

NYC and the “test score scam”:  
effects of high-stakes accountability systems

CBID presentation

Leonie Haimson, Class Size Matters

October 17, 2010

# High stakes testing in NYC

- 2001, NCLB mandated annual testing in math and reading for grades 3-8<sup>th</sup> across country
- Schools labeled low-performing faced sanctions if failed to make “adequate yearly progress” on test scores.
- Federal accountability system made worse by even higher stakes put on schools, students & teachers by Bloomberg/Klein.
- In 2003, Bloomberg proposed holding back 3<sup>rd</sup> graders based on test scores; now extended to all grades 3-8<sup>th</sup>;
- In 2007, all NYC schools given grades, based 85% on test scores;
- Also in 2007, bonuses given to principals & teachers based on test scores;
- Schools threatened with closure and teachers with loss of jobs on basis of scores.

# NYC school “progress” reports or grades

- All schools given grades “A” to “F”, depending 85% on state test scores;
- 60% of grade based on “progress” or “value-added” (change in student test scores from previous year)
- 25% on level of current year’s scores
- 15% on the results of surveys and attendance
- Each school’s measure in above categories compared to a bunch of “peer schools”
- Low scores can mean closure.

# High school grades

- Depend primarily on the change and level of credit accumulation of students (course passing rates);
- Student Regents exam scores and passing rates;
- Graduation rates;
- Again, compared to “peer groups”.

# So what's wrong with this?

- 34% - 80% of the annual fluctuations in school's scores is random, or due to one-time factors alone, leading to huge amount of volatility.\*
- The fact that 60% of NYC grade based on “progress” (one year’s change in student test scores) makes it inherently unreliable;
- In 2007, many high-achieving schools got failing grades – including some recognized by federal govt. for exemplary work with high needs students.
- In 2008, 77% of schools that had received an F previous year got an A or a B, with little or no change in teachers or overall program.
- **There was NO relationship between the progress scores that schools received in 2007 and 2008.**

• Source: Thomas J. Kane, Douglas O. Staiger, “The Promise and Pitfalls of Using Imprecise School Accountability Measures,” *The Journal of Economic Perspectives*, Vol. 16, No. 4).

# Same problems with DOE's teacher data reports, used to evaluate teachers for tenure

- Rely on complex models to estimate their “value-added” (before and after student test scores) but these models are unreliable. Why?
- Students are not randomly assigned to teachers or schools; meaning models have to try to control for many factors, including student background;
- Also difficult to account for other classroom conditions, including class size and peer effects, out of teacher's control;
- Teachers whose students score in the top category on one type of exam often rank in lowest category on another exam;
- Sean Corcoran of NYU found uncertainty range of 34 to 61 percentage points (out of 100) in ranking of teachers in NYC teacher data reports.
- Recent national study found 25-35% error rate in value-added methods to identify best or worse teachers.

# Prof. Bruce Baker of Rutgers:

- ***“If the “best available” automobile burst into flames on every fifth start, I think I’d walk or stay home instead.”***

# Campbell's law

- Coined by sociologist Donald Campbell in 1975
- Effect widely observed in fields of medicine, industry, and education.
- *“...the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and...to distort and corrupt the social processes it is intended to monitor.”*



# What does Campbell's law mean about high stakes testing?

- High stakes leads to excessive test prep and cheating, with little or no oversight;
- Many or most NYC cheating allegations not followed up by DOE or state; often teacher whistleblowers end up in the “rubber room”.
- Since 2002, questions on the NY state exams got much easier and narrower in focus -- and cut scores for “proficiency” were lowered each year;
- In some grades/exams, student could pass (or get a level #2) by randomly answering multiple choice questions
- City made big jumps in state test scores, which Bloomberg used in his campaign to renew mayoral control and to run for re-election.

In 2009, this “test score inflation”  
was reflected in school grades”

