# DIFFERENTIATED ACCOUNTABILITY SESSION ONE: NEW APPROACHES TO MEASURING SCHOOL & DISTRICT PERFORMANCE

10:30-12:15



# DIFFERENTIATED ACCOUNTABILITY SESSION ONE: NEW APPROACHES TO MEASURING SCHOOL & DISTRICT PERFORMANCE

### Presenting Experts:

Jim Liebman, Columbia University

Chris Domaleski, National Center for the Improvement of Educational Assessment

Key Discussants:

Eric Smith

Kati Haycock, The Education Trust

Martha Thurlow, National Center for Educational Outcomes

Facilitated by:

Brad Jupp, U.S. Department of Education

# **ESEA Flexibility Forum**

September 30, 2011 Washington, D.C.

Chris Domaleski National Center for the Improvement of Educational Assessment

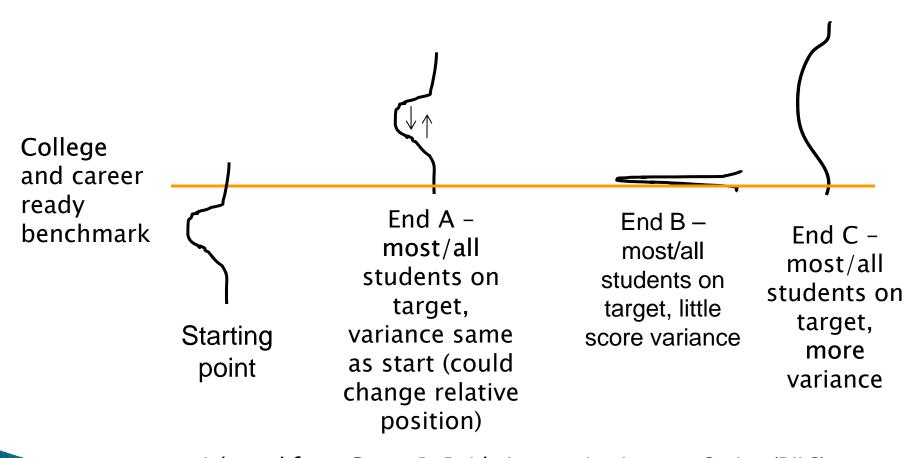
# Framing Questions

- What are the valued indicators of school performance?
- What are the prioritized outcomes?
- What evidence will bolster the credibility of the system?
- How will states create coherent accountability systems?

# Selecting Valued Indicators

- Focus on **broadening** the indicators, especially to account for **college and career readiness**.
- This may include:
  - Along-The-Way Indicators: "Is the student making progress toward readiness standards?" (e.g. course credit)
  - Attainment Indicators: "Has the student met readiness expectations?" (e.g. diploma or career credential)
  - Post Secondary Indicators: "What evidence certifies that the student is achieving post-secondary success?" (e.g. enrollment/performance in credit bearing course)
- Consider role of 'non-traditional' measures of effectiveness (e.g. academic behaviors, contextual skills, see Conley, D.T.)
- Importantly, indicators signal what is valued and have a strong connection to the theory of action for the system.

# Prioritized Outcomes - What Does Ideal Performance Look Like?

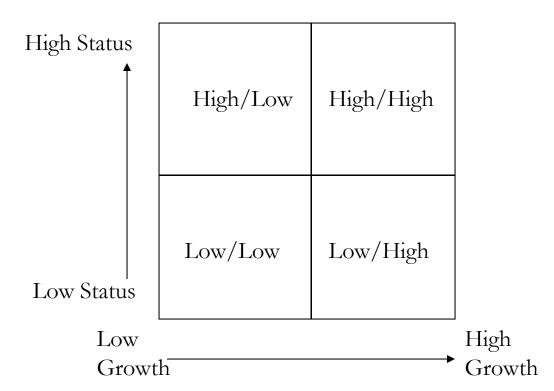


Adapted from Gong, B. Reidy Interactive Lecture Series (RILS), September, 2011.

# **Growth and School Type**

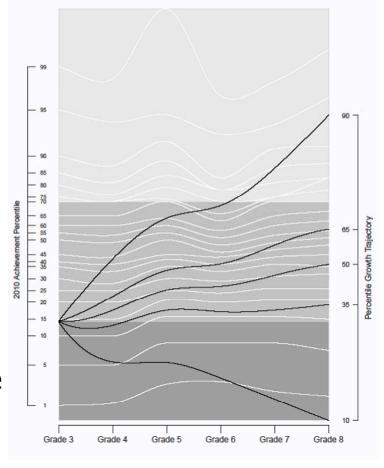
A thoughtful approach to growth targets should consider how standards interact with status.

