

Test Score Growth Among Chicago Public School Students, 2009-2014

AUTHORS

Sean F. Reardon

Stanford University

Rebecca Hinze-Pifer

Stanford University

Executive Summary

A comparison of Chicago public school students' standardized test scores in 2009-2014 with those of public students across the U.S. reveals two striking patterns. First, Chicago students' scores improved dramatically more, on average, between third and eighth grade than those of the average student in the U.S. This is true for students of all racial/ethnic groups. The average Chicago student's test scores improved by roughly 6 grade-level equivalents in the 5 years from third to eighth grade. Second, at each grade level in grades three through eight, Chicago students' scores improved more from 2009 to 2014 than did the average scores of all students in the U.S. Test scores rose in Chicago by roughly two-thirds of a grade level from 2009 to 2014, compared to an increase of one-sixth of a grade level nationally. Again, this was equally true for black, Hispanic, and white students. These patterns do not appear to result from increasingly test-aligned instruction or from changing city demographics and enrollment patterns.

VERSION

November 2017

Suggested citation: Reardon, S.F., & Hinze-Pifer, R. (2017). Test Score Growth Among Public School Students in Chicago, 2009-2014. Retrieved from Stanford Center for Education Policy Analysis: <https://cepa.stanford.edu/content/test-score-growth-among-chicago-public-school-students-2009-2014>

Test Score Growth Among Chicago Public School Students, 2009-2014

Sean F. Reardon

Rebecca Hinze-Pifer

Stanford University

November 1, 2017

Direct correspondence to Sean F. Reardon, Stanford University, 520 Galvez Mall, #536, Stanford, CA 94305. Email: sean.reardon@stanford.edu. The research reported here was supported by grants from the Institute of Education Sciences (R305D110018 and R305B090016), the Spencer Foundation (Award #201500058), the William T. Grant Foundation (Award #186173), the Bill and Melinda Gates Foundation and the Overdeck Family Foundation to Stanford University (Sean F. Reardon, Principal Investigator). Some of the data used in this report was provided by the National Center for Education Statistics (NCES). The opinions expressed here are ours and do not represent views of the Chicago Public Schools, NCES, the Institute of Education Sciences, or the U.S. Department of Education.

Test Score Growth Among Chicago Public School Students, 2009-2014

Sean F. Reardon
Rebecca Hinze-Pifer
Stanford University

Executive Summary

A comparison of Chicago public school students' standardized test scores in 2009-2014 with those of public students across the U.S. reveals two striking patterns. First, Chicago students' scores improved dramatically more, on average, between third and eighth grade than those of the average student in the U.S. This is true for students of all racial/ethnic groups. The average Chicago student's test scores improved by roughly 6 grade-level equivalents in the 5 years from third to eighth grade. Second, at each grade level in grades three through eight, Chicago students' scores improved more from 2009 to 2014 than did the average scores of all students in the U.S. Test scores rose in Chicago by roughly two-thirds of a grade level from 2009 to 2014, compared to an increase of one-sixth of a grade level nationally. Again, this was equally true for black, Hispanic, and white students. These patterns do not appear to result from increasingly test-aligned instruction or from changing city demographics and enrollment patterns.

