriculum by grade in the districts and schools.

☐ The district had a curriculum that adequately translated the state's standards into an explicit guide for what students are to be taught by grade.

- The curriculum clearly defined the knowledge and skills that students would be taught and how they would be expected to demonstrate that knowledge or skill.
- Curriculum documents or guides were explicit enough to tell teachers what level of depth or rigor the content was to be taught and how it was to be paced across the year.
- Teachers were using the district's curriculum and programs appropriately and were teaching at a level that would build student comprehension.
- District made it clear what level of mastery students were to acquire by the end of the school year in core courses.
- District had a uniform program in reading, math, and science at the lower elementary grades or used an overarching curricular framework for its instructional system.
- Materials, whether purchased commercially or developed, were explicitly aligned to the curriculum, state standards, and assessments.
- Gaps between the state standards and the curriculum were explicitly identified for teachers and filled with supplemental materials.
- Materials were up-to-date for that time period and reflected the best research on effectiveness.
- District had a specified and adequate time each day for reading, math, and science instruction.
- District curriculum, programs, and supplemental materials were aligned and sequenced to build comprehension skills, vocabulary acquisition, and literacy skills as students approached the mid- and late elementary grades
- District had an explicit pacing system to ensure that teachers covered skills before they were assessed.
- District had articulated a clear set of tiered intervention tools and procedures for when students were falling behind and a clear set of guidelines to indicate when interventions were to be used.

Professional Development, Teacher Quality, Principal Capacity

Opening Questions:

1. How much time was allotted for professional development?
2. How was professional development content determined?
3. How was professional development in the content areas delivered?
4. If the district used internal or external coaches, what were their responsibilities, and how did the district coordinate and monitor their work?
5. How was the district's professional development infused with the elements of NAEP?
6. How did professional development address the academic needs of English language learners and other special populations?
7. What kind of professional development did central office and principals receive about curriculum and instructional leadership?
8. What types of professional development took place at the school level? Who set the agenda? How were participation and success monitored?
9. How was the success of professional development evaluated?
10. What kind of induction programs existed for new teachers, experienced teachers new to the district, and new principals?
11. What was the district's teacher mobility rate?
12. How did the district deploy teachers and principals in schools?

Listen for:

- Extent of focus on academic content and how students learned
- Duration of a particular area of professional development focus
- Collective participation of principals and teachers
- Coherence of the program; differentiation for various audiences
- Type of activities (presentation lecture vs. active learning, use of student work)
- Sense of who owns professional development
- Impact of the professional development on actual classroom practice and student achievement

- District’s program of professional development was coherent and explicitly tied to the curriculum and state standards.
- The district had adequate time in the calendar or daily schedule in order to conduct necessary professional development.
- Participation in professional development was required when it was tied to implementation of the curriculum.
- Principals and academic coaches participated in the same professional development as teachers but had their own professional development to strengthen their particular roles.
- Professional development was differentiated by teacher experience, prior professional development, content area, and grade.
- Professional development was tailored explicitly to skill levels of students and where they needed to be strengthened.
- Professional development was ongoing and followed by technical assistance.
- Professional development and participation in it was centrally tracked by teacher.
- Professional development was evaluated for how well it was implemented in the classroom and what its impact was on student achievement.
- Professional development at the district level had a feedback loop by which teachers were able to critique the training they had received.
- The professional development offered in the district was explicitly differentiated between what schools provided and what the school system itself provided.
- Policies and strategies were in place to create ongoing professional learning communities among teachers.
- Teachers were provided structured time to collaborate for such activities as planning lessons, analyzing data, observing other classrooms, and/or analyzing student work.
- District had an adequate and timely recruitment strategy.
- District had a new teacher mentoring and induction program.
- District had a mechanism to ensure an equitable distribution of quality teachers.
- Teacher mobility rate was not inordinately large.

Reform Press/ Quality of Implementation

Opening Questions:

1. What were the district's key reform strategies at the elementary and middle school levels?

2. What did the district perceive as key to a successful implementation?

3. How frequently were classrooms visited and by whom? What did they look for? How were observations used?

4. How did the district ensure that its reforms were implemented in classrooms?

5. If the district had instructional coaches, how were they selected and trained? What were their roles, and how was their success monitored and evaluated? (unless answered in professional development section)

6. How did involvement in NAEP affect the district's reform efforts?

Listen for:

- Coherence of the program(s)
- Alignment of resources
- Sense of urgency
- Level of quality
- Transparency of data and its use in instruction
- Program leadership at central and school level

- [ ] District had a clear way to ensure that its reforms were reflected in the classrooms and were not solely seen at the central-office level.
- [ ] The district had uniform or well-coordinated "walkthrough" forms and procedures.
- [ ] Classroom monitoring and walk through procedures were not used mainly for personnel evaluation purposes but to strengthen classroom instructional practice.
- [ ] Results of classroom walkthroughs were aggregated at the school level and used during common planning time or professional learning community sessions.
- [ ] Principals were held accountable for conducting and using walkthroughs.
- [ ] Academic coaches were trained to provide technical assistance, model teaching, and instructional support and not to serve as substitute teachers or in other auxiliary functions.

Assessments and Data

Opening Questions

1. Describe the district's assessment system during the period of this study. What did the district measure?

2. Who were end users of data? How were data accessed?

3. What kind of training did people receive in the use of data?

4. How were data used to track the progress of the general student population and subgroups during the school year?

5. What changes were made in curriculum/instruction as a result of NAEP scores?

6. What kinds of assessment were done in schools and classrooms?

Listen for:

- Sense of urgency around student achievement and student success
- Focus on student learning
- Level of expectations for students
- Coherency of expectations across layers of district staff and community
- Other emerging themes

□ District regularly assessed student knowledge and skills over the course of the school year to ensure students are on track.

□ Interim assessments were explicitly linked to state and/or other assessments on which the district was gauged.

□ Interim assessments demonstrated predictive validity with state assessments.

□ State assessment and interim assessment data were returned to schools and teachers in a timely and useful manner.

□ School staff and teachers were provided appropriate and ongoing professional development on the interpretation, analysis, and use of assessment date in order to make necessary instructional decisions.

□ District results were used to decide where and how to alter curriculum, shape professional development, and target interventions.

□ Data from student assessments and other sources were used at central office, school, and classroom levels to improve professional development and strengthen the implementation and placement of instructional interventions.

- District disaggregated interim and end-of-year data by school and subgroup.
- District collected and used an array of data on student performance to inform instruction and make systemwide adjustments beyond those needed to demonstrate accountability on state and federal sanction systems.

Low-performing Schools and Low-performing Students

Opening Questions:

1. How was curriculum/instruction for low-performing students designed to attain higher student gains?

2. How were teachers assigned at low-performing schools?

3. What types of programs were put into place at the lower performing schools? How were they like or different from programs in all of the other schools?

4. How were low-performing students exposed to NAEP standards?

Listen for:

- Defined strategy to improve lowest-performing schools

- Clear interventions to identify and address low student performance

- Systems to reinforce positive student behavior

- Teacher and principal placement in areas of greatest need

- [ ] District had a clear strategy for addressing the academic performance of its lowest-performing schools and students.

- [ ] Teachers were clear about the tiered interventions they were to use and how to use them if and when there were signs that students were not keeping pace.

- [ ] Teachers also had clear enrichment strategies in place.

- [ ] District's strategy for improving its lowest-performing schools was measurably different than the strategy the district used in other schools.

- [ ] District had a strategy for reconstituting and supporting low-performing schools.

- [ ] District's extended time programs were clearly articulated with regular-day instruction and gaps in it.

- [ ] District had a definable and developmental positive behavior support program for all students.

- [ ] District had an explicit process and strategy for differentiating classroom instruction for low-performing students, English language learners, and students with disabilities.

- [ ] Instructional programs for English language learners had adequate time devoted to English acquisition and may have used native language skills to build content knowledge.

- School improvement plans were developed to improve student achievement, based on a careful review of data. The plans were seriously reviewed for potential effectiveness and were monitored.

- District had incentives for the best teachers to work in its lowest-performing and hardest-to-staff schools.

- District provided additional resources to its lowest-performing schools.

Early Childhood Education and Elementary Schools

Opening Questions:

1. Describe the district's early childhood program during the period of this study, starting from 2002-03. (Follow with additional questions regarding focus on academics, outreach to community programs, alignment with kindergarten and first grade, etc.)
2. What changed at the elementary school level as a result of NAEP? (new programs, new professional development, monitoring, etc.)
3. How do you account for NAEP gains (losses) at the elementary school level?

Listen for:

- Cohesiveness of reforms
- Academic focus and vertical alignment of early childhood programs
- Outreach of the district to enroll students from the community
- Longitudinal progress monitoring by program types
- Screening systems for gifted programs

- [ ] District had a clear sequence of reforms starting in the elementary grades and working up.
- [ ] District's early childhood program served a substantial number of children who were eventually served in the district's kindergarten and first grades.
- [ ] District's early childhood program had a definable literacy and cognitive development component that was aligned to the kindergarten and first grade curriculum.
- [ ] District was able to track the academic progress of early childhood pupils through the early elementary grades to determine the longer-range impact of the early childhood program.
- [ ] District had a definable gifted/talented program and screened everyone for participation using an appropriate measure.

Grades 6-8

Opening Questions:

1. What was the district's improvement strategy at the secondary level?
2. What changed at the middle school level as a result of NAEP?
3. How do you account for NAEP gains (declines) at the eighth-grade level?

Look for:

- Middle school course alignment to advanced courses in high school
- Intervention programs for students who performed below grade level
- Monitoring student achievement progress

- ☐ District explicitly aligned its secondary curriculum in core courses to college entry requirements or better.
- ☐ District had a strategy (e.g., double-blocking) to assist students who had arrived in secondary school a year or more behind academically.
- ☐ District high schools had AP and other similar courses available in all high schools, with a focus on middle school classes that would prepare students for advanced high school courses.
- ☐ District could track and act on student course-taking patterns, grades, and absences to prevent dropouts and encourage more rigorous course work and greater attendance

Final Question

Since the period of study, what changes has the district undergone that are likely to have an impact on NAEP?

Addendum A

Data We Need Prior to Site Visit

- Select schools based on disaggregated state achievement data
- District gains compared to state gains on NAEP composite and subscales

Documents to be Reviewed

Documents listed are to be from the period of study: 2002-03 and 2006-07

a. Organization structure (org chart) for academics during the period of study and members of academic departments serving on the Superintendent's Cabinet.
b. Copy of the district's strategic plan
c. Copy of the evaluation of the district's strategic plan from that time period
d. Information about the district's choice plan, if applicable
e. Board agendas and minutes from three 2002-03 and three 2006-07 board meetings
f. Description of process used to evaluate principals during that time period, with appropriate forms
g. Description of process used to evaluate teachers during that time period, with appropriate forms.
h. District vision of teaching and learning during the time period
i. An annotated list of school-level reform projects that were in place
j. Annual state report for district achievement 2003-04, 2005-06, and 2007-08
k. Copy of any instructional study of the district during that time period, if available
l. Samples of communicating district progress on goals to the public during that time period
m. Copies of a sample of the district's grades 3-5 and 7-8 language arts, science, and math curriculum guides, with pacing guides (previously received)
n. Samples of (short cycle) tests in those grade levels and content areas, if they existed during that time period
o. Description literacy instructional approach and names of textbooks/programs/interventions at pre-kindergarten through grade 8 during that time period
p. District approach to the teaching of writing during that period
q. Description of mathematics instructional approach and names of textbooks/programs/interventions at pre-kindergarten through grade 8 during that time period
r. Description of science instructional approach, time allocation, and names of textbooks/programs/interventions at pre-kindergarten through grade 8 during that time period
s. How lab materials were provided in elementary schools
t. Copies of the district's professional development plans from that time period
u. A description of how the district supported low-performing schools and students during that time period
v. Number and percentages of students participating in the district's special education programs, per school by race/ethnicity (if available)
w. Number and percentages of students participating in the district's gifted and talented programs, per school with racial/ethnic, English language learner, and gender data
x. Number and percentages of students participating in the district's bilingual or English language learner programs, per school with racial/ethnic and gender data

y. A description of the philosophy and time requirements of the district's programs for English language learners
z. Course pass rates in grade 9 mathematics, English, and science
aa. High school graduation requirements compared to state graduation requirements during the study period
bb. List of high schools and the AP courses offered at each (indicator of college-bound focus), distribution of AP courses and participation rates
cc. Number of AP tests taken and exam grades earned by school and district and subgroup (if available)

Addendum B

Persons to be interviewed (preferably working in the district during the period of study)
It is possible to have both the current person in charge and a person from that time period at the same interview session.

a) Supervisor of Curriculum and Instruction – 60 minutes
b) Two or three board members – 45 minutes
c) Person in charge of curriculum – 60 minutes
d) Person in charge of professional development – 60 minutes
e) Person in charge of gifted and talented – 45 minutes
f) Person in charge of language arts/literacy – 45 minutes
g) Person in charge of mathematics – 45 minutes
h) Person in charge of science – 45 minutes
i) Assistant Superintendents for the regions – 60 minutes
j) Person in charge of research, testing, & evaluation – 45 minutes
k) Persons in charge early childhood and special education – 45 minutes
l) Person in charge of English language learners – 45 minutes
m) President of teachers' union – 45 minutes
n) Person in charge of NCLB, legislation, state & federal projects – 45 minutes
o) Three elementary, and three middle school instructional coaches – 45 minutes
p) Representatives of any external group/organizations that work closely with the district e.g., university, community-based organization, business organization, religious leaders, etc. – 60 minutes
q) Parent representatives (total of seven to nine) from **local** parent-teacher school associations selected from a mix of schools (list would be included)* – 45 minutes
r) Focus group of eight principals from the schools listed*– 60 minutes
s) Focus group of 12 teachers: (Please select four math, five reading/language arts, and three science **classroom** teachers from the schools listed.* Do not select support staff or coaches.) – 60 minutes
t) Person in charge of HR to discuss teacher placement and turnover during the period of study

*The team will use data to select schools that are high-, medium-, and low-performing on state tests in 2006-07.

APPENDIX F

ATLANTA CASE STUDY

Introduction -- District Context

Atlanta, Georgia, is one of the largest city school districts in the United States and the largest city school district in Georgia. Atlanta Public Schools (APS) serves approximately 49,000 students in 103 schools. Roughly 98 percent of these students are students of color, and 76 percent qualify for the National School Lunch Program.

