CATO

Research Note

Questioning the "Chicago Miracle"

Did Arne Duncan's policies result in student improvement superior to that of other cities?

By Andrew J. Coulson[1]
Wednesday July 15, 2009

Introduction

Federal Education Secretary Arne Duncan is often credited with producing a "Chicago miracle" of improved student achievement during his tenure at the helm of Chicago's public school system. As evidence of that success, Duncan and his supporters point to a dramatic rise in the pass rate of Chicago students taking Illinois' statewide assessment test, the ISAT. That improvement has been used to justify the secretary's chosen policies, and the "Chicago miracle" has been widely accepted and uncritically reported in the media.[2]

The improvement in Chicago's ISAT results is remarkable, with pass rates rising from 38 percent to 67 percent in just seven years. But as Secretary Duncan himself has observed, state assessments like the

[1] Andrew J. Coulson directs the Cato Institute's Center for Educational Freedom and is author of the recent *Journal of School Choice* study: "Comparing Public, Private, and Market Schools." He blogs at www.Cato-at-Liberty.org.

[2] Barack Obama, press conference nominating Arne Duncan for U.S. secretary of education, December 16, 2008, found online at http://www.scribd.com/doc/9028999/Obama-Press-Conference-Announcing-Arne-Duncan-for-Education-Secretary. This claim was widely reprinted. For example, see: *Seattle Times*, editorial, December 17, 2008, found online at http://seattletimes.nwsource.com/html/editorialsopinion/2008532391_edit18educa.html.
While still CEO of Chicago Public Schools, Duncan made similar claims on his own behalf, as in his testimony to the U.S. House Subcommittee on Education Reform, August 28, 2006, found online at http://republicans.edlabor.house.gov/archive/hearings/109th/edr/nclb082806/duncan.htm

ISAT are notoriously unreliable indicators of actual academic improvement. Their scores can be improperly affected by "teaching to the test," passing rates can be lowered, and the tests themselves can be made less difficult.

To escape those problems, many education experts, and Secretary Duncan himself, often verify states' claims of improvement using the results of the National Assessment of Educational Progress (NAEP), a standardized test administered by the federal Department of Education that is less subject to manipulation.[3] Speaking in June of 2009 to the Department's Institute of Education Sciences, Duncan noted that:

> When states lower [their own academic] standards, they are lying to children and they are lying to parents. Those standards don't prepare our students for the world of college or the world of work. When we match NAEP scores and state tests, we see the difference. Some states, like Massachusetts compare very well. Unfortunately, the disparities between most state tests and NAEP results are staggeringly large.[4]

Following the secretary's approach, this study uses NAEP scores to determine whether Chicago's ISAT gains accurately reflect the changes in Chicago student achievement. In addition to assessing the absolute magnitude of Chicago's NAEP gains, it also compares them to the gains made by students in large central cities (LCCs) around the country. The U.S. Department of Education provides the LCC scores specifically as a benchmark for evaluating urban districts like Chicago.[5]

Arne Duncan served as CEO of Chicago Public Schools from June 2001 through December 2008. NAEP results for the city first became available in 2002 (for 4th and 8th grade reading) and 2003 (for 4th and 8th grade math), with the most recent results in both subjects having been collected in 2007.

Findings

NAEP scores are reported on a scale of 0 to 500. The results of recent NAEP tests in Chicago and LCCs nationwide are charted in Figures 1 through 4, and the test scores themselves appear in Appendix A. Table 1 summarizes the differences in the score gains between Chicago students and those in large central cities around the country.

