

Children First and Student Outcomes: 2003-2010

James J. Kemple, Research Alliance for New York City Schools,<sup>1</sup> New York University

Prepared for the New York City Education Reform Retrospective Project

November 2010

This paper will be published in a slightly edited form in

*Education Reform in New York City:*

*Ambitious Change in the Nation's Most Complex School System,*

to be published by Harvard Education Press in spring 2011

**DRAFT FOR CONFERENCE INVITEES ONLY**

**EMBARGOED UNTIL 9 AM NOVEMBER 10, 2010**

## INTRODUCTION

As discussed in other chapters in this volume, the New York City Department of Education’s (DOE’s) Children First reforms have evolved over time into a complex set of strategies aimed explicitly at changing the dynamics of school leadership, instructional empowerment, and accountability. The cornerstones of the reforms include new organizational structures, internal and external support mechanisms, and a performance management system that did not exist prior to 2003. The resulting web of initiatives has placed the New York City school system on the cutting edge of educational reform, eventually drawing attention and even imitation from districts across the state and the nation.

The primary bases for assessing the progress of the system at all levels have been measures of student performance in the form of the New York State mathematics and English language arts (ELA) tests for elementary and middle schools and graduation rates and Regents examination results for high schools. Most notably, 85 percent of the annual school progress report score is derived from student achievement measures.<sup>2</sup> Each year the DOE releases its analysis of these indicators and maps trends over time. Recent reports from the DOE have indicated that, on average, the city’s schools have made significant progress both on test score measures and on high school completion rates.<sup>3</sup> The DOE often cites these indicators as evidence that Children First reforms have had a positive effect on the quality of education provided to the city’s children.

To date, however, the field has lacked a rigorous and independent analysis of the extent to which these improved test scores and graduation rates reflect Children First effects *per se* or are instead artifacts of prior reform efforts or of other influences during the Children First era (2003-2010). Indeed, the decade of 2001–2010 saw an unprecedented array of educational improvement initiatives at the national, state, and local levels. Nationally, the No Child Left Behind Act of 2002 (NCLB) generated an intense focus on state, district, and school accountability. New York State and New York City helped frame this renewed focus on standards-based accountability with the development of more rigorous statewide standards and assessments during the 1990s and early 2000s (prior to Children First) and by instituting reforms aimed at improving opportunities for students to learn and to meet these standards during the Children First years.

To explore the distinctive effects of Children First reforms on student outcomes, this chapter addresses four sets of questions:

- How do student test score trends in New York City since 2003 compare to trends prior to that point? How have New York City test score trends changed compared to other school districts in New York State during the same period? To what extent do these comparisons provide evidence of Children First effects on student test scores?
- How have test score trends changed for key subgroups of students and schools, including students with disabilities and schools serving a high percentage of low-income families? What do these trends suggest about the effect of Children First reforms on high-need populations?
- What is the relationship between students’ performance on the New York State assessments and their likelihood of graduating from a New York City high school within four years? How has this changed over time, and what does it suggest about whether these test scores are robust indicators of students’ preparation for high school?

- How have high school graduation rates changed in New York City for cohorts of students who began high school after 2002 compared to graduation rates for students who entered high school prior to the Children First reforms? How have graduation rates in New York City changed compared to those of other New York State school districts during the same period? To what extent do these comparisons provide evidence of Children First effects on high school graduation rates?

The key findings in this chapter are derived from what is known as a comparative interrupted time series analysis, a method used widely in education research and evaluation to assess the impact of broad systemic policies and interventions on student outcomes.<sup>4</sup> The methodology provides an assessment of the degree to which the unique constellation of reform activities and conditions established in New York City during the Children First era (2003–2010) are associated with systematic changes in test score or graduation trends compared to trends prior to 2003 and compared to trends in similar districts and schools in New York State during the same period.<sup>5</sup> This approach allows us to reasonably attribute systematic changes in outcome trends during the Children First era to elements of the reform initiative that mark important breaks with the recent past and that set New York City apart from other urban school districts across the state.

#### TEST SCORE TRENDS IN GRADES 4 AND 8: EVIDENCE OF REFORM EFFECTS

This section focuses primarily on trends in New York State assessments of English language arts (ELA) and mathematics from 1999 through 2010. This analysis will be supplemented by a summary of findings from the National Assessment of Education Progress (NAEP) for New York City between 2003 and 2009.

##### *Data Sources, Analytic Approach, and Analytic Issues*

The analyses of test score trends utilize school-level data compiled from the New York State Education Department (NYSED), the DOE, and the Common Core of Data (CCD) from the National Center for Education Statistics (NCES) at the U.S. Department of Education.<sup>6</sup> NYSED data on grade 4 and grade 8 ELA and math test scores are available for the school years 1998–99 through 2009–10 and include both the average scale score for each school and the percentage of students in each school that scored at designated performance levels aligned with statewide performance standards (NYSED, 2010).<sup>7</sup>

Consistent with the accountability requirements of NCLB and New York State, the analyses of test score trends in this paper focus on the percentages of students who scored at Level 3 or 4 on the state assessments, the level at which the New York State Board of Regents considers a student to have met the state’s proficiency standards.<sup>8</sup> In addition to their policy relevance, proficiency levels provide a transparent and intuitive benchmark for describing and assessing test score performance trends. However, proficiency rates can be misleading because they focus on a single point in the distribution of test scores. They may thus mask trends in overall performance or neglect increases or decreases in test scores at other points in the distribution. To address this important shortcoming, a future supplement to this paper will include an analysis of trends in scale scores on the state tests; preliminary findings from this analysis are consistent with those discussed in the current paper that focus on the percentage of students who scored at Level 3 or 4 proficiency levels.<sup>9</sup>

The sample of schools for this analysis includes 2,141 schools in New York State that tested students in grade 4 between 1999 and 2009, and 922 schools that tested students in grade 8 during the same period. This sample is divided into three groups: 1) schools in New York City; 2) schools in the four next-

largest districts in New York State (Buffalo, Rochester, Syracuse, and Yonkers), referred to as the Big Four districts; and 3) schools from all other districts in New York State. For the analysis of grade 4 test scores, the sample encompasses 609 schools from New York City, 119 schools from the Big Four, and 1,413 schools from the rest of New York State. For the analysis of grade 8 test scores, the sample encompasses 204 schools from New York City, 51 schools from the Big Four, and 667 schools from the rest of New York State. Only schools with test score data for a minimum of three years prior to the 2002–03 school year and a minimum of three years during and after the 2002–03 school year were included in this sample.<sup>10</sup>

Table 1 provides a summary of characteristics of the schools in the test score analysis sample and illustrates similarities and differences among the three groups of schools. As discussed below, the analysis of test score trends must account for the observed differences in order to provide a valid comparison of test score trends.

DRAFT

Table 1  
 Characteristics of Schools with Tested Students  
 Averaged Across School Years 1998-1999 to 2001-2002<sup>a</sup>

Characteristic	4th Grade			8th Grade		
	New York City	Big Four	Rest of New York State	New York City	Big Four	Rest of New York State
<b>Student Demographic Characteristics</b>						
Race/ethnicity (%)						
African American	34.1	51.9	8.1	34.3	51.4	7.7
Hispanic	37.9	19.2	6.1	34.9	17.9	5.5
White	16.1	25.8	82.6	17.9	27.8	83.8
Asian or Native American	11.9	3.1	3.2	12.9	2.9	3.0
Gender (%)						
Male	50.8	50.7	51.4	51.1	50.5	51.4
Female	49.2	49.3	48.6	48.9	49.5	48.6
<b>Student Free/Reduced-Price Lunch Eligibility</b>						
Students Eligible (%)	76.9	75.0	23.9	74.8	74.1	20.3
Schools with:						
81-100 Percent of Students Eligible (%)	59.1	43.2	1.5	57.4	47.3	0.9
50-80 Percent of Students Eligible (%)	26.2	47.5	12.1	27.8	38.9	5.4
0-49 Percent of Students Eligible (%)	14.7	9.3	86.4	14.9	13.8	93.6
<b>Average Student Enrollment and Students Tested</b>						
Enrollment: All Grades	831	574	494	1069	759	640
Enrollment: Tested Grade (4 <sup>th</sup> or 8 <sup>th</sup> ) <sup>b</sup>	122	76	87	294	134	176
Tested Students (4 <sup>th</sup> or 8 <sup>th</sup> )	118	75	86	275	128	170
Tested Students Classified for Special Education (%)	12.7	17.6	12.0	12.2	18.0	13.4
<b>NYS English Language Arts Test Performance</b>						
Average Scale Score	638.8	635.5	662.4	689.6	685.0	704.5
Students Scoring at Level 3 or 4 (%)	43.7	39.2	68.5	35.0	25.8	52.6

(continued)

Table 1 (continued)

Characteristic	4 <sup>th</sup> Grade			8 <sup>th</sup> Grade		
	New York City	Big Four	Rest of New York State	New York City	Big Four	Rest of New York State
<b>NYS Mathematics Test Performance</b>						
Average Scale Score	637.8	634.7	663.0	690.7	683.1	716.2
Students Scoring at Level 3 or 4 (%)	52.6	48.6	80.0	26.8	19.9	51.4
<b>NYS Composite Test Performance</b>						
Average Scale Score	638.3	635.1	662.7	690.1	684.0	710.4
Percent of Schools with: <sup>c</sup>						
Low Proficiency Rate (%)	25.6	26.0	0.4	36.6	58.6	2.3
Middle Proficiency Rate (%)	45.2	59.7	9.6	36.8	29.6	15.0
High Proficiency Rate (%)	29.2	14.3	90.0	26.6	11.8	82.7
<b>Sample size</b>	609	119	1,413	204	51	667

SOURCE: Research Alliance calculations from New York State Education Department test score reports and National Center for Education Statistics Common Core of Data.

NOTES: Sample includes all schools in New York State with at least three years of test score data prior to 2003 and at least three years of test score data after 2003.

Student demographic characteristics, average scale scores, and percent of students scoring at Level 3 or 4 are weighted by the number of students tested in the grade.

Missing values for percent free and reduced-price lunch, number male, and number female for individual schools were imputed using values from surrounding years and taking the average. In cases where there was only one surrounding year (e.g., a missing value for 1998-1999), the value for the non-missing year replaced the missing value.

<sup>a</sup>Characteristics are averaged for schools in New York City, the Big Four, and the rest of New York State, respectively; the characteristics have been averaged across the four academic school years 1998-1999 to 2001-2002.

<sup>b</sup>The number of students tested and percent of schools with students scoring at Level 3 or 4 come from the mean of both the English language arts and mathematics tests.

<sup>c</sup>Proficiency categories were determined by examining the distribution of the percent of students scoring at Level 3 or 4 for New York City schools and the Big Four schools in 1999 to 2002. The proficiency categories for 4th grade are as follows: Lowest Proficiency Rate = ≤ 35% of students scoring at Level 3 or 4; Middle Proficiency Rate = 36-60% of students scoring at Level 3 or 4; and Highest Proficiency Rate = > 60% of students scoring at Level 3 or 4. The proficiency categories for 8th grade are as follows: Lowest Proficiency Rate = ≤ 20% of students scoring at Level 3 or 4; Middle Proficiency Rate = 20-40% of students scoring at Level 3 or 4; and Highest Proficiency Rate = > 40% of students scoring at Level 3 or 4.

