

# **New York is Not Working**

**An Analysis of Student Test Results in New York City**

**Trevor Cobbold**

**SOS Research Paper**

**Save Our Schools**

**November 2008**

**<http://www.soscanberra.com/>**

## Key Findings

Two aspects of student achievement are considered in this paper: average test scores and gaps in average test scores between different groups of students.

### Average test scores

**Average scores in reading and mathematics in Grades 4 and 8 in New York City have mostly stagnated since 2003, with virtually no improvements for Black, Hispanic and low income students.**

1. National tests show:
  - No statistically significant change in average student scores for reading in Grades 4 and 8 and for mathematics in Grade 8 between 2003 and 2007. There was a small improvement in Grade 4 mathematics;
  - No improvement in average reading scores for low income, Black and Hispanic students in either Grade 4 or 8;
  - Small improvements in average mathematics scores in Grade 4 for low income, Black and Hispanic students;
  - No improvement for Black and Hispanic students in Grade 8 mathematics, but a slight improvement for low income students.
2. State tests show:
  - A mixture of large and small increases and declines in average scores in reading and mathematics in Grades 3 to 8 in New York City since 2003;
  - A similar pattern of increases and declines in average scores between 1999 and 2003 and between 2003 and 2007/8.

### Achievement gaps

**There has been little or no change in the difference in average scores between Black and White students, Hispanic and White students and low income and other students in New York City since 2003. The achievement gaps remain large.**

1. National tests show:
  - A large gap in average reading and mathematics scores between White students and low income, Black and Hispanic students in New York City;
  - No reduction in the achievement gaps between Black and White students, Hispanic and White students and between low income and other students in reading and mathematics in Grades 4 and 8 between 2003 and 2007;
  - The achievement gaps between Black and White students, Hispanic and White students and low income and other students in Grade 8 in 2007 were not statistically different from those for the same cohort in Grade 4 in 2003.
2. State tests show:
  - Small increases and declines in the gaps in average scores for reading and mathematics between Black and White students and between Hispanic and White students from 2003 to 2007/8.

## **Introduction**

Reforms to the New York City public education system have drawn praise from the Federal Minister of Education, Julia Gillard. She says that she is “impressed” by the reforms introduced by New York’s Schools Chancellor, Joel Klein, and that they are “working” and have produced “remarkable outcomes”. She says that there has been continual improvement in student achievement in New York City under Klein.

However, these claims are refuted by an analysis of national and state test results for New York City. They show no general improvement in average student achievement or reductions in achievement gaps since reforms to the public education system were implemented by Klein.

## **No improvement in average student achievement**

### **National test results**

National tests in reading and mathematics show that average student achievement in New York City schools has mostly stagnated since Klein took charge while state tests show a mixture of small increases and declines.

The National Assessment of Education Progress (NAEP) tests conducted by the US Department of Education show no statistically significant change in average student scores for reading in Grades 4 and 8 between 2003 and 2007 in New York City [Lutkus, Grigg & Donahue 2007; Lutkus, Grigg & Dion 2007].<sup>1</sup> They show a small improvement in Grade 4 mathematics but no improvement in Grade 8.

Despite Klein’s claims, there was no general improvement for disadvantaged students. There was no improvement in average reading scores for low income, Black and Hispanic students in either Grade 4 or 8. There were small improvements in average mathematics scores in Grade 4 for low income, Black and Hispanic students. In Grade 8 mathematics there was no improvement for Black and Hispanic students, but a slight improvement for low income students.

### **State test results**

Scores for New York City students on the New York State Department of Education tests are just as unconvincing about improved achievement as the national tests. Average scores for English Language Arts across Grades 3-8 show a mixture of large and small increases and declines between 2003 and 2008 [NYC 2008a]. For example, average Grade 3 scores increased significantly while average Grade 8 scores declined substantially. Small increases occurred in Grades 4 and 5 while Grade 5 and 6 scores declined slightly.