APS has shown significant and consistent growth in performance on NAEP—especially in reading—between 2003 and 2009.[1] In fourth-grade reading, the percentage of students performing at or above proficient increased 10 percentage points, from 14 percent to 24 percent. In eighth-grade reading, the percentage of students at or above proficient increased 6 percentage points, from 11 percent to 17 percent.

Performance in math also increased between 2003 and 2009. In fourth-grade math, the percentage of students performing at or above proficient increased 8 percentage points from 13 percent to 21 percent. In eighth-grade math, the percentage of students at or above proficient increased 5 percentage points from 6 percent to 11 percent.

To further explore the political and instructional context in which these achievement gains occurred, this chapter is divided into three sections.

1. The first section, Setting the Stage for Reform, examines the broader, strategic foundations of reform in the district, focusing on the role played by district leadership and the district's approach to setting goals and holding people accountable for progress.

2. The second section, Key Policies/Strategies in Implementing Reform, details the tactical decisions made by the district in the areas of curriculum and instruction, teacher quality and professional development, support for program implementation, and the use of data and assessments.

3. The third section delves more deeply into the district's NAEP achievement results and trends.

I. Setting the Stage for Reform

Leadership and Reform Vision

The two main factors that appear to have made reform possible in Atlanta Public Schools were the leadership of a strong new superintendent and school board, and the sustained support of and attention to a clear plan of action for raising student achievement across the board.

---

[1] 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.

After many years of frustration with low student achievement, a groundswell of community support for improvement and reform in Atlanta's public schools led to the election of a new school board in 1999, as well as to important changes in the board charter and governance policies. In the process of electing a new school board, the city's business community played an important role in training board candidates, offering prospective board members a professional development course in educational leadership, board ethics, and policy.

One of the first steps the city and new school board took was to select an experienced superintendent, Beverly Hall, who came to the city steeped in the reforms of other major urban school districts. She, too, had a vision of how the board should be reorganized to best support the work of the district, and as part of contract negotiations, she changed the fragmented, committee-specific operating structure to a "committee of the whole" structure, wherein all functions are overseen by the entire school board. This structure of governance proved successful in reining in what one interviewee referred to as “the contentious, micro-managing tendencies of smaller, function-specific committees.” During this time the board evolved into a unified body committed to the larger mission of supporting sound district governance and instructional reform.

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.

The new superintendent's central focus, however, was on instruction, and she brought with her a clear vision and a plan for districtwide improvement based on establishing goals that reflected high expectations for student achievement, not just minimum standards for meeting Adequate Yearly Progress (AYP). The district aspired to compete at the national level—one reason behind its decision to volunteer for the NAEP Trial Urban District Assessment. Moreover, the district worked hard to sustain its commitment to this vision for reform and its implementation throughout the jurisdiction. Despite initial pushback from teachers who disliked the systematic approach of the reading program, the district pressed forward with the implementation of its literacy reforms and gained and sustained teacher support over a number of years.

Along the way, the district developed and strengthened what the site-visit team found to be an extremely strong and deep cadre of school leaders and central office staff members with considerable expertise in instructional programming, including Kathy Augustine, the deputy superintendent for instruction, Robin Hall, director of reading, Dottie Whitlow, director of mathematics, and others. These key staff leaders formed the core of the instructional team that the superintendent used to implement and drive reforms.

Accountability

To support its high expectations for student growth and its ambitious reform agenda, APS enacted a two-tiered goal system aimed not only at increasing the number of students reaching proficiency, but also at driving improvements across the achievement spectrum. These achievement goals and standards of performance were clear, measurable, and communicated consistently throughout the district. Each school had specific achievement targets calculated by the district and based on a formula tied to districtwide goals of improvement.

In fact, Atlanta had one of the most explicit accountability systems observed by the site visit team. These measures drove performance evaluations from the district leadership level down to the school and teacher levels. The superintendent and all district senior staff, including executive directors of the regional School Reform Teams (SRTs), are on performance contracts tied to the attainment of these districtwide academic targets, and the school board receives quarterly reports on district initiatives and progress toward these goals. Principals, meanwhile, are on performance contracts wherein 30 percent of their evaluation is based on progress toward school targets. Bonuses to staff and schools are also based on the attainment of goals, and through growth and performance, schools earn greater freedom and latitude over such school governance decisions as the hiring (and removal) of teachers.

The transparency of these goals helped create widespread buy-in for the new accountability structure, as well as a culture of ownership for student achievement at all levels of the organization.

II. Key Policies/Strategies in Implementing Reform

Curriculum and Instruction

The introduction of a standards-based core curriculum in reading and math was one of the first steps in Atlanta's reforms. Early in her tenure, the new superintendent initiated a series of school audits to identify effective practices and instructional problem areas school by school. The audit cited as the main problem areas (1) a widespread lack of teacher training, (2) inconsistencies in instruction both between and within schools, (3) a failure to sufficiently serve all student groups, and (4) a lack of direct, research-based instruction in reading across the curriculum. Based on these audits, the district provided schools with comprehensive feedback on the specific steps they needed to take to improve instruction. They also identified literacy as the cornerstone of their reform efforts, basing their strategy for improving student achievement on establishing and sustaining the use of common, research-based literacy practices. From that point, however, schools were given wide latitude to choose among a list of district approved Comprehensive School Reform Models (CSRM), as long as they consistently met their site-specific growth targets.[2]

To create coherence in the district's instructional program amid these various models, the district worked to ensure that all instruction in the district was aligned to the newly introduced Georgia Performance Standards (GPS). In fact, senior staff members attribute achievement gains in the district more to this move toward standards-based instruction than to any one given instructional reform model. The district identified any gaps that existed between instructional models and state standards, working with publishers to tailor or supplement programs to best meet both district and state learning objectives. Comprehensive curriculum and framework documents were created and disseminated to unpack the Georgia Performance Standards and guide instruction, although the element of districtwide pacing took a little longer to fully articulate.

In line with the district's focus on reading as the first step to instructional reform, the central office clearly laid out research-based strategies for how literacy would be taught throughout the system. They enlisted the Consortium on Reading Excellence (CORE) to help define and drive high quality standards and literacy practices, and they used professional development, assessment, and monitoring strategies to ensure even implementation, irrespective of the unique programs within each school. Around this time the district also began to emphasize writing and

[2] CSRDs used in Atlanta included Success for All, Modern Red School House, CONNECT, Project Grad, Direct Instruction, Middle Schools that Matter, High Schools that Work, and other programs.

the development of content area literacy skills. By most accounts, this focus yielded benefits for student performance across the curriculum.

Math reforms, on the other hand, lagged behind literacy reforms in Atlanta by a number of years. The district began to phase in math standards in the 2005-06 school year, adopting a districtwide math program, "Move it Math," in all but five schools. To bolster the new program, the district developed scope and sequence documents by unit and grade level, as well as instructional guides. The district also provided strong math coaching support and extensive professional development widely credited with improving math instruction and addressing student performance in key areas, including number sense, operations, and problem solving.

Professional Development and Teacher Quality

By most accounts, both the quality and intensity of professional development and instructional capacity-building in APS increased during the study period. District- and school-level staff reported that the district was very clear about what quality instruction should look like, and what was expected of teachers and administrators. Atlanta based 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 throughout the district. Built around national standards of literacy development, CORE training was conducted over a three-year period and focused on the how and why of literacy instruction. Moreover, CORE training focused on narrative and informational texts, as well as strategies including questioning, graphic organizers, and student progress monitoring. 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 NAEP data analysis, some of the largest reading gains in Atlanta came on subscales that were a strong focus of CORE training, particularly reading for information.

In addition, APS developed a two-step process for assuring effective teaching. The first step involved laying out a set of "26 expectations for effective teaching" that were reflected in teacher evaluations and monitored through regular school walkthroughs. The second step laid out the process that principals were to use to determine if expectations were being met and to provide accurate and meaningful feedback. Extensive professional development and the hands-on support of School Reform Teams helped build staff capacity to meet these expectations, and each SRT developed internal trainers to provide professional development to schools that picked specialized programs. These trainers themselves received extensive training before they were allowed into the schools to work with teachers. Coaches were also on hand to monitor instruction and assist teachers with data interpretation, working an extended school year of 220 days. These coaches were available to model lessons and co-teach classes, and they pushed teachers to tailor instruction to student needs and to begin asking higher-level questions.

This focus on building capacity and on providing clear guidance regarding what is expected of teachers and administrators allowed APS to recruit, develop, and retain excellent staff over the years. In addition to offering a regionally competitive teacher pay scale, the district's site-based support structure offers teachers an opportunity for professional growth and advancement. Teachers who are involved in training can become model teachers and even go on to become principals. The district has also made an effort to support and encourage specialized training in key instructional areas, providing endorsements for English as a Second Language (ESL) teachers and gifted and talented teachers, as well as in reading.

Support for Implementation

Many urban school districts will adopt an instructional program and abandon it after a year or so when it fails to get immediate results. In contrast, Atlanta adopted a practice of staying with its programs over extended periods and supporting, refining, and augmenting them as the data dictated rather than replacing them. The district also strategically deployed its staff to support its instructional programming at the school and classroom levels in ways that one does not often see. This led to consistency of program development and implementation districtwide.

For example, in 2000-2001 APS developed a network of five School Reform Teams (SRTs), which served about seven to fourteen schools each and were headed by executive directors, to support schools in their efforts to meet performance targets. This organizational structure, which was based in regions throughout the city, was unique in that it moved a majority of district-level staff out of the central office and created a school-based, "direct service model" of support in a process the district termed "Flipping the Script." Each SRT has tapped about fourteen exemplary teachers, designated "teacher leaders," to work with the SRTs and help support and train their peers. The SRTs are also staffed with human resource generalists, maintenance liaisons, curriculum and instructional experts, and an English language learner (ELL) and special education (SPED) representatives to build capacity within schools for supporting all students.

In addition to reinforcing teachers in the classroom with cross-functional experts who could provide comprehensive feedback on instructional needs and strengths, this support structure promoted close collaboration between SRTs and school staff around district academic goals, and helped shift the district's focus from compliance to shared responsibility for the implementation and success of instructional reforms. SRTs and executive directors are held explicitly accountable for progress in their assigned schools. This structure was also designed to transform the role of principals from building managers to instructional leaders, now responsible for improving student achievement.

Low-performing schools were given even more support and direction from the district. These schools were mandated to implement the program Project Grad and to support struggling students. Teachers stayed after school for 60 to 90 minutes every Wednesday to conduct tutoring. There was also an increased emphasis on differentiated instruction in professional development for teachers and staff in these schools.

The district's school-based support network also included coaches, model reading teacher leaders (MRTLs), both language arts *and* reading coordinators, and instructional liaison specialist (ILS) teams trained by the district to support school staff with instruction and with the use of student data. Interestingly, these various layers of support personnel did not appear to interfere with each other, functioning instead as a seamless support network. In fact, having multiple lines of communication between school sites and the central office helped the district refine and maintain strong support for the implementation of instructional programming and reforms. The superintendent and deputy continue to meet regularly with staff and principals in schools, and as one district-level staff member said, "When something is good, we try to plan ways to support and sustain it."

Data and Assessments

The use of data and assessments was another key initiative in Atlanta. During the study period, the district changed its focus from simply the provision of instruction and services to the continuous assessment of student performance. Central office staff reported that "we are no longer a surprise district," and it was apparent in talking to district- and school-level staff that student data were used to both gauge progress and chart the organization’s direction.

APS employed both benchmark (formative) and summative assessments as part of this assessment system, and developed a protocol (Fishbone) to allow for interpretation and use of test results and data. Schools administered unit and end-of-course tests, as well as benchmark tests in September and January that were aligned to the district's instructional program and mirrored state assessments to diagnose student needs and assess progress. These benchmark tests were also infused with NAEP-like items in order to increase the overall level of rigor.

At the district-level, quadrant analysis of these data was used to identify schools with low performance or lack of growth in reading, math, and science; to target resources; and to refine and revise the curriculum based on student- and school-specific needs. The district's analysis of data also extended to item-level analysis, including the identification of “distracters" and student weaknesses by subject and topic area. Conversations with school-level staff also revealed a strong familiarity with the use of data to inform instruction and identify students’ academic strengths and weaknesses. School leaders were charged with constantly reinforcing the use of data, and teachers reported using data to group and regroup students and to differentiate instruction based on the needs of their students.

To further support this growth of a data culture in the district, there was extensive and strategic training in the use of data at all levels, and the district developed well-defined protocols to help with interpretation and use of test results and data.

III. NAEP Results and Trends

This section of Atlanta’s profile examines student performance on the National Assessment of Educational Progress (NAEP) in grades 4 and 8. Data are analyzed by comparing Atlanta’s scale scores over time, (2003 compared to 2009) and comparing Atlanta’s 2009 scale scores to student performance in large cities and in national public schools. (See tables F.1 through F.4.)

Reading Grades 4 and 8

In Atlanta, 2009 reading scales scores compared to 2003 scores

- Fourth graders made significant double-digit gains on their reading composite score and on both reading subscales—reading for literacy experience and reading to gain information.

- African American fourth graders significantly increased their reading scale score.

- Eighth graders made significant double-digit gains on their composite scale score and on one reading subscale, reading to gain information.

- Eighth grade African American and National School Lunch Program (NSLP)-eligible students showed significant increases.

Atlanta’s 2009 reading scale scores compared to students in large cities/national public schools

- Atlanta’s fourth- and eighth-grade White students achieved significantly higher reading scale scores than White students in large cities.

- Atlanta's fourth- and eighth-grade White students earned significantly higher scale scores than White students in national public schools

.

Mathematics Grades 4 and 8

In Atlanta, 2009 mathematics scale scores compared to 2003 scores

- Fourth and eighth graders in Atlanta made significant gains on their mathematics composite score and, on four of the five subscales—algebra, geometry, measurement, and number.

- In both fourth and eighth grades, African American and NSLP-eligible students achieved significantly higher composite scale scores.

Atlanta's 2009 mathematics scale scores compared to students in large cities/national public schools

- Fourth-grade White students in Atlanta achieved significantly higher scale scores than fourth-grade White students in large cities.

- Fourth-grade White students in Atlanta achieved significantly higher scale scores than fourth-grade White students in national public schools.

