



Pieces of the Puzzle: Factors in Improving Achievement of Urban School Districts

By Michael Casserly

In one of the first large-scale analyses of urban trends on the National Assessment of Educational Progress (NAEP), the Council of the Great City Schools and the American Institutes for Research identified urban school systems that demonstrated high achievement or significant achievement gains on the NAEP, and examined possible factors behind these gains. The overarching goal was to identify variables that might be contributing to improvement in urban education nationwide, and to explore what is needed to accelerate these gains. The results demonstrated that improvement was less related to how well state and local standards aligned with NAEP frameworks, and more so to the instructional and organizational practices in the participating districts. These findings suggest broad implications for urban education reform, as well as the importance of strong instructional programming, leadership, and support for implementation of the new Common Core State Standards.

America's urban schools are under more pressure to improve than any other institution—public or private—in the nation. But instead of folding under the pressure of increasing demands and mounting criticism, many urban school systems and their leaders are rising to the occasion. They are innovating with new approaches, learning from each other's successes and failures, and aggressively pursuing reforms that will boost students' academic performance.

There is fresh evidence that the efforts of these urban school systems are beginning to pay off. Results from the National Assessment of Educational Progress (NAEP) on large-city schools indicate that, between 2003 and 2009, public schools in major urban areas in the United States made statistically significant gains in both reading and mathematics in grades four and eight.¹

Moreover, an analysis of differences between the rates of improvement of large cities and those

of the nation between 2003 and 2009 shows that gains in fourth- and eighth-grade reading and

Key points in this Outlook:

- There are encouraging signs that US urban school districts are effectively innovating to improve students' performance—between 2003 and 2009, data showed gains in math and reading in grades four and eight—yet more reform efforts are needed to raise student achievement in these districts.
- The key to improved student performance in major cities is reforming urban districts' instructional and organizational practices and supporting effective implementation of new Common Core State Standards.
- Urban school districts can boost math and reading performance via six reform strategies: strong leadership, robust accountability mechanisms, a coherent curriculum, professional development programming, districtwide support, and quantitative assessment.

Michael Casserly (mcasserly@cgcs.org) is executive director of the Council of the Great City Schools.

mathematics were significantly greater in large cities as compared to the national sample.² This same overall pattern was also seen in the 2011 NAEP testing results. Large-city schools and the districts participating in the Trial Urban District Assessment (TUDA)—which allows them to be oversampled as part of the regular NAEP testing to yield district-specific results—continue to lag behind national averages for the most part, but they are making progress over and above what is occurring at the national level.³

An analysis of differences between the rates of improvement of large cities and those of the nation between 2003 and 2009 shows that gains in fourth- and eighth-grade reading and mathematics were significantly greater in large cities as compared to the national sample.

A Closer Look

The Council of the Great City Schools and the American Institutes for Research sought to examine these emergent patterns—and the factors that might be driving improvement—in greater detail. In a report released late last year titled *Pieces of the Puzzle: Factors in the Improvement of Urban School Districts on the National Assessment of Educational Progress*, we set out to present new data related to urban school district achievement on NAEP reading and mathematics assessments in grades four and eight.⁴

The available data allowed us to report patterns and trends between 2003 and 2009 in a variety of ways. We examined how TUDA districts were performing overall; how they performed relative to each other, large cities in general, and the nation; how they performed in specific academic areas (subscales); how they performed across the distribution of student achievement scores and levels; and, finally, achievement trends among specific student groups, such as African American students or economically disadvantaged students.⁵ We were able to identify districts that were showing—or failing to show—significant and consistent gains, as well as districts with performance levels that were higher or lower than what was predictable given their student background characteristics.

To get to the heart of what was driving these achievement patterns, we selected four districts for

intensive study—one district with consistently high overall performance (Charlotte-Mecklenburg), one demonstrating significant and consistent improvements in reading (Atlanta Public Schools), one that showed such improvements in mathematics (Boston Public Schools), and one district that lacked improvements overall (Cleveland Metropolitan School District).⁶

Our project team set out to determine what role alignment between state standards and NAEP frameworks may be playing in the achievement patterns we observed. In particular, we were interested in discovering whether a close adherence to state standards—and whether these state standards were more or less closely matched to NAEP frameworks— hindered or helped the four districts make larger achievement gains as measured by the NAEP, the results of which we would impart to the districts. To do this, we investigated the alignment between NAEP frameworks and various state- and district-level standards. We then examined the relationship between this alignment and a district’s performance on the NAEP over time.