Test Score Growth Among Chicago Public School Students, 2009-2014

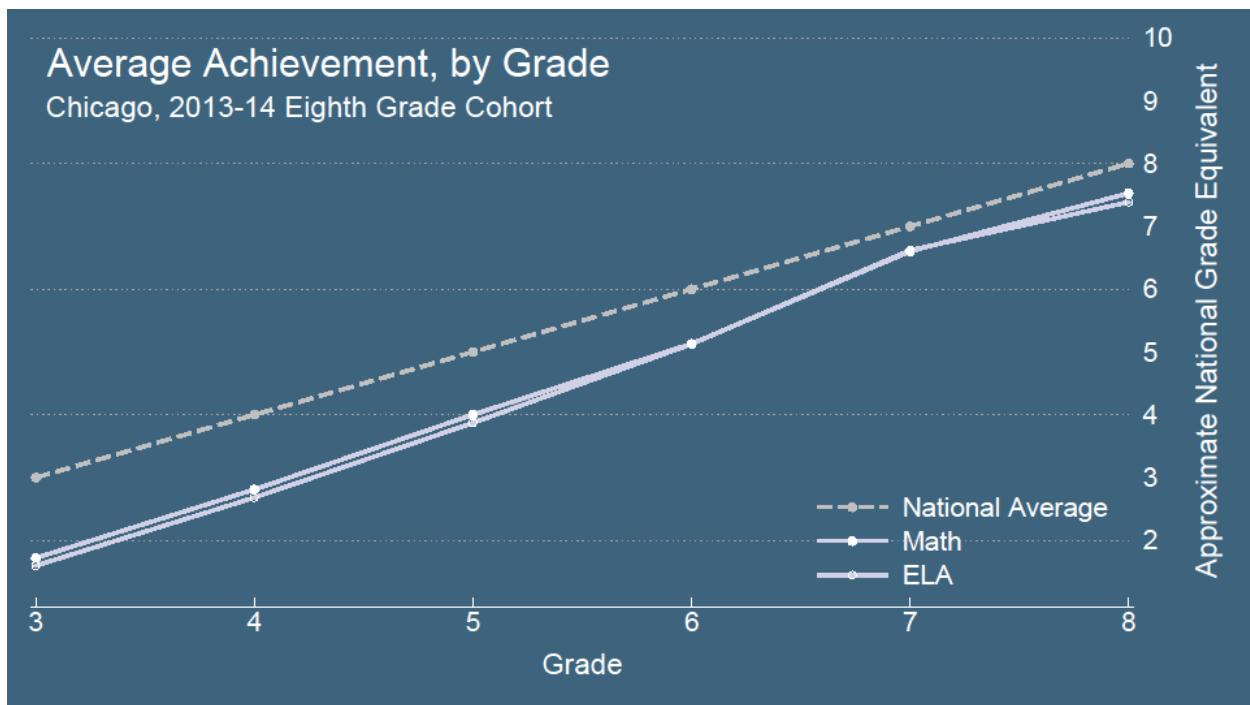
Sean F. Reardon
Rebecca Hinze-Pifer
Stanford University

In this brief report, we compare the standardized test scores of students enrolled in public schools in Chicago (including those in charter schools), to those of students in other school districts in the U.S. We use the scores from roughly 2.1 million math and English Language Arts (ELA) tests taken by Chicago public school students in grades three to eight in the school years from 2008-09 to 2013-14.¹ We compare these Chicago² students' scores to those of third- to eighth-grade public school students in every school district in the U.S. during that time same time period (using some 250 million test scores in total).³ The test scores are placed on a common scale, which is normed to the average score among U.S. students in each grade.⁴ This enables us to answer three questions:

- 1) How does the academic performance of Chicago public school students compare to the national average?
- 2) How does the growth of test scores within each cohort of Chicago students (as they progress from third to eighth grade) compare to that in other districts and to the national average?
- 3) How has academic performance in Chicago changed over time?

We start by examining the average test scores among Chicago public school students in the cohort of students who were in third grade in 2008-09 and in eighth grade in 2013-14. Their average scores, relative to the national average for students in that same cohort, are displayed in Figure 1.