Table F.1 Average scale score of grade 4 Atlanta Public School students in 2003-2009 NAEP reading assessment, overall, by subscale and by selected characteristics, compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Reading Composite					
Atlanta	197	201	207	209**	12***
Georgia	214	214	219	218	4***
Large City	204	206	208	210**	6***
National Public	216	217	220	220*	3***
Reading for Literary Experience Scale					
Atlanta	201	204	210	212**	12***
Georgia	217	217	220	220	2
Large City	208	209	211	212**	4***
National Public	220	220	221	221*	1***
Reading for Information Scale					
Atlanta	192	197	204	206**	14***
Georgia	209	211	217	216	7***
Large City	200	202	205	207**	8***
National Public	213	214	217	218*	5***
African American Students (composite)					
Atlanta	191	194	200	201	11***
Georgia	199	199	205	204	6***
Large City	193	196	199	201**	8***
National Public	197	199	203	204*	7***
White Students (composite)					
Atlanta	250	253	253	253*,**	4
Georgia	226	226	230	229	3
Large City	226	228	231	233**	7***
National Public	227	228	230	229*	2***
Hispanic Students (composite)					
Atlanta	‡	‡	‡	‡	‡
Georgia	201	203	212	208	8
Large City	197	198	199	202**	4***
National Public	199	201	204	204*	5***
Asian/Pacific Islander Students (composite)					
Atlanta	‡	‡	‡	‡	‡
Georgia	233	243	232	238	5
Large City	223	223	228	228**	5
National Public	225	227	231	234*	10***
National School Lunch Program-Eligible Students (composite)					
Atlanta	189	191	198	199**	10***
Georgia	200	201	207	207	7***
Large City	196	198	200	202**	6***
National Public	201	203	205	206*	5***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

Table F.2 Average scale score of grade 4 Atlanta Public School students in 2003-2009 NAEP mathematics assessment, overall, by subscale, and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Mathematics Composite					
Atlanta	216	221	224	225*,**	10***
Georgia	230	234	235	236**	6***
Large City	224	228	230	231**	7***
National Public	234	237	239	239*	5***
Algebra Scale					
Atlanta	223	229	231	234**	11***
Georgia	236	238	239	241**	5***
Large City	231	235	236	237**	6***
National Public	240	243	244	244*	4***
Data Analysis, Statistics, and Probability Scale					
Atlanta	220	228	229	224*,**	5
Georgia	232	237	239	235**	3***
Large City	227	231	233	233**	6***
National Public	237	241	243	242*	5***
Geometry					
Atlanta	216	221	227	229**	12***
Georgia	229	232	234	238	10***
Large City	225	227	230	232**	7***
National Public	233	236	238	239*	6***
Measurement Scale					
Atlanta	209	215	216	221*,**	12***
Georgia	229	231	233	232**	3
Large City	220	225	226	228**	8***
National Public	233	236	238	238*	5***
Number Scale					
Atlanta	215	219	223	224*,**	8***
Georgia	229	233	235	236	7***
Large City	222	226	228	230**	8***
National Public	232	235	237	237*	6***
African American (composite)					
Atlanta	211	215	217	218**	7***
Georgia	217	221	222	221	4***
Large City	212	217	219	219**	7***
National Public	216	220	222	222*	6***
White Students (composite)					
Atlanta	258	263	266	266*,**	9
Georgia	241	243	246	247	6***
Large City	243	247	249	250**	8***
National Public	243	246	248	248*	5***
Hispanic Students (composite)					
Atlanta	‡	‡	223	222	‡
Georgia	219	229	229	231**	12***
Large City	219	223	224	226	7***
National Public	221	225	227	227	6***

Asian/Pacific Islander Students (composite)					
Atlanta	‡	‡	‡	‡	‡
Georgia	248	255	255	256	8
Large City	246	247	251	253	8
National Public	246	251	254	255	9***
National School Lunch Program-Eligible Students (composite)					
Atlanta	209	213	216	216*,**	6***
Georgia	219	224	224	225**	7***
Large City	217	221	223	225**	8***
National Public	222	225	227	228*	6***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

Table F.3 Average scale score of grade 8 Atlanta Public School students in 2003-2009 NAEP reading assessment, overall, by subscale and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Reading Composite					
Atlanta	240	240	245	250**	10***
Georgia	258	257	259	260	3
Large City	249	250	250	252**	4***
National Public	261	260	261	262*	1***
Reading for Literary Experience Scale					
Atlanta	239	240	243	246*,**	7
Georgia	256	256	258	258**	2
Large City	249	250	249	251**	3***
National Public	260	260	260	261*	1
Reading to Perform a Task Scale					
Atlanta	238	239	245	—	
Georgia	258	258	259	—	
Large City	245	248	247	—	
National Public	261	260	260	—	
Reading for Information Scale					
Atlanta	241	240	247	253**	12***
Georgia	260	258	259	263	3
Large City	250	252	252	254**	3***
National Public	262	261	262	264*	2***
African American Students (composite)					
Atlanta	237	237	242	246	9***
Georgia	244	241	246	249**	5***
Large City	241	240	240	243**	2***
National Public	244	242	244	245*	2***
White Students (composite)					
Atlanta	‡	‡	‡	292*,**	‡
Georgia	268	268	271	268	0
Large City	268	270	271	272	4***
National Public	270	269	270	271	1***
Hispanic Students (composite)					
Atlanta	‡	‡	‡	‡	‡
Georgia	245	247	250	254	9
Large City	241	243	243	245**	4***
National Public	244	245	246	248*	4***
Asian/Pacific Islander Students (composite)					
Atlanta	‡	‡	‡	‡	‡
Georgia	265	275	‡	286**	22***
Large City	260	266	263	268**	8***
National Public	268	270	269	273*	5***

National School Lunch Program-Eligible Students (composite)					
Atlanta	235	234	240	244**	8***
Georgia	243	243	247	249	6***
Large City	241	243	242	244**	3***
National Public	246	247	247	249*	3***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

Table F.4 Average scale score of grade 8 Atlanta Public School students in 2003-2009 NAEP mathematics assessment, overall, by subscale, and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Mathematics Composite					
Atlanta	244	245	256	259*,**	15***
Georgia	270	272	275	278**	8***
Large City	262	265	269	271**	9***
National Public	276	278	280	282*	6***
Algebra Scale					
Atlanta	252	255	262	267*,**	16***
Georgia	274	277	280	286	12***
Large City	266	270	274	276**	11***
National Public	279	281	284	286*	8***
Data Analysis, Statistics, and Probability Scale					
Atlanta	248	244	260	260**	12
Georgia	272	275	277	279**	7***
Large City	263	266	270	270**	8***
National Public	279	280	283	283*	5***
Geometry Scale					
Atlanta	243	244	254	261*,**	17***
Georgia	267	270	270	274**	7***
Large City	261	263	268	270**	9***
National Public	274	275	277	279*	5***
Measurement Scale					
Atlanta	225	228	247	245*,**	21***
Georgia	261	265	268	269**	8***
Large City	254	258	261	266**	12***
National Public	274	274	276	278*	5***
Number Properties Scale					
Atlanta	247	245	255	256*,**	9***
Georgia	271	271	274	274**	3
Large City	263	264	266	269**	6***
National Public	276	276	278	279*	3***
African American Students (composite)					
Atlanta	241	242	253	255**	14***
Georgia	250	255	261	262	12***
Large City	247	250	254	256**	9***
National Public	252	254	259	260*	9***
White Students (composite)					
Atlanta	298	‡	‡	‡	‡
Georgia	284	284	288	289	6***
Large City	285	288	292	294	8***
National Public	287	288	290	292	5***

Hispanic Students (composite)					
Atlanta	‡	‡	‡	‡	‡
Georgia	262	258	266	270	8
Large City	256	258	261	264	9***
National Public	258	261	264	266	8***
Asian/Pacific Islander Students (composite)					
Atlanta	‡	‡	‡	‡	‡
Georgia	286	301	‡	300	14
Large City	281	289	291	299	17***
National Public	289	294	296	300	11***
National School Lunch Program-Eligible Students (composite)					
Atlanta	239	240	251	253*,**	14***
Georgia	253	257	262	265	12***
Large City	252	256	260	262**	10***
National Public	258	261	265	266*	8***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

APPENDIX G
BOSTON CASE STUDY

Introduction -- District Context

The oldest public school system in America, Boston Public Schools (BPS) serves a diverse population of more than 56,000 pre-kindergarten through grade 12 students in 140 schools. Eighty-five (85) percent of these students are children of color, almost 20 percent are identified as English language learners, and 71 percent qualify for the National School Lunch Program.

Once the epicenter of bitter race politics and struggles over governance, Boston has recently gained nationwide recognition for educational reform and strong student achievement. A Broad Prize winner in 2006, Boston has also shown notable gains on the National Assessment of Educational Progress (NAEP), particularly in mathematics. From 2003 to 2009, the percentage of fourth-grade students scoring at or above proficient on NAEP climbed from 12 percent to 30 percent. In the eighth grade, the percentage students scoring at or above proficient on the NAEP math test climbed from 18 percent to 32 percent.

Boston has also made gains in reading, although to a lesser extent. The proficiency rate among fourth-grade students increased from 15 percent of students scoring at or above proficient in 2003 to 24 percent in 2009. In the eighth grade, the percentage of students scoring at or above proficient in reading increased only slightly from 22 percent in 2003 to 23 percent in 2009.

To further explore the political and instructional context in which these achievement gains occurred, this chapter is divided into four sections.

1. The first section, Setting the Stage for Reform, examines the broader, strategic foundations of reform in the district, focusing on the role played by district leadership and the district's approach to setting goals and holding people accountable for progress.

2. The second section, Key Policies/Strategies in Implementing Reform in Mathematics Instruction, details the tactical decisions made by the district in the areas of curriculum and instruction, teacher quality and professional development, support for program implementation, and the use of data and assessments.

3. The third section, A Study in Contrasts: Reading/Literacy Reforms in Boston, compares and contrasts Boston’s math initiatives with those in reading/language arts.

4. The fourth section delves more deeply into the district’s NAEP achievement results and trends.

I. Setting the Stage for Reform

Leadership and Reform Vision

Sustained development and broad collaboration around clearly articulated district goals, coupled with strong and stable leadership focused on student achievement, appear to have made reform possible in Boston Public Schools (BPS).

BPS benefited from the consensus and support of a strong, mayor-appointed school board led by a board president (Elizabeth Reilinger and now Gregory Groover) who had strong working relations with the former and current superintendents—Tom Payzant and Carol Johnson, respectively. The board used its mandate for improvement to spearhead a comprehensive school improvement plan in 1996 that focused on student achievement in math and literacy and advanced data-driven, standards-based instructional practice. Developed with significant input from parents, teachers, administrators and community partners, these reforms had the strong support of the city's mayor and benefited from the collaborative relationship between the school board and the superintendent. In fact, much of the original plan remained intact, though with substantial enhancements in reading, under the leadership of the next superintendent, Carol Johnson. No doubt, the leadership of the district was also spurred by state action in 1998 to require students to pass the Massachusetts exams in order to graduate.

The strategic hiring and placement of instructional leaders in key roles at the district level also helped to drive reforms. Most notably, the district hired experts skilled at building partnerships and overseeing instructional reform, including a former principal—Sid Smith—to lead curriculum and instruction and a strong math leader, Linda Davenport, to oversee the strategic rollout of the new math program, paying particular attention to the management of change in the implementation process. By most accounts, this leadership team was open to and eager for change and innovation, and staff members at all levels were unified and passionate about improving student achievement.

Yet beyond the consensus on the need to improve student achievement, there was a sustained commitment to supporting the initiatives adopted, particularly in mathematics. Although literacy was the district's first area of focus for reform, a lack of a common, coherent instructional philosophy and curriculum seems to have stunted the district's efforts in this area. The math program benefited from this experience, and pursued a very different course in its reforms. An Urban Systemic Grant from the National Science Foundation (NSF) focused mainly on math and helped jump-start the district's overhaul of its math program. This 2001 grant was aligned with district goals and was strategically employed to help the district focus on professional development and implement a common, cohesive K-8 mathematics curriculum.

The new math program was met with considerable resistance from schools initially, but the school board resisted efforts to change course and abandon the new math program. Instead, the district pursued a well-planned, thoughtful process for (1) rolling out math reform, (2) engaging and communicating with schools and the community regarding the strategic plan, and (3) building broad-based ownership for the success of the city's public schools. Based on interviews with school level staff, parents, and community members, it was clear to the study team that the district worked hard to communicate the instructional goals of its math program throughout the organization and community. District leadership developed a clear list of specific steps and initiatives to be pursued and used whole school improvement plans to articulate goals and priorities.

Accountability

Accountability for results in Boston during this period was defined more by mutual ownership of results than by a traditional system of data-driven accountability. Although the Office of Research set school performance targets for both student performance and growth toward proficiency, personnel evaluations in Boston were not tied to student scores per se, except, in part, for the superintendent's evaluation. But the review and analysis of student performance data from state assessments reportedly drove 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—a practice that may have contributed to Boston's math gains among all subgroups and across all quintiles.

Accountability was also pursued through district monitoring of schools, as well as through a network of school-based coaches. As the link between the central office and the school site, coaches put subtle pressure on principals and teachers to implement district programs and reforms to increase student achievement, as well as offering support and guidance in addressing areas of student and teacher weakness.

II. Key Policies/Strategies in Implementing Reform in Mathematics Instruction

Curriculum and Instruction

BPS began the process of improving the district's instructional program in math by "seeking out the warm pockets of reform in the district" according to one interviewee-- surveying schools to figure out what was working, talking to key staff and administrators, and incorporating what they were learning into a new district math plan. During this time, "learning site schools" were identified, and the district worked to provide opportunities for school visits to observe exemplary sites.

Based on this careful study of the programs and strategies employed by the most successful schools, the district's leadership decided to adopt TERC's (Technical Education Research Centers) *Investigations* as the districtwide math curriculum at the elementary level and the Connected Mathematics Program (CMP) at the secondary level. They also emphasized the importance of providing teachers with professional development and support to implement this curriculum, as well as the potential value of using formative assessments and professional learning communities to gauge progress and reinforce instructional priorities.