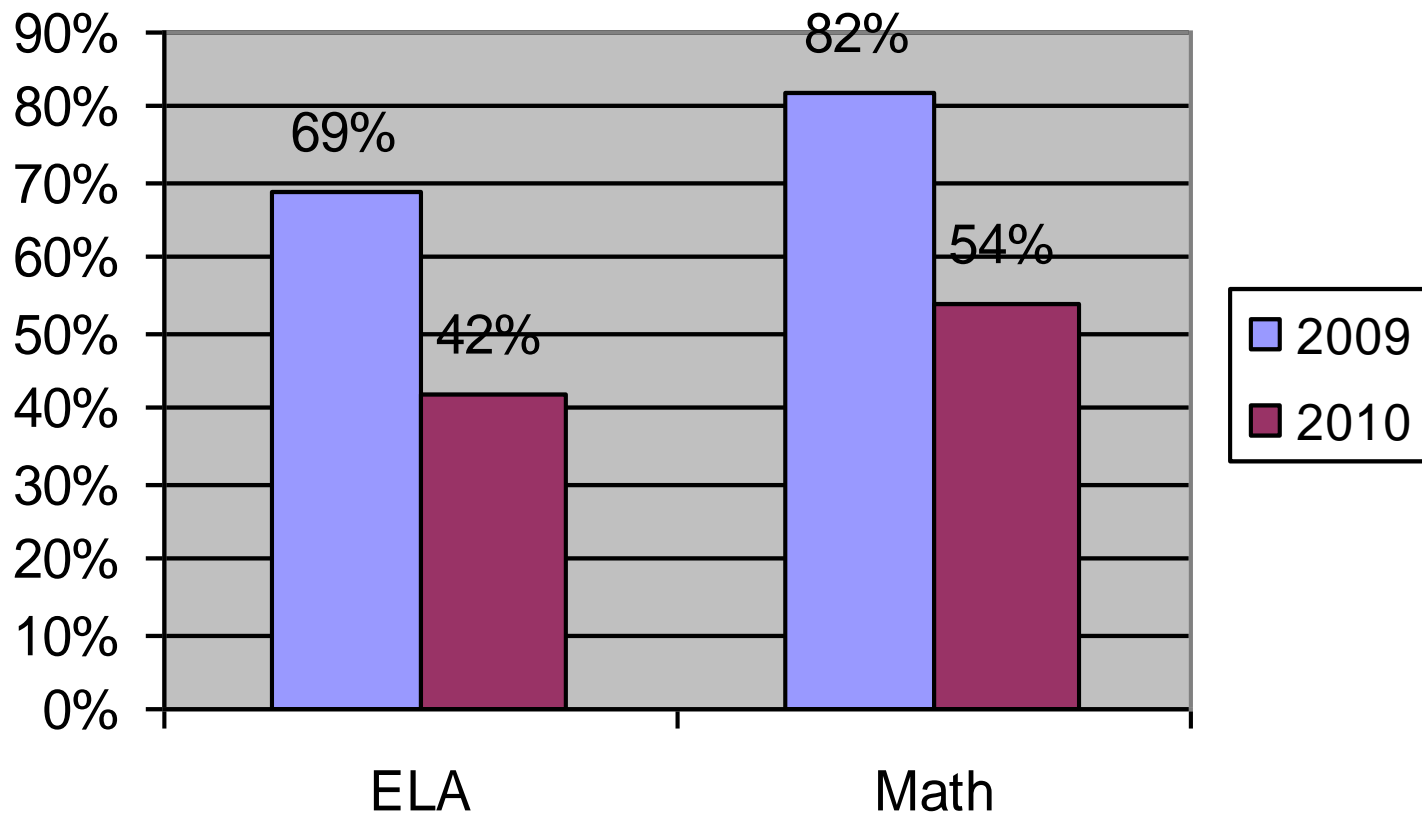
- 84% of NYC elementary and middle schools received a letter grade of A, and 13% of schools received a B.
- Only two schools out of 1,058 received an F, and just five were awarded a D.

# What happened July 2010?

## The test score bubble burst!



### 3rd-8th graders "proficient"



*There are large number of schools with huge number of low-performing students*

- Students who tested below basic (Level 1) on the ELA exam increased nearly six fold: from 12,000 to 63,000 citywide.
- 369 K-8 schools where at least two-thirds of students are not meeting standards in ELA, with at least 20 percent below basic; .
- This is 36% of the elementary and middle schools in NYC.

# Myth and reality

Claim: *“In recent weeks, there has been some controversy and confusion stemming from the state's decision to raise the standards for proficiency on its math and English tests.”* (Joel Klein, NY Post oped, 8/20/10)

Reality: Actually, the state just attempted to reverse the lowering of standards that started as far back as 2002.

# So what evidence can we look at to gauge NYC achievement levels?

- The state exams are still completely unreliable;
- Even though they have raised cut scores, but not the exams themselves; which are still too easy and narrow in focus;
- Only semi-reliable source of info on achievement in NYC are its results on the national exams known as NAEPS.

# Why are NAEPs more reliable?

- These are “low stakes” tests, given only to 4<sup>th</sup> and 8<sup>th</sup> graders in reading and math, every two years, with no consequences for schools or students;
- Only a statistical sample of students take NAEPs each year; with little or no test-prep :
- They are very carefully “scaled,” meaning the difficulty level is maintained from year to year.
- Allow reliable tracking of trends over time.



# More myth and reality

- *“In fourth grade, NYC’s performance [on the NAEPs] now matches that of the nation as a whole, even though NYC serves a much more challenging population.”* (Joel Klein, letter to principals, Sept. 3, 2010)
- ***“That’s called “closing the achievement gap.”***
- Neither of these claims are true.

# NYC 4<sup>th</sup> grade NAEP scores do not match nation as a whole.

- Average NYC score of 217 for reading in 2009 was at the 44<sup>th</sup> percentile for the nation.
- Average score of 237 for math was at the 46<sup>th</sup> percentile for the nation.