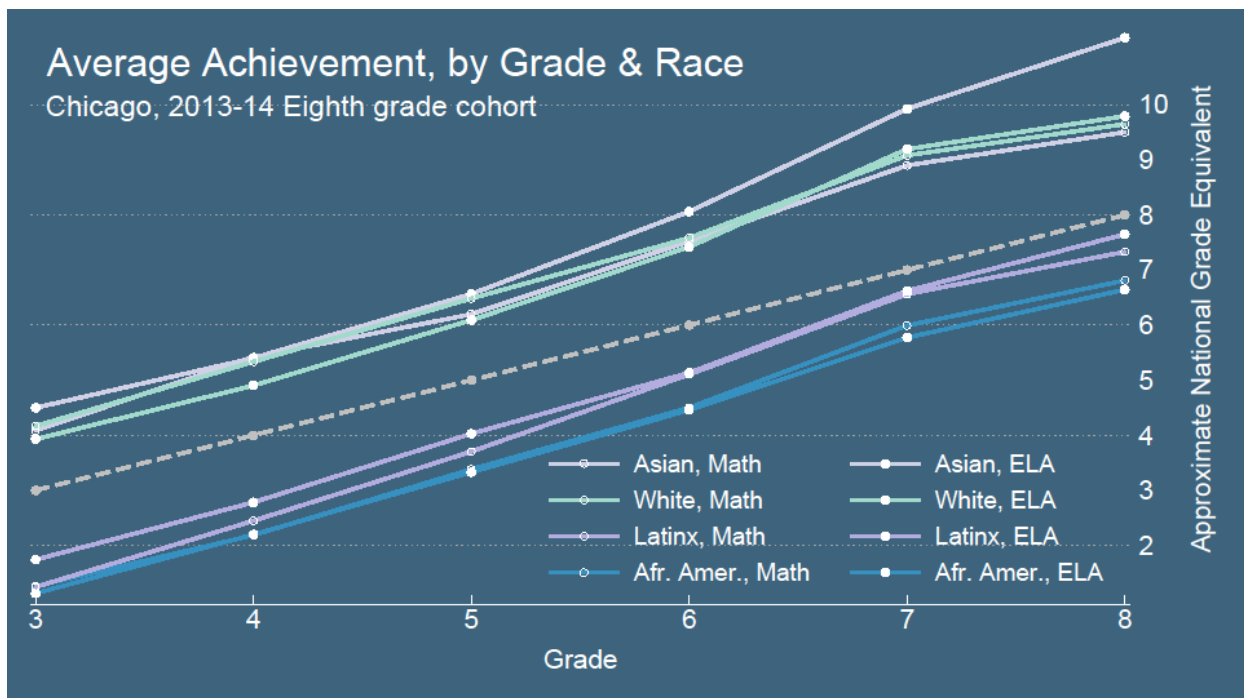
Figure 1



In third grade in 2008-09, Chicago students scored, on average, about 1.4 grade levels below the national average in both math and ELA. Five years later, when that same cohort of students was in eighth grade, Chicago students scored, on average, roughly 0.4 grade levels below the national average. In other words, students in this cohort (the cohort that is scheduled to graduate from high school in 2018) improved their performance by roughly 6 grade levels over the course of 5 years (20% more than the national average).

This same pattern is evident in other cohorts of students, and is similar for male and female students and for those in each racial/ethnic subgroup. As shown in Figure 2, the learning rates for African American, Latinx, white, and Asian students were all higher than the national average. The rate of learning was particularly rapid among Latinx students (45% of students in this cohort), who gained 1.2 grade equivalents more than the national average during third through eighth grade. African American students (41% of the cohort) grew from performing 1.9 grade levels below the national average in third grade to 1.2 grade levels below in 8th grade – equivalent to 0.7 grade levels of additional achievement. White students (8% of Chicago students) gained an extra 0.8 grade levels as they progressed from third to eighth grade. Because all white and black students’ scores grew at relatively similar rates (and faster than the national average), the white-black achievement gap remained roughly constant in size—and large—across third to eighth grade. The white-Latinx gap in Chicago, however, was roughly 0.4 grade levels smaller in eighth grade than in third grade, because Latinx students’ scores increased even faster than white students from grade three to eight.