Large improvements in average scores occurred in mathematics in Grades 3, 4 and 5 between 2003 and 2007 while there was a large decline in Grade 8 [NYC 2008b]. There was a small decline in Grade 7 and no change for Grade 6.

---

<sup>1</sup> See Attachment for further details.

The pattern of increases and decreases in scaled scores for 2003-2007/8 is not dissimilar to that prior to 2003. Between 1999 and 2003, there were instances of large improvements in average scores as well as small improvements. Consequently, it is difficult to claim any substantial improvement in state test scores during 2003-2007/8 compared to the previous period.

It should be noted that there are issues about the comparability of average scaled scores on the state tests over time (see below). Also, in contrast to the NAEP tests, statistical uncertainty intervals are not reported for state tests so it is difficult to distinguish what are real improvements or declines except in those few instances of very large changes.

### **No reduction in achievement gaps**

National and state tests also show that the achievement gaps in average scores between Black and White students, between Hispanic and White and between low income and other students have remained as large as ever under Klein's reign.

This evidence contradicts the recent testimony to the US Congress by Mayor Bloomberg and Chancellor Klein claiming to have reduced New York City's achievement gap [Bloomberg 2008; Klein 2008]. Indeed, Bloomberg claimed the achievement gap has been reduced by half. It contradicts claims made by the New York City Department of Education [NYC 2008c]. As one testing expert who advises the New York State Department of Education and the US Department of Education says:

This is not strong evidence that the gap is closing...The only thing you can say is that they're relatively flat, that the gap is relatively stable. [cited in Green 2008]

### **National test results**

The NAEP reports show that there was no reduction in the gaps between Blacks and Whites, Hispanics and Whites and between low income and other students in reading and mathematics in Grades 4 and 8 between 2003 and 2007 [Lutkus, Grigg & Donahue 2007; Lutkus, Grigg & Dion 2007].<sup>2</sup> While there were small reductions and increases in the point score gaps none were statistically significant.

The gap between all White students and low income, Black and Hispanic students in New York City remains very large. The average score of 4<sup>th</sup> grade White students in reading in 2007 was equal to or better than 61% of students nationwide while the average scores of low income, Black and Hispanic students were equal to or better than only 35%, 33% and 30% of students nationwide, respectively. In 8<sup>th</sup> grade reading the corresponding figures were 58% for White students compared to 30%, 25% and 26% for low income, Black and Hispanic students respectively. Similar gaps exist for mathematics in grades 4 and 8.

Jennings [2008a, 2008b] found that, in 2007, 77% of White 4<sup>th</sup> grade students in New York City scored above the average NAEP reading score of Black students in the City and 79% of White 8<sup>th</sup> grade students were above the average score for Blacks. In mathematics, 80% of White 8<sup>th</sup> grade students performed above the average score for

---

<sup>2</sup> See Attachment for further details.

Black students. The gap was similar for Grade 4 students and there were similar gaps between Hispanic and White students in reading and mathematics in grades 4 and 8.

Achievement gaps can also be considered for the same cohort of students over time. Students in grade 4 in 2003 were in grade 8 in 2007, so the achievement gaps can be compared for a similar group of students at the two points in time to see if any progress has been made in reducing them.

The NAEP reports show that the achievement gaps between Black and White students, Hispanic and White students and low income and other students who were in 8<sup>th</sup> grade in 2007 were not statistically different from those for the same cohort in 4<sup>th</sup> grade in 2003 [Lutkus, Grigg & Donahue 2007; Lutkus, Grigg & Dion 2007].

### **State test results**

State test data on achievement gaps tends to support the national data. A recent study by Jennings & Dorn [2008; see also Jennings 2008c] found small changes in the achievement gaps in New York City on state test scores between 2003 and 2008. In most cases, the gaps in average test scores increased between Blacks and Whites and between Hispanics and Whites.

The achievement gap between Blacks and Whites in reading increased in Grade 4 but declined in Grade 8 and increased in mathematics in both Grades 4 and 8. For example, the achievement gap between Blacks and Whites in 4<sup>th</sup> grade increased by 13% in reading and 7% in mathematics. The gap for 8<sup>th</sup> grade mathematics increased by 4%, but the reduction in the gap for 8<sup>th</sup> grade reading of 12% was due to a fall in the average score for Whites.