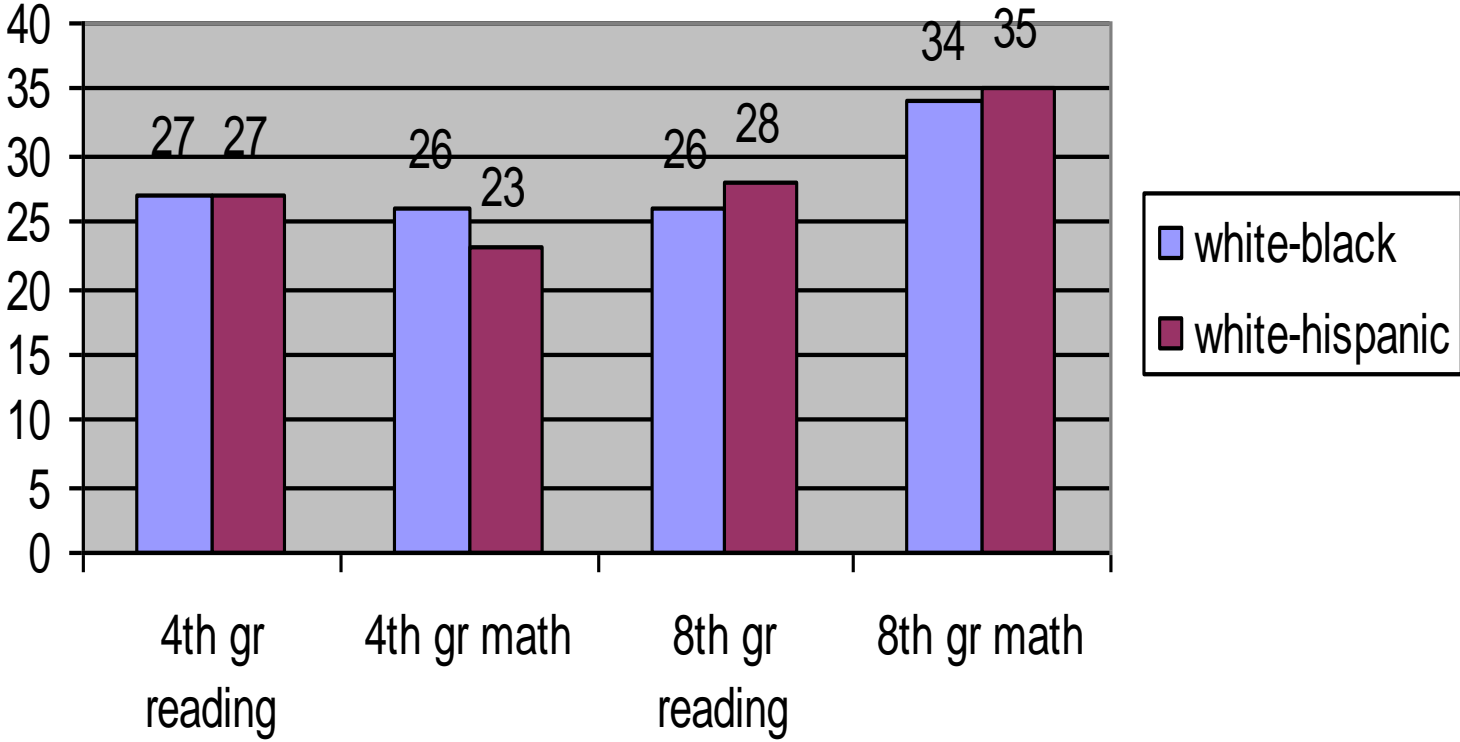
(Source: ***The Nation's report card; Reading and Math 2009, Trial Urban District Assessment, National Center for Education Statistics.***

“Achievement gap” means gap in test scores between ethnic and racial groups

***Reality:*** According to the NAEPs, NYC has **not significantly narrowed** the achievement gap in any category or grade since 2003, either Black/white or Hispanic/white, when the Chancellor’s policies were first implemented.