The findings presented in this section of the chapter are based comparative interrupted time series analyses. The central strength of this methodology is that it accounts for many factors that may have produced changes in test scores in New York City instead of or in addition to the Children First reforms that began in 2002–2003. The goal of accounting for these factors is to identify the most viable alternative to what actually occurred in New York City, that is, to construct the best estimate of test score trends that were likely to have occurred for New York City schools in the absence of reforms that were instituted during the Children First era. This alternative trend is known as a counterfactual. The analyses conducted for this chapter are based on a particularly strong counterfactual in that it accounts for many important alternative influences on student test scores that may have been present over and above the reforms instituted in New York City during the Children First era. A strong counterfactual increases confidence that the findings from the analyses constitute rigorous evidence of effects, or lack of effects, from these reforms. These effects are reflected in the differences between New York City test score outcomes and the estimated counterfactual test score outcomes.

There are several potential influences on test scores that must be controlled for by the comparative interrupted time series analysis:

- **Reforms that were underway prior to Children First.** Test scores in New York City (and across New York State) were improving, virtually linearly, even before the Children First reforms. Thus, it is possible that these trends would have continued even if the Children First reforms had never been developed or implemented. The interrupted time series analysis isolates changes in test score trends that occurred New York City from 2003–2010 over and above what would have occurred had the pre-2002 trends continued.
- **National or state policies and reforms, such as No Child Left Behind (NCLB), aimed at improving schools across New York State and the country as a whole.** It is possible that the accountability mandates and school improvement initiatives required under NCLB beginning in 2002 produced improvements in student test scores independently of the reforms initiated under Children First. Similarly, there may be other federal or state policies aimed at school improvement that caused test scores in New York City during the period from 2003 through 2010. The comparative interrupted time series analysis isolates changes in test scores that occurred in New York City during this period over and above those that occurred in other districts that were subject to the same federal and state policies, mandates, and reform initiatives.
- **Changes in the state test, scoring methods, or performance criteria and increasing familiarity with the assessments and their frameworks.** It is possible that the state assessments in ELA and math became easier over time or that teachers and students became increasingly familiar with test content, scoring methods, and performance criteria. By comparing test score trends in New York City schools with those of other schools in New York State (schools that used the same ELA and math tests for students in grades 4 and 8 and were subject to the same scoring methods and standards over time), the method can hold constant the independent effect of changes in the test or scoring criteria.<sup>11</sup>

- Differences in school or district characteristics between New York City and other districts that may influence test scores trends independent of Children First reforms.** As illustrated in Table 1, there were several differences in the characteristics of schools from New York City and the Big Four districts prior to the introduction of Children First reforms in 2003. These differences may have enhanced or limited the capacity of schools to make progress during subsequent years. A higher concentration of especially challenging schools or of schools highly susceptible to change in New York City could account for some or all the differences in test score performance. The comparative interrupted time series analysis accounts for many of these potential influences by including school characteristics in the statistical models as control variables.

In short, the counterfactual for this analysis is the estimated New York City test score trend for 2003–2009 controlling for: 1) the continuation of test score trends underway in New York City schools prior to 2003; 2) changes in test score trends before and after 2003 in the Big Four districts; and 3) differences in school characteristics between New York City and the Big Four. This counterfactual represents the best estimate of test score trends that were likely to have occurred for New York City schools in the absence of reforms that were instituted during the Children First era. Thus, the best evidence of effects from these reforms is derived from the difference between the test score trends that actually occurred in New York City and the estimated counterfactual trends.

Even with a particularly strong counterfactual, the comparative interrupted time series analysis has some limitations. First, it is possible that other events and initiatives, unique to New York City and completely unrelated to Children First reforms, produced changes in student outcomes during this period. For example, as discussed elsewhere in this volume, New York City experienced a substantial influx of additional education funding throughout this period, particularly after 2007 when Fair Student Funding formulas were being implemented.<sup>12</sup> However, a large body of prior research finds that increases in school spending, in and of themselves, have little impact on student achievement.<sup>13</sup> Much of this research does suggest, however, that specific uses of additional funding in schools may have an effect on student achievement. In the context of this analysis, therefore, it is more likely that the interaction between Children First reforms and the manner in which they influenced specific uses of additional funding could account for changes in student achievement, rather than just the funding itself. Thus, the influence of additional funding may be seen, at least in part, as an artifact of Children First rather than an independent factor that would have occurred without Children First.

A second limitation is that although the analysis can shed light on the overall effect of reforms that occurred during the Children First era, it cannot isolate the specific features of the Children First reforms that may have been most responsible for these effects. As discussed elsewhere in this volume, the reforms have had many components, and these were designed to interact with each other and with other policies and school conditions. By exploring the variation in Children First effects across schools within New York City and by collecting data on the implementation of specific reform activities, further research can expound on the mechanisms that enhanced or limited the impact of Children First reforms; studies of this nature are being undertaken by a number of researchers, including the Research Alliance for New York City Schools.<sup>14</sup>

As a final note before discussing the findings, the paper also presents differences and similarities between New York City and the remaining New York State districts in an effort to provide a further context for interpreting test score trends in New York City. These comparisons control statistically for test score trends prior to 2003 and for differences in school demographic characteristics. However, there are dramatic demographic and prior-performance differences between New York City and the rest



of the state; greater caution should be exercised when using these comparisons to infer evidence about effects of reforms undertaken during the Children First era.

*Effects on New York State Assessment Test Scores: 2003–2010*

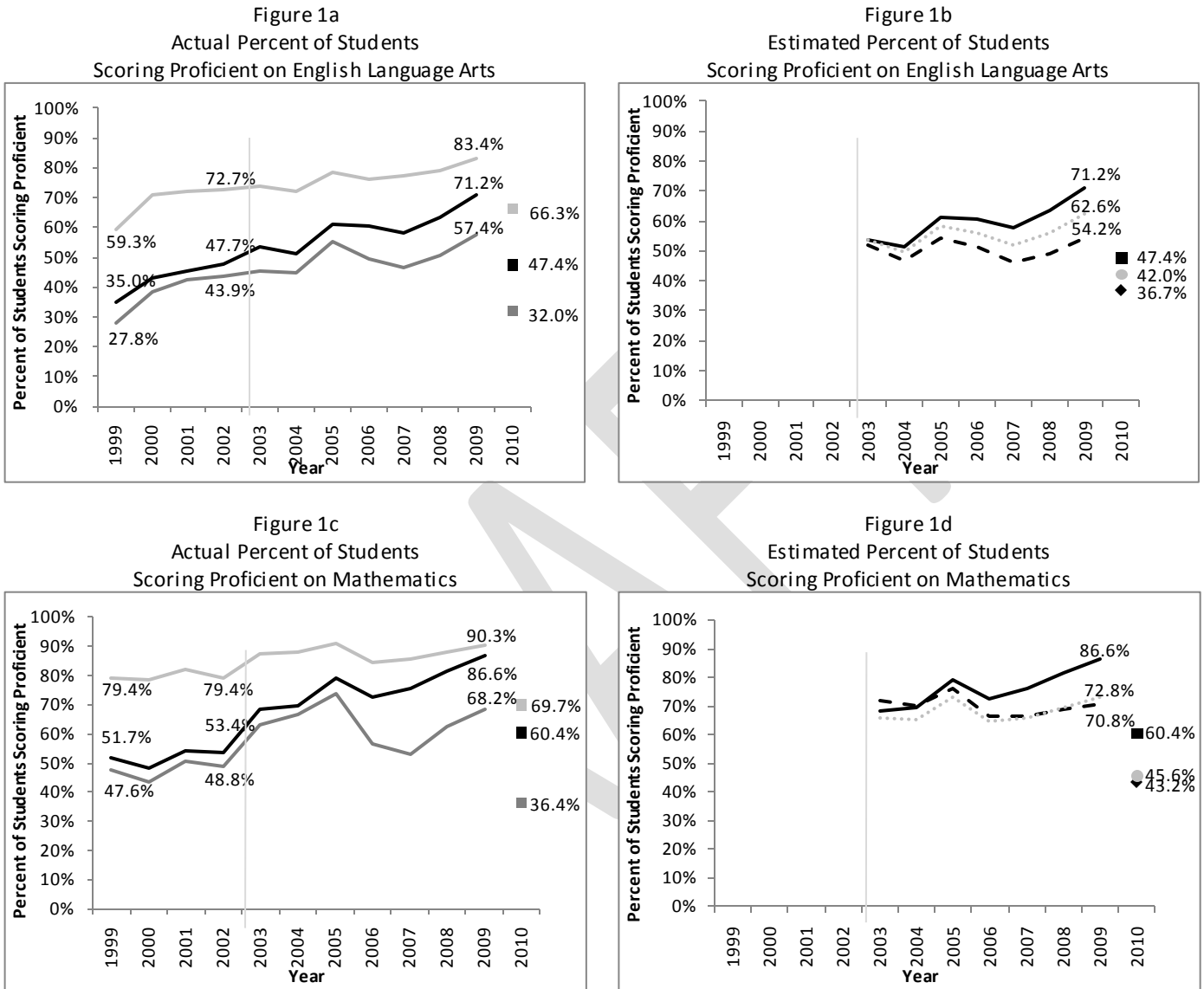
In 2010 the New York State Board of Regents made significant changes in the performance levels used to classify students as meeting the state’s proficiency standards. Therefore, we treat the scores for school year 2009–2010 as independent from those of prior years rather than as a continuation of preceding trends.<sup>15</sup>

School years 2002–03 through 2008–09

Student proficiency rates in New York City increased substantially from 2002–03 to 2008–2009. However, some of this increase was likely to have been due to influences outside the reforms instituted as part of Children First, such as prior reforms, reforms being initiated at the state and federal level, and a growing familiarity with the tests and scoring rubrics. After accounting for these factors and for the influence of school demographic characteristics, however, the analysis yields persuasive evidence that the reforms instituted during the Children First era did produce improvements in student test scores. These effects persisted and increased throughout the period.

Figures 1a and 1b illustrate the key elements of the comparative interrupted time series analysis and present detailed findings for grade 4 ELA test scores. Figure 1a shows the average percentage of grade 4 students who scored at Level 3 or Level 4 (defined as proficient by state standards) among New York City schools, among schools in the Big Four districts, and among schools in the remaining school districts in New York State. The figure presents these averages for 1998–99 through 2009–10. (This figure is nearly identical to those presented by the DOE in its report on test score trends between 2002 and 2009 (DOE, 2010a, slides 10 and 11)). The proficiency rate increased from an average of 48% in 2002 to 71% in 2009 (an average increase of over 3 percentage points per year over seven years).

Figure 1  
Percent of Students Scoring at Level 3 or 4  
Grade 4, English Language Arts and Mathematics  
School Year 1998-1999 to School Year 2009-2010



SOURCE: See Table 2.