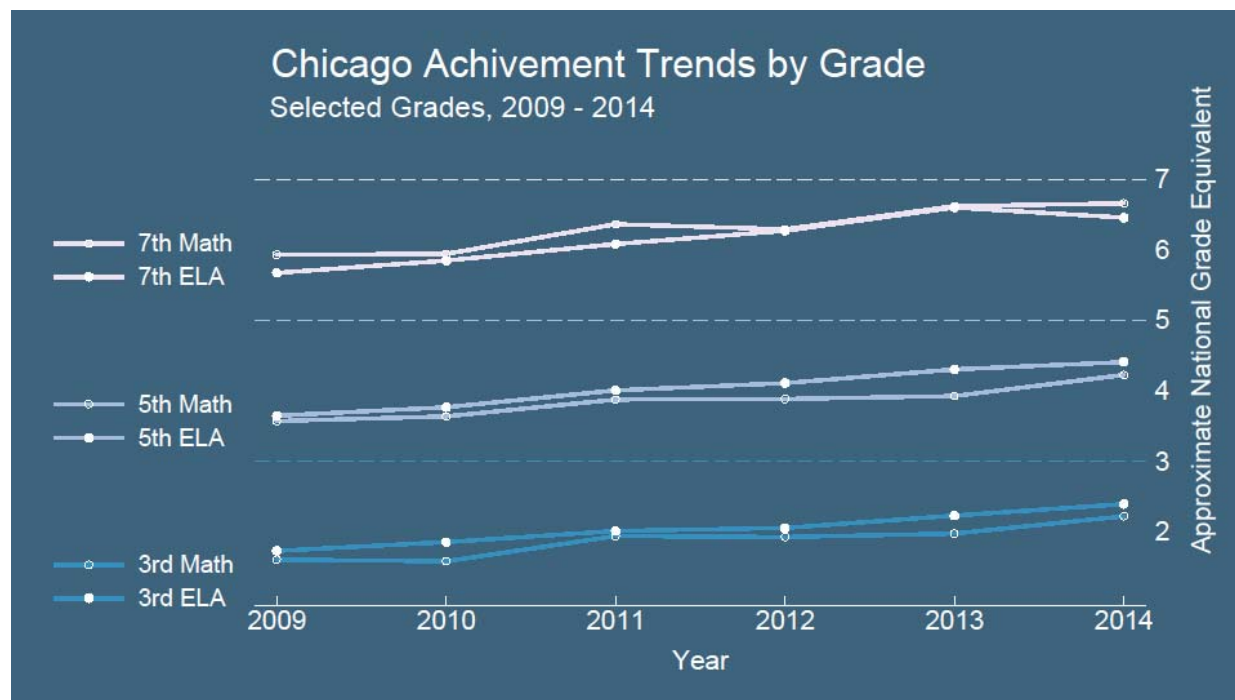
Figure 2



The other notable pattern in Chicago is that student test scores have been increasing across cohorts (see Figure 3). Chicago third-graders’ test scores in 2014 were significantly higher than third-graders’ scores in 2009, increasing by two-thirds (0.66) of a grade level (compared with an increase of 0.18 grade levels in

the U.S. overall). The same was true in each grade from third to eighth. Figure 3 shows representative trends from selected grades. This upward trend is roughly the same for black, Latinx, and white students (not shown here). While average test scores in the U.S. generally increased over the same period, the increase in Chicago has been substantially faster than the national average.

Figure 3



Summarizing the trends

Figures 1-3 above show patterns of test scores for one specific cohort or for specific grades. We describe these patterns more generally by pooling the data and computing the average rates of change (across grades, within cohorts; and across cohorts, within grades) in Chicago and comparing them to other districts in the US. As above, test scores are measured in grade level equivalent units.

When we pool the data from all cohorts of students who were in third through eighth grade at some point from 2008-09 through 2013-14 (the cohorts of students scheduled to graduate high school from 2013 through 2023), the average annual growth rate among Chicago students was 1.19 grade equivalents per year (19% faster than the national average).⁵ This growth rate is higher than 96% of all districts in the US. Among the 100 largest districts in the country, the average growth rate from third to eighth grade is 0.95 grade equivalents per year; Chicago has the highest growth rate between third and eighth grade of any large district in the US.

Likewise, pooling data from 2008-09 to 2013-14 indicates that average test scores within a given grade improved by 0.13 grades per year (or by 0.65 grade equivalents from 2008-09 to 2013-14).⁶ This was the 5th largest improvement rate among the 100 largest school districts in the US, and faster than all but 8%

of school districts nationwide. The average improvement rate from 2008-09 to 2013-14 nationally was 0.026 grade levels per year (or about 0.13 grade levels over the 5-year period, about a half-grade level less than in Chicago).

In other words, average test scores among Chicago students show much larger than average improvement in two dimensions. First, test scores improve from third to eighth grade in Chicago much faster than in most districts in the US. Second, Chicago test scores are higher in more recent student cohorts than in older cohorts. This latter improvement appears to be driven by the fact that third-grade students in recent cohorts have higher math and ELA skills than third-graders in earlier cohorts, perhaps because of improvements in the quality of early elementary grade schooling or increased school readiness and skills when children enter kindergarten. It does not appear to be driven by improvements in the quality of schooling during third through eighth grade (since the average annual test score growth rate does not appear to have increased in the years for which we have test score data). That is, the second trend in test scores arises due to forces operating prior to the end of third grade.

Are these patterns real?