# NYC achievement gap

## 2009 NAEP scale scores



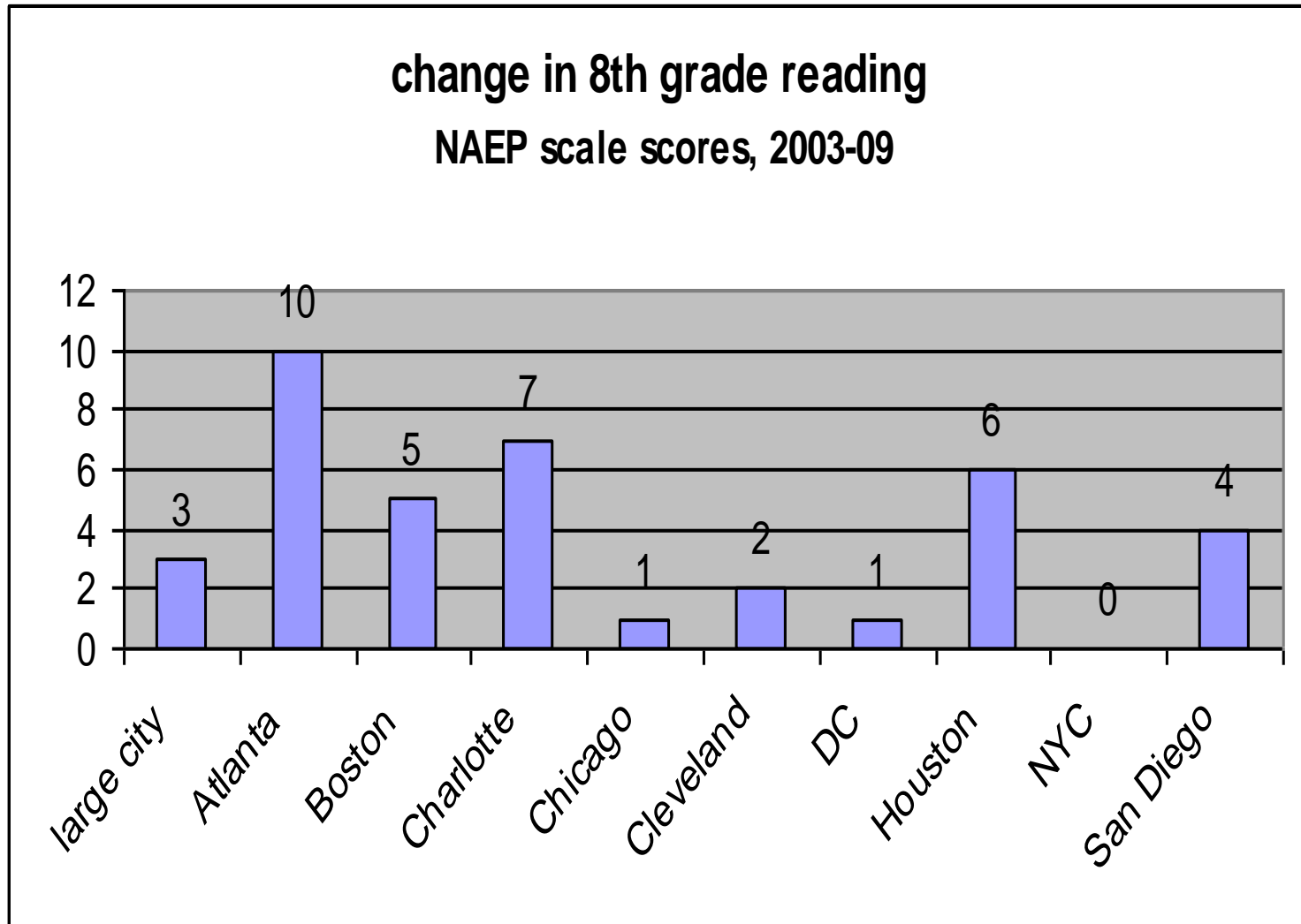
# NAEP Score Growth from 2003-2009: NYC vs. nine other urban districts

	4th Grade Reading	8th Grade Reading	4th Grade Math	8th Grade Math
NYC Change	+7	+1	+11	+7
Atlanta	●	↓	●	↓
Boston	●	●	↓	↓
Charlotte	●	●	↑	●
Chicago	●	●	●	●
Cleveland	↑	●	↑	●
District of Columbia	↓	●	↓	●
Houston	●	●	●	↓
Los Angeles	●	↓	↑	↓
San Diego	●	●	●	↓

 NYC grew sign. less ten times  
 NYC grew sign. more four times  
 No sign. difference 22 times

Slide adapted from one provided by Prof. Aaron Pallas

NYC only city among ten tested over this period to make NO progress in 8<sup>th</sup> grade reading.



***According to NAEPs, relative ranking of average NYC black student scores in every grade and subject fell behind their peers from 2003-2009, compared to the nine other cities tested over this period.***

- 4th grade math; NYC black students were in 2nd place in 2003; by 2009, were tied for third place.
- 4th grade reading; NYC black students were tied for 3rd place in 2003, fell to 4th place in 2009.
- 8th grade math: NYC black students were in 3rd place in 2003, fell to 5th place by 2009.
- 8th grade reading: NYC black students tied for 2nd place in 2003; tied for 3rd place by 2009

***For NYC Hispanic students, relative ranking fell in every grade and subject, from 2003-2009, compared to peers in other cities.***

- 4th grade math: in 2003, NYC Hispanic students were tied for third place among large cities; in 2009, they fell to 4th place.
- 4th grade reading, NYC Hispanics were in 1st place in 2003, fell to 3rd place.
- 8th grade math: Hispanics were in 3rd place in 2003; fell to 7th place.
- 8th grade reading: 2nd place and fell to 6th place.



*DOE claims great progress for low-income students but the reality is different...*

- NYC free-lunch students **already ranked #1 in average NAEP scale scores** in 2003 among peers compared all cities tested;
- By 2009, in 4<sup>th</sup> gr. reading, NYC free lunch students still #1 among cities tested since 2003, but had made smaller gains than DC, Charlotte & Atlanta;
- In 8<sup>th</sup> grade reading, NYC fell from #1 to #2; and made smaller gains than LA, Houston, Boston, & Atlanta;
- In 4<sup>th</sup> grade math, NYC still #1, but made smaller gains than Boston;
- In 8<sup>th</sup> grade math, NYC fell from #1 to #3, and made smaller gains than SD, LA, Houston, Charlotte, Boston, and Atlanta

In reading, ***NYC also has by far worst record of any city w/ non-free lunch students between 2003-2009***

- In 4<sup>th</sup> gr. reading, NYC was ***only city*** in which non-poor students had ***lower average scores in 2009***; and fell from 1st to 3<sup>rd</sup> place in this category;
- In 8<sup>th</sup> gr. reading, NYC was ***only city*** in which non-poor students had ***lower average scores in 2009***, and had fallen from #1 to sixth place!