The decision to adopt a cohesive, districtwide math program was based on BPS's stated priorities to move students beyond memorizing math procedures toward a deeper conceptual understanding of the material. Importantly, district leadership and the math department approached implementation of the program as a gradual, multistep process which allowed a stronger, more thoughtful phase-in period. The program was first piloted at selected campuses to build model sites. These pilot schools were asked to name Math Leadership Teams of three to six teachers and principals that would start learning and adapting the curriculum, as well as aligning it to various school-based supports and professional development opportunities. Eventually, the numbers of teachers on each of these teams in each building were expanded 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 math 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 in their schools 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 by the Spring of 2001, the program was expanded to all remaining schools, and all teachers were using the program and participating in professional development on the program's implementation, including ELL and special education teachers.

The district also worked to ensure close alignment between this curriculum and state standards and frameworks and developed formative assessments to help gauge both implementation and student progress toward these standards. The district strengthened the program with supplemental materials, including additional instruction in math language, scope and sequence pacing guides, and daily calendars to ensure instructional consistency, ten-minute math sections 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 ten additional minutes devoted to reviewing math facts and procedures.

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 math professional development for teachers.

In addition to the districtwide math plan, each school developed its own comprehensive, seven-year school math plan, which helped focus and then sustain the effort. The math department also strategically used built-in structures as "leverage points" for supporting the new program. For example, principals had discretion over one of the five planning periods allocated for teachers, and the district made sure that this planning period was shared by grade-level teams and used for reviewing the math curriculum and student data.

Professional Development and Teacher Quality

Another important component of the roll-out of math reforms in Boston was the district's approach to professional development. Built around implementation of the Investigations and Connected Math programs, there was a clear emphasis on using professional development to change how teachers and administrators approached math instruction, ensuring that teachers were teaching math concepts and not blindly following the textbook. Teachers received extensive professional development in math content as well as in the workshop model of pedagogy. Training included on-site workshops, meetings with grade-level teams, monthly professional development sessions with principals, and training around the use of data. Subject and topic-specific professional development in the pacing of classroom instruction was rolled out in advance of upcoming areas. In addition to 30 hours of mandated professional development, math teachers were required to take three topic-specific courses (24 hours each) in math over the five-year plan period, which they could take after school or over weekends or summers.

Every school had to develop a plan for professional development. Principals had some flexibility in this process but received considerable support and oversight. With this approach, the central office helped cultivate clarity regarding both what quality instruction looked like and the expectations the district set for teachers. Moreover, this multi-faceted approach to professional development in Boston was designed to augment the limited number of formal professional development days provided for in the collective bargaining agreement.

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.) This 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 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.

Another critical layer of this professional development was the extensive training (50 hrs) provided for administrators (Lenses on Learning), focusing on developing instructional leadership and expertise through enhanced math knowledge and knowledge of program implementation. The professional development for principals also covered the use of “learning walk” procedures.

In addition to training in math content and pedagogy, the district focused on putting structures into place to support teachers and principals and promote collaboration. The district developed lesson and unit planning templates that could be used during team planning sessions, and teachers and administrators were encouraged to constructively share the time slated for professional development, working in teams to preview upcoming lessons and units in math, to look at data, and to develop the ability to predict student needs and areas of confusion. The central office provided protocols, as well as funds, for teachers to participate in structured visits to colleagues' classrooms, creating opportunities to reflect on practice together. Principals were also encouraged to collaborate with one another, attending principal breakfasts and leadership seminars together.

The math department also built strong collaborative relationships with the English language learner (ELL) and special education (SPED) units to ensure shared, consistent professional development opportunities for all teachers. ELL and SPED teachers were involved in professional development related to the math curriculum, and general and special education teachers alike were provided professional development for differentiated instruction to meet SPED and ELL student needs. In addition, the SPED director identified the schools with gaps between general education and special education teacher resources and offered specialized training for the teachers at those sites.

Despite initial growing pains, this push to improve teachers' math knowledge and promote collaboration paid off in terms of building capacity at the school level and a sense of shared ownership in the new math program. By many accounts, it was through strong, content-based professional development that the math department was able to overcome resistance to the new math program. Teachers generally gave high marks to the district's professional development program, reporting that they appreciated the flexibility and the confidence it gave them when it came to math instruction.

Support for Implementation

The new math program in Boston was sustained by considerable guidance and oversight from the math department and district leadership. The district developed a series of “walkthroughs” or “learning walks” in 2002 and 2003 to track math program implementation and 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. While district staff reported that there certainly remained unevenness from school to school, they “worked hard to bring everyone on board.”

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 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. Most coaches came to the district 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.

This strategy of building buy-in through broad-based knowledge about the program even extended to the district's outreach efforts to parents. Because Investigations and Connected Mathematics present a concept-based approach to math instruction, the district designed math content seminars held at schools and libraries so that parents could understand the curriculum their children were learning in the classroom. Demand for these seminars was surprisingly strong, and parents reported that gaining this understanding of the "new math" helped them support their kids with the assignments they brought home.

Throughout the study period, these support structures and lines of communication helped the district make continuous adjustments to the math plan and refinements to the curriculum based on feedback from school sites. In approaching math reform as an iterative process, the district built capacity and ownership of the program within schools and, over time, a culture of math reform developed.

Data and Assessments

With the new reform initiative came a push to examine achievement data down to the school and teacher levels and to analyze what the data were showing about student performance and needs, as well as about the performance and needs of teachers. In an effort to reach this detailed understanding of student progress, the math department spearheaded the development and systematic implementation of formative assessments aligned with both state standards and the district's instructional program. These assessments used released items from the state test (not NAEP), which research staff indicated helped focus instructional strategies around results. Elementary- and secondary-level math directors worked together to oversee not only the development, but implementation of the assessments, distributing districtwide testing calendars and overseeing administration of the assessments, as well as collecting and disseminating the results.

At school sites, these data drove discussions among teachers and principals and helped guide instructional planning. Principals were pushed to become consumers of data, and several school- and district-level staff interviewed talked about the rise of the "data principal" during this time. Principals reported that the data helped them gain a clear picture of where they were at any given point and of how to target extra support and professional development to address areas of need at their site. These formative assessments in math were also reported by district staff as being critical in assessing curriculum implementation.

Furthermore, the district promoted the use of state assessment data by teachers and administrators to inform instruction, particularly in math. In the 2002-2003 school year, BPS rolled out the MYBPS online data system. To support and promote use of this new system, the Office of Research conducted data systems training and sent support staff to school sites to work with and train school staff. Importantly, this system was specifically designed for teacher use, giving them access to data on students in both in the current year and the previous year, as well as disaggregated MCAS results by group and item. The district understood that the system was doomed to failure unless it was easy for teachers to use. The district explicitly focused on teachers to promote the use of data and to inform and guide their classroom instruction.

III. A Study in Contrasts: Reading/Literacy Reforms in Boston

Prior to the launch of the math plan, Boston had already mounted reforms in the area of literacy instruction, reforms that took a different approach and appeared to have lacked the same focus and results. In an effort to implement a more authentic model for the teaching of literacy, the district moved from a basal reading textbook approach to a Reading/Writing Workshop model (RWW). In fact, it was noted that while the workshop model of pedagogy supplemented the math program, in literacy it *was* the program.

By most accounts implementation of this workshop model represented innovative thinking at the time, but it did not evolve to reflect the onset of the standards movement to the same extent that the math plan did. For example, there was no consistent set of learning goals and objectives in literacy, as was the case in math, where learning objectives were clearly laid out in pacing guides and tied to benchmark assessments. In fact, the district's literacy work was not even organized inside the curriculum unit for much of the study period.

The reading program (2002-2007) did use off-the-shelf assessment materials such as Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and Scholastic Reading Inventory (SRI). But these instruments, while they are diagnostic and can be used to show student growth over time, are not tied to the reading curriculum and cannot be used to predict progress toward meeting state standards as measured on the MCAS test. In fact, the district's first application for Reading First funding was denied because the reading/writing workshop model was deemed inconsistent with the requirements of the No Child Left Behind legislation.

Another critical distinction between the math and literacy reforms in Boston during the 2002-2007 period was that, while the district adopted one common core curriculum in math, the literacy program lacked uniformity and consistency across the district. It was noted in interviews that a "bifurcation," characterized by the existence of "two reading camps," contributed to a lack of successful implementation of the reading/writing workshop approach across the district. In 2004, the district began to phase in a basal reading program, Harcourt *Trophies,* in 12 of its

Reading First Schools. By 2006, a total of 34 schools were using Harcourt *Trophies.* Yet attempts at integrating Harcourt *Trophies* into the reading/writing workshop format were often unsuccessful, and most elementary schools (50+) using the reading/writing workshop model wrote their own curriculum. As a result, schools did not use the same instructional materials and students moving from one school to another within a school year were greatly disadvantaged.

To support the workshop model, the district provided considerable professional development in collaborative coaching and learning (CCL), and literacy specialists and instructional coaches were available in the schools to work with teachers. However, the focus of this support and training was on approach rather than content, emphasizing strategies designed to promote structured collaboration and the analysis of classroom practice. It was noted, however, that implementing and managing reading/writing workshop in the classroom was not a straightforward or easy proposition. One interviewee commented: "A well-taught reading/writing workshop is a thing of elegance; I do not believe that every teacher can become a good workshop teacher."

IV. NAEP Results and Trends

This section of Boston's profile examines student performance on the National Assessment of Educational Progress (NAEP) in grades 4 and 8. Data are analyzed by comparing Boston's scale scores over time – 2003 compared to 2009 and comparing Boston's 2009 scale scores to student performance in large cities and in national public schools. (See tables G.1 through G.4.)

Reading, Grades 4 and 8

In Boston, 2009 reading scale scores compared to 2003 scores

- Composite NAEP reading scale scores increased significantly
- Scale scores increased significantly on both reading subscales, *reading for literacy experience* and *reading to gain information.*
- Three student groups—African American, Hispanic, and National School Lunch Program (NSLP)-eligible—showed significant increases.

Boston's 2009 reading scale scores compared to students in large cities

- Both fourth and eighth graders in Boston earned significantly higher reading composite scale scores and higher scores on one reading subscale, *reading for literary experience,* than their peers in large cities.
- In both grades 4 and 8, Boston's Hispanic students and NSLP-eligible students scored significantly higher than those student groups in large cities.
- In grade 8, Boston's White students, scored significantly higher than their counterparts in large cities.

Boston's 2009 reading scale scores compared to students in national public schools

- Boston's African American, Hispanic, and NSLP-eligible students achieved significantly higher scale scores than their peers in national public schools.

Mathematics, grades4 and 8

In Boston, 2009 mathematics scores compared to 2003 scores

- Fourth and eighth graders in Boston had significant double-digit gains in their NAEP composite scale scores and for all mathematics subscales: algebra; data, analysis, statistics and probability; geometry; measurement; and number.

- Fourth and eighth graders in five student groups—African American, White, Hispanic, Asian, and NSLP-eligible—achieved significant double-digit scale score gains.

Boston's 2009 mathematics scale scores compared to students in large cities

- Fourth graders in Boston earned significantly higher scale scores in their NAEP composite scale scores and in two of the five mathematics subscales, geometry and measurement than their peers in large cities.

- Eighth graders in Boston had significantly higher composite scale scores and higher scale scores on all of the mathematics subscales--algebra; data analysis, statistics and probability; geometry; measurement; and number—than their peers in large cities.

- Eighth graders in five student groups—African American, White, Hispanic, Asian, and NSLP-eligible—scored significantly higher than their peers in large cities.

Boston's 2009 mathematics scale scores compared to students in national public schools

- Although significant gains have been made Boston's fourth- and eighth-grade students continue to score below national averages.

Table G.1 Average scale score of grade 4 Boston Public School students in 2003-2009 NAEP reading assessment, overall, by subscale and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Reading Composite					
Boston	206	207	210	215*,**	9***
Massachusetts	228	231	236	234**	6***
Large City	204	206	208	210**	6***
National Public	216	217	220	220*	3***
Reading for Literary Experience Scale					
Boston	210	211	213	219*	8***
Massachusetts	230	234	238	235**	5***
Large City	208	209	211	212**	4***
National Public	220	220	221	221*	1***
Reading for Information Scale					
Boston	201	203	206	211**	10***
Massachusetts	225	228	233	232**	8***
Large City	200	202	205	207**	8***
National Public	213	214	217	218*	5***
African American Students (composite)					
Boston	202	203	204	212*,**	10***
Massachusetts	207	211	211	216**	9***
Large City	193	196	199	201**	8***
National Public	197	199	203	204*	7***
White Students (composite)					
Boston	225	230	230	231	6
Massachusetts	234	237	241	241**	6***
Large City	226	228	231	233**	7***
National Public	227	228	230	229*	2***
Hispanic Students (composite)					
Boston	201	200	204	209*,**	8***
Massachusetts	202	203	209	211**	9***
Large City	197	198	199	202**	4***
National Public	199	201	204	204*	5***
Asian/Pacific Islander Students (composite)					
Boston	223	224	229	231	8
Massachusetts	229	234	241	241	12***
Large City	223	223	228	228**	5
National Public	225	227	231	234*	10***

National School Lunch Program-Eligible Students (composite)					
Boston	204	205	207	211*,**	8***
Massachusetts	210	211	214	215**	5***
Large City	196	198	200	202**	6***
National Public	201	203	205	206*	5***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

Table G.2 Average scale score of grade 4 Boston Public School students in 2003-2009 NAEP mathematics assessment, overall, by subscale and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Mathematics Composite					
Boston	220	229	233	236*,**	16***
Massachusetts	242	247	252	252**	11***
Large City	224	228	230	231**	7***
National Public	234	237	239	239*	5***
Algebra Scale					
Boston	227	234	236	237**	10***
Massachusetts	247	252	256	255**	9***
Large City	231	235	236	237**	6***
National Public	240	243	244	244*	4***
Data Analysis, Statistics, and Probability Scale					
Boston	222	234	233	234**	12***
Massachusetts	245	250	254	254**	9***
Large City	227	231	233	233**	6***
National Public	237	241	243	242*	5***
Geometry Scale					
Boston	221	229	233	240*	19***
Massachusetts	240	245	249	251**	11***
Large City	225	227	230	232**	7***
National Public	233	236	238	239*	6***
Measurement Scale					
Boston	216	230	230	234*	19***
Massachusetts	241	248	252	252**	11***
Large City	220	225	226	228**	8***
National Public	233	236	238	238*	5***
Number Scale					
Boston	218	226	233	236*	18***
Massachusetts	240	245	252	251**	11***
Large City	222	226	228	230**	8***
National Public	232	235	237	237*	6***
African American Students (composite)					
Boston	216	223	226	231*,**	15***
Massachusetts	222	228	232	236**	15***
Large City	212	217	219	219**	7***
National Public	216	220	222	222*	6***
White Students (composite)					
Boston	234	244	250	251	16***
Massachusetts	247	252	257	258**	11***
Large City	243	247	249	250**	8***
National Public	243	246	248	248*	5***