NOTES: See Table 2.

Actual percent of students scoring proficient is weighted by the number of students tested in the grade.

The Estimated Counterfactual is the regression-adjusted proficiency rate in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the Big Four school districts, and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is City.

The Adjusted New York State is the regression-adjusted test score trend in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the remaining New York State districts (other than New York City and the Big Four), and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education.

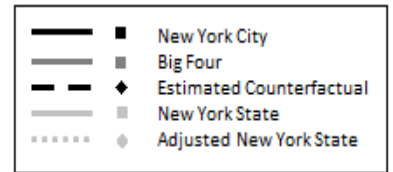


Figure 1a also illustrates several of the factors that may account for the increases that occurred in New York City independent of the reforms that were implemented as part of Children First starting in 2003. Most notably, test scores in New York City schools were already on the rise prior to 2003. The figure shows that the proficiency rate increased from 35% in 1999 to 48% in 2002 (an average increase of just over 3 percentage points per year over four years). This rate of growth during the pre-Children First era is consistent with the increases that occurred in subsequent years. Among the Big Four districts, the proficiency rate increased from 28% in 1999 to 44% in 2002 (an average increase of 4 percentage points per year over four years), and among the remaining districts the rate increased from 59% to 73% (3.5 percentage points per year). These annual growth rates are the same as or larger than those exhibited for New York City Schools during the same period. The figure also suggests that the growth rates in other districts slowed somewhat after 2002 relative to those for New York City Schools (almost 2.5 percentage points per year for New York City versus 1.7 percentage points per year in the Big Four districts and 1.4 percentage points per year in the remaining New York State districts).

The estimated counterfactual for this analysis is shown in Figure 1b as the dashed line. The differences between the grade 4 ELA test score trend for New York City schools and the estimated counterfactual trend provide evidence of effects on test scores from the reforms instituted during the Children First era. The figure shows a widening gap between the actual proficiency rate among New York City schools and the estimated counterfactual. By 2009, the proficiency rate for New York City schools was estimated to be an average of 17 percentage points higher than the average for estimated counterfactual. (This difference is statistically significant at the  $p < 0.001$  level.) Although this difference is smaller than the overall increase in proficiency rates in New York City between 2002 and 2009, it provides a more rigorous indication of the increases associated with Children First reforms specifically.

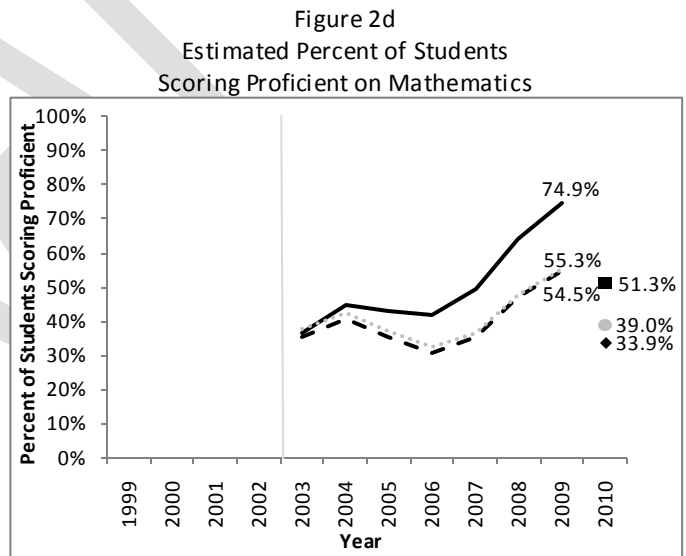
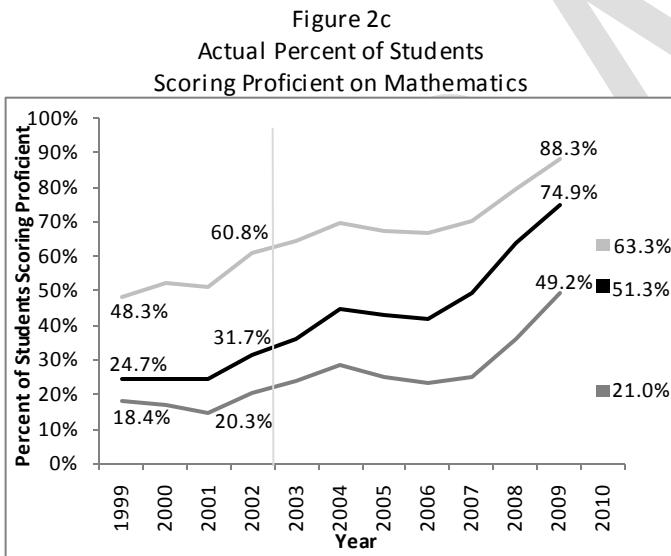
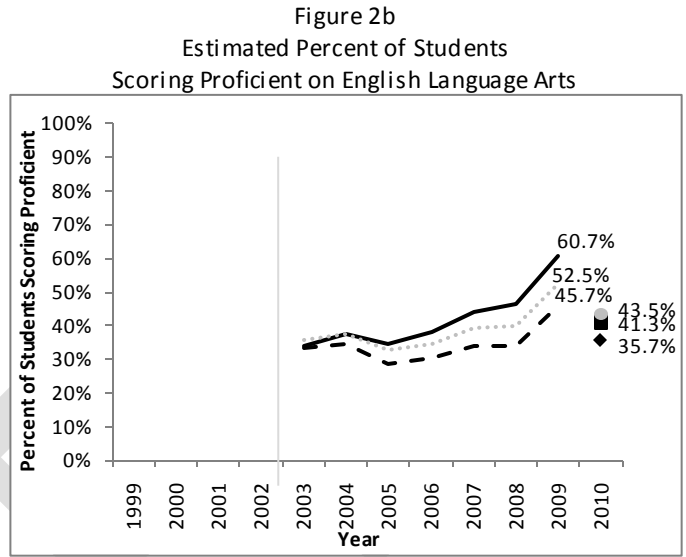
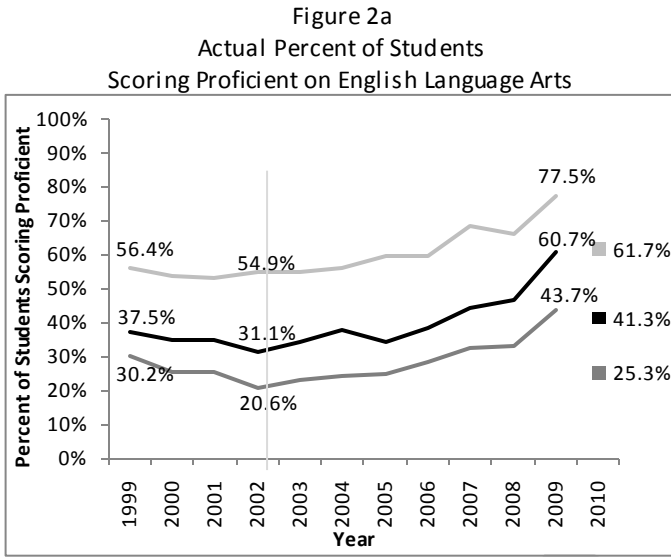
The dotted line represents the remaining New York State districts after controlling for prior trends and demographic differences. It shows that grade 4 ELA proficiency rates in New York City outpaced the statistically adjusted New York State rates. By 2009, New York City proficiency rates were estimated to be about 9 percentage points higher than the statistically adjusted rates for schools from the remaining New York State districts. (This difference is statistically significant at the  $p < 0.001$  level.)

Figures 1c and 1d show a similar pattern for grade 4 math proficiency rates. By 2009, the grade 4 math proficiency rate for New York City schools was estimated to be an average of 16 percentage points higher than the estimated counterfactual. This difference is statistically significant at the  $p < 0.001$  level and provides evidence of positive effects on grade 4 math proficiency rates from the reforms instituted during the Children First era.

Grade 4 math proficiency rates for New York City schools also outpaced the adjusted rates for schools from the remaining New York State districts. By 2009, the grade 4 proficiency rate for New York City schools was estimated to be 14 percentage points higher than the adjusted rates for schools from the remaining New York State districts (statistically significant at the  $p < 0.001$  level).

In general, the findings for grade 8 ELA and math test scores were similar to those for grade 4. Proficiency rates in both subjects were increasing steadily in New York City between 2003 and 2009. Although some of the increase was likely do to influences outside the Children First reforms, the analysis provides compelling evidence of positive effects on grade 8 ELA and math proficiency rates from the reforms instituted during the Children First era.

Figure 2  
Percent of Students Scoring at Level 3 or 4  
Grade 8, English Language Arts and Mathematics  
School Year 1998-1999 to School Year 2009-2010



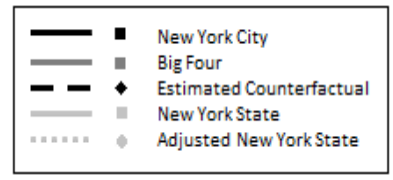
SOURCE: See Table 2.

NOTES: See Table 2.

Actual percent of students scoring proficient is weighted by the number of students tested in the grade.

The Estimated Counterfactual is the regression-adjusted proficiency rate in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the Big Four school districts, and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is City.

The Adjusted New York State is the regression-adjusted test score trend in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the remaining New York State districts (other than New York City and the Big Four), and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education.



Figures 2a and 2c show that grade 8 ELA and math proficiency rates were climbing steadily for New York City schools, schools from the Big Four districts, and schools from the other districts in New York State. At the same time, Figure 2b indicates that the grade 8 ELA proficiency rates for New York City schools outpaced the estimated counterfactual (represented by the dashed line) and that this difference increased systematically over time. By 2009, the grade 8 ELA proficiency rate for New York City schools was 15 percentage points higher than the estimated counterfactual (statistically significant at the  $p < 0.01$  level).

Figure 2d also shows evidence of Children First effects on grade 8 math proficiency rates. By 2009, the grade 8 math proficiency rate for New York City schools was estimated to be an average of 20 percentage points higher than the estimated counterfactual (statistically significant at the  $p < 0.001$  level). Figures 2b and 2d show an increasing difference between New York City proficiency rates and the adjusted rates for the remaining New York State districts after controlling for prior trends and school demographic characteristics.

#### School year 2009–10

For 2010, the New York State Board of Regents raised the score that students needed to achieve in order to be classified as meeting the state’s proficiency standards (from Level 3 to Level 4). As a result, many fewer students across New York State were classified as proficient, even though they achieved scale scores that were the same or higher than those that would have met the 2008–09 standards. Because of this change, we analyzed 2009–10 test scores independently from the prior school years. The goal of the analysis is to assess whether the effects of Children First reforms were sustained despite dramatic changes in the state testing standards.

Table 2 shows estimated differences in proficiency rates in both 2008–09 and 2009–10 between New York City schools and the estimated counterfactual. The results in Table 2 indicate that for grade 4, the effect of Children First reforms on proficiency rates persisted through 2010 despite the drop in rates across the state. The effects on grade 4 ELA proficiency rates were somewhat smaller in 2010 than they were in 2009, while the differences for grade 4 math were somewhat larger.