# ***For non-free lunch students, NYC also has worst record of any city in math 2003-2009***

- In 4<sup>th</sup> grade math, non-poor students gained in average score, but fell from #2<sup>nd</sup> to 5<sup>th</sup> place by 2009;
- NYC was the **only city in the country** in which 8<sup>th</sup> grade math scores fell among non-poor students between 2003- 2009.
- In 8<sup>th</sup> grade math, started out as #1, but fell to #5 by 2009;
- And their average score fell 10 points!

# City also claims rising graduation rates

Here too *Campbell's law in action*

- Questions on state Regents exams and passing scores needed for graduation have become far easier, just like K-8 exams
- High school teachers score Regents exams of own schools; the practice of “scrubbing” and principals changing scores allowed and even encouraged by DOE.
- The more students teacher passes, the higher the school grade; the more like you are to receive a bonus and/or keep your job;
- Incentive system clear: pass as many students as possible!

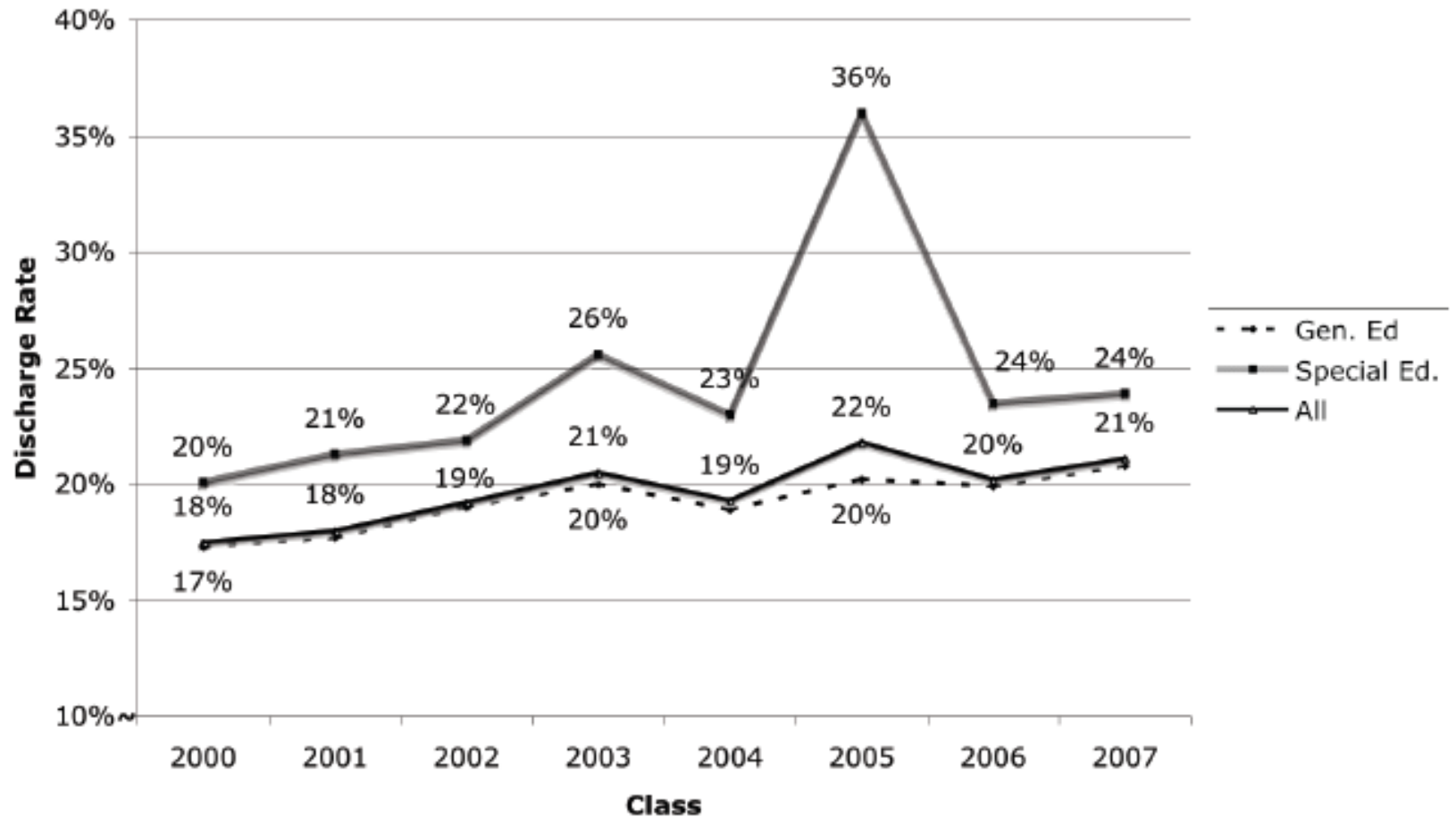
# If that's not enough...