The achievement gap between Hispanics and Whites increased in Grade 4 reading and mathematics but narrowed in Grade 8 reading and mathematics. In 4<sup>th</sup> grade, the gaps increased by 6% in reading and by 5% in mathematics. In 8<sup>th</sup> grade, the gaps narrowed by 3% in reading and by 6% in mathematics.

### **Misleading and inaccurate data**

Klein and his Department resort to several artifices to claim that student achievement has increased in New York City since he took over the public school system. These are:

- Use of the 2002 results as the comparison benchmark instead of 2003;
- Failure to report the margins of statistical error on test results;
- Use of proficiency gaps rather than achievement gaps;
- Use of less reliable state test data instead of national results.

### **Misleading comparisons**

First, Klein and the New York City Department of Education often compare the 2007/08 results with those in 2002, which pre-dates the changes he implemented to the New York City public school system [Bloomberg & Klein 2008; NYC 2008a, 2008b, 2008c]. Klein became head of the New York City schools in August 2002, but his reforms were not implemented until September 2003. Thus, the appropriate starting point for comparisons on the impact of these reforms is the NAEP tests

conducted in early 2003, not those conducted in early 2002. The 2002 tests were conducted some 18 months before Klein's reforms began to be implemented.

Using 2002 as the starting point for comparing changes in student test scores instead of 2003 exaggerates improvements in student achievement. The NAEP and state test results show significant improvement between 2002 and 2003. By comparing the results for the later years with those of 2002, Bloomberg and Klein incorrectly include improvements that occurred before he made any changes to the New York City public school system and create the perception that their reforms are working.

### **Failure to report statistical error**

The second artifice is not to report statistical margins of error or uncertainty intervals within which the average scores and percentages can be considered to reliably lie. Reporting margins of error is necessary in order to indicate whether a change in test scores is likely to indicate a real change.

When Klein and his Department cite NAEP results they do so without reference to the reported measurement errors or statistical confidence limits. They compare changes in average scores or percentages without regard to whether they are statistically distinguishable. The New York City Education Department has re-produced the results of the NAEP tests in a special publication [NYC 2007]. It claims significant improvements in student achievement, in contrast to the actual NAEP reports. However, it fails to report measurement errors and fails to take account of the fact that several of the results for different grades in 2007 are statistically indistinguishable from 2005 and 2003. Instead, the changes are presented as improvements without qualification.

The National Center on Education Statistics has reportedly criticised this practice, saying that Klein's conclusions about progress on student achievement are "incomplete" and "do not take into account whether the changes or differences are statistically significant" [Green 2008]. Klein's response was that confidence limits do not matter and that statistical significance is "playing something of a game".

### **Use of proficiency gap instead of achievement gap**

Third, Klein and his Department generally measure achievement gaps by figures on the proportion of students achieving certain standards, often called 'proficiency rates' [Bloomberg & Klein 2008; NYC 2007, 2008a, 2008b, 2008c]. While the percentage of students achieving different standards can convey useful information it is a misleading and inaccurate way to measure achievement gaps because it does not distinguish between students who achieve just above the standard from those achieving well above the standard. Trend comparisons are very dependent on the choice of scores for the cut-off points of different standards [Thiessen et.al. 2008: 3-4].

Large increases in proficiency rates can occur even while achievement gaps are increasing. For example, a number of low income students who have been slightly below a given standard may improve sufficiently to score slightly above a standard, thus significantly increasing the percentage of these students above standard. However, the gap between the average scores of low income and other students may

increase if the average scores of other students increase by more than for the low income students.