Table 2  
 Percent of Students Scoring at Level 3 or 4  
 Grade 4 and Grade 8, English Language Arts and Mathematics  
 School Year 2008-2009 and School Year 2009-2010

Grade, Subject, and Year	New York City		Estimated Counterfactual <sup>a</sup>		Adjusted New York State <sup>b</sup>	
	Average	Average	Difference	Average	Difference	
<b>Grade 4</b>						
<b>English Language Arts</b>						
2009	71.2	54.2	17.0 ***	62.7	8.5 ***	
2010	47.4	36.7	10.8 ***	42.0	5.4 ***	
<b>Mathematics</b>						
2009	86.6	70.8	15.8 ***	72.8	13.8 ***	
2010	60.4	43.2	17.2 ***	45.6	14.8 ***	
Sample Size	609		119		1,413	
<b>Grade 8</b>						
<b>English Language Arts</b>						
2009	60.7	45.7	15.0 **	52.5	8.2 ***	
2010	41.3	35.7	5.6 *	43.5	-2.2 *	
<b>Mathematics</b>						
2009	74.9	54.5	20.4 ***	55.3	19.6 ***	
2010	51.3	33.9	17.4 ***	39.0	12.2 ***	
Sample Size	204		51		667	

SOURCE: Research Alliance calculations from New York State Education Department test score reports and National Center for Education Statistics Common Core of Data.

NOTES: Sample includes all schools in New York State with at least three years of test score data prior to 2003 and at least three years of test score data after 2003.

All values are weighted by the number of students tested in the grade.

A two-tailed t-test was applied to the estimated differences for each year. Robust standard errors are used to account for serial autocorrelation. Statistical significance levels are indicated as: ~ = p-value < 0.10; \* = p-value < 0.05; \*\* = p-value < 0.01 percent; \*\*\* = p-value < 0.001.

<sup>a</sup>The Estimated Counterfactual is the regression-adjusted proficiency rate in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the Big Four school districts, and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education. The difference between the actual New York City proficiency rate and the estimated counterfactual is presented as evidence of the effect of reforms instituted during the Children First era (2003-2010) in New York City.

<sup>b</sup>The Adjusted New York State is the regression-adjusted test score trend in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the remaining New York State districts (other than New York City and the Big Four), and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education.

The results for grade 8 are somewhat more mixed in 2010, particularly for ELA. While Children First effects persisted through 2010 in both grade 8 ELA and math, the effects for ELA were considerably smaller in 2010 than in 2009. The New York City proficiency rates in grade 8 ELA and math were higher in 2010 than the estimated counterfactual by 5.6 and 17.4 percentage points, respectively. Both of these differences were statistically significant at the  $p < 0.05$  or lower. However, the grade 8 ELA difference was nearly 10 percentage points smaller in 2010 than in 2009, and the grade 8 math difference was three percentage points lower.

Estimated differences in grade 8 ELA proficiency rates between New York City schools and the adjusted rates for the remaining New York State districts also declined from 2009 to 2010—and in the 2010 New York City rates were lower than those for the rest of New York State. The difference in grade 8 math proficiency rates between New York City schools and the adjusted rates for the remaining New York State districts also declined but continued to be positive and statistically significant.

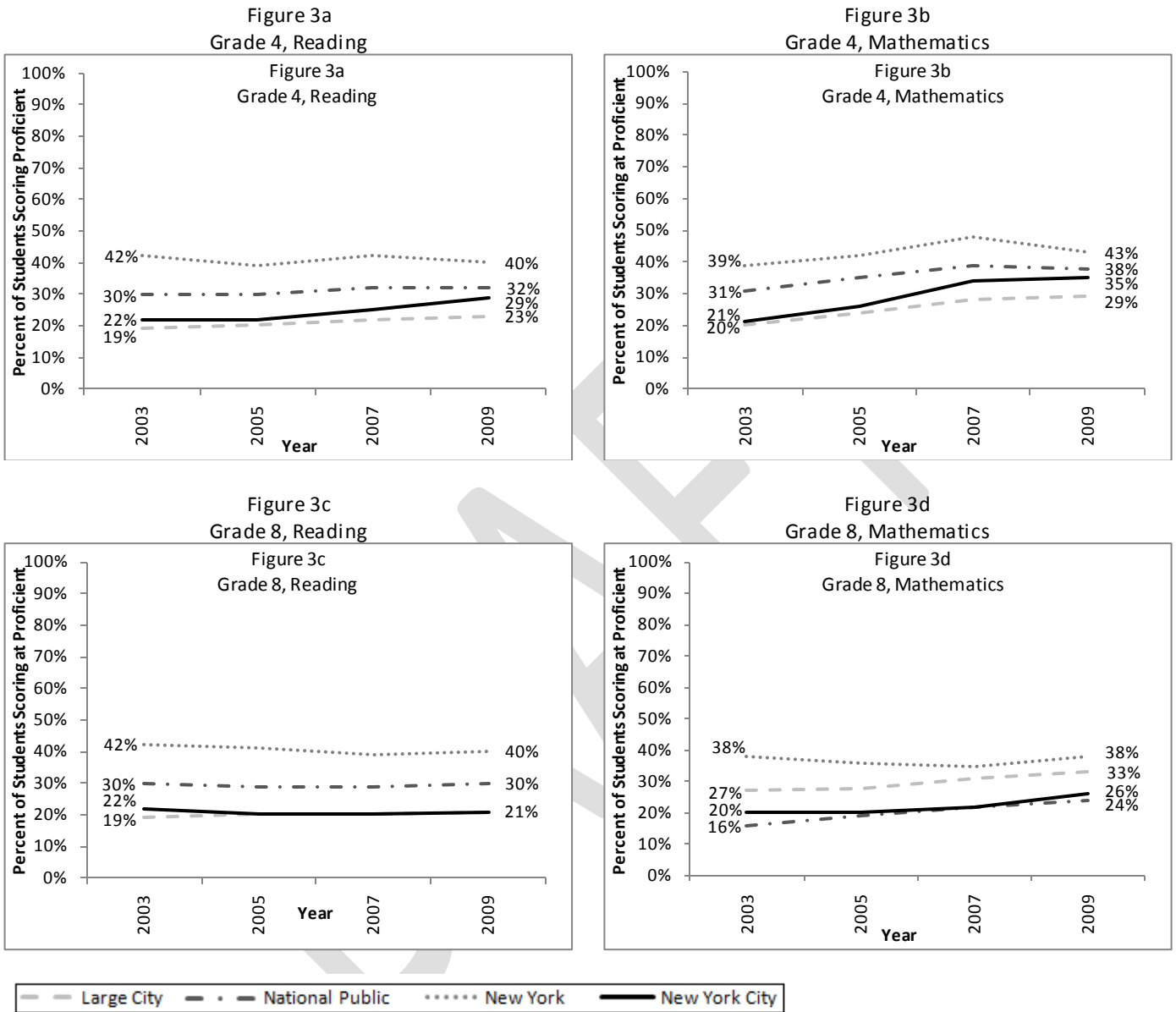
Thus, the findings for 2010 suggest that, although Children First effects appear to have persisted through the change in proficiency standards, the gains made by New York City schools up through 2009, relative to both the Big Four and the remaining districts throughout the state, were diminished to varying degrees by the move to higher standards. Because the score required to be classified as proficient increased, the smaller differences in 2010 suggest that the effects of Children First reforms may not have been as robust at the higher ranges of the test score distribution compared to the effects for 2009 when the proficiency standard was lower. As noted earlier, additional analyses are being conducted to examine effects across the distribution of scores and at other points in the distribution throughout the Children First era. Preliminary findings from these analyses indicate that the patterns of differences and similarities are similar to those presented here.

#### NAEP results for New York City: 2003 to 2009

The National Assessment of Educational Progress (NAEP) is the only national assessment of student achievement in the United States. Initiated in 1969, NAEP assessments are conducted periodically in reading, mathematics, science, writing, U.S. history, civics, geography, and other subjects. In addition to testing in statewide samples of schools the NAEP Trial Urban District Assessment (TUDA) has administered assessments in reading and math in selected large urban school districts including New York City. TUDA has been conducted every two years since 2003 and allows scores for New York City to be compared with scores from the rest of New York State as well as with other urban school districts and the nation as a whole.

Figures 3a and 3b show NAEP results in grade 4 reading and math from 2003 through 2009 for New York City, New York State, all urban districts in the TUDA sample, and all of the nation's public schools. Figures 3c and 3d show the NAEP results for grade 8. The figures illustrate trends in the percentage of students scoring at or above proficient.<sup>16</sup>

Figure 3  
Percent of Students Scoring Proficient on the National Assessment of Educational Progress (NAEP)  
Grade 4 and 8, Reading and Mathematics, 2003-2009



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.

The figures show that the NAEP results for New York City have been somewhat mixed. In general, New York City has performed below the national average and below the average for the rest of New York State in both reading and math and in both grade 4 and grade 8. However, New York City has scored above the average for urban districts in grade 4 reading and math, and has scored at or slightly above the average for urban districts in grade 8 reading and math. Grade 4 students in New York City have



shown significant improvements on both reading and math NAEP assessments since 2002. In grade 8, New York City proficiency levels have improved only in math. Grade 8 reading scores have remained flat over this seven-year period, which is consistent with the rest of the country and other urban districts.

Because NAEP data do not exist for New York City prior to 2002, they do not allow a comparative interrupted time series analysis.<sup>17</sup> However, NAEP provided us with statistics on differences in changes in NAEP proficiency rates over time between New York City and the rest of New York State, allowing a systematic comparison between the two.<sup>18</sup> This information has the effect of controlling for initial differences between New York City and the rest of New York State in 2003 and then assessing the statistical significance of remaining differences in subsequent years.

Some important cautions should be noted. First and most importantly, NAEP is not aligned with the standards and curriculum frameworks developed by the New York State Board of Regents and so may not be measuring the same underlying constructs as the New York State assessments. In addition, NAEP does not carry negative or positive consequences either for schools and systems or for students, and therefore, schools and students might not take the NAEP as seriously as the state tests. NAEP is also administered only to a small stratified random sample of students within a small stratified random sample of schools within New York City and across New York State. As with assessments based on any sampling strategy, as opposed to a universally administered assessment, the results may be sensitive to sampling variation or to relatively low or high representations of subpopulations of schools or students.

For these reasons and because the NAEP-based comparisons cannot incorporate the features of the comparative interrupted time series analysis, it is not possible to draw any conclusions about the effect (or lack of effects) of Children First reforms on NAEP test scores. Because of the very high profile of NAEP and its nationwide testing program, however, they are presented here to offer an additional framework for understanding test score trends in New York City and how they are situated in both a statewide and national context.

Two notable patterns stand out. First, the overall proficiency rates on NAEP are much lower than on the New York State assessments (compare Figures 1 and 2 with Figure 3). This difference in results is likely due to a combination of differences in the content of the tests, in the standards that are used to designate proficiency levels, and in the consequences attached to student performance. To the extent that the NAEP assessments capture important aspects of what students should know and be able to do, Figure 3 indicates that in 2009, less than half of New York City students possessed this knowledge and these skills.