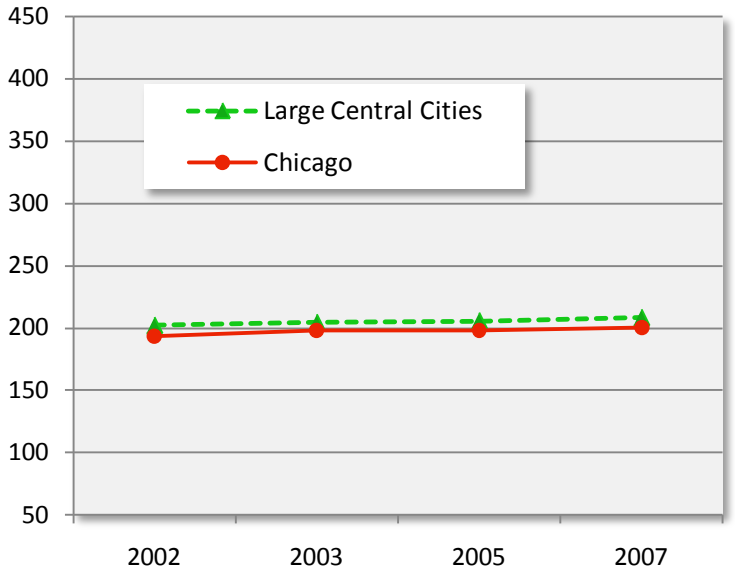
---

[3] NAEP is the best available metric for testing state and district claims of improvement, because it is administered in every state and its content and grading are outside the direct control of state and district officials. Its results do not affect public school employees' salaries or job security in the way that high-stakes state-administered tests do. There is thus little incentive for districts to try to "teach to the [NAEP] test."

[4] Arne Duncan, Speech to the Fourth Annual IES Research Conference, June 8, 2009. http://www.ed.gov/news/speeches/2009/06/06082009.html

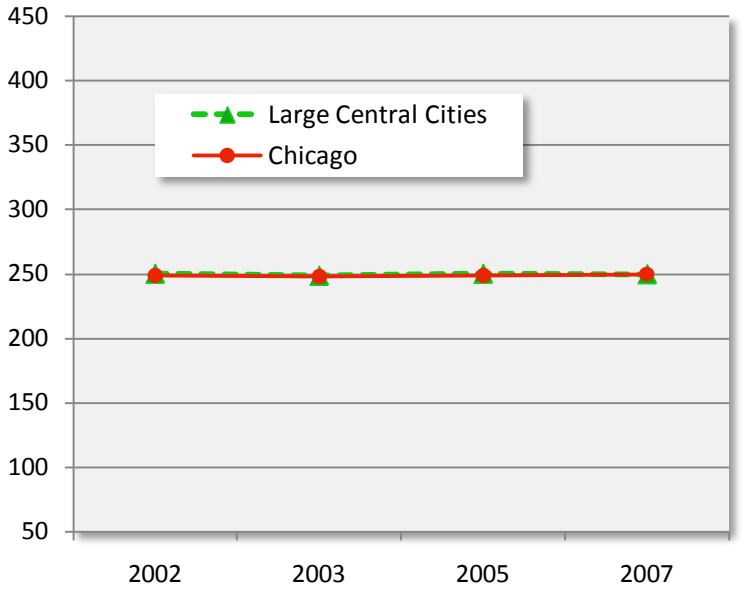
[5] The Department of Education's Institute for Education Sciences defines LCCs as "Public schools from all participating urban districts and non-participating, sampled large central cities. [This category is t]ypically used as a benchmark for urban districts in the same way that ['']national public['] is used as a benchmark for states." NAEP Data Explorer. http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx

Figure 1. NAEP scores, 4th Grade Reading, Chicago and Large Central Cities



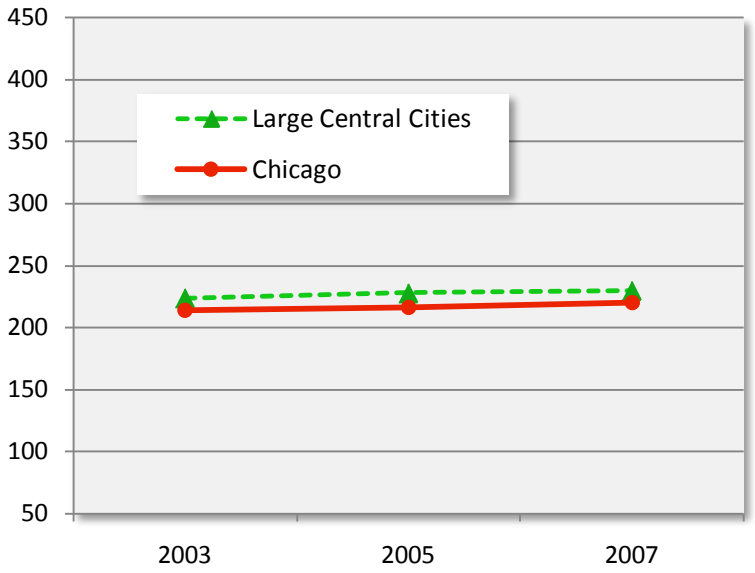
Data source: NAEP Data Explorer, http://nces.ed.gov/nationsreportcard/naepdata/

Figure 2. NAEP scores, 8th Grade Reading, Chicago and Large Central Cities



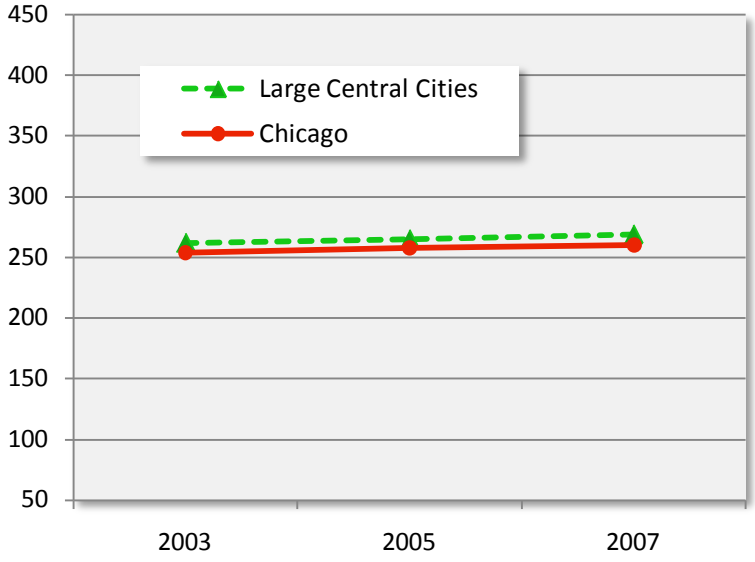
Data source: NAEP Data Explorer, http://nces.ed.gov/nationsreportcard/naepdata/

Figure 3. NAEP scores, 4th Grade Mathematics, Chicago and Large Central Cities



Data source: NAEP Data Explorer, http://nces.ed.gov/nationsreportcard/naepdata/

Figure 4. NAEP scores, 8th Grade Mathematics, Chicago and Large Central Cities



Data source: NAEP Data Explorer, http://nces.ed.gov/nationsreportcard/naepdata/

Table 1. NAEP Score Gains, Differences, and P-Values

Subject	Grade	Chicago gain	LCC gain	Difference	p-value
Reading	4	7.2	6.2	1.0	.67
Reading	8	0.3	-0.4	0.7	.82
Math	4	5.5	5.8	0.3	.89
Math	8	6.0	6.8	-0.8	.76

Table 1 shows that scores improved slightly over the time period examined for both Chicago and LCCs as a whole, but the magnitude of these improvements was small — about 1 percent of the 500-point NAEP score range. In three of the four categories — all but 8th grade math — Chicago public school students experienced a greater gain in their NAEP scores than the broader group of LCC students.

But since NAEP scores are drawn from samples of students, not from the entire student population, there is a level of statistical sampling error associated with them. In order to be confident that Chicago's gains really were greater than those of LCC students, we must be able to reject the possibility that the difference in score gains is a statistical illusion caused by sampling error. It is only when we can reject that possibility that we can call the difference in score gains statistically significant.