Our analysis above relies on scores on the ISAT tests, high-stakes assessments used for accountability purposes in Illinois. In some cases, scores on high stakes tests can distort actual patterns of academic achievement. This can happen because schools and teachers may “teach to the test,” because teachers and students may become increasingly familiar with the content and format of the tests over time, because schools may manipulate the tested population through selective suspension, or because of outright cheating.⁷ One might reasonably ask, then, do the impressive trends in Chicago students’ test scores and the apparent gains in students’ skills from third to eighth grade reflect real changes in student abilities? Might they be inflated as a result of the test score improvement strategies driven by the high-stakes nature of the tests?

One way to answer this is to examine Chicago students’ performance on the National Assessment of Educational Progress (NAEP) tests. Administered by the U.S. Department of Education, NAEP is a set of nationally representative, low-stakes math and reading assessments designed to allow comparisons over time, across grade levels, and between states. Because students’ scores on the NAEP assessments are not used for any accountability purposes (and are never reported for individual schools or teachers), and because only a relatively small random sample of students are tested by NAEP, the NAEP assessments carry few incentives for teachers or schools to focus their instruction on improving scores on the test at the expense of deeper learning. The NAEP assessments are typically administered every two years in 4th and 8th grades. In recent years, NAEP has drawn representative samples within a group of large urban districts – including Chicago.⁸ This means we can examine the average scores of Chicago students in 4th and 8th grade, from 2003 to 2015, on a low-stakes test that is comparable over time and allows comparisons with other large districts and the US as a whole. If the trends in NAEP scores are similar to those on the ISAT, that would suggest that the Chicago ISAT score patterns reflect real patterns of academic achievement.

Figure 4

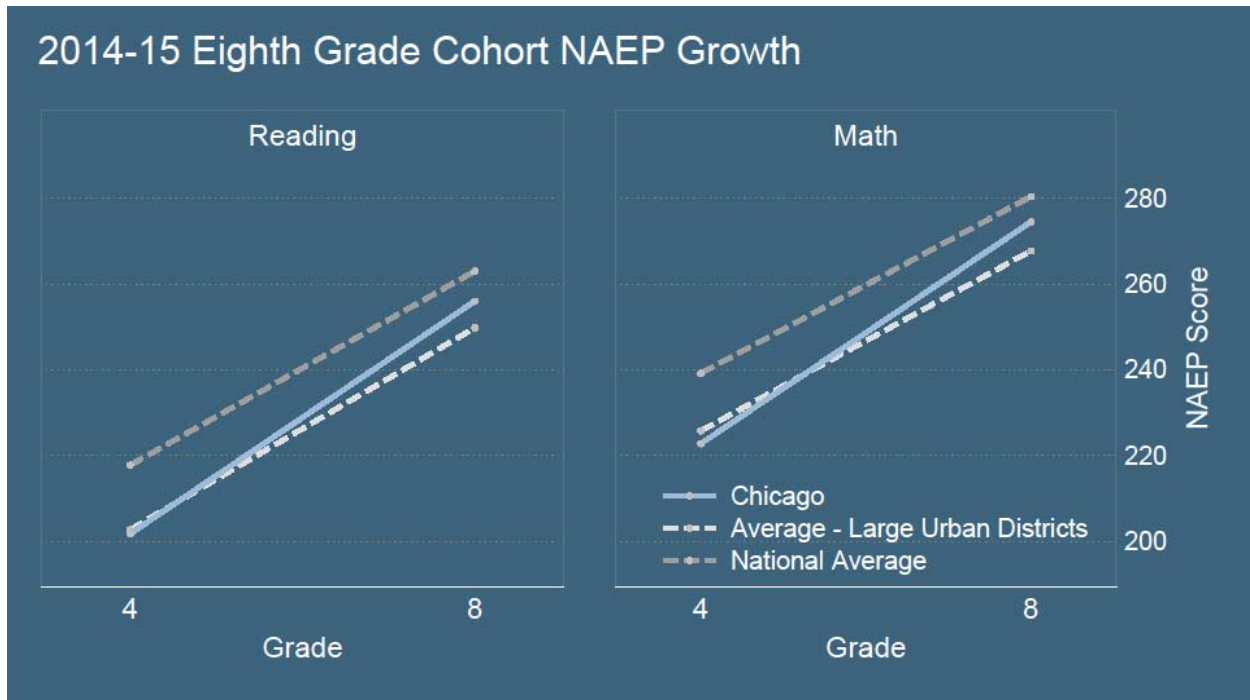
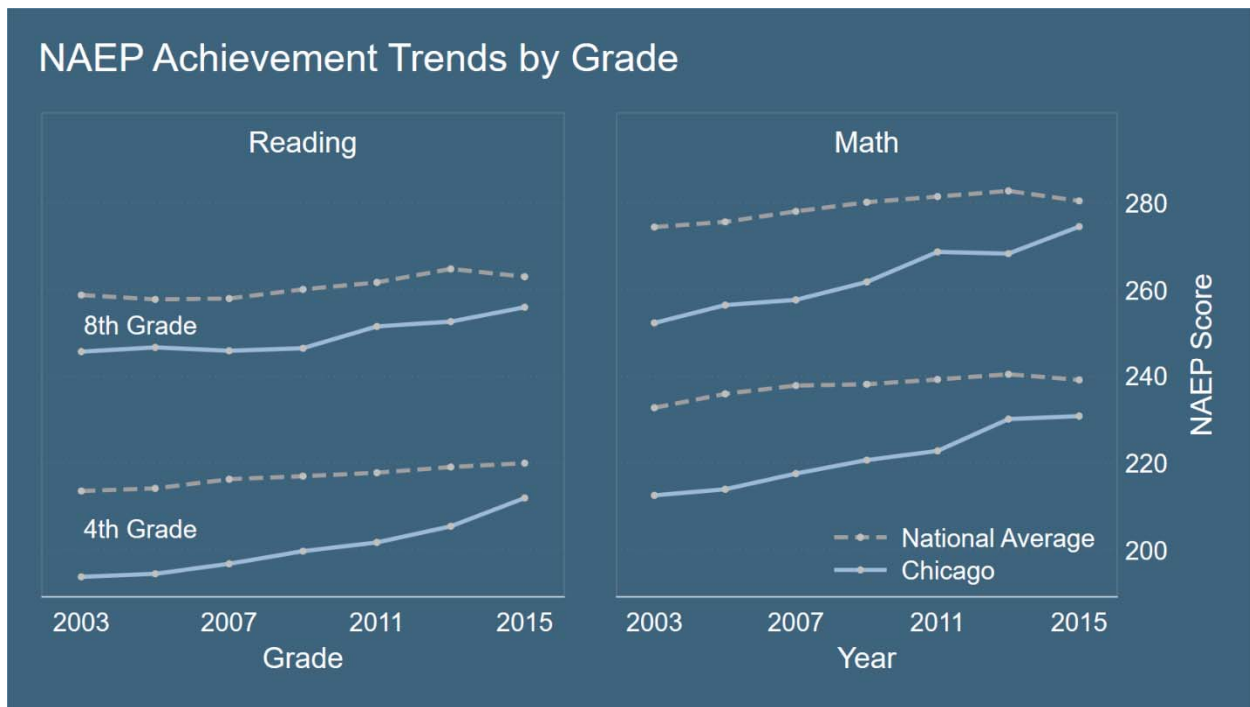