The proficiency gap is closing even as the achievement gap stays essentially the same because each gap represents a different kind of improvement. Proficiency rates detect movements across the proficiency bar, rising when students who had been below it learn enough knowledge and skills to reach the standard, but registering no change if students who were already meeting the standard surge even further above it. The achievement gap, on the other hand, is sensitive to changes both above and below the proficiency bar. [Green 2008]

A comparison of average scores of different student group is a better indication of the achievement gap than differences in the proportions of students meeting a given standard(s). It reflects changes in the achievement of students across the spectrum and not just those who moved above or below the proficiency standard.

In any case, Klein's claims are largely contradicted by the NAEP reports which show no improvement in the proportion of all students at or above the "Basic" standard for reading in Grades 4 and 8 and no improvement in the percentage of students at or above the "Basic" standard in Grade 8 mathematics [Lutkus, Grigg & Donahue 2007; Lutkus, Grigg & Dion 2007]. However, there was a significant improvement in the percentage of Grade 4 students at or above the "Basic" mathematics standard.

There was no improvement in the percentage of Hispanic and low income students above the "Basic" standard in Grade 4 reading, but a small improvement for Black students. There was no improvement in the percentage of Black, Hispanic and low income students above "Basic" in Grade 8 reading and mathematics. However, there was a large increase in the percentage of Black, Hispanic and low income students above "Basic" in Grade 4 mathematics.

### **National tests more reliable than state tests**

A further issue is that Klein and other officials frequently base their claims for improvement in student achievement on state test results. There are several reasons to consider the state data to be less reliable than the national test data.

State tests are prone to "test score inflation" because of the high stakes attached to them as measures of school performance and teacher accountability [Koretz 2008]. The NAEP results are not used to make high-stakes decisions regarding the performance of individual educators or schools and there is also less room for schools to manipulate results by excluding some students, cheating or other means [Thiessen et.al. 2008].

A number of studies have found considerable discrepancies between student performance trends on state assessments and those on the NAEP [for example Jacob 2007, NCES 2007; Thiessen et.al. 2008]. There are significant discrepancies in student achievement in New York City on the state and national tests [Medina 2007, 2008; Stern 2008].

In contrast to the NAEP, the published results of New York state tests do not report measurement and sampling errors, so that it is not possible to determine whether results are statistically different over given periods or between different groups of students.

Another problem with the state test data is that it is not strictly comparable between 2003 and 2007/8. In 2006, the State Department of Education changed the way it calculated the scale scores so that they are calculated differently for 2007/8 and the years prior to 2006. This makes it difficult to determine with any confidence whether the absolute scores of any particular groups of students increased or declined over the period [VerBruggen 2008].

## **Conclusions**

Despite the claims of the New York City Schools Chancellor, Joel Klein, there has been little to no progress in student achievement in New York City schools since 2003 when Klein's reforms began to be implemented. In particular:

- Average scores in reading and mathematics in Grades 4 and 8 in New York City have mostly stagnated since 2003, with virtually no improvements for Black, Hispanic and low income students;
- There has been little or no change in the achievement gaps between Black and White students, Hispanic and White students and low income and other students in New York City since 2003. The achievement gaps remain large.

Klein and his Department have resorted to several artifices to claim that student achievement has increased and that achievement gaps have been reduced.

First, they often use the 2002 results as the comparison benchmark instead of 2003. The 2002 tests were conducted some 18 months prior to the implementation of the Klein reforms. The reason this benchmark is used is because there were significant increases in student achievement from 2002 to 2003, so this comparison exaggerates the impact of the reforms.

Second, they fail to report the margins of statistical error on test results. Reporting margins of error is necessary in order to indicate whether a change in test scores is likely to indicate a real change.

Third, they use proficiency gaps rather than achievement gaps to indicate changes in the gaps between groups of students. Proficiency gaps can give a false perception of differences between groups of students because they can be declining while gaps in average scores are increasing.

Fourth, they tend to cite state test data which is less reliable than the results of national tests. There are significant discrepancies between state and national test results.