Second, New York City experienced a greater increase in proficiency rates than did the rest of New York State. The differences in increases were statistically significant for grade 4 reading and math and for grade 8 math. The difference was smaller and not statistically significant for grade 8 reading.

Between 2003 and 2009, in grade 4 reading, New York City proficiency rates increased by 7 percentage points, compared to a 2 percentage point drop for the rest of New York State. This difference of 9 percentage points is statistically significant at the  $p < 0.001$  level. For grade 4 math, New York City proficiency rates increased by 14 percentage points compared to a 4 percentage point increase for the rest of New York State (a difference of 10 percentage points, statistically significant at the  $p < 0.001$  level).

Figure 3 shows that proficiency rates in grade 8 reading actually declined somewhat for both New York City and the rest of New York State between 2003 and 2009; the difference is not statistically significant.

By contrast, increases in New York City proficiency rates in grade 4 math (6 percentage points between 2003 and 2009) did outpace the flat trend for the rest of New York State (statistically significant at the  $p < .01$  level).

#### VARIATION IN TEST SCORE TRENDS AND EVIDENCE OF EFFECTS FOR SUBGROUPS OF STUDENTS AND SCHOOLS

This section of the paper extends the comparative interrupted time series analysis of test scores based on the New York State assessments to explore evidence of Children First effects for subgroups of students and schools in New York City. The first set of analyses considers general education students and students with disabilities. The second set looks at subgroups of schools based on their concentration of students from low-income families. The goal of these analyses is to determine whether effects were widespread or were concentrated among certain subpopulations.

##### *Findings for General Education Students and Students with Disabilities*

Since 1999, NYSED has reported state test scores separately for students in general education programs and students enrolled in special education (students with disabilities).<sup>19</sup> Table 3 presents results from analysis of the test scores for these two subgroups of students. The table shows the difference in proficiency, as of 2009, between New York City schools and the estimated counterfactual.

The findings in Table 3 provide evidence of positive Children First effects both for grade 4 general education students and for grade 4 students with disabilities. The effects appear to be substantially larger, however, among students with disabilities. For example, by 2009 ELA proficiency rates for students with disabilities in New York City were an average of 35 percentage points higher than the estimated counterfactual. The difference for general education students was 11 percentage points. The differences in effects for math were even larger: 47 percentage points for students with disabilities compared to 12 percentage points for general education students.

Table 3  
Percent of Students Scoring at Level 3 or 4  
Grade 4 and Grade 8, English Language Arts and Mathematics  
General Education and Special Education Students  
School Year 2008-2009

Student Subgroup, Grade, and Subject	New York City		Estimated Counterfactual <sup>a</sup>		Adjusted New York State <sup>b</sup>	
	Average		Average	Difference	Average	Difference
<b>General Education Students</b>						
<b>Grade 4</b>						
English Language Arts	77.9		66.6	11.2 **	71.1	6.7 ***
Mathematics	90.7		79.1	11.6 *	75.0	15.7 ***
Sample Size	609		119		1,413	
<b>Grade 8</b>						
English Language Arts	65.8		52.4	13.4 *	57.1	8.7 ***
Mathematics	79.4		63.7	15.7 **	60.5	18.8 ***
Sample Size	204		51		667	
<b>Special Education Students</b>						
<b>Grade 4</b>						
English Language Arts	36.9		2.2	34.7 ***	9.6	27.3 ***
Mathematics	65.6		18.8	46.8 ***	44.0	21.5 ***
Sample Size	609		119		1,413	
<b>Grade 8</b>						
English Language Arts	21.3		10.1	11.2 **	20.0	1.3
Mathematics	41.9		8.0	33.9 ***	23.2	18.8 ***
Sample Size	204		51		667	

SOURCE: See Table 2.

NOTES: See Table 2.

<sup>a</sup>The Estimated Counterfactual is the regression-adjusted proficiency rate in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the Big Four school districts, and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education. The difference between the actual New York City proficiency rate and the estimated counterfactual is presented as evidence of the effect of reforms instituted during the Children First era (2003-2010) in New York City.

<sup>b</sup>The Adjusted New York State is the regression-adjusted test score trend in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the remaining New York State districts (other than New York City and the Big Four), and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education.

Children First also appears to have had positive effects on grade 8 proficiency rates in each subject for both general education students and students with disabilities. The effects on grade 8 ELA proficiency rates were similar for general education students and students with disabilities. The effects on grade 8 math proficiency rates were more pronounced for students with disabilities.

It is not clear what explains the larger effects for students with disabilities. Unlike previous efforts to track student progress in New York City, the accountability framework that formed a central element of Children First reforms explicitly incorporated test scores for special education students into its school progress reports. However, this focus was also incorporated into the NCLB accountability requirements that were being implemented throughout the state during this period. Further research is needed to explore ways in which Children First reforms may have placed additional emphasis on services for students with disabilities.

#### *Findings for High- and Low-Poverty Schools*

An important goal of the Children First reforms has been to improve schools serving large proportions of traditionally low-performing and high-need under-served students. One well-documented indicator of these populations is economic status. For this reason, we analyzed differences in achievement trends for subgroups of schools based on the percentage of students eligible for free or reduced-price lunch before the Children First era. As with the previous findings, the results for the school subgroups are based on the comparative interrupted time series method that accounts for pre-2003 test score trends and other demographic differences across schools.

For the purposes of this analysis, schools from across New York State were categorized into three groups:

- Highest poverty rate: Schools with an average of more than 80% of their students eligible for free or reduced-price lunch between 1999 and 2002;
- Middle poverty rate: Schools with an average of between 51% and 80% of students eligible; and
- Lowest poverty rate: Schools with an average of less than or equal to 50% eligible.

Tables 4 and 5 present findings for grade 4 and grade 8, respectively, for the three subgroups of schools and for both 2009 and 2010.

Table 4  
 Percent of Students Scoring at Level 3 or 4  
 Grade 4, English Language Arts and Mathematics  
 School Year 2008-2009 and School Year 2009-2010  
 for Subgroups of Schools Defined by Percent of Students Eligible for Free or Reduced-Price Lunch

Percent of Students Eligible, Subject, and Year	New York City		Estimated Counterfactual <sup>a</sup>		Adjusted New York State <sup>b</sup>	
	Average	Average	Difference	Average	Difference	
<b>81-100% of Students Eligible</b>						
<b>English Language Arts</b>						
2009	65.1	49.2	16.0 **	30.5	34.6 *	
2010	39.3	28.6	10.7 ***	21.1	18.2 **	
<b>Mathematics</b>						
2009	83.3	65.4	17.8 *	37.1	46.2 ***	
2010	52.9	35.7	17.2 ***	23.3	29.6 ***	
Sample Size	360		51		20	
<b>50-80% of Students Eligible</b>						
<b>English Language Arts</b>						
2009	75.1	58.0	17.1 **	54.4	20.7 ***	
2010	52.1	40.8	11.3 ***	41.1	10.9 ***	
<b>Mathematics</b>						
2009	89.1	72.3	16.8 *	75.7	13.4 **	
2010	65.0	46.5	18.5 ***	47.9	17.1 ***	
Sample Size	159		56		164	
<b>0-49% of Students Eligible</b>						
<b>English Language Arts</b>						
2009	89.4	70.2	19.3	92.7	-3.2	
2010	72.3	60.7	11.6 ~	69.8	2.5 ~	
<b>Mathematics</b>						
2009	96.0	85.6	10.4	94.7	1.3	
2010	82.9	68.8	14.0 **	70.0	12.8 ***	
Sample Size	90		11		1,175	

SOURCE: See Table 2.

NOTES: See Table 2.

The percent of students eligible categories were determined by examining the distribution of the percent of students eligible for free and reduced-price lunch for New York City schools and the Big Four schools in 1999 to 2002.

<sup>a</sup>The Estimated Counterfactual is the regression-adjusted proficiency rate in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the Big Four school districts, and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education. The difference between the actual New York City proficiency rate and the estimated counterfactual is presented as evidence of the effect of reforms instituted during the Children First era (2003-2010) in New York City.

<sup>b</sup>The Adjusted New York State is the regression-adjusted test score trend in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the remaining New York State districts (other than New York City and the Big Four), and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education.

Table 5  
 Percent of Students Scoring at Level 3 or 4  
 Grade 8, English Language Arts and Mathematics  
 School Year 2008-2009 and School Year 2009-2010

for Subgroups of Schools Defined by Percent of Students Eligible for Free or Reduced-Price Lunch

Percent of Students Eligible, Subject, and Year	New York City		Estimated Counterfactual <sup>a</sup>		Adjusted New York State <sup>b</sup>	
	Average		Average	Difference	Average	Difference
<b>81-100% of Students Eligible</b>						
<b>English Language Arts</b>						
2009	51.4		37.9	13.5 *	n/a	n/a
2010	31.0		25.3	5.6 *	n/a	n/a
<b>Mathematics</b>						
2009	68.3		46.7	21.5 ***	n/a	n/a
2010	40.4		24.3	16.2 ***	n/a	n/a
Sample Size	117		24		6	
<b>50-80% of Students Eligible</b>						
<b>English Language Arts</b>						
2009	65.2		38.8	26.4 **	45.7	19.6 *
2010	45.9		37.6	8.3	40.6	5.2 ~
<b>Mathematics</b>						
2009	77.6		48.2	29.4 ***	52.9	24.7 *
2010	56.2		34.8	21.5 ***	35.9	20.4 ***
Sample Size	57		20		35	
<b>0-49% of Students Eligible</b>						
<b>English Language Arts</b>						
2009	77.6		n/a	n/a	74.1	3.5
2010	59.5		n/a	n/a	62.9	-3.4 ~
<b>Mathematics</b>						
2009	87.7		n/a	n/a	73.8	13.9 **
2010	70.1		n/a	n/a	56.5	13.6 ***
Sample Size	30		7		607	

SOURCE: See Table 2.

NOTES: See Table 2.

The subgroups of schools were defined based on the average percent of students eligible for free and reduced-price lunch between 1999 and 2002.

"n/a" denotes that sample sizes were less than 10 and judged to be too small to calculate reliable estimates.

<sup>a</sup>The Estimated Counterfactual is the regression-adjusted proficiency rate in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the Big Four school districts, and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education. The difference between the actual New York City proficiency rate and the estimated counterfactual is presented as evidence of the effect of reforms instituted during the Children First era (2003-2010) in New York City.

<sup>b</sup>The Adjusted New York State is the regression-adjusted test score trend in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the remaining New York State districts (other than New York City and the Big Four), and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education.