Figure 4 displays average NAEP scores for the cohort of students in 4th grade during the 2010-11 school year and in 8th grade in 2014-15, comparing fourth- to eighth-grade growth in scores in Chicago to the national average and to the set of 20 large urban districts where NAEP was administered. Compared with other large urban districts, Chicago students perform at roughly the same level during 4th grade, but improve more by 8th grade. The increase in test scores in Chicago is about 20 percent faster than in the other large districts or the national average, the same pattern seen in the ISAT scores. The same pattern is true in other recent cohorts (e.g., those in 8th grade in 2010-11, and in 2012-13), though we don't show those data here.

Figure 5 (below) shows Chicago students' scores on the NAEP assessments from 2003 to 2015. The Figure shows that test scores in Chicago, in both 4th and 8th grade in both math and reading, have improved faster than the national averages, particularly since 2007.

The NAEP data in Figure 4 and 5 are very similar to the patterns seen in the ISAT data, suggesting that Chicago's impressive growth and improvement rates on the ISAT accurately reflect growth and improvement in Chicago students' academic achievement.

Figure 5



Are these patterns due to changing district demographics or high rates of grade retention?

Comparing average test scores of students in third grade in one year to those of students in fourth grade (as we do above) the next year may produce a misleading estimate of how much the average student learns during that time period, because the students tested in one grade may not be the same set of students tested in the following year and grade. For example, some students may move in or out of the school district, may transfer in or out of private schools, may be absent on the testing date in one of the years, or may be retained in grade. If these natural student turnover processes substitute students with systematically different average academic skills from the prior-year students they replace, the trends in average scores from one grade or year to the next may not accurately reflect the average increase in individual students' academic skills.