We then explored 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 as well as determining how the practices of these school systems might differ in critical ways from those of districts that were not showing substantial progress.

The Role of Alignment between State Standards and the NAEP

For each of the four districts, we determined alignment of state standards and NAEP frameworks by looking at NAEP content specifications in each subject area—reading, math, and science—and by comparing them to state (and district) standards that were in place in reading and math in 2007 and in science in 2005. We created alignment charts for each of the four districts selected for in-depth analysis. Each chart included actual NAEP specification language and displayed the degree to which each respective state and/or district’s content standards matched those specifications, either completely or partially, in content, rigor, and at grade level. Matches were determined by at least three independent “coders” specially trained in reliably conducting

the comparisons, and were then reviewed by senior content experts.

The content-matching verification process entailed assigning each NAEP specification in fourth- and eighth-grade reading, math, and science one of the following codes:

- *N (not matched)*, meaning there was no explicit or implicit match or reference to the NAEP content in the state or district standards;
- *P (partial)*, meaning there was some—even minimal—explicit or implicit match or reference to the NAEP content in the state or district standards; or
- *C (complete)*, meaning there was a complete match of the NAEP content and the state or district standards. That is, a reasonable person, with relatively strong content knowledge would say that the NAEP specification and the relevant state or district standard refer to essentially the same content or skills.

The level of alignment was deemed “high” when at least 80 percent of NAEP specifications were completely matched by the state or district standards and “low” when 50 percent or less of the NAEP specifications were completely matched. Anything in between was identified as a “moderate level” of alignment.

Following this process, we found that the extent of content alignment between NAEP specifications and the respective district or state standards of the four selected TUDA cities varied somewhat from math to reading, from fourth grade to eighth grade, and from district to district. Yet, overall, content alignment with NAEP frameworks in fourth- and eighth-grade reading and math was low to moderate, while fourth- and eighth-grade science matches were low.

We then looked for a connection between a district’s relative degree of alignment in a specific subject and its improvement on the NAEP between 2003 and 2007. Though the sample size was too small to be generalized, the results of this analysis revealed no apparent relationship between student gains on the NAEP and the degree of total content alignment between NAEP specifications and state or district standards. Some districts made significant improvements on the NAEP even when their state standards were not well-aligned with the national

assessment. At the same time, high alignment did not guarantee better results or more gains.

Atlanta was the only one of the four selected districts that had a significant increase in fourth-grade reading levels, yet it had the same percentage of complete content matches with the NAEP as Boston and Cleveland (39 percent). Atlanta and Cleveland both saw significant increases in eighth-grade reading scores (although Cleveland did not see increases once exclusions were taken into account through “full-population” estimating).⁷ However, the degree of content matches in Atlanta appeared similar to those in Boston, which saw no significant score increases over the study period. Meanwhile, Cleveland had content matches that appeared similar to Charlotte, which saw no reading increases.

Atlanta and Boston were the only selected districts to see significant increases in NAEP scores on fourth-grade math, yet both districts had lower complete content matches than Charlotte and Cleveland, which saw no significant increases in NAEP math scores. Atlanta, Boston, and Charlotte saw significant increases in eighth-grade math scores, but the districts had varying complete content matches, ranging from 24 percent in Charlotte to 45 percent in Boston. In addition, Cleveland, which saw no gain in math, had the highest level of complete content matches.

What Drove Results?

While there seemed to be no connection between standards alignment and NAEP performance, the factors that did appear to drive a school system’s ability to improve on the NAEP were a comprehensive set of instructional policies and practices, as well as strong leadership and accountability.

For example, findings from our team’s site visit to Atlanta suggested that the district benefited from a comprehensive, districtwide literacy initiative, which was launched in 2000 and rolled out in 2001. The initiative was well-defined, sustained over a long period of time, and bolstered by a system of regionally based school reform teams (SRTs) deployed to provide services directly to schools and to 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 flexibility to build ownership for reforms at the building level.