## **Attachment**

### **National Assessment of Educational Progress 2003-07**

#### **4<sup>th</sup> Grade:**

##### **Reading**

- No improvement in average reading scores for all students
- No improvement in average reading scores for low income students
- No improvement in average readings scores for Whites, Asian/Pacific or Hispanics but small improvement for Blacks
- No reduction in achievement gap between Whites and Blacks, Whites and Hispanics and low income and other students
- No improvement in the percentage of all students at or above basic or proficiency levels, no improvement for Hispanic and low income students, but a small improvement for Black students

##### **Mathematics**

- Small improvement in average mathematics scores for all students
- Small improvement in average mathematics scores for low income students
- Small improvements in average mathematics scores for Whites, Asian/Pacific, Blacks and Hispanics
- No reduction in achievement gap between Whites and Blacks, Whites and Hispanics and low income and other students
- Improvement in the percentage of all students at or above basic or proficiency levels and large improvement for Blacks, Hispanics and low income students

#### **8<sup>th</sup> Grade:**

##### **Reading**

- No improvement in average reading scores for all students
- No improvement in average reading scores for low income students
- No improvement in average readings scores for Whites, Asian/Pacific, Blacks or Hispanics
- No reduction in achievement gap between Whites and Blacks, Whites and Hispanics and low income and other students
- No improvement in the percentage of all students at or above basic proficiency level and no improvement for Blacks, Hispanics and low income students

##### **Mathematics**

- No improvement in average mathematics scores for all students
- A small improvement in average mathematics scores for low income students
- No improvement in average mathematics scores for Whites, Asian/Pacific, Blacks or Hispanics
- No reduction in achievement gap between Whites and Blacks, Whites and Hispanics and low income and other students
- No improvement in the percentage of all students at or above basic proficiency level and no improvement for Blacks, Hispanics or low income students

*Sources:* Lutkus, Grigg & Donahue 2007; Lutkus, Grigg & Dion 2007

## References

Bloomberg, Michael R. 2008. Testimony on Mayor and Superintendent Partnerships in Education: Closing the Achievement Gap. House Committee on Education and Labor, US Congress, 17 July. Available at: <http://edlabor.house.gov/hearings/fc-2008-07-17.shtml>

Bloomberg, Michael R. and Klein, Joel I 2008. Mayor Bloomberg and Chancellor Klein Announce Across-the-Board Gains on State Math and Reading Exams. Media Release, 23 June. Available at: [http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2007-2008/20080623\\_test\\_scores.htm](http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2007-2008/20080623_test_scores.htm)

Green, Elizabeth 2008. 'Achievement Gap' in City Schools is Scrutinized. *New York Sun*, 5 August. Available at: <http://www.nysun.com/new-york/achievement-gap-in-city-schools-is-scrutinized/83215/>

Jacob, Brian A. 2007. Test-Based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments, Working Paper No. 12817, National Bureau of Economic Research, January. Available at: <http://www.nber.org/papers/w12817>

Jennings, Jennifer 2008a. In New York City, A Long Wait Ahead to Close the Math Achievement Gap. Eduwonkette blog, 22 July. Available at: [http://blogs.edweek.org/edweek/eduwonkette/2008/07/in\\_new\\_york\\_city\\_a\\_long\\_wait\\_a\\_1.html](http://blogs.edweek.org/edweek/eduwonkette/2008/07/in_new_york_city_a_long_wait_a_1.html)

\_\_\_ 2008b. More Bad News on the Reading Achievement Gap in New York City. Eduwonkette blog, 24 July. Available at: [http://blogs.edweek.org/edweek/eduwonkette/2008/07/more\\_bad\\_news\\_on\\_the\\_reading\\_a\\_1.html](http://blogs.edweek.org/edweek/eduwonkette/2008/07/more_bad_news_on_the_reading_a_1.html)

\_\_\_ 2008c. On New York State Tests, A Growing Achievement Gap Between White/Asian and Black/Hispanic New York City Students. Eduwonkette blog, 30 July. Available at: [http://blogs.edweek.org/edweek/eduwonkette/2008/07/on\\_new\\_york\\_state\\_tests\\_a\\_grow.html](http://blogs.edweek.org/edweek/eduwonkette/2008/07/on_new_york_state_tests_a_grow.html)