Chicago public school enrollment has been declining for over a decade. Overall enrollment dropped by roughly five percent from 2009 to 2014. One might reasonably wonder, then, whether average test scores in Chicago improved because of changes in the composition of the student population. If most of the students who left the system were students with low test scores, average scores would improve simply because many of the lowest achieving students are no longer enrolled in the system. Or if low-achieving students were often held back a grade, they would effectively have two years in which to learn the same material (and more than five years between third and eighth grade), which could artificially inflate growth rates between third and eighth grades.

The best way to investigate whether these factors affect our estimates of the average grade-to-grade growth rates would be to use individual student-level test score data to estimate each student's annual growth rate during the time she is enrolled in Chicago public schools, and to compare the average of these individual student growth rates in Chicago to the national average. We do not have individual student academic records, so we cannot do that.

Instead, we examined retention rate data and changes in the demographics of students across time and grade in Chicago. Data on the CPS website show that, during the period of our study, roughly 14% of CPS students were retained in grade sometime between third grade and eighth grade.⁹ This means that roughly 14% of students in Chicago took six years to move from third to eighth grade. Nationally, about 9% of students are retained sometime between third and eighth grade.¹⁰ If each Chicago student learned at the national average learning rate, we would expect the average Chicago student to improve her scores by 0.05 grade levels faster than the national average from third to eighth grade. But the improvement in average scores was 0.95 grade levels more in Chicago than the national average. This is vastly more than can be accounted for by grade retention policies. At best, grade retention might explain one-twentieth of the apparent rapid growth rates in Chicago.

The demographic composition of the Chicago public schools changed relatively little during the six-year period we examine (though it has become increasingly Hispanic over the last 15 years). In 2008-09 the district was 46% black, 41% Hispanic, and 9% white; in 2013-14, it was 40% black, 46% Hispanic, and 9% white. In both years 86-87% of students were eligible for free lunch. The district enrollment did decline by roughly 5% during that period, however. A rough estimate based on publicly available data suggests that each year 3-5% of elementary school students leave the district, and a slightly smaller number of new students enter the district. If the replacement students have test scores an average of 1 grade level higher than those who leave, for example, we would expect average scores to go up by roughly 0.03-0.05 grade levels per year. In fact, test scores in Chicago have increased at a rate three to four times faster than that from 2008-09 to 2013-14 (0.13 grade levels per year). And the change in scores from third to eighth grade has been four to six times faster than that. It therefore seems unlikely that in-and out-migration from the district accounts for much of the apparent improvements in test scores.

This sort of back-of-the-envelope calculation is not definitive. We cannot completely rule out the possibility that factors such as in- and out-migration play some role in the patterns of improvement we observe. Without analyzing longitudinal student-level data, we cannot determine how much of the improvements are driven by changes in the student population and how much are driven by the quality of the Chicago public schools and their improvement over time. Such an analysis—carried out by CPS or the University of Chicago Consortium on School Research perhaps—would be very useful.¹¹

Conclusion

With over 370,000 enrolled students, Chicago is the third-largest school district in the US. Its students' test scores in grades three through eight are generally one-half to one-and-a-half grade levels below the national average, and racial/ethnic and socioeconomic achievement gaps remain large, a pattern that

many would interpret as evidence of sub-standard public schools. But those below average test scores—while a significant concern—are not the whole story. They mask, as we have shown, two patterns of significant growth and improvement. Chicago students’ scores, well below average in third grade, improve from third to eighth grade in Chicago at a rate faster than 96 percent of districts in the US. And Chicago third-graders’ test scores are higher in more recent student cohorts than in older cohorts. These patterns are evident both on the Illinois Standards Achievement Test results and the NAEP assessments, which suggests that the patterns of test score growth reflect real improvements in students’ academic skills. The improvements do not appear to result from increased teaching to the test, test familiarity, or any cheating or gaming of the tests. Moreover, these improvements appear too large to have been caused by patterns of grade retention, demographic change, or in- and out-migration from the Chicago public school system.