At the same time, Atlanta, which closed approximately twenty mostly low-performing schools during

the study period, laid out clear, research-based strategies and “best practices” for how literacy would be taught throughout the school system, creating a common vocabulary for reading instruction and providing extensive site-based and cross-functional support through literacy coaches and professional development. Atlanta also began to emphasize writing and the development of literacy skills across the curriculum beginning in the early years of its literacy initiative (around 2003).

Atlanta, which closed approximately twenty mostly low-performing schools during the study period, laid out clear, research-based strategies and “best practices” for how literacy would be taught throughout the school system, creating a common vocabulary for reading instruction and providing extensive site-based and cross-functional support through literacy coaches and professional development.

Although this initiative was not designed to establish causality between specific district strategies and initiatives and gains in student achievement, the overall strength of the district’s instructional programming appeared to be an important factor in Atlanta’s improved performance on the NAEP reading assessment. Among the TUDA districts selected for further examination, Atlanta had the most consistent overall gains in reading between 2003 and 2007, and those gains continued through 2011. The district also instituted a universal, sustained professional development effort that emphasized “reading for information” in fourth grade and “reading to perform a task” in eighth grade, areas in which the district showed the greatest gains on the NAEP.

Similar to Atlanta’s reading initiative, Boston began implementing a common, challenging, concept-rich math program in 2000 that has been kept in place since then. The city pursued a multistaged, centrally defined, and well-managed rollout while providing sustained support and oversight for implementation of its math reforms despite a lack of immediate, systemwide improvements. Boston was successful in these rollouts

despite the fact that—according to Council of the Great City Schools staff members who have tracked efforts in many urban school systems—these programs have proven difficult to implement in other cities.

Also like Atlanta, Boston kept its math program in place for many years, supporting it with extensive, sustained professional development and coaching assistance for teachers rather than constantly changing programs. Unlike Atlanta, however, Boston had a “softer” accountability system, but the city was able to create a strong culture in support of results that served many of the same purposes as Atlanta’s more formal system. Boston also had stable leadership at the school board, superintendent, and program director levels. Not surprisingly, Boston showed the most consistent gains in math on the NAEP during our study period.

Charlotte-Mecklenburg, meanwhile, was one of the first districts in the nation to develop and institute academic standards. The district also pioneered an instructional management theory of action that it moved away from in 2007 in favor of less centralized instructional control. During the period of our study, 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.

While Charlotte-Mecklenburg did not demonstrate the same gains as Atlanta or Boston in NAEP reading and math over the course of our study period, the district maintained consistently high performance at or above national averages from 2003 to 2007—a standing the district maintained through 2011 testing. Charlotte-Mecklenburg was selected for study because, after controlling for student and family background characteristics such as poverty and English language learner status, the district outperformed all other TUDA districts in reading and math in 2007.

Finally, Cleveland was selected as a case study district because it showed few gains on the NAEP between 2003 and 2007 and maintained weak gains through the 2011 NAEP testing. Until 2006, there was no functional curriculum to guide instruction. The school district’s instructional program remained poorly defined and the system had little room to build upon school capacity and

the teachers' ability to deliver quality instruction. Interestingly, according to school system officials, the district used the same math program as Boston but never expanded its use after the program showed results.

The greater rigor embedded in the new Common Core State Standards is likely to be squandered if the standards themselves are not well-implemented, and if the content of the curriculum, instructional materials, classroom instruction, and professional development are not top-notch, integrated, and consistent with the standards.

Cleveland 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 eyes of our team, the outcome was a weak sense of ownership for results and little capacity to advance achievement on rigorous assessments such as the NAEP. The district also endured substantial budget cuts in 2005 that resulted in the dramatic reassignment of teachers according to seniority, a move that left many instructors teaching subjects and grades for which they were unprepared, with few central office staff members to assist. By 2007, the district had fewer teachers for a school system of its enrollment than any of the other TUDA districts.

Common Themes

Despite their differences, there were a number of shared traits and themes among the improving and high-performing districts—Atlanta, Boston, and Charlotte-Mecklenburg—which clearly contrasted with the experiences and practices documented in Cleveland. These characteristics fell under six broad categories:

1. *Leadership and Reform Vision:* Atlanta, Boston, and Charlotte each benefited from the strong leadership of their school boards, superintendents, and curriculum directors. These leaders were able to unify the districts in promoting and sustaining a vision for instructional reform.