Hispanic Students (composite)					
Boston	215	225	230	232*,**	17***
Massachusetts	222	225	231	232**	10***
Large City	219	223	224	226	7***
National Public	221	225	227	227	6***
Asian/Pacific Islander Students (composite)					
Boston	243	256	255	260	18***
Massachusetts	248	258	259	264**	16***
Large City	246	247	251	253	8
National Public	246	251	254	255	9***
National School Lunch Program-Eligible Students (composite)					
Boston	218	227	231	233*,**	15***
Massachusetts	226	231	237	237**	11***
Large City	217	221	223	225**	8***
National Public	222	225	227	228*	6***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

Table G.3 Average scale score of grade 8 Boston Public School students in 2003-2009 NAEP reading assessment, overall, by subscale and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Reading Composite					
Boston	252	253	254	257*,**	5***
Massachusetts	273	274	273	274**	1
Large City	249	250	250	252**	4***
National Public	261	260	261	262*	1***
Reading for Literary Experience Scale					
Boston	254	252	252	257*,**	3
Massachusetts	271	272	271	272**	1
Large City	249	250	249	251**	3***
National Public	260	260	260	261*	1
Reading to Perform a Task Scale					
Boston	247	251	252	—	
Massachusetts	273	273	272	—	
Large City	245	248	247	—	
National Public	261	260	260	—	
Reading for Information Scale					
Boston	254	255	257	258*,**	5***
Massachusetts	274	276	276	276**	2
Large City	250	252	252	254**	3***
National Public	262	261	262	264*	2***
African American Students (composite)					
Boston	245	244	250	248	4
Massachusetts	252	253	253	251	0
Large City	241	240	240	243**	2***
National Public	244	242	244	245*	2***
White Students (composite)					
Boston	273	274	275	282*,**	9
Massachusetts	278	279	278	279**	1
Large City	268	270	271	272	4***
National Public	270	269	270	271	1***
Hispanic Students (composite)					
Boston	245	248	241	251*	7
Massachusetts	246	246	251	250	4
Large City	241	243	243	245**	4***
National Public	244	245	246	248*	4***
Asian/Pacific Islander Students (composite)					
Boston	274	280	275	276*	2
Massachusetts	281	282	281	281**	0
Large City	260	266	263	268**	8***
National Public	268	270	269	273*	5***

National School Lunch Program-Eligible Students (composite)					
Boston	247	247	249	251*	5***
Massachusetts	251	256	256	254**	4
Large City	241	243	242	244**	3***
National Public	246	247	247	249*	3***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

Table G.4 Average scale score of grade 8 Boston Public School students in 2003-2009 NAEP mathematics assessment, overall, by subscale and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Mathematics Composite					
Boston	262	270	276	279*	18***
Massachusetts	287	292	298	299**	12***
Large City	262	265	269	271**	9***
National Public	276	278	280	282*	6***
Algebra Scale					
Boston	264	273	281	282*	18***
Massachusetts	288	294	301	302**	14***
Large City	266	270	274	276**	11***
National Public	279	281	284	286*	8***
Data Analysis, Statistics, and Probability Scale					
Boston	263	273	278	284*	21***
Massachusetts	292	297	305	304**	12***
Large City	263	266	270	270**	8***
National Public	279	280	283	283*	5***
Geometry Scale					
Boston	262	268	274	275*	14***
Massachusetts	282	287	292	294**	12***
Large City	261	263	268	270**	9***
National Public	274	275	277	279*	5***
Measurement Scale					
Boston	256	267	272	279*	23***
Massachusetts	287	292	297	299**	12***
Large City	254	258	261	266**	12***
National Public	274	274	276	278*	5***
Number Properties Scale					
Boston	263	268	274	277*	14***
Massachusetts	286	288	294	296**	10***
Large City	263	264	266	269**	6***
National Public	276	276	278	279*	3***
African American Students (composite)					
Boston	251	256	263	268*,**	17***
Massachusetts	260	263	264	272**	12***
Large City	247	250	254	256**	9***
National Public	252	254	259	260*	9***
White Students (composite)					
Boston	289	299	305	311*,**	22***
Massachusetts	292	297	305	305**	13***
Large City	285	288	292	294	8***
National Public	287	288	290	292	5***

Hispanic Students (composite)					
Boston	252	261	270	269*	17***
Massachusetts	255	265	270	271	16***
Large City	256	258	261	264	9***
National Public	258	261	264	266	8***
Asian/Pacific Islander Students (composite)					
Boston	300	309	305	312*,**	12***
Massachusetts	304	314	315	314**	10
Large City	281	289	291	299	17***
National Public	289	294	296	300	11***
National School Lunch Program-Eligible Students (composite)					
Boston	256	264	271	273*,**	16***
Massachusetts	261	273	275	278**	17***
Large City	252	256	260	262**	10***
National Public	258	261	265	266*	8***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

APPENDIX H

CHARLOTTE-MECKLENBURG CASE STUDY

Introduction -- District Context

Charlotte Mecklenburg Public Schools (CMS) enrolls over 130,000 students and employs over 9,000 full-time teachers within 160 schools in the district. Forty-five percent of the student population are African American and 15 percent are Hispanic. Nearly 50 percent of students enrolled in Charlotte Mecklenburg Schools are eligible for the National School Lunch Program (NSLP), and 12 percent are English language learners. In addition, a little over 10 percent of the student population is identified as having a disability.

CMS shows consistent performance on NAEP assessments of reading and math, with scores at or above national averages from 2003 to 2009. In reading, the percentage of fourth-grade students performing at or above proficient level rose from 31 percent in 2003 to 36 percent in 2009; however in the eighth grade, student performance dropped from 30 percent in 2003 to 28 percent in 2009. In fourth-grade math, the percentage of students performing at or above proficient rose from 41 percent in 2003 to 45 percent in 2009; in eighth grade, the percentage was 32 percent in 2003 and 33 percent in 2009. At the same time, NAEP data show that the performance of Charlotte's fourth and eighth graders was either comparable to or significantly higher than that of students in North Carolina, large cities, and national public schools in both reading and math.

To further explore the political and instructional context for this consistently high level of achievement, this chapter is divided into three sections.

1. The first section, Setting the Stage for Reform, examines the broader, strategic foundations of reform in the district, focusing on the role played by district leadership and the district's approach to setting goals and holding people accountable for progress.

2. The second section, Key Policies/Strategies in Implementing Reform, details the tactical decisions made by the district in the areas of curriculum and instruction, teacher quality and professional development, support for program implementation, and the use of data and assessments.

3. The third section delves more deeply into the district's NAEP achievement results and trends.

I. Setting the Stage for Reform

Leadership and Reform Vision

Amidst heated political battles over school assignment after the landmark *Swann* case was overturned in 2000,[1] the leaders of CMS sought to redirect the focus of district staff, schools, and the community back to student achievement. The school board at that time was led by a long-serving president, Arthur Griffin, who worked with a series of strong superintendents-- Eric Smith, James Pughsley, Frances Haithcock, and Pete Gorman—to ensure support on the board to move forward on an aggressive instructional reform agenda, even when the board was not always unified on other issues.

[1] In 2000, the *Swann v Charlotte-Mecklenburg Board of Education* desegregation case ruling was overturned, ending the use of busing to ensure racial balance in CMS schools. In the fall of 2002, the city adopted the "School Choice Plan," which divided the city into four large attendance zones based on neighborhoods. As many neighborhoods are predominantly white or predominantly African American, opponents pointed out that this new assignment policy essentially reinstated racial segregation in the school system.

This shared commitment to instructional improvement allowed APS to develop and pursue a strong reform vision and clear, measurable objectives for systemwide improvement. These reform efforts began by surveying the landscape of academic programs in place throughout the district, as well as studying the reforms of other urban school districts and borrowing strategically from these other models, modifying programs and approaches to fit their own culture and needs.

The district decided to replace its site-based management approach 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 support and oversight for its lowest-performing schools, mandating the implementation of prescriptive reading and math programs and offering incentives for teachers and staff to move to these sites to ensure the highest quality of education was provided to their students. At the same time, the district was careful to implement programs that met the needs of students along the continuum of achievement. This meant pressing schools to set high standards and ensuring that students were placed in academically rigorous courses, including Advanced Placement (AP) courses when specific test score thresholds were attained. In fact, district-level staff reported that enrollment in AP courses was as carefully monitored as the identification of struggling students for intervention programs.

To support these reforms, district leadership systematically selected central office staff they felt were committed to student achievement and had a record of success in ensuring that the right people were in the right places to advance the district's reform agenda. According to one central office staff leader, Atlanta's approach to reform was guided by the core belief that “people more than programs made the difference.” The district took proactive steps to build a culture of collective responsibility and collaboration among these staff members, housing instructional departments near each other to facilitate cross-functional planning. In fact, CMS's desire to promote shared accountability and change how the district was viewed led to their willingness to volunteer for the NAEP Trial Urban District Assessment (TUDA), a decision that brought transparency to their work and allowed them to compare their progress to that of other large urban districts and to public schools across the nation. Their results showed that the district sustained high achievement across the board at every level of achievement.

Accountability

At the heart of all district initiatives and policies was a strong and explicit accountability system. In the early 1990s, Charlotte became one of the nation’s early leaders and innovators in the standards movement, implementing "balanced scorecards" as a management tool to outline school- and department-specific goals aligned to systemwide goals. As one of the first districts to implement this system, these balanced score cards were strategically used to track activities and results based on a multi-year plan, with explicit assignment of responsibilities and detailed action plans. This system helped define district expectations of central office staff, administrators, and teachers, and to create a roadmap to achieving specific, measurable goals related to student achievement. Moreover, the district employed the acronym SMART (Stretching, Measurable, Aspiring, Rigorous, and with a Timeline) as a guide for setting these goals, further clarifying the message that everything that staff did should be measurable, systematic, and ultimately aimed at upholding high academic standards for all students.

Although the balanced scorecard 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 throughout the district. Senior staff met with the superintendent on a regular basis to monitor progress, and these conversations revolved around student data. It was noted in interviews with district staff that the superintendent was not impressed by the number of professional development sessions held, but rather by whether they had an impact on student achievement. The district was also among the first to establish locally developed quarterly exams and mini-assessments to track student progress through the year, and the central office was charged with monitoring these scores and providing specific support on site through Rapid Response Teams—teams that were deployed to schools that were falling behind in order to address instructional weaknesses identified in the data. The presence of these Rapid Response 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. Rapid Response Teams also conducted audits of schools that measured school progress toward district-defined objectives.

Charlotte also built community and student feedback into their accountability system. Family and student surveys focused on gauging not only academic progress but also progress toward the district's goal of providing a safe and orderly learning environment and community collaboration.

II. Key Policies/Strategies in Implementing Reform

Curriculum and Instruction

During the study period CMS designed and successfully implemented a comprehensive literacy plan for the teaching of reading and writing throughout the district. The core curriculum was based mainly on the North Carolina Standard Course of study and the *Open Court* reading program and was supplemented with a strong writing component, 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 employ basal texts and supplemental materials designed to meet the full range of students' literacy needs. Moreover, benchmark assessments closely tied to this program helped monitor student progress and identify areas where students needed additional support. Despite objections from teachers who disliked the prescriptive, systematic approach of *Open Court*, the district pressed forward with its implementation of the program and sustained this support over a number of years, asserting that the curriculum was well aligned with state standards and well equipped to advance reading and literacy development.

In fact, throughout CMS, literacy was considered the cornerstone for improvement in other content areas. As one district-level staff member pointed out, there was a strong belief that, "as we increase reading skills, we increase thinking." Within this framework of "literacy first," other core content areas were merged with the district's literacy plan. In order to embed reading within the math curriculum, for example, there was a strong emphasis on building academic math vocabulary, which included the use of interactive, changing "word walls" and the assignment of "math journals," where students were expected to demonstrate their understanding of math by writing down the steps involved in working through various math problems. This reading- and writing-enriched approach to math instruction was widely believed to have helped improve math problem solving and increased math scores on both the state assessment and NAEP.

CMS also implemented an innovative universal early childhood program to build the foundation for later academic achievement: *Bright Beginnings* and *More at Four* for four-year-olds. Both of these programs fostered early language and literacy and development and school-readiness, assessing students' performance by asking the questions, "Who is learning and how do we know?" Aside from its strong focus on student achievement, *Bright Beginnings* also encouraged parental involvement through its parent literacy programs and parent education programs. In fact, *Bright Beginnings* was so popular that it was sold commercially. Moreover, targeted and intensive interventions were made available for all elementary school students needing additional support, and literacy facilitators were on site to provide assistance and support for teachers and principals.

Professional Development and Teacher Quality

CMS's professional development efforts to recruit, develop, and retain high-quality staff was closely aligned with the district's goal to build school and district capacity to support a rigorous program of managed instruction. Central office staff, principals, and teachers were carefully selected by district leadership and the human resources department. The district sought to attract effective teachers to the district by offering competitive monetary packages based on experience and certification, as well as bonuses for working in low performing schools. At the same time, to maintain the goals of each school and help principals cultivate a school culture, principals were given the power to interview and select their own staff and teachers.

The district developed and mandated professional development and training for these teachers, including a week of professional development prior to every school year. This professional development was defined around student assessment results and district instructional priorities. In fact, the rollout of the district's reading initiative was infused with intensive initial training, and ongoing site-based support for instruction continued to focus on literacy and writing. The ELL and special education departments were also included in these professional development efforts to maintain alignment and coherence in reading initiatives.

Professional development courses generally 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.

Moreover, 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 survey 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.

Support for Implementation

From the beginning the district set clear, non-negotiable expectations to ensure fidelity to the new districtwide instructional program. These expectations included the mandatory use of the adopted reading textbooks and curriculum, the use of pacing calendars, and the administration of quarterly short-cycle assessments.