- Practice of “credit recovery” spreading like wildfire in city schools, esp. online credit recovery;
- Online courses where students can gain credits needed to graduate in a few weeks, despite failing all their courses or not attending class,
- Students commonly answer multiple choice questions while looking up responses online; and/or cutting and pasting answers into a form.
- No “seat time” required and can be done at home w/ no adult supervision.
- Effect: many NYC high schools becoming “diploma mills”

***Also... thousands of students are discharged each year from NYC schools and not counted as dropouts***

- Discharge category the “black hole” of DOE accounting.
- The higher the discharge rate, the higher the school’s graduation rate, as these students removed from cohort.
- Number and rate of students “discharged” to other schools and/or GED programs (or perhaps nowhere at all) has been rising.
- Under this administration, the percent of students discharged in their first year of HS has **doubled**.
- The last year for which we have complete data (the class that should have graduated in 2007) there were 20,488 students discharged from NYC high schools.

## New York City High School Discharge Rate, Class of 2000-2007



Source: **HIGH SCHOOL DISCHARGES REVISITED:  
TRENDS IN NEW YORK CITY'S DISCHARGE RATES, 2000-2007**  
JENNIFER L. JENNINGS and LEONIE HAIMSON, APRIL 2009.

Clearly test based accountability systems do not work to improve schools;  
What should we do instead?

- Reduce class size;
- Lessen teacher attrition by improving working conditions, leading to more experienced and effective workforce;
- Increase parental involvement and input at the school level;
- Unfortunately, DOE is implementing not any of these reforms.