2. *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.
3. *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 districts.
4. *Professional Development and Teaching Quality:* Atlanta, Boston, and Charlotte each supported their programs with well-defined professional development or coaching tied to instructional programming. In doing so, these districts set direction, built capacity, and enhanced teacher and staff skills in priority areas.
5. *Support for Implementation and Monitoring of Progress:* Each of the three improving or high-performing districts designed specific strategies and structures to ensure that reforms were supported and implemented districtwide, and to deploy staff to support instructional programming at the school and classroom levels.
6. *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.

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.

Implications of these Findings

The results of this exploratory study are encouraging because they indicate that urban schools are making significant academic progress in reading and math and may be catching up with national averages. More

importantly, our findings suggest some explanations for this progress and reveal steps that might be required to accelerate this headway, particularly as the new Common Core State Standards are implemented.

In particular, the finding that student improvement on the NAEP was less related to content alignment than to the strength or weakness of a district's instructional programming has significant implications. Many educators—and the public in general—assume that more demanding standards alone will improve student achievement. Our study, however, suggests that the greater rigor embedded in the new Common Core State Standards is likely to be squandered—with little effect on student achievement—if the standards themselves are not well-implemented, and if the content of the curriculum, instructional materials, classroom instruction, and professional development are not top-notch, integrated, and consistent with the standards.

This finding also has implications for a variety of high-profile reform strategies and governance models. The city school systems examined throughout this project included a mixture of governance models, ranging from mayor-controlled systems to more traditional district structures. Yet, what appears significant in these varied organizational models has less to do with who controls the systems and more to do with the actions those individuals take 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 (if to differing extents) in both improving and nonimproving districts. The broader lesson may be that these reforms or conditions are not likely to improve student achievement unless they directly serve the instructional program of the students.

What may have also emerged from this study is further evidence that progress is possible when districts act systematically at scale rather than trying to improve one school at a time. Moreover, our study clarified 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 at the tactical level, with programs and practices adopted in the pursuit of higher academic achievement.

Finally, each urban school system had its own history of reforms, with differing cultures, politics, capacities,

and personalities shaping the sometimes erratic nature of urban school reform. It became apparent that a district's ability to accurately and objectively gauge its own place in the reform process, its capacities, and when and how to transition to new approaches or theories of action is critical to continuous improvement in student achievement.

Conclusion

This study suggests that there is increasing reason to be optimistic about the future of urban public education, not simply because large-city schools are making significant progress (which they are), but because this progress appears to be the result of purposeful and coherent reforms.

In the long run, we will need to do more than explain post hoc why urban school systems did or did not improve. We will need to be able to predict these improvements and then produce them. The study detailed in this *Outlook* brings 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 to avoid forecasting improvement for its own sake, and to instead have confidence that we are identifying and acting on the appropriate levers in raising student achievement in large-city school districts. If we lack this confidence, there may be reason to think that gains are being driven by unknown reasons and that large-city school systems may be pursuing the wrong reforms. As NAEP trend lines grow longer, as more urban districts participate in the TUDA program, and as the research base expands, our understanding of the factors that will spur better performance in urban districts will improve.

Notes

1. Michael Casserly et al., *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 American Institutes for Research, Fall 2011), www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/4/Pieces%20of%20the%20Puzzle_FullReport.pdf (accessed June 29, 2012).

2. Ibid.

3. Ibid.

4. Ibid.

5. For subscale trends, we looked at the period from 2003 to 2007.

6. A recent state investigation of the Atlanta Public Schools found evidence of cheating on Georgia's Criterion-Referenced Competency Tests, but the investigative report presented no evidence of tampering with the NAEP and made no mention of Atlanta's progress on the 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, and collecting and safeguarding NAEP assessment data to guarantee the accuracy and integrity of results. In addition, an internal investigation by the National Center for Education Statistics found no evidence that NAEP procedures in Atlanta had been tampered with.

7. Full population estimates are statistical projections used by the National Center for Education Statistics to estimate what NAEP reading and math scores would look like if all students were included in the sample.