To help build the capacity of schools to meet these expectations and successfully implement and sustain reforms, the district 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 were 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. The district also established 90 minute common planning periods designed to help teachers and staff identify and focus in depth on areas of student academic needs.

In addition, CMS 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. The key idea was to build capacity at the campus level.

These literacy and academic facilitators also provided a critical line of communication 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.

In addition to the staff and support provided to all schools, CMS provided intensive support to struggling school sites through Rapid Response Teams. 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. Moreover, this focus on interventions in low-performing schools may have contributed to the relatively high performance of various student groups on NAEP assessments. At the same time, the special education and English language learner departments were intimately involved in the instructional planning process, and the district's overall culture of collaboration and shared accountability ensured that the district's instructional program was accessible and designed to advance achievement among all students.

Data and Assessments

CMS relied heavily on the use of student data and assessments to measure progress toward districtwide goals. The district conducted three types of assessments throughout the school year—state assessments, quarterly assessments, and mini-assessments. The mini-assessments were based on focused lessons, created by teachers with the help of literacy and academic facilitators, to address areas of weaknesses and assess skill mastery after teaching. Quarterly or benchmark assessments were aligned with state standards and the district's curriculum, and included questions similar to the state assessments, as well as constructed response items similar to those found on NAEP.

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 emphasis on these regular assessments of student progress helped create a culture of data use throughout the district. The district set clear expectations for all staff using the balanced scorecard system, and it was clear that, in order to meet both individual and systemwide objectives, every central office member, principal, and teacher needed to continually review his/her data and use it to make well-informed decisions about instruction and planning.

Principals and academic facilitators, for example, used data to help target support and professional development to ensure that their teachers were equipped to meet student needs. Common planning periods 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 addition to serving as diagnostic measures of student learning, benchmark assessments were used as a monitoring system for the central office. As described earlier, Rapid Response Teams were deployed to schools whose benchmark data revealed that they were falling behind. CMS also examined disaggregated data to target support to low-performing and high-poverty schools. For example, the percentage of students eligible for the National School Lunch Program was used to justify lowering the student-teacher ratio in schools with higher poverty levels. Through intensive programs like Equity Plus and A-Plus, these schools received funding to reduce class sizes and provide classroom training and support, as well as incentives to enhance and stabilize the teaching staff.

In addition, district staff used data to measure equity in the quality of teachers and resources across schools in the district. To promote equal access to higher-level and AP courses, the district identified candidates for accelerated classes—particularly among low-income students and students of color—by examining PSAT scores. The district also regularly conducted program evaluations and expanded or eliminated programs based on student results.

Aside from the consistent use of data by CMS staff, student data and test results were also clearly communicated to parents and the community. These stakeholders were well informed on the district's overall performance, and parents were aware of the performance on state assessments of both their child and their child's school.

III. NAEP Results and Trends

This section of Charlotte's profile examines student performance on the National Assessment of Education Progress (NAEP) in grades 4 and 8. Data are analyzed by comparing Charlotte's scale scores over time—2003 compared to 2009—and comparing Charlotte's 2009 scale scores to student performance in large cities and in national public schools. (See tables H.1 through H.4.)

Reading, Grades 4 and 8

In Charlotte, 2009 reading scale scores compared to 2003 scores

- Fourth graders made significant gains in their NAEP composite scale score.
- Fourth graders made significant gains in the *reading to gain information* subscale.
- Fourth graders *eligible for the National School Lunch Program (NSLP) showed a* significant increase on their NAEP reading composite scale scores.

Charlotte's 2009 reading scale scores compared to students in large cities

- Charlotte's fourth and eighth graders had significantly higher NAEP reading composite scale scores than students in large cities.
- Scores for Charlotte's fourth and eighth graders were significantly higher than those of students in large cities for each reading subscale: *Reading for literary experience, reading to gain information and reading to perform a task* (grade 8 only).
- Charlotte's Hispanic, African American, and NSLP-eligible *students*, at both grades 4 and 8, also significantly outscored their counterparts in large cities.

Charlotte's 2009 reading scale scores compared to students in national public schools

- Charlotte's fourth and eighth graders had significantly higher NAEP composite scale scores than students in national public schools.
- Charlotte's fourth and eighth graders scored significantly higher than their national public school peers in all of the reading subscales: *Reading for literary experience, reading to gain information and reading to perform a task* (grade 8 only).

- African American fourth and eighth graders in Charlotte had significantly higher NAEP scale scores than their national public school peers.

Mathematics, Grades 4 and 8

In Charlotte, 2009 math scale scores compared to 2003 scores

- White students at great four made significant gains on their NAEP composite scale score.
- Fourth graders made significant gains on the geometry subscale.
- Eighth-grade students made significant gains on their NAEP composite scale scores.
- Eighth-grade scale scores increased significantly on the measurement subscale.
- Eighth-grade composite scale scores increased significantly among NSLP-eligible students.

Charlotte's 2009 mathematics scale scores compared to students in large cities

- Charlotte's fourth and eighth graders achieved significantly higher NAEP composite mathematics scale scores and higher scores on all mathematics subscales–algebra; data analysis, statistics, and probability; geometry; measurement; and number—than fourth graders in large cities.
- Charlotte's fourth and eighth graders in four student groups—African American, White, Hispanic, and NSLP-eligible—scored significantly higher than their counterparts in large cities.

Charlotte's 2009 mathematics scale scores compared to students in national public schools

- Fourth graders in Charlotte achieved a significantly higher mathematics composite scale score and higher scores on four out of five mathematics subscales--algebra; data analysis, statistics, and probability; geometry; and number—than fourth graders in national public schools.
- Charlotte's fourth graders in four student groups—African American, White, Hispanic, and NSLP-eligible—had significantly higher mathematics composite scale scores than their counterparts in national public schools.
- At the eighth grade, African American and White students in Charlotte scored significantly higher than their counterparts in national public schools.

Table H.1 Average scale score of grade 4 Charlotte Public School students in 2003-2009 NAEP reading assessment, overall, by subscale and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Reading Composite					
Charlotte	219	221	222	225*,**	6***
North Carolina	221	217	218	219	-2
Large City	204	206	208	210**	6***
National Public	216	217	220	220*	3***
Reading for Literary Experience Scale					
Charlotte	222	224	223	226*,**	4
North Carolina	225	219	219	221	-4***
Large City	208	209	211	212**	4***
National Public	220	220	221	221*	1***
Reading for Information Scale					
Charlotte	215	218	221	223*,**	8***
North Carolina	216	214	217	217	1
Large City	200	202	205	207**	8***
National Public	213	214	217	218*	5***
African American Students (composite)					
Charlotte	205	206	206	211*,**	6
North Carolina	203	200	202	204	1
Large City	193	196	199	201**	8***
National Public	197	199	203	204*	7***
White Students (composite)					
Charlotte	237	240	244	243*,**	6
North Carolina	232	227	228	230	2
Large City	226	228	231	233**	7***
National Public	227	228	230	229*	2***
Hispanic Students (composite)					
Charlotte	202	209	207	212*,**	10
North Carolina	212	204	205	204	-8
Large City	197	198	199	202**	4***
National Public	199	201	204	204*	5***
Asian/Pacific Islander Students (composite)					
Charlotte	218	‡	235	233	16
North Carolina	227	221	228	241	13
Large City	223	223	228	228**	5
National Public	225	227	231	234*	10***

National School Lunch Program-Eligible Students (composite)					
Charlotte	200	206	205	210*,**	10***
North Carolina	206	202	205	205	-1
Large City	196	198	200	202**	6***
National Public	201	203	205	206*	5***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05
*** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

Table H.2 Average scale score of grade 4 Charlotte Public School students in 2003-2009 NAEP mathematics assessment, overall, by subscale and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Mathematics Composite					
Charlotte	242	244	244	245*,**	3
North Carolina	242	241	242	244**	2
Large City	224	228	230	231**	7***
National Public	234	237	239	239*	5***
Algebra Scale					
Charlotte	249	250	251	252*,**	2
North Carolina	248	248	248	250**	2
Large City	231	235	236	237**	6***
National Public	240	243	244	244*	4***
Data Analysis, Statistics, and Probability Scale					
Charlotte	245	246	246	251*.**	5
North Carolina	245	245	247	249**	4***
Large City	227	231	233	233**	6***
National Public	237	241	243	242*	5***
Geometry					
Charlotte	237	244	245	243*.**	7***
North Carolina	240	239	241	243**	2
Large City	225	227	230	232**	7***
National Public	233	236	238	239*	6***
Measurement Scale					
Charlotte	240	244	239	239*	1
North Carolina	240	240	239	239	1
Large City	220	225	226	228**	8***
National Public	233	236	238	238*	5***
Number Scale					
Charlotte	241	243	242	245*,**	4
North Carolina	240	239	240	243**	2
Large City	222	226	228	230**	8***
National Public	232	235	237	237*	6***
African American Students (composite)					
Charlotte	229	230	230	231*,**	2
North Carolina	225	225	224	226**	1
Large City	212	217	219	219**	7***
National Public	216	220	222	222*	6***
White Students (composite)					
Charlotte	257	261	261	263*,**	6***
North Carolina	251	250	251	254**	3
Large City	243	247	249	250**	8***
National Public	243	246	248	248*	5***
Hispanic Students (composite)					
Charlotte	233	234	234	235*,**	2
North Carolina	235	234	235	236**	1
Large City	219	223	224	226	7***
National Public	221	225	227	227	6***

Asian/Pacific Islander Students (composite)					
Charlotte	252	256	263	257	5
North Carolina	255	256	253	259	4
Large City	246	247	251	253	8
National Public	246	251	254	255	9***
National School Lunch Program-Eligible Students (composite)					
Charlotte	229	230	231	232*,**	3
North Carolina	229	229	231	232**	3***
Large City	217	221	223	225**	8***
National Public	222	225	227	228*	6***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05; *** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

Table H.3 Average scale score of grade 8 Charlotte Public School students in 2003-2009 NAEP reading assessment, overall, by subscale and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Reading Composite					
Charlotte	262	259	260	259*,**	-2
North Carolina	262	258	259	260**	-2
Large City	249	250	250	252**	4***
National Public	261	260	261	262*	1***
Reading for Literary Experience Scale					
Charlotte	262	258	260	258*,**	-4***
North Carolina	261	259	260	258**	-3
Large City	249	250	249	251**	3***
National Public	260	260	260	261*	1
Reading to Perform a Task Scale					
Charlotte	265	260	258	—	
North Carolina	264	258	259	—	
Large City	245	248	247	—	
National Public	261	260	260	—	
Reading for Information Scale					
Charlotte	260	261	260	261*,**	1
North Carolina	261	257	259	262	0
Large City	250	252	252	254**	3***
National Public	262	261	262	264*	2***
African American Students (composite)					
Charlotte	247	244	246	249*,**	1
North Carolina	247	240	241	243	-4
Large City	241	240	240	243**	2***
National Public	244	242	244	245*	2***
White Students (composite)					
Charlotte	278	278	279	276	-2
North Carolina	271	267	270	270	-1
Large City	268	270	271	272	4***
National Public	270	269	270	271	1***
Hispanic Students (composite)					
Charlotte	244	248	251	254*	9
North Carolina	244	248	246	249	5
Large City	241	243	243	245**	4***
National Public	244	245	246	248*	4***
Asian/Pacific Islander Students (composite)					
Charlotte	‡	‡	‡	‡	‡
North Carolina	267	275	265	272	5
Large City	260	266	263	268**	8***
National Public	268	270	269	273*	5***

National School Lunch Program-Eligible Students (composite)					
Charlotte	244	242	245	248*	4
North Carolina	247	244	246	245**	-1
Large City	241	243	242	244**	3***
National Public	246	247	247	249*	3***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05
*** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

Table H.4 Average scale score of grade 8 Charlotte Public School students in 2003-2009 NAEP mathematics assessment, overall, by subscale and by selected characteristics compared with state, large city, and national public

	2003	2005	2007	2009	Difference 2003 to 2009
Mathematics Composite					
Charlotte	279	281	283	283*	4***
North Carolina	281	282	284	284**	3
Large City	262	265	269	271**	9***
National Public	276	278	280	282*	6***
Algebra Scale					
Charlotte	286	287	289	286*	0
North Carolina	285	286	289	288	3
Large City	266	270	274	276**	11***
National Public	279	281	284	286*	8***
Data Analysis, Statistics, and Probability Scale					
Charlotte	280	282	285	286*	5
North Carolina	283	282	285	287**	4
Large City	263	266	270	270**	8***
National Public	279	280	283	283*	5***
Geometry Scale					
Charlotte	279	280	282	284*	5
North Carolina	281	281	282	283**	2
Large City	261	263	268	270**	9***
National Public	274	275	277	279*	5***
Measurement Scale					
Charlotte	272	276	278	280*	8***
North Carolina	277	279	280	282	5
Large City	254	258	261	266**	12***
National Public	274	274	276	278*	5***
Number Properties Scale					
Charlotte	274	274	276	276*	2
North Carolina	279	278	280	279**	1
Large City	263	264	266	269**	6***
National Public	276	276	278	279*	3***
African American Students (composite)					
Charlotte	258	264	267	270*,**	11***
North Carolina	260	263	266	262	2
Large City	247	250	254	256**	9***
National Public	252	254	259	260*	9***
White Students (composite)					
Charlotte	301	304	308	304*,**	3
North Carolina	294	292	295	297**	3
Large City	285	288	292	294	8***
National Public	287	288	290	292	5***
Hispanic Students (composite)					
Charlotte	262	262	264	272*	10
North Carolina	263	265	273	274**	11***
Large City	256	258	261	264	9***
National Public	258	261	264	266	8***

Asian/Pacific Islander Students (composite)					
Charlotte	293	‡	305	‡	‡
North Carolina	297	303	299	311	14
Large City	281	289	291	299	17***
National Public	289	294	296	300	11***
National School Lunch Program-Eligible Students (composite)					
Charlotte	256	261	265	268*	12***
North Carolina	263	266	268	268	4***
Large City	252	256	260	262**	10***
National Public	258	261	265	266*	8***

*Statistically different from large cities at p <.05; ** Statistically different from national public at p <.05
*** Statistically different from 2003 at p <.05; ‡ Reporting standards not met. Note: Some differences may appear larger or smaller due to rounding that occurs when differences between scale scores are calculated.