With the data available to us we cannot tell whether the improvements in test scores across cohorts are due to improvements in out-of-school factors (quality of preschools, the safety of neighborhoods, after school programs) or to improvements in elementary schools (school leadership, teachers’ skills, instructional methods, school climate, etc.). We are hopeful that closer examination of the underlying mechanisms may produce broad lessons applicable to districts nationwide. Specifically, we recommend that CPS and its partners work to answer a few remaining questions:

1. How much, if any, of the patterns of improvement are due to changing district demographics, in- and out-migration, or patterns of grade retention? While our analysis suggests that these are not the primary drivers of the evident improvement, a more detailed analysis using student-level data would allow direct estimation of the average rate of learning between third and eighth grade.
2. To what extent do between-cohort improvements in third grade test scores arise due to increased learning occurring during early elementary compared with increased skills and school readiness at school entry? How much are the improvements in test scores attributable to different family investments in children, changes in neighborhoods, increased access to high quality child care and preschools, and/or improvements in elementary schools?
3. Is the improvement in scores and their growth from third to eighth grade equal among all schools in the district, or are there specific schools or types of schools producing particularly rapid student learning? To the extent that improvements are concentrated in specific types of schools, can we identify educational practices or school characteristics that account for the growth rates?

Answering these questions—determining when and where and why Chicago children’s opportunities to learn are improving—is crucial if other large school districts are to learn from Chicago’s successes.

In sum, Chicago schools seem to have produced a real and sustained pattern of above average learning rates and performance improvement from 2008-09 through 2013-14. These trends are important not only for students in Chicago, but for those in other large districts, because they provide an existence proof that it is possible for large urban districts to produce rapid and substantial learning gains, and to do so in ways that benefit students of all racial and ethnic groups equally. There are clearly lessons to be learned in Chicago.

Endnotes

¹ We use the test scores used for accountability purposes as required under the No Child Left Behind (NCLB) Act. During this time period, Chicago and other Illinois districts used the Illinois Standards Achievement Test (ISAT) tests for third- to eighth-grade grade accountability purposes.

² We include all third- to eighth-grade students enrolled in both Chicago Public Schools (CPS) and those enrolled in public charter schools within the boundaries of CPS.

³ The data we use are assembled in the Stanford Education Data Archive (SEDA): <https://seda.stanford.edu>. SEDA includes nationally-comparable district-level standardized math and English Language Arts (ELA) test scores in grades 3-8 from 2008-09 through 2014-15 for every public school district in the US.

⁴ See Reardon, Kalogrides, & Ho (2017) for technical details. <https://cepa.stanford.edu/sites/default/files/wp16-09-v201706.pdf>

⁵ The standard error of the growth rate per grade in Chicago is 0.034, meaning that we can be 95% certain that the true growth rate was between 1.12 and 1.26.

⁶ The standard error of the annual improvement rate in Chicago is 0.024, meaning that we can be 95% certain that the true annual improvement rate was between 0.08 and 0.18 (and the 5-year improvement was between 0.40 and 0.90 grade level), considerably higher than the average national improvement.

⁷ See, for example, Figlio, D. (2006). Testing, crime and punishment. *Journal of Public Economics*, 90(5), 837-851; Jacob, B. A. & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, 118(3), 843-877; Jennings, J. L. & Bearak, J. M. (2014). Teaching to the test in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher*, 43(8), 381-389; Neal, D. & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92(2), 263-283.

⁸ See <https://www.nationsreportcard.gov/tuda.aspx> for additional information about the NAEP Trial Urban District Assessments.

⁹ <http://www.cps.edu/SchoolData/Pages/SchoolData.aspx>.

¹⁰ John Robert Warren, Emily Hoffman, Megan Andrew (2014). Patterns and Trends in Grade Retention Rates in the United States, 1995–2010. *Educational Researcher* 43(9): 433-443.

¹¹ One analysis of student-level data in Chicago was conducted by the University of Chicago Consortium on School Research in 2015. Data in that report (see tables 7 and 8) indicates that annual student growth rates on the ISAT math tests from 2011-2013 was 25% higher among Chicago students than students in the rest of Illinois in those years (<https://consortium.uchicago.edu/sites/default/files/publications/Statewide%205E%20Report.pdf>). This report does not contain a thorough analysis of student growth rates in Chicago, but its limited analysis of growth rates is consistent with what we see in the aggregated ISAT and NAEP scores.