To test for statistical significance, we can look at the p-values presented in Table 1.[6] These values represent the probability that the real difference in score gains is in fact zero — that Chicago students progressed at the same rate as students in other large cities. Social scientists typically insist on p-values of .05 or less before they will reject the possibility that an observed difference is merely the result of sampling error. As the p-values in Table 1 make clear, none of the score gain differences between Chicago students and those in the nation's other large central cities comes close to that threshold for statistical significance.

So, judging from NAEP scores during his tenure as Chicago Public Schools CEO, Secretary Duncan's policies resulted in very modest gains that cannot be distinguished from those made in other urban centers around the country.

Discussion and Conclusions

Chicago students made only very modest progress on the NAEP test during Secretary Duncan's tenure at the helm of Chicago Public Schools. Moreover, those modest gains were statistically indistinguishable from those made by students in other large central cities around the country. The NAEP results indicate that there was no "Chicago miracle" and that Secretary Duncan performed no better than the average superintendent of schools in the nation's big cities.

[6] Calculated using Welch's *t*-test.

How can we explain the difference between these sober findings and the much rosier picture painted by Chicago's rising ISAT scores? The answer can be found in a recent study by the Civic Committee of the Commercial Club of Chicago. Their conclusion:

> [M]ost of the improvement in Chicago's elementary school [ISAT] scores over the past decade appears not to be due to real improvement in student performance. It appears to be due to changes in the tests[,] ... new formats and test substance, and lower cut scores (in 8th grade math), along with new testing procedures.... State and local school officials knew that the new tests and procedures made it easier for students throughout the state — and throughout Chicago — to obtain higher marks.[7]

Taken together, the findings of the present study and the Civic Committee report suggest that Secretary Duncan has fallen prey to the same score-inflation for which he has rightly faulted many states. As the secretary has said, this inflation creates a false impression of dramatic improvement in student performance by relying on state tests that greatly exaggerate students' real achievement gains.

Appendix A. NAEP Scores for Chicago and Large Central Cities

All data in this section were obtained using the NAEP Data Explorer.[8]

Table A1. NAEP scores for 4th Grade Reading

Year	Jurisdiction	Average scale score	Standard error
2007	Large Central City	208	(0.7)
2007	Chicago	201	(1.5)
2005	Large Central City	206	(0.6)
2005	Chicago	198	(2.1)
2003	Large Central City	204	(0.7)
2003	Chicago	198	(1.4)
2002	Large Central City	202	(0.8)
2002	Chicago	193	(1.6)

[7] Civic Committee of The Commercial Club of Chicago, "Still Left Behind: Student Learning in Chicago's Public Schools," June 2009, p. 6. http://www.chicagobusiness.com/downloads/CPS.pdf.

[8] http://nces.ed.gov/nationsreportcard/naepdata/

Table A2. NAEP scores for 8th Grade Reading

Year	Jurisdiction	Average scale score	Standard error
2007	Large Central City	250	(0.7)
2007	Chicago	250	(1.5)
2005	Large Central City	250	(0.6)
2005	Chicago	249	(1.2)
2003	Large Central City	249	(0.7)
2003	Chicago	248	(1.3)
2002	Large Central City	250	(1.1)
2002	Chicago	249	(2.2)

Table A3. NAEP scores for 4th Grade Math

Year	Jurisdiction	Average scale score	Standard error
2007	Large Central City	230	(0.5)
2007	Chicago	220	(1.0)
2005	Large Central City	228	(0.4)
2005	Chicago	216	(1.7)
2003	Large Central City	224	(0.8)
2003	Chicago	214	(1.2)

Table A4. NAEP scores for 8th Grade Math

Year	Jurisdiction	Average scale score	Standard error
2007	Large Central City	269	(0.7)
2007	Chicago	260	(1.9)
2005	Large Central City	265	(0.6)
2005	Chicago	258	(1.4)
2003	Large Central City	262	(0.7)
2003	Chicago	254	(1.5)