In general, the findings for grade 4 ELA and math suggest that while there is solid evidence of Children First effects across schools with different poverty rates, the reforms do not appear to have substantially narrowed the gap between the highest- and lowest-poverty schools. Table 4 indicates that Children First effects on grade 4 ELA and math proficiency rates are generally consistent across the three subgroups of schools. For example, in 2009, the difference in grade 4 ELA proficiency rates between New York City and the estimated counterfactual ranged from 16 percentage points for the highest-poverty schools to 19 percentage points for the lowest-poverty schools. Although somewhat smaller than the differences in 2009, the differences in 2010 remained positive. For the highest- and middle-poverty schools, all of the grade 4 ELA differences in proficiency rates were statistically significant in both 2009 and 2010. The small sample sizes for the estimated counterfactual in the lowest-poverty group of schools contributed to the lack of statistical significance even though the differences were similar to those in the other groups.

The pattern of findings for grade 8 is more complex and mixed.<sup>20</sup> Table 5 indicates that effects on grade 8 ELA and math proficiency rates were concentrated among the middle- and high-poverty schools. A comparison of the estimated counterfactual proficiency rates suggests that, in the absence of Children First reforms, there was likely to be very little difference in proficiency rates between the middle- and high-poverty schools in New York City. Thus, because the effects for the middle-poverty schools were somewhat larger than those for the high-poverty schools, the achievement gap between these types of schools may have increased during the Children First era. In general, the effects for both ELA and math in grade 8 declined between 2009 and 2010. Given the smaller sample sizes in the subgroups of schools and the changes in effects over time, further analysis is needed to better understand the nature of these findings.

#### STUDENT TEST SCORES AND HIGH SCHOOL GRADUATION RATES

A critical concern regarding the use of test scores to assess student achievement or to gauge school improvement is that test scores do not necessarily capture the skills and knowledge that students need to succeed in other areas of their education and development. For example, test score improvements may signal that students are becoming more skilled at answering specific types of questions or that teachers' familiarity with the tests enables them to prepare students for specific items rather than give them the underlying knowledge and skills. To the extent such factors drive test performance, increased scores are unlikely to reflect true learning gains. Similarly, they may be unlikely to predict subsequent achievement or school completion.

This section provides an analysis of the relationship between students' grade 8 test scores and their likelihood of graduating from high school within four years. This analysis utilizes student-level data to shed light on the extent to which students' performance on the New York State assessments of ELA and math indicates their preparation for high school work and their likelihood of earning a high school diploma. This analysis also examines the extent to which the test score-graduation relationship changed over the period before and after Children First reforms began to be implemented.

#### *Data Sources, Samples, and Analysis*

The data for these analyses cover the period from October 2001 through August 2009 and include information about individual students' grade 8 test score performance and about their high school completion status four years later. The analysis focuses on two samples of students: 1) those enrolled in a New York City high school for the first time as a ninth grade student in October 2001 (referred to as the 2001 cohort) and 2) those enrolled in a New York City high school for the first time as a ninth grade

student in October 2005 (referred to as the 2005 cohort). Students from the 2001 cohort were scheduled to graduate in June 2005, and students from the 2005 cohort were scheduled to graduate in June 2009. The key outcomes for this analysis are whether students had earned a local diploma or a New York State Regents diploma by August of 2005 (2001 cohort) and August of 2009 (2005 cohort).<sup>21</sup>

*Findings*

Figures 4a and 4b show the high school completion rates for the 2001 and 2005 ninth grade cohorts, respectively. High school completion rates are shown for the range of scale scores that these students achieved on the New York State grade 8 ELA and math assessments. The scale scores are the average of each student’s ELA and math scale scores. Table 6 presents a summary of the distribution of test scores and graduation rates for both cohorts.

Table 6  
Four-Year High School Completion Status  
by Score on Grade 8 ELA and Mathematics Examinations

4-year Completion Status	Proficiency Level 1		Proficiency Level 2		Proficiency Level 3		Proficiency Level 3 or 4	
	2001 Cohort	2005 Cohort	2001 Cohort	2005 Cohort	2001 Cohort	2005 Cohort	2001 Cohort	2005 Cohort
Percent in Level (%)	35.9	7.6	28.3	47.8	20.6	37.4	35.8	44.6
4-year Completion Status (%)								
Graduated	30.2	28.1	60.1	57.6	78.7	85.5	84.9	87.3
Regents Diploma	7.8	6.8	33.4	32.6	64.7	77.8	75.1	80.7
Local Diploma	22.4	21.3	26.7	25.0	14.0	7.7	9.8	6.6
GED	2.0	0.6	3.9	2.4	3.1	1.9	2.2	1.7
Still Enrolled	45.7	42.9	26.2	27.4	13.9	9.6	10.0	8.4
Dropped Out	21.4	25.2	9.6	12.2	4.2	2.9	2.9	2.5
Number of Students	16,462	4,070	13,002	25,714	9,468	20,085	16,416	23,986

SOURCE: Research Alliance calculations from New York City Department of Education test score and graduation reports.

NOTES: Proficiency levels were determined based on the average of a student’s 8<sup>th</sup> grade ELA and math scale scores. Cut points for each level were based on the average of the ELA and math cut points for each year.

There are several notable patterns. First, there is a strong positive relationship between grade 8 test scores and the probability of earning a high school diploma within four years after entering high school. Figures 4a and 4b show that across both cohorts of first-time ninth grade students, those with higher test scores were more likely to have earned either a local diploma or a Regents diploma within four years than those with lower test scores. For example, for the 2001 cohort, the average graduation rate for students scoring at Level 3 was 79%, compared to only 30% for students scoring at Level 1. The difference in Regents diploma rates was even larger, with 65% of those at Level 3 receiving a Regents diploma compared to only 8% of those at Level 1. Similar differences across proficiency levels can be seen for the 2005 cohort. This suggests that the ELA and math assessments provide a useful indicator of students’ preparation for high school work, particularly at the extremes of the distribution of test scores.<sup>22</sup>

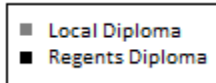
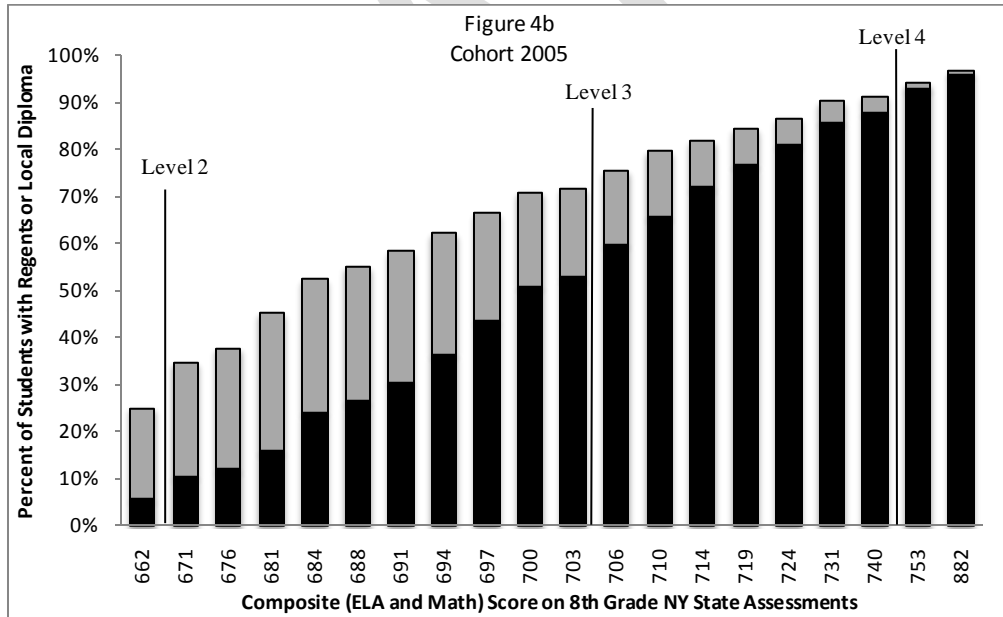
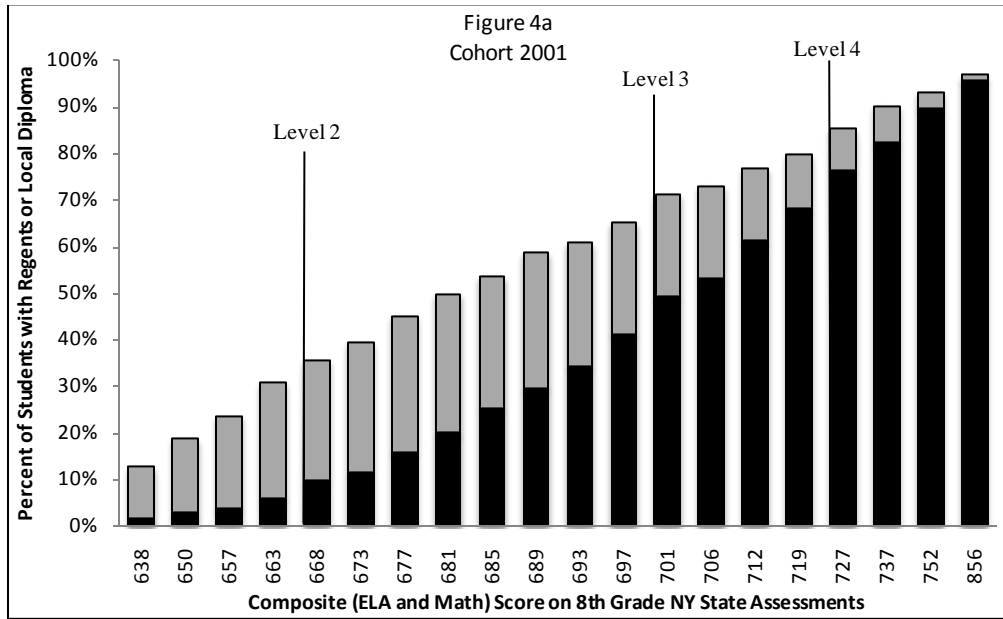
Second, many students who met the state proficiency standards in the eighth grade were unlikely to earn a Regents diploma. The New York State Education Department designates students who score at



Level 3 or 4 on the eighth grade assessments as having reached the state’s proficiency standards. Yet a substantial proportion of students who meet the state standards do not graduate from high school, and many more do not receive a Regents diploma. For example, Figure 4a shows that for the 2001 cohort, less than 60% of students who scored at the lower end of Level 3 graduated from high school with a Regents diploma within four years, and less than 50% received a Regents diploma. This finding suggests that the proficiency standards and the distinctions among performance levels in the middle of the distribution may not be the best indicators of students’ preparation for Regents-level high school work. Changes in the proficiency standards in 2010 were designed to address this shortcoming directly by setting the grade 8 proficiency standards at levels that are associated with passing scores on the high school Regents exams.

DRAFT

Figure 4  
Percent of Students Who Graduate  
by 8<sup>th</sup> Grade Composite Test Scores



SOURCE: Research Alliance calculations from New York City Department of Education test score and graduation reports.

NOTES: Students in the sample include first-time 9th graders in October 2001 and 2005, respectively.

Each bar represents 5% of all students in the sample.

The composite score beneath each bar represents the upper-bound of the range of scores.

Proficiency levels were determined based on the average of a student's 8th grade ELA and math scale scores. Cut points for each level were based on the average of the ELA and math cut points for each year.