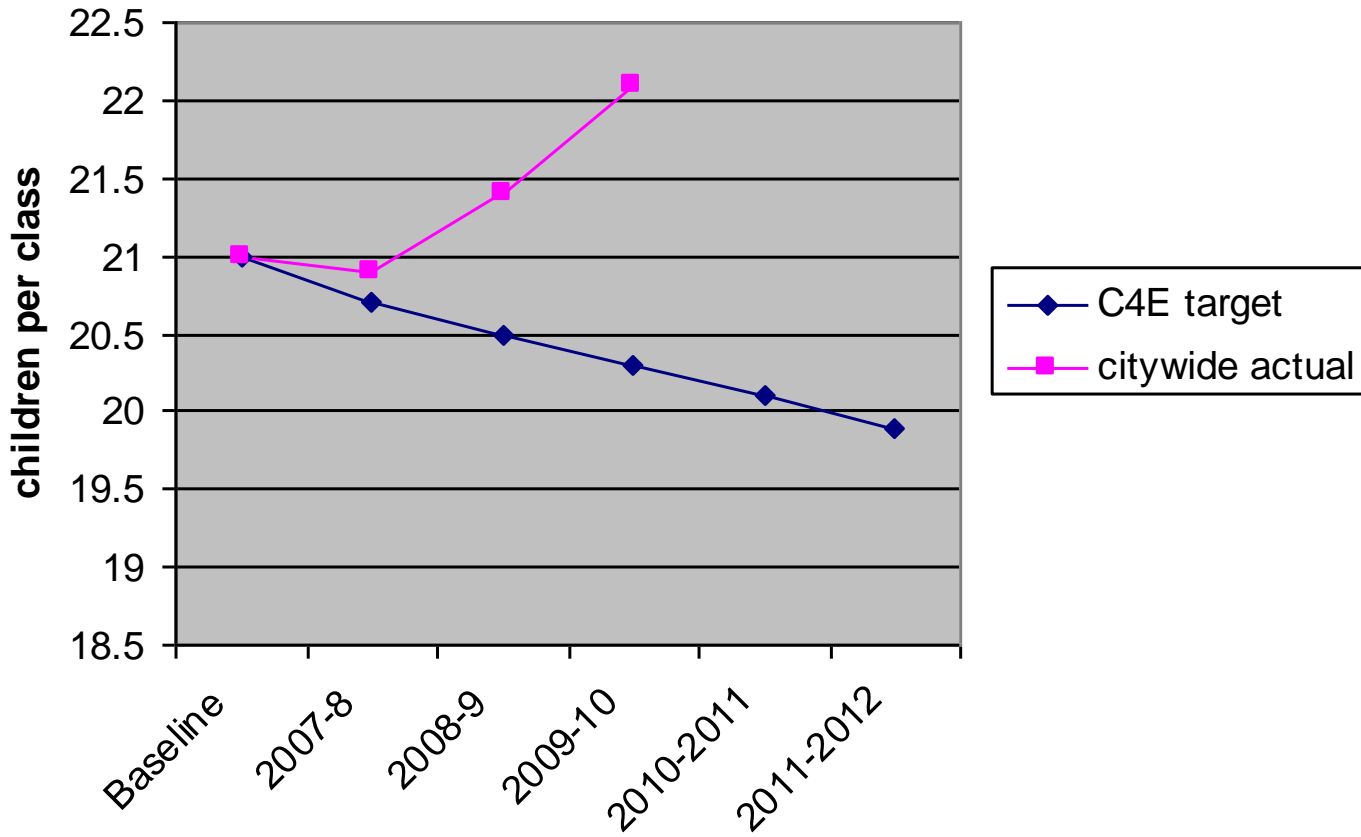


# Class Size

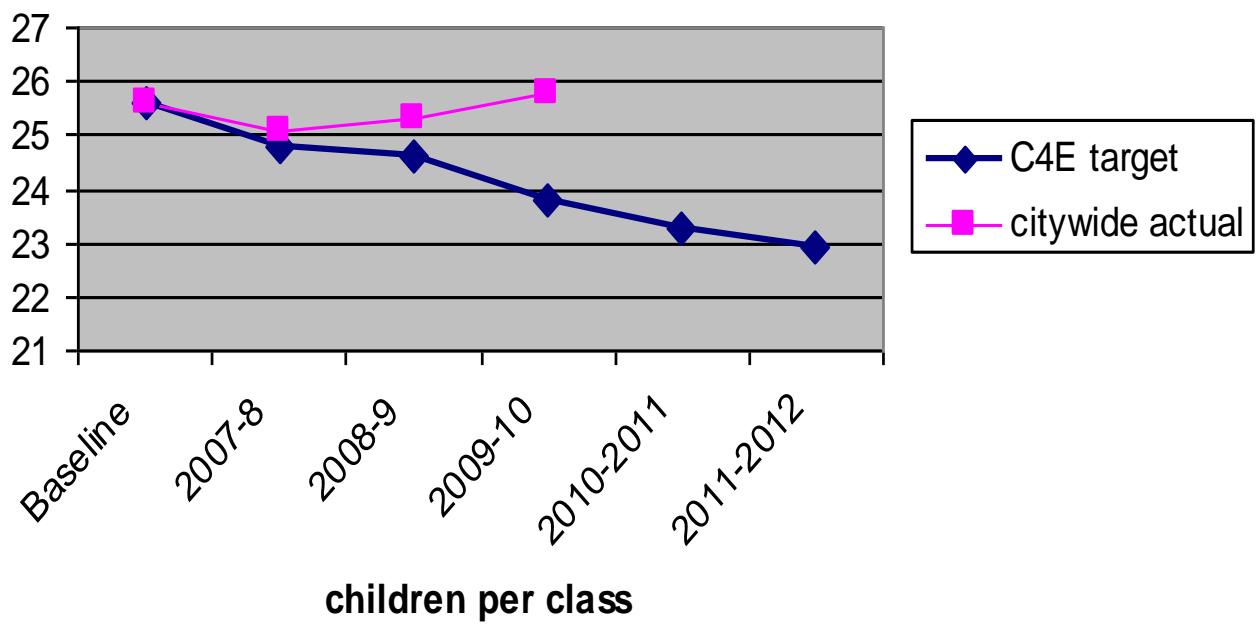
- Reducing class size one of only four K-12 reforms backed by “rigorous evidence,” acc. to Institute of Education Sciences, research arm of US Dept. of Education.\*
- Proven effect in narrowing the achievement gap;
- NYC has by the far the largest class sizes in state: in CFE case, Court of Appeals said that NYC students deprived of constitutional right to adequate education because of excessive class sizes;
- Between 2002-2009, while the number of out-of-classroom positions grew by over 10,000, number of general ed teachers shrunk by more than 1600. (*NY Times*, “With More Money, City Schools Added Jobs,” June 30, 2009).
- Over last two years, class size increased at an unprecedented rate; this year is expected to jump even higher, due to 2,000 lost teaching positions and 18,000 more students.

***\* Other three K-12 evidence-based reforms, include one-on-one tutoring by qualified tutors for at-risk readers in grades 1-3, Life-Skills training for junior high students, and instruction for early readers in phonemic awareness and phonics.***