APPENDIX I

INDIVIDUALS INTERVIEWED ON SITE VISITS AND MATERIALS REVIEWED

I.1 Individuals Interviewed on Site Visits

Atlanta

- LaChandra Butler-Burks, Chair, School Board
- Cecily Harsch-Kinnane, Vice-Chair, School Board
- Emmett Johnson, School Board (At-Large, Seat 9)
- Beverly Hall, Superintendent
- Kathy Augustine, Deputy Superintendent
- Crystal Lottig, Executive Director, Language Arts/Literacy
- Monishee Mosley-O'Neil, Language Arts/Literacy
- Anita Lawrence, Coordinator, World Languages
- Mary Mohead, Director of External Programs
- Mary Bailey, Director of External Programs (former)
- Randolph Bynum, Associate Superintendent
- Constance Goodson, Interim Director, PEC
- Arlene Snowden, Principal, Capitol View
- Betty Green, Principal, Dunbar
- Betty Tinsley, Principal, Herndon
- Rebecca Pruitt, Principal, Stanton DH
- Patricia Lavant, Principal, Whitefoord
- Donnell Underdue, Jr., Principal, Brown
- Christopher Waller, Principal, Parks
- Carla Petis, Principal
- Pat Plavant, Principal
- Brian Mitchell, Principal
- Will Davenport, Principal
- Tammy Kirshtein, Director, Professional Development
- Gina Glymph, Coordinator, Early Childhood
- Cynthia Terry, Director, Fine Arts
- Althea Bolton, Gifted and Talented
- Millicent Few, Chief Human Resources Officer
- Aaron Fernandez, Executive Director, Student Programs and Services
- Lester McKee, Executive Director, Research, Planning and Accountability
- Dottie Whitlow, Executive Director, Mathematics and Science
- Isiah Faggins, Math Facilitator
- Nazur Buck, Math Facilitator
- Robin Hall, Executive Director, SRT-3
- Sharon Davis Williams, Executive Director, SRT-1
- Tamara Cotman, Executive Director, K-8 schools
- Michael Pitts, Executive Director, K-8 schools
- Stephen Fowler, Executive Director
- Dihanne Hayes, Program Manager
- Janet Johnson, Specialist

- Caroline Brown, Instructional Coach, SRT-3
- Tiffany Momin, Math Coach
- Gwendolyn Alston, Reading Coach
- Barbara Dremeny, Math Coach
- Deborah Dixon, SFA Coach
- Zsa Boykin, Literacy Coach
- Sharon Green, Instructional Specialist
- Tomeka Alexander, TLS
- Patrick Crabtree, Atlanta Association of Educators
- Verdalia Turner, Atlanta Federation of Teachers
- Thelma Mumford Glover, Links
- Gwen Benson, Georgia State University
- John Grant
- Yolanda Watson, Project GRAD
- Rev. Darrell Elligan, Concerned Black Clergy
- Ovella Roberts, Teacher
- Tanisha Johnson, Teacher
- Susan Dunn, Teacher
- Carla Thomas, Teacher
- Chavon Kirkland, Teacher
- Jocelyn Daniels, Teacher
- Lataura Gregory, Teacher
- Michelin Taylor, Teacher
- Cherl Jones-Ali, Teacher
- Deborah Mitchell, Teacher
- Janice Edmonds, Teacher

Boston

- Liz Reilinger, former President, School Committee
- Marchelle Raynor, Member, School Committee
- Carol Johnson, Superintendent
- Marilyn Decker, former Director, Science
- Pam Pelietier, Director, Science
- Ann Deveney, former Director, Language Arts/Literacy
- Barbara McLaughlin, Director, Language Arts/Literacy
- Judith Berkowitz, Director, Gifted and Talented
- Maryellen Donahue, former Director, Research, Testing and Evaluation
- Kamal Chavda, Director, Research, Testing and Evaluation
- Linda Davenport, Director, Mathematics
- Chris Coxon, former Deputy Superintendent, Teaching and Learning
- Janet Palmer Owens, Chief Academic Officer
- Sid Smith, former Director, Curriculum and Instruction
- Shonda Huery, Director, Curriculum and Instruction
- Rachel Curtis, former Director, Professional Development
- Casel Walker, Director, Professional Development

- Vickie Megias Batista, Assistant Superintendent, Elementary Schools
- Mary Nash, Assistant Superintendent, Elementary Schools
- Jeff Riley, Assistant Superintendent, Middle Schools
- Maryann Martinelli, former Director, Early Childhood Education
- Jason Sachs, Director, Early Childhood Education
- Carolyn Riley, Director, Special Education
- Eileen de los Reyes, Director, English Language Learners
- Michelle Boyer, former Director, Human Resources and Teacher Placement
- Charlotte Harris, former Director, Federal and State Programs
- Monica Harris, Director, Federal and State Programs
- Michelle Carpinteri, Coach, Middle School Language Arts
- Cassandre Felix, Coach, Middle School Math
- Eileen Cronin, Coach, Elementary Schools
- Clair Jones, Coach, Elementary Schools
- Jana Sunkle, Coach, Math
- Liz MacDonald, Coach, Language Arts
- Margarita Ruiz, Principal, Adams
- Ann Garafalo, Principal, Condon
- Ron Jackson, Principal, Grew
- Marice Diakite, Principal, PJ Kennedy
- Joy Salesman, Principal, Higginson/Lewis K-8
- Bak Fun Wong, Principal, Quincy Upper
- Richard Stutman, President, Boston Federation of Teachers
- Betty Smith, Teacher, Clap Elementary
- Michael Wilkinson, Special Education Teacher, East Greenwood Elementary School
- Hector Soto, Teacher, JF Kennedy Elementary School
- Susan Ashton, Teacher, Taylor Elementary School
- Eloise Biscoe, Teacher, Hernandez K-8 School
- Kelly Keady, Teacher, Murphy K-8 School
- Delores Martinez, Teacher, Lyon School
- J. Thomas, Teacher, Harbor Elementary School
- Maria Ciampa, Teacher, Perry K-8 School
- Filiberto Santiago-Lizardi, Teacher, Timilty Middle School
- Mark Rukavina, Parent, Haley Elementary School
- Angela Veale, Parent, Otis Elementary School
- Mike Lewis, Parent, Lee Academy
- Irma Gomes, Parent, Winthrop Elementary School
- Mary Tamer, Kilmer K-8 School
- Warren Prescott, Parent, Dearborn Middle School
- Neil Sullivan, Private Industry Council
- Ellen Guiney, Boston Partners in Education
- Klare Shaw, Barr Foundation
- Chris Smith, Boston and Beyond
- Dania Vasquez, Center for Collaborative Education
- Janet Anderson, Ed Vestors

- Abby Weiss, Full Schools Service Roundtable
- John Mudd, Massachusetts Advocacy

Charlotte-Mecklenburg

- Muffet Garber, Supervisor of Curriculum and Instruction
- John Fries, Curriculum and Instruction (former)
- Katy Dula, Language Arts/Literacy Specialist
- Ann Clark, Curriculum and Instruction
- Ron Dixon, Assistant Superintendent for Curriculum and Instruction
- George Dunlap, School Board Member (former)
- Molly Griffin, School Board Member (former)
- Elva Cooper, Regional Superintendent
- Bev Moore, Regional Superintendent (former)
- Ron Thompson, Regional Superintendent
- Gwen Bradford, Human Resources
- Jan Richardson, Human Resources (former)
- Tekle Ayano, Research, Testing and Evaluation
- Chris Cobitz, Research, Testing and Evaluation (former)
- Jason Schoeneberger, Research, Testing and Evaluation (former)
- Gloria Cox, Talent Development
- Cathy Capps, Instructional Coach
- Ormond Cottle, Instructional Coach
- Ann Ganzert, Instructional Coach
- Angie Larner, Instructional Coach
- Susan Patterson, Instructional Coach
- Judy Goins, Language Arts/Literacy Specialist
- Michelle Bogan, Barringer Elementary School Parent
- Ellen Cotton, Hawk Ridge Elementary School Parent
- Shawna Coulter, Reid Park Elementary School Parent
- Cynette Edwards, Newell Elementary School Parent
- Dianne Elliott, Ballantyne Elementary School Parent
- Margary Massey, Tuckaseegee Elementary School Parent
- Michele Price, Lake Wylie Elementary School Parent
- Melissa Walker, Cotswold Elementary School Parent
- Snowden Littlejohn, Quail Hollow Middle School Parent
- Robbin Stackhouse, Coulwood Middle School Parent
- Kim Graham, Community Member
- Bill Anderson, Community Member
- Sharon Starks, Community Member
- Julie Babb, Early Childhood
- Kim Foxworth, Early Childhood (former)
- Jane Meyer, Early Childhood (former)
- Jane Rhyne, Exceptional Children
- Gina Smith, Exceptional Children (former)
- Laura Hamby, Exceptional Children (former)

- Tracie Lynn Zakas, Exceptional Children (former)
- Kathy Meads, English Language Learners
- Jennifer Pearsall, English Language Learners (former)
- Regina Boyd, English Language Learners (former)
- Diane Adams, Principal, Providence Springs Elementary School
- Penni Beth Crisp, Principal, Torrence Creek Elementary School
- Leah Davis, Principal, Montclaire Elementary School
- Kathy Elling, Principal, Croft Community Elementary School
- Pam Frederick, Principal, Huntingtowne Farms Elementary School
- Maria Petrea, Principal, Collinswood Elementary School
- Jennifer Dean, Principal, Bailey Middle School
- Terri Cockerham, Principal, Eastway Middle School
- Jackie Menser, Principal, Randolph Middle School
- Tony Bucci, State and Federal Programs
- Kelly Price, State and Federal Programs (former)
- Dot Cromwell, President, North Carolina Association of Educators
- Barb Temple, Professional Development
- Mary Webb, Professional Development (former)
- Barb Bissell, Math and Science
- Cindy Moss, Math and Science (former)
- Bill Scott, Math and Science (former)
- Kathleen Koch, Math and Science (former)
- Ormond Cottle, Math and Science (former)
- Stacey Wood, Math and Science (former)
- Kat Eaker, Talent Development
- Shirley Kohl, Talent Development (former)
- Stephanie Range, Talent Development (former)
- Carol Abritton, Teacher, Walter G. Byers Elementary School
- Cathey Cooper, Teacher, Piney Grove Elementary School
- Pam Darcey, Teacher, Blythe Elementary School
- Yolanda Parsons, Teacher, Nathaniel Alexander Elementary School
- Suzie Rose, Teacher, Devonshire Elementary School
- Marcy Sanders, Teacher, Highland Creek Elementary School
- Mary Torkildson, Teacher, Lansdowne Elementary School
- Farncine Carr, Teacher, Smith Academy of International Languages
- Margaret Kohlmeyer, Teacher, Alexander Middle School
- Derek Shoup, Teacher, Randolph Middle School
- Jennifer Snyder, Teacher, Randolph Middle School
- Susan Sweet, Teacher, Alexander Middle School

Cleveland

- Eric Gordon, Chief Academic Officer
- Russ Brown, Director, Research, Testing and Evaluation
- Karen Thomson, Deputy Chief Curriculum and Instruction
- David Quolke, President, Cleveland Federation of Teachers

- Thea Wilson, Executive Director for Early Childhood
- Shirley Arnold, Manager for Early Childhood
- Robert Walsh, Executive Director for Special Education and Psychological Services
- Beverly Weccia, Manager for Gifted and Talented and Advanced Placement
- Clara Hayes, Manager for English language arts (Prek-8)
- Gayle Philpot, Manager for English language arts (Grades 7-12)
- Theon Jone, Coach
- Elizabeth Nelson, Manager for mathematics (Prek-8)
- Ovella McIntyre, Manager for mathematics (Grades 7-17)
- Juanita Holt, NCLB
- Margariete Hunt-Smith, NCLB
- William Badders, Math Science Partnership
- Cheryl Shelton, Director of Office of Professional Development
- Cliff Haynes, Regional Superintendent
- Regina Paris, Regional Superintendent
- Francine Watson, Regional Superintendent
- Laura Prenell, Regional Superintendent
- Bruce Thomas, Regional Superintendent
- Erica James, Parent
- Michael Herrson, Parent
- Tyrone Parker, Parent
- Mirian & Elliot Crews, Parents
- Amanda Wood, Parent
- Izetta Grayer, Parent
- Nahshonda Cundiff, Parent
- Loreal Buckner, Parent
- Amanda Gielink, Parent
- Valencia Washington, Parent
- Kanika Davis, Parent
- Marwa Ibrahim, Principal
- Janet Moore, Principal
- Mike Morowsky, Principal
- Dakota Williams, Principal
- Amy Peck, Principal
- Sandra Bullazqeuz, Principal
- Julie Shepphard, Principal
- Hearther Grant, Principal
- Charles Burden, Principal
- Denise Welsh, Teacher
- Welvina Buffington, Teacher
- Bob Stan, Teacher
- Meghan Mets, Teacher
- Tersa Baker, Teacher
- Tish Henry, Teacher
- Maurine Eagle, Teacher

- Nancy Salvoka, Teacher
- Rasa Wade, Teacher
- Omega Brown, Human Resources
- Natavid Pagan, English language learners
- Margaret Frye, English language learners
- Ron Soeder, Boys & Girls Club
- Robin Martin, Family & Children First Council
- Terry Butler, College Pathways Programs, Cuyahoga Community College
- Karen Butler, Cleveland Department of Public Health
- Susan Wentz, Case Western Reserve University
- Lynn Lotas, Case Western Reserve University
- Barbara Byrd-Bennett, former CEO