Finally, over time, a higher proportion of eighth grade students were meeting the state proficiency standards, and a higher percentage of these students were earning a Regents diploma. Table 6 shows that the percentage of students scoring at Level 3 or Level 4 increased from 36% for the 2001 cohort to 45% for the 2005 cohort. At the same time, Regents diplomas for students scoring at Level 3 or 4 increased from 75% to 81%. This trend may suggest that the ELA and math proficiency standards became better indicators of student preparation for high school work. During this same period, however, the high schools in New York City were evolving, and their improvement may account for some or all of the improvement in graduation rates across the spectrum of grade 8 ELA and math performance levels.

#### TRENDS IN HIGH SCHOOL GRADUATION RATES

This section examines trends in high school graduation rates for students across New York State who began high school between school year 2000–01 and school year 2004–05. These five cohorts of entering high school students span the period during which Children First reforms were initiated in New York City and can offer some insight into whether trends in the city’s graduation rates began to change as the reforms were being implemented and whether they were deviating from trends in other school districts across the state.

##### *Data Sources, Samples, and Analysis*

The findings presented here are based on a modified application of the comparative interrupted time series analysis that was used for the earlier test score analysis. As with the test score analysis, the counterfactual for this analysis is estimated as the New York City graduation rate trend for post-2003 cohorts of first-time ninth graders controlling for: 1) continuation of graduation rate trends underway in New York City schools for cohorts prior to 2003; 2) changes in graduation rate trends before and after 2003 in the Big Four districts and schools; and 3) differences in school characteristics between New York City and the Big Four. This counterfactual represents the best estimate of graduation rate trends that were likely to have occurred in New York City schools in the absence of reforms that were instituted during the Children First era. Thus, the best evidence of effects from these reforms is derived from the difference between the graduation rate trends that actually occurred in New York City and the estimated counterfactual trends.

However, two important features of the data for the graduation trend analysis make it a weaker analysis and require that more caution be exercised in drawing inferences from this analysis about the effect of Children First reforms on graduation rates.

First, data for this section are only available for five cohorts of students: those who were first-time ninth graders in October 2001 (and scheduled to graduate in 2005) through October 2005 (and scheduled to graduate in 2009). As a result, the analysis will only be able to include two years of graduation trends prior to 2003 and three years for the Children First era, from 2003 to 2005. Consequently, the comparison with pre-Children First graduation trends will be less robust than the test score analysis, and there will be less information about subsequent trends. Also, the early cohorts of students were progressing through high schools during the period when Children First reforms were first being implemented, yet the analysis treats their graduation rates as a function of pre-Children First circumstances. To the extent that the early features of Children First reforms produced positive outcomes for students who started high school before 2003, the differences underestimate Children First effects on graduation rates.

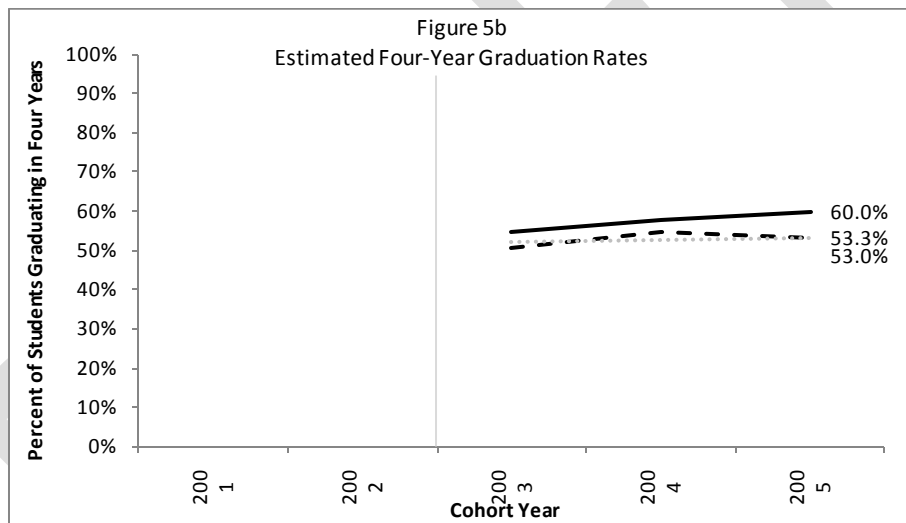
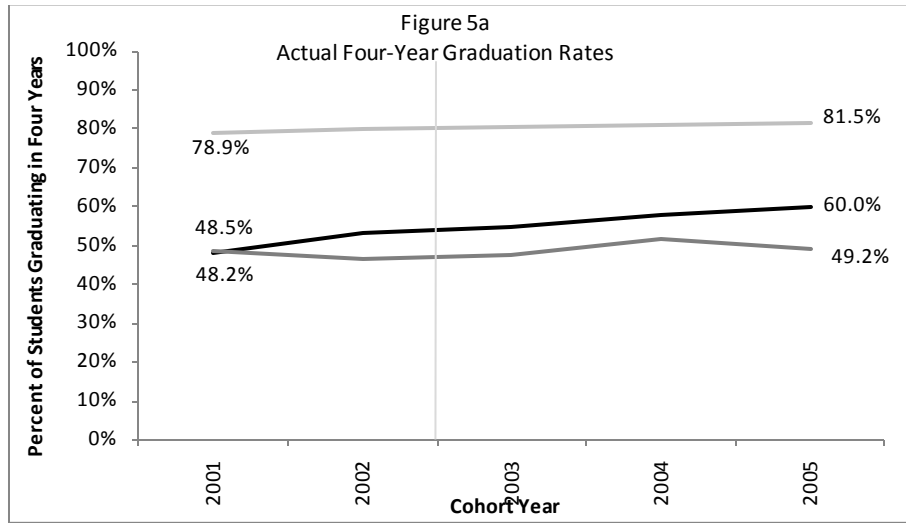
Second, the unit of measurement for graduation rates is the district rather than the individual school. A major feature of those Children First reforms aimed at high schools involved the closing of low-performing schools and the opening of new small schools, typically in the same or a nearby location. Thus, some schools will have no post-2002 data because they were closed down. Other schools will not have any pre-2002 data because they were opened later. Using the district as the unit of analysis, rather than individual schools, diminishes the statistical power somewhat. Also, because of the closing and opening of schools in New York City and because of the open choice process that was implemented during this period, there were changes in the composition of the cohorts of students within and across the New York City districts. This analysis attempts to account for these changes by controlling for demographic characteristics of the districts.

### *Findings*

Figure 5a shows that New York City graduation rates improved by nearly 12 percentage points between the 2001 cohort of first-time ninth graders and the 2005 cohort. However, just under half of this improvement occurred for the cohorts that entered ninth grade prior to the start of Children First reforms (the 2001 and 2002 cohorts). At the same time, graduation rates for the Big Four districts also improved somewhat over these five years, although they declined slightly from the 2001 cohort to the 2002 cohort. In short, some of the improvement in New York City graduation rates is likely to have been an artifact of prior reforms and some may be the result of other policies and reforms that resulted in improvements in other urban districts and across the state.

After accounting for these factors and for the influence of district-level demographic characteristics, the analysis yields evidence that the reforms instituted during the Children First era did produce improvements in graduation rates for the 2005 cohort. Figure 5b shows the difference between New York City graduation rates for the 2003 through 2005 cohorts and the estimated counterfactual graduation rates for the same period. Although graduation rates for New York City outpaced those of the estimated counterfactual, the difference was statistically significant only for the 2005 cohort (a 7 percentage point difference, statistically significant at the  $p < 0.001$  level).

Figure 5  
Percent of Students Who Graduate in Four Years  
9<sup>th</sup> Grade 2001 and 2005 Cohorts



SOURCE: See Table 2.

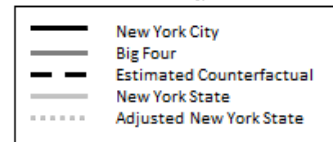
NOTES: See Table 2.

Graduation rates reflect students who graduate with a regents or local diploma. It does not include students who transfer out of the New York City school system.

Actual percent of students who graduate is weighted by the number of students enrolled in the grade.

The Estimated Counterfactual is the regression-adjusted proficiency rate in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the Big Four school districts, and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education. The difference between the actual New York City proficiency rate and the estimated counterfactual is presented as evidence of the effect of reforms instituted during the Children First era (2003-2010) in New York City.

The Adjusted New York State is the regression-adjusted test score trend in New York City in 2009 and 2010 after controlling for the continuation of test score trends underway in New York City schools prior to 2003, changes in test score trends before and after 2003 in the remaining New York State districts (other than New York City and the Big Four), and differences in school characteristics between New York City and the Big Four. School characteristics include total school enrollment; percent of total enrollment that is Black, White, Hispanic, or other race; percent of total enrollment that is eligible for free or reduced-price lunch; and percent of tested students who are classified for special education.



## CONCLUSION

New York City schools experienced substantial improvements in proficiency rates on the New York State ELA and math assessments and on graduation rates during the Children First reform era. Some amount of this improvement is likely an artifact of reforms and trends that were underway before the implementation of Children First reforms, some is likely due to other reform initiatives at the federal and state level, and some is likely due to a growing familiarity with the assessments and testing strategies across the state.

There is compelling evidence that the constellation of reforms being instated in New York City from 2003-2009 had a positive effect on ELA and math proficiency rates in grades 4 and 8 and on graduation rates, over and above continuing effects of prior reforms or conditions shared by other districts. These findings are based on a comparative interrupted time series analysis, a rigorous evaluation method designed to account for these other influences on student outcomes. The evidence suggests that the effects on proficiency rates extend both to general education students and to students with disabilities, with especially large effects for the latter. The evidence also suggests that the effects of Children First persisted through 2010 when the New York State Board of Regents substantially increased the threshold required for students to be classified as meeting the state's proficiency standards.

Still, there are a number of important questions that these analyses are not able to address:

- How have Children First reforms affected the performance of other important subgroups of students including English language learners, recent immigrants, students who enter their grade already behind or with limited skills, and other students with characteristics often associated with low performance on achievement tests?

Test score and graduation data for these subgroups are available from the New York State Education Department only after 2002; the comparative interrupted time series analysis requires several years of information prior to the Children First reforms. Nonetheless, a key goal of the Children First reforms has been to close the gaps from groups of students with characteristics associated with a history of low performance to their peers. A scientifically rigorous analysis of Children First effects on these gaps will require a creative application of other statistical methods that can control for alternative influences on test score trends.

- Which features of the Children First reform are most likely to account for the positive effects on student test scores and graduation rates and which features are likely to need strengthening in order to bring about further improvements?

At the school level, Children First manifests itself in structures, leadership and teaching capacities, and supports that vary widely across the New York City education system. This variation is likely to have proliferated even further with the introduction in 2007 of the school-level autonomy and empowerment that have become cornerstones of the second phase of Children First reforms. Thus, it will be impossible to attribute the effects of this complex and interconnected set of reforms to any specific features or groups of features. It is essential, however, that additional research investigates the conditions that have led to improvements in student outcomes, particularly in schools that have experienced the steepest gains during the Children First era. The Research Alliance for New York City Schools has begun this process by identifying groups of historically low-performing schools and collecting data to determine how and why some of these schools have experienced dramatic improvements in student

performance while others have remained stagnant or declined further. This effort includes both qualitative inquiry and additional quantitative analysis.