Jennings, Jennifer and Dorn, Sherman 2008. The Proficiency Trap: New York City's Achievement Gap Revisited. *Teachers College Record*, 8 September. Available at: <http://www.tcrecord.org/Content.asp?ContentId=15366>

Klein, Joel I 2007. NYC Results, E-mail message to teachers and principals, 20 November. Available at: <http://nycpublicschoolparents.blogspot.com/2008/08/more-statistical-malpractice-from-tweed.html>

\_\_\_ 2008. Testimony on Mayor and Superintendent Partnerships in Education: Closing the Achievement Gap. House Committee on Education and Labor, US Congress, 17 July. Available at: <http://edlabor.house.gov/hearings/fc-2008-07-17.shtml>

Koretz, Daniel 2008. *Measuring Up: What Educational Testing Really Tells Us*. Harvard University Press, Harvard.

Lutkus, A.; Grigg, W. and Donahue, P. 2007. *The Nation's Report Card: Trial Urban District Assessment Reading 2007* (NCES 2008-455). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.  
<http://nces.ed.gov/nationsreportcard/pubs/dst2007/2008455.asp>

Lutkus, A.; Grigg, W. and Dion, G. 2007. *The Nation's Report Card: Trial Urban District Assessment Mathematics 2007* (NCES 2008-452). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.  
<http://nces.ed.gov/nationsreportcard/pubs/dst2007/2008452.asp>

Medina, Jennifer 2007. Little Progress for City Schools on National Test. *New York Times*, 16 November. Available at:  
[http://www.nytimes.com/2007/11/16/education/16scores.html?\\_r=1&scp=1&sq=%22Little%20Progress%20for%20City%20Schools%22%20&st=cse&oref=slogin](http://www.nytimes.com/2007/11/16/education/16scores.html?_r=1&scp=1&sq=%22Little%20Progress%20for%20City%20Schools%22%20&st=cse&oref=slogin)

\_\_\_ 2008. Reading and Math Scores Rise Sharply Across N.Y. *New York Times*, 24 June. Available at:  
<http://www.nytimes.com/2008/06/24/education/24scores.html?partner=rssnyt&emc=rss>

National Center for Education Statistics (NCES) 2007. *Mapping 2005 State Proficiency Standards On To the NAEP Scales*. Institute of Education Sciences, U.S. Department of Education, Washington DC, June. Available at:  
<http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007482>

New York City Department of Education 2007. 2007 Trial Urban District Assessment New York City Highlights. National Assessment of Educational Progress. Available at: <http://schools.nyc.gov/Accountability/YearlyTesting/TestResults/default.htm>

\_\_\_ 2008a. 1999-2008 English Language Arts Test Results by Grade, School and District. Available at:  
<http://schools.nyc.gov/Accountability/YearlyTesting/TestResults/ELATestResults/default.htm>

\_\_\_ 2008b. 1999-2008 Mathematics Test Results by Grade, School and District. Available at:  
<http://schools.nyc.gov/Accountability/YearlyTesting/TestResults/MathTestResults/default.htm>

\_\_\_ 2008c. New York City Achievement Gap Results. Available at:  
<http://www.nysun.com/files/doesppt.pdf>

Stern, Sol 2008. New York's Lake Wobegon Effect. *City Journal*, 26 June. Available at: <http://www.city-journal.org/2008/eon0626ss.html>

Thiessen, Brad; Magda, Tracey and Ho, Andrew 2008. Nonparametric Comparisons of High-Stakes and Low-Stakes Trends: 2003 – 2007. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, March 24-28. Available at: <http://homepage.mac.com/bradthiessen/presentpaper.pdf>

VerBruggen, Robert 2008. Has NYC Discovered the Trick for Closing the Racial Achievement Gap? *National Review Online*, 31 July. Available at: <http://phibetacons.nationalreview.com/post/?q=MjZjZGQ5NmI3NGFjNjYyM2MyZTdjMjhiMmViMzM2ZmY>