- **I.2 Materials Reviewed**

Atlanta

- Organization structure (org chart) for academics during the period of study and members of academic departments serving on the Superintendent's Cabinet.
- Description of process used to evaluate principals during that time period, with appropriate forms
- Copy of the district's strategic plan (chose the one that you think is the best) – also copy of evaluation of district's strategic plan from that time period
- . Information about the district's choice plan
- Board agendas and minutes from three (3) 2002-03 and three (3) 2006-07 board meetings
- Description of process used to evaluate teachers during that time period, with appropriate forms
- District vision of teaching and learning during the time period
- An annotated list of school level reform projects that were in place
- Description literacy instructional approach and names of textbooks/programs/interventions at pre-kindergarten through Grad**e** 8 during that time period
- District approach to the teaching of writing during that period
- Copies of the district's professional development plans from that time period
- A description of the philosophy and time requirements of the district's programs for English language learners
- Copies of a sample of the district's Grades 3-5 and 7- 8 language arts, curriculum guides with pacing guides
- Number and percentages of students participating in the district's gifted and talented programs, per school by racial/ethnic, English language learners, and gender data
- Number and percentages of students participating in the district's bilingual or English Language Learner programs, per school with racial/ethnic and gender data
- Course pass rates in Grade 9 mathematics, English, and science
- High school graduation requirements compared to state graduation requirements during the study period
- Copy of any instructional study of the district during that time period, if available
- Annual state report for district achievement 2003-2004, 2005-2006, and 2007-2008
- Samples of (short cycle) tests in those grade levels and content areas, if they exited during that time period
- List of high schools and the AP courses offered at each (indicator of college-bound focus) distribution of AP courses and participation rates
- Number of AP tests taken and exam grades earned by school and district and subgroup
- Samples of communicating district progress on goals to the public during that time period
- Description of mathematics instructional approach and names of textbooks/programs/interventions at pre-K through Grade 8 during that time period
- Description of science instructional approach, time allocation, and names of textbooks/programs/interventions at pre-K through Grade I during that time period
- Copies of a sample of the district's Grades 3-5 and 7-8 science and math curriculum guides with pacing guides
- A description of how the district supported low-performing schools and students during that time period
- Number and percentages of students participating in the district's special education programs, per school by racial/ethnicity

Boston

- Organizational chart for curriculum and instruction and executive staff (2004-2005)
- Copy of district's strategic plan, *Focus on Children 2, Boston's Education Reform Plan 2001-2006*
- Copy of the evaluation of the district's strategic plan, along with its vision of teaching and learning: *Transforming Boston's Schools, A Decade of Focus on Children and the Challenges of the Future*, December 2005
- Various articles and informational pamphlets about the district's choice plan, 2004-05
- Board agendas and minutes from three 2002-2003 and three 2006-2007 board meetings
- BPS Guide for Principals, Headmasters and Directors on Performance Evaluation Process 2004 2005 (Description of process used to evaluate principals during that time period)
- Superintendent's Circular. *Performance evaluation of teachers, School Year 2005-2006* (description of process used to evaluate teachers during that time period) + Teacher Performance Evaluations: Detailed Procedures and Timetable
- Materials from various school-level reform projects that were in place during that time period, including:
  -- Boston Teacher Residency 2004
  --The 6 Essentials--Boston Public Schools Whole-School Improvement Plan 2004
  --Strong Foundation-Evolving Challenges: A Case Study to Support Leadership Transition in BPS February 2006
  --Professional development spending in Boston Public Schools, December 2005
  --Introduction to CCL: Collaborative Coaching and Learning 2002
  --Workshop Instruction—Boston's Schools 2002
  --A Decade of Boston School Reform, 2007
- Annual MCAS reports for district achievement for 2002-2003, 2005-2006, and 2007-2008
- Copies of various instructional studies of the district during that time period, including *Student Stability and Mobility in Boston Public Schools, School Year 2003-2004*, and *Report on Adequate Yearly Progress: 2005 Mid-Cycle IV AYP Determinations for Boston Public Schools*
- Samples of community and media outreach, including articles and announcements in *The Boston Educator* and *Focus* (newsletter for Boston teachers)
- Sample ELA, science, and math curriculum and pacing guides for grades 3-5 and 7-8
- Description of literacy instructional approach and names of textbooks/programs/interventions at pre-kindergarten through grade 8 during that time period *(CCL, Balanced Literacy approach, etc.)*
- Materials outlining the district approach to the teaching of writing, such as teaching guide *Four Kinds of Writing/Four Levels of Support*
- Materials describing the district's mathematics instructional approach and names of textbooks/programs/interventions for pre-k through grade 8 and math adoption plan 2003-2007
- Materials describing the district's science instructional approach, time allocation, and names of textbooks/programs/interventions for pre-k through grade 8
- Copies of the Boston Math & Science Plan annual progress reports, 2001-2002, 2002-2003

- Copies of the district's professional development plans from that time period, including: *Introduction to CCL: Collaborative Coaching and Learning* in the *Boston Plan for Excellence, September 2002*
- Materials detailing how the district supported low-performing schools and students, including documents on transition and support programs, a memo from the superintendent to the Boston School Committee on Additional Resources for Low-Performing Schools, a memo from the superintendent to principals and headmasters laying out the district's Individual Student Success Plan (ISSP), and a listing of Reading Intervention Programs and Supports for Schools in 2005-2006.
- Boston Public Schools Policy for English Language Learners, modified January, 2004
- Revised Boston Public Schools Graduation Policy for High School Students + Massachusetts Department of Education Time & Learning Questions and Answers, August 1999 (Outlines state requirements)
- BPS Advanced Placement Data Review, March 11,2008 + Advanced Placement Research Summary, Updated April 2, 2007
- Chart showing the number of SAT, SAT subject tests, PSAT, and AP tests taken by school in 2006
- Boston Public Schools Budget, Fiscal Years 2001, 2002, 20003-2004, 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2009

Charlotte-Mecklenburg

- Organization structure (org chart) for academics during study period and members of academic departments serving on the Superintendent's Cabinet.
  --Charlotte Mecklenburg Schools- TEAMING FOR EXCELLENCE
- The district's strategic plan, 2006-2010
- Evaluation of the district's strategic plan between 2006-2010
- Project Charter 2007--
  --K-3 Intensive Reading; A Design for Academic Success
  --Behavior Support Model
  --Professional Development
  --LEP Charter
  --Achievement Zone
  --Expanded Day Charter
  --Science and Math
  --Inclusive Practices, 2007
  --Eight-Plus Programs
  --Accountability Plan Charter
  --Inclusive Practices, 2008
- Information about the district's choice plan
- Pupil Assignment Plan: Choice Plan 2006-2007
- Board agendas and minutes from three 2002-03 and three 2006-07 meetings
- Board agenda and minutes – July11, 2006-June 26,2007
- Description of process used to evaluate principals during study period, with appropriate forms: Principal Evaluation Process At a Glance, 2007
- Description of process used to evaluate teachers during study period, with appropriate forms**:** Charlotte-Mecklenburg Public Schools Evaluation Guide, 2006-2007

- District vision of teaching and learning during study period
- An annotated list of school-level reform projects that were in place
- Annual state report for district achievement in 2003-04, 2005-06, and 2007-08
- 2004-2008 Adequate Yearly Progress (AYP) Results
- Copy of instructional study of the district during study period
- Samples of communicating district progress on goals to the public during study period
- CMS media releases
- Copies of the district's grades 3-5 and 7-8 language arts, science, and math curriculum guides with pacing guides
- Proposed Mathematics Pacing Chart, grade four
- Proposed Mathematics Pacing Chart, grade eight
- CMSD Eight Grade Mathematics Pacing Calendar
- Samples of (short cycle) tests in fourth and eighth grades in reading and math
- How to Employ Short Cycle Assessments
- Description literacy instructional approach and names of textbooks/programs/interventions at pre-kindergarten through grade 8 during that time period.
  --Pre K-12 Comprehensive Reading Model
  --Reading/language arts and English Initiatives as of July 2007
  --2006-07 Project charter – Elementary
  --Elementary school literacy
  --Middle school literacy
  --High school literacy
  --Accomplishments and challenges
  --2006-07 Elementary Needs Assessment Form
  --Elementary team strategies
  --2006-2007 PMOC Presentation
  --2007 Board Report k-2 Interventions
  --PLATO
- District approach to the teaching of writing during study period
  --CMS K-12 writing program overview (same as 02-03)
  --K-12 writing plan model (same as 02-03)
  --Balanced writing
  --Components of effective writing instruction
  --K-5 comprehensive writing plan
  --K-2 writing assessment
  --Elementary writing project schedule
- Description of mathematics instructional approach and names of textbooks/ programs/interventions at pre-kindergarten through grade 8 during study period
  --Mathematics Infrastructure in CMS for 2006-2007
  --Vision for mathematics teaching and learning
  --Elementary math pacing guide for NC 2003-2008 Standard Course of Study
  --Quarterly assessment, Grade 6 math, Quarter 3, 2006-2007
- Description of science instructional approach, time allocation, and names of textbooks/programs/interventions at pre-kindergarten through grade 8 during study period
  --Vision for Science Teaching and Learning
  --Big Ideas of Science
  --Comprehensive science model K-12
  --End-of-year assessment, grade 5 Assessment, 2006-2007

--Science infrastructure in CMS for 2006-2007

- How lab materials were provided in elementary schools
- Copies of the district's professional development plans during study period: Annual Performance Report, Improving Teacher Quality
- A description of how the district supported low-performing schools and students during study period
- Number and percentages of students participating in the district's special education programs, per school by race/ethnicity: Exceptional Children Data, 2006-2007
- Number and percentages of students participating in the district's gifted and talented programs, per school with racial/ethnic, English language learner, and gender data
- Students participating in district bilingual or ELL programs, 2006-2007
- Number and percentages of students participating in the district's bilingual or English language learner programs, per school with racial/ethnic and gender data
- A description of the philosophy and time requirements of the district's programs for English language learners
  --Elementary ESL curriculum model
  --Pull-out program of services for students in grades 1 and 2 at ESL sites
  --Pull-out program of services for students in grades 3-5 at ESL sites
  --Pull-out program of services for students for middle school students at ESL sites
  --Programs of services for high school ESL students
  --Textbook adoption
- Course pass rates in grade 9 mathematics, English, and science: ninth-grade pass rates 2002-2003 and 2006,2007
- High school graduation requirements compared to state graduation requirements during study period: Charlotte graduation requirements
- List of high schools and the AP courses offered at each (indicator of college-bound focus)
- Distribution of AP courses and participation rates.
- Number of AP tests taken and exam grades earned by school and district and subgroup
- District Integrated Summary, 2006-2007
- CMS students served by gifted program, 2006-2007
- Teacher turnover by school

Cleveland

- NCLS Task Force Org Chart SY 2002-2003
- CMSD Org Chart May 2003
- CMSD Org Chart SY 2006-07
- Strategic Plan/Executive Summary 2007-2012
- CMSD Audits, Reports, and Investigations Summary - December 2006
- School Choice Summary SY 2002-2003
- CMSD Choice Schools SY 2006-2007
- Building Capacity
- Status of school choice applications
- Application status
- School choice application summaries
- Board agenda and minutes - January 17, 2002
- Board agenda and minutes - June 6, 2002

- Board agenda and minutes - April 15, 2003
- Board agenda and minutes - March 14, 2006
- Board agenda and minutes - November 9, 2006
- Board agenda and minutes - March 27, 2007
- Principal's performance review SY 2007-2008
- Teacher evaluation 2002-2003
- CMSD Individual Visit Evaluation SY 2002-2008
- CMSD Principal Composite Evaluation SY 2002-2008
- Vision/Mission Statement SY2003-2004
- Weekly status report - February 11, 2002
- Weekly status report - March 3,2003
- State report cards (2003-04, 2005-06, and 2007-08)
- Results Engineering Case Study (reeng.com) cruse
- Keeping Learning on Track
- (http://www.utdanacenter.org/pwoa/downloads/cleveland.pdf)
- Cleveland Metropolitan School District Human Ware Audit: Finding
- (www/air.org/news/documents/AIR_Cleveland_8-20-08.pdf)
- Proposed mathematics pacing chart grade four
- Proposed mathematics pacing chart grade eight
- CMSD Eight Grade Mathematics Pacing Calendar
- How to Employ Short Cycle Assessments
- CMSD - A comprehensive Approach to Changing Instructional Practice
- www.cgcs.org/images/pastconference-pdfs/AGAI O.pdf
- CMSD K-8 textbook inventory
- Pre-k-5 math textbook adoption
- Pre-k-8 math textbook adoption
- Pre-k-8 science description and materials list 2003-2007
- How Lab Materials Were Provided in Elementary Schools 2003-2007
- CMSD Office of Professional Development Focus 2002-2003
- CMSD Office of Professional Development Focus Creates Electronic Professional Development Plan 2006-2007
- 2002-2003 Supplemental Educational Services Summary
- CMSD Supplemental Educational Services Provider Summary ST 2003-2007
- SPED codes
- SPED ethnic 2002-2003 and 2006-2007
- Summary page for gifted enrollment
- 2002-2003 school year enrollment broken down by gifted sites
- 2006-2007 school year enrollment broken down by gifted sites
- Student enrollment data for English language learners 2002-2003 and 2006-2007
- Information in box LED Procedure Manuals 1999 and 2008-2009
- AP: Performance and participation overview SY2008-2009
- AP: Exam participation and performance SY2008-2009
- AP: Number of examinations and number of examinations with grades of 3, 4, or 5 SY 2008-2009

- AP: Participation by ethnic groups taking one or more exams SY2008-2009
- AP: Participation by ethnic groups with grades 3, 4, or 5
- A Premier Future: Dept. of Arts Education Strategic Plan 2008-2009
- Cleveland- A City. A Community. A Home.
- Opening Day SY 2002 (Blue Folder)
- SY2002-2003 Annual Calendars
- Facilities Update May 2003
- The New Standards-Based Report Card and New Promotion Policy Flyer
- Educating Cleveland's Children (significant first-year accomplishments) Pamphlets
- Educating Cleveland's Children Newspapers
- Keeping Cleveland's Children Safe and Secure
- Engaging Families and the Community to Support Cleveland's Children
- A 21st Century School Building on a Proud Past John Hay
- We are Making Progress, also in Spanish
- Gun Violence Prevention
- School Social Work
- Student Internship
- Putting the Pieces Together for Cleveland's Children
- Step up to Victory
- Publications Division: Creating and Maintaining Effective Communications
- Building Quality Schools by Investing in Our Teachers
- Report of the Facilities Assessment Commission
- From the Ground Up: Building Schools, Building Community
- Vision to Victory: Opportunity Schools of Choice
- Designing Student Success
- Comprehensive Health Plan for the CMSD
- Educating Cleveland's Children, Getting the Job Done
- Evaluation of Responsible Sexual Behavior Education
- English Language Arts Standards - K,1,3,4,5,6,7,8,11,12
- Academic offices status reports May 26, 2006 (Binder)

APPENDIX J

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.

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