In addition, nearly 20 years of research conducted by the Consortium on Chicago School Research has yielded useful insights into the conditions necessary to bring about significant improvement in student achievement. A recent summary of this research points to “five essential supports” for school improvement: leadership, parent-community ties, professional capacity, student-centered learning climate, and ambitious instruction.<sup>23</sup> The Research Alliance is embarking on a similar line of research to measure these and other supports and to determine the extent to which they are associated with school improvement in New York City.

Other recent studies have also begun to examine key elements of the reforms instated during the Children First era. For example, a study of the many new small schools of choice established during the Children First era shows that these schools have had substantial effects on student performance and progress toward graduation.<sup>24</sup> Building on a naturally occurring randomized controlled trial that results from the New York City high school admissions process, this study found that the small schools increased the likelihood that students would stay on track to graduation and increased graduation rates by nearly 7 percentage points. Further research is planned to identify the features of these small schools that make them more effective and to learn how these conditions can be created in more high schools across the city. Other studies have focused on leadership development (Corcoran, 2010) and the changes in grade-to-grade promotion standards (RAND, 2010). Extensions of this work would also shed valuable light on the sources of effects discussed in this chapter.

- What are the best early indicators of student preparation for and progress toward a high school diploma and ultimately a successful transition to college and work?

The findings presented here show that student performance on the grade 8 ELA and math assessments are positively associated with the likelihood that students will graduate within four years of entering high school. The association with earning a Regents diploma is especially strong. The analysis also suggests, however, that these assessments are likely to be incomplete indicators of student preparation for high-level high school work, as many students who are classified as proficient are unlikely to graduate or earn a Regents diploma. In addition, little is known about the degree to which these assessments provide useful insights into students’ likely achievement trajectories as they progress toward high school, particularly at the critical transitions into and through the middle grades and pre-adolescence.

The Research Alliance for New York City Schools is conducting studies of these issues by first identifying individual and school factors associated with a student’s likelihood of achieving performance levels in the middle grades that are associated with a high probability of success in their first year of high school. This work will provide information for both policymakers and practitioners about early warning indicators that can be used to identify students for special supports and help during the middle grades and during their initial transition into high school. In related work, the Research Alliance is also building on work conducted by the Consortium on Chicago School Research that identified indicators of whether students are getting off track for graduation and preparation for college and careers.<sup>25</sup> This analysis will expand the range of performance and engagement indicators beyond ELA and math test scores and will result in tools that school leaders and counselors can use to target students for special supports and intervention. This analysis will also build on work conducted by the DOE through its Achievement Reporting and Innovation System (ARIS) and its development of graduation “on track” and college readiness indicators.

## ENDNOTES

---

<sup>1</sup> The author is especially grateful to Jessica Lent who assisted with the analysis and prepared all of the tables and figures and reviewed multiple drafts of the paper. Janet Brand played a central role in constructing the dataset for this paper. Jennifer O’Day, Jennifer Bell-Elwanger, and Hans Bos provided helpful comments on earlier drafts.

<sup>2</sup> New York City Department of Education, *NYC 2010 mathematics and English language arts citywide test results grades 3-8* [PowerPoint Slides], 2010, [http://schools.nyc.gov/accountability/Reports/Data/TestResults/2010\\_MATH\\_ELA\\_NYC\\_FULL%20DECK.pdf](http://schools.nyc.gov/accountability/Reports/Data/TestResults/2010_MATH_ELA_NYC_FULL%20DECK.pdf)

<sup>3</sup> New York City Department of Education. *NYC graduation rates class of 2009 (2005 Cohort)* [PowerPoint Slides], 2010, [http://schools.nyc.gov/Accountability/Reports/Data/Graduation/GRAD\\_RATES\\_2009\\_HIGHLIGHTS.pdf](http://schools.nyc.gov/Accountability/Reports/Data/Graduation/GRAD_RATES_2009_HIGHLIGHTS.pdf); New York City Department of Education, *School Accountability Tools – Progress Report*, 2010, <http://schools.nyc.gov/Accountability/tools/report/default.htm>

<sup>4</sup> This statistical methodology has been used widely in education research and evaluation. See Howard S. Bloom, *Estimating program impacts on student achievement using "short" interrupted time series* MDRC Working Papers on Research Methodology (New York, NY: MDRC, 1999), <http://www.mdrc.org/publications/82/full.pdf>; and William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin, 2002).

As in this paper, comparative interrupted time series analyses have been applied primarily to study:

- Broad systemic policies and interventions such as the federal No Child Left Behind Act of 2002: See Thomas Dee and Brian Jacob, *The impact of No Child Left Behind on student achievement*, NBER Working Paper 15531 (Cambridge, MA: National Bureau of Economic Research, 2009), <http://www.nber.org/papers/w15531.pdf>; and Manyee Wong, Thomas D. Cook, and Peter M. Steiner, *No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series*, IPR Working Paper WP-09-11 (Evanston, IL: Institute for Policy Research, Northwestern University, 2010), <http://www.northwestern.edu/ipr/publications/papers/2009/wp0911.pdf>;
- Accountability systems: See Brian A. Jacob, “Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools,” *Journal of Public Economics*, 89 no. 5-6 (2005): 761–796, [doi:10.1016/j.jpubeco.2004.08.004](https://doi.org/10.1016/j.jpubeco.2004.08.004);
- Comprehensive school reforms such as Accelerated Schools: See Howard S. Bloom, *Measuring the Impacts of Whole-School Reforms: Methodological Lessons from an Evaluation of Accelerated Schools* (New York, NY: MDRC, 2001), <http://www.mdrc.org/publications/76/full.pdf>; and
- Talent Development High Schools: See (see James J. Kemple, Corinne M. Herlihy, and Thomas J. Smith, *Making Progress toward Graduation: Evidence from the Talent Development High School Model*, 2005, <http://www.mdrc.org/publications/408/full.pdf>).

<sup>5</sup> For the purposes of the analysis presented in this chapter, effects of Children First reforms are assessed beginning with the 2002-2003 school year.

<sup>6</sup> National Center for Education Statistics, *Common Core of Data (CCD)*, 2010, <http://nces.ed.gov/ccd/>



<sup>7</sup> Statewide test scores from grades 3, 5, 6, and 7 were not available prior to 2002. As a result, they could not be included in this analysis. Also, scale scores are not available for 2005. New York State Education Department, *The New York State School Report Card* (1998-2009), 2010, <http://www.emsc.nysed.gov/irrts/reportcard>

<sup>8</sup> Note that the New York State Board of Regents made dramatic changes in the state testing standards in the 2009-2010 school year. Most notably, it raised the score that students need to reach in order to be classified as meeting the state's proficiency standards (Level 3 or Level 4). This change is discussed later in the paper.

<sup>9</sup> Research Alliance for New York City Schools, *Analysis of Trends in Scaled Scores* (working title), forthcoming, to be posted on [http://steinhardt.nyu.edu/research\\_alliance](http://steinhardt.nyu.edu/research_alliance)

<sup>10</sup> The sample of schools used for these analyses represents 81% of all schools that conducted testing of 4<sup>th</sup> or 8<sup>th</sup> grade students at any time during the period from 1999 through 2009.

<sup>11</sup> In controlling for these factors, it is important that the comparison districts and schools be as similar as possible to New York City in their test score trends prior to 2003. The primary findings in this paper are based on analyses that focus on test score trends for New York State's Big Four school districts. As shown in Table 1, these four districts exhibited quite similar test scores prior to 2003. As urban school districts with high percentages of students with similar characteristics, these districts also experienced many of the same challenges and opportunities presented by national and state education policy initiatives and secular trends.

<sup>12</sup> See Schwartz and Stiefel, this volume.

<sup>13</sup> For a general review of research in this area, see Larry V. Hedges, Richard D. Laine, and Rob Greenwald, "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," *Educational Researcher*, 23 no. 3 (1994):5-14.

<sup>14</sup> Research Alliance for New York City Schools, *Research Agenda*, 2010, [http://steinhardt.nyu.edu/research\\_alliance/agenda](http://steinhardt.nyu.edu/research_alliance/agenda)

<sup>15</sup> Analyses were conducted to test the sensitivity of the 2009-2010 findings to a continuous specification. Although the pattern of findings is very similar to that presented here, the steep increases in test scores through 2008-2009 inflates the performance levels that were estimated for 2009-2010. Treating 2009-2010 as a discrete and independent break in the test score trend provides a more accurate indication of changes in Children First effects after the change in state proficiency standards.

<sup>16</sup> The figure and discussion focus on proficiency rates in an effort to be consistent with the analysis of proficiency rates from New York State assessments. The pattern of results comparing New York City and New York State NAEP scale scores are similar to those using proficiency rates.

<sup>17</sup> The comparisons presented here also do not include controls for demographic differences between New York City and the rest of New York State. These data are not available for NAEP schools.

<sup>18</sup> Calculations provided to the author by the U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, through the New York City Department of Education.

<sup>19</sup> Test score data for 2010 were not yet available for these subgroups of students at the time this paper was being written. Also, NYSED does not provide test score results for subgroups of students defined by race/ethnicity, socioeconomic status, or English Language Learning status prior to 2002. Thus, it was not possible to assess Children First effects on test scores for these subgroups using the comparative interrupted time series method.

<sup>20</sup> The sample size of schools with lowest poverty rates (N=7) was too small to generate a reliable estimated counterfactual.

<sup>21</sup> Note that students who transferred to private schools or to public schools outside of New York City are classified as discharged from the system and are not included in calculations of graduation rates for the purposes of these analyses. Supplemental analyses using graduation rates that do include students who were discharged produce results that are very similar to those presented here. Students who transfer to alternative high schools or credit recovery programs and students who drop out of high school altogether are included in calculations of graduation rates for the analysis presented in this paper.

<sup>22</sup> It should be noted that both test scores and graduation rates are likely to be associated with other student characteristics such as race and ethnicity, English language learning status, special education status, and family income. Thus, one should not conclude that simply increasing test scores would result in a direct and proportional increase in the likelihood of graduating from high school. Further analysis can provide better evidence about the causal relationship between test scores and graduation rates and, more importantly, about strategies that minimize the negative influences that some of these other characteristics may have on educational attainment.

<sup>23</sup> Anthony S. Bryk, Penny B. Seabring, Elaine Allensworth, and Stuart Luppescu, *Organizing Schools for Improvement: Lessons from Chicago* (Chicago: University of Chicago, 2010).

<sup>24</sup> Howard S. Bloom, Saskia L. Thompson, and Rebecca Unterman, *Transforming the High School Experience: How New York City's New Small Schools are Boosting Student Achievement and Graduation Rates* (New York: MDRC, 2010).

<sup>25</sup> Elaine Allensworth and John Q. Easton, *The On-Track Indicator as a Predictor of High School Graduation* (Chicago: University of Chicago, 2005).