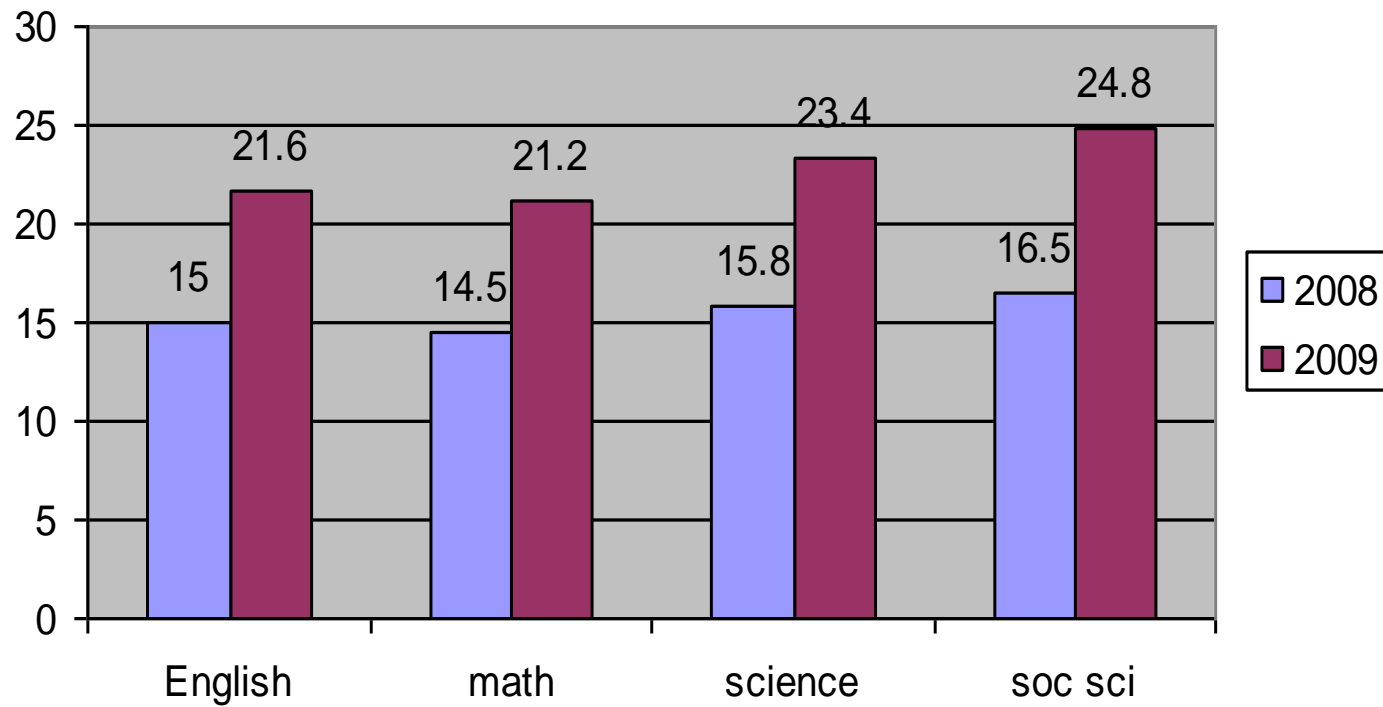
### Class size grades K-3



### Class size grades 4-8



## Percent of HS classes at 34 or more



# What's happening this year?

- This year, some Kindergartens are up to 30 students per class; some 1<sup>st</sup> grades are up to 34;
- Even so, DOE refusing to use \$200 million in federal funding awarded NYC to address hire teachers;
- Also DOE has wasted nearly \$1 billion in state CFE funds, since 2007, provided in exchange for promise to reduce class size in all grades.

# Nationally, results of high stakes testing

- Narrowing of curriculum and less time spent on art, music, gym, and science;
- Teaching to the test and less emphasis on creative thinking and problem solving;
- More cheating and “gaming” of system;
- More stress and anxiety for kids and their teachers;
- Suspending or “pushing out” low achievers to improve results;
- Fleeing of high-quality teachers from schools labeled as “low performing”
- Test score results become less reliable;
- And, as we have seen in NYC, does not lead to more learning!

# Yet Obama administration wants to impose even more high stakes testing across nation

- Requiring states receiving federal grants to impose teacher evaluation systems based in large part on test scores;
- Funding teacher merit pay schemes, based on test scores, that have not worked to improve learning;
- Requiring states to adopt “common core standards” and develop new computer-based assessments in every grade & subject.
- Forcing “low-performing” schools to be closed, and/or fire half their staff, and/or turned over to charter school operators.

# Eerie parallels between federal policies leading to financial crisis and now school reform

- **“Deregulation” in banking industry like risky federal policies forcing expansion of charter school sector;**
- **Idea that “free market” and competition will lead inevitably to better results, with minimal oversight, regulation and little or no quality control;**
- **Only accountability system is based on “results” in education, unreliable test scores;**
- **After financial deregulation, rating agencies inflated value and stability of risky mortgage-backed derivatives;**
- **One of these agencies (Standard and Poors) owned by McGraw Hill, which designed NY State exams w/inflated results;( also NYC interim assessments & exams for many other states);**
- **Financial deregulation and educational deregulation enthusiastically embraced by both parties, w/ little or no debate, because of influence of corporate world and the Billionaire’s boys club.**
- **Like financial crisis, these policies may end up further destabilizing our public school system, with taxpayers (and children) paying the price for failure and/or corruption in unregulated charter expansion.**



# Creeping privatization

## *example: Brooklyn*

- Robin Hood Foundation, large NYC foundation started by hedge fund operator
- Raised \$88 million at gala in one night last year;
- Used to fund major initiatives in NYC public schools to build libraries etc.
- Now shifting funding primarily to charter schools and affiliated causes;
- Emma Bloomberg, mayor's daughter, foundation program officer;
- In August, she proposed to city officials that foundation pay for renovation of Brooklyn public park (St. Andrews) in return for exclusive use during day by two charters, Uncommon Schools and Achievement First.