#### **Status/Growth Combinations**



# **Growth Expectations**

- Two ways (at least) to consider growth:
  - Normative
  - Criterion (or growth to standard)
- Both are important and complimentary
- Ideally, the are used iteratively
  - Examine patterns of performance for schools and subgroups to set initial expectations for what is possible and what is reasonable
  - For students that grow at specified rates, what is the probability of attaining or maintaining target status?



Example depiction of norm and criterion referenced growth based on SGP. Betebenner, D. W.

# Evidence the System is Working

- To what degree are outcomes stable for schools/ groups of various types and size? (reliability)
- To what extent are the results associated with variables not related to effectiveness? (e.g. percent ED)
- What evidence bolsters the claim that classifications are credible? (e.g. related to other valued quantitative and qualitative indicators not modeled)
- Are the results useful for improvement?
- Are negative consequences mitigated?

Research is ongoing and should shape both initial design and ongoing refinements

# Coherence

- As states consider multiple accountability / performance management systems (i.e. student, educator, school) coherence across systems is critical.
- Systems should encourage mutually supportive actions.
- Consider:
  - Components
  - Consequences/ Supports
  - Performance Targets

# New Approaches to Measuring School and District Performance: Lessons from New York City

James S. Liebman
Simon H. Rifkind Professor
Columbia Law School

U.S. Education Department September 30, 2011

### **Main Points**

- The examples given here are <u>a</u>, not <u>the</u>, way to measure district and school performance. They show what is possible, not what exactly to do.
- These measures were developed over years, becoming better (more complex) over time. Start simple and improve. Listen to complaints by educators about what you're missing.
- There is vast technical flexibility to measure what you want. First, decide what you want to measure, then ask how to measure it, not the reverse.
- Base <u>what</u> you measure on an <u>overall strategy</u> for using information about student results to motivate educators to accelerate student learning.
- One overall strategy (~NYC's) is to identify the highest achieving schools, given their populations; make their results the benchmark against which other schools are measured; empower all schools to meet or beat the benchmark; as schools "beat" it, raise the bar.

# **Evaluation Embedded in Broader Theory of Action**

#### **EMPOWER**

**Devolve Authority to Districts and Schools, Minimizing Mandates and Categorical Grants** 

Fund Districts and Schools Per Pupil, Weighted to Need

**Replace Supervision with Accountable Facilitation** 

**Encourage Districts and Schools to Self-Affiliate into Clusters Based on Common Problems** 

**Manage Portfolio of Districts and Schools** 

#### **EVALUATE**

Rate Districts, Schools, Educators Based on Student Learning, Given Student Challenge (Lagging)

**Qualitatively Review Districts, Schools, Educators Based on Strategic Use of Available Tools (Leading)** 

**Survey Parents/Students/Teachers on School's Learning Conditions & Central Support (Leading)** 

**Align to State and Federal Metrics** 

#### **ENABLE**

Frequently Assess Student Learning, Strengths and Weaknesses

Provide Data to Educators and Parents, with Multiple Diagnostic Comparisons

Train Educators to Work in Teams to Diagnose and Cure Instructional Failure

Distribute Knowledge Horizontally (Small Innovation) and Conduct R&D (Big Innovation)

#### **ENFORCE CONSEQUENCES**

Reward Districts, Schools, Principals, Teams and Teachers Based on Student Learning Gains and Strategic Use of Available Resources

Close or Redesign Districts, Schools, Counsel Out Teachers, and Deny Tenure Based on Evidence of Chronic Failure

Align All Measures of Success and Forms of Recognition to Single Set of Evaluation Measures

# **Assumptions**

- State has College and Career Ready Expectations for all students
- State has High-Quality Assessments in reading and math aligned to those expectations that can measure student growth)
  - Short-term (modify existing assessments and data systems for these purposes)
     vs. long-term (adopt better assessments and data systems)
- State wants to develop Annual Measurable Objectives (AMOs) and use existing assessments to measure whether districts and schools meet them.
  - Applies to <u>all</u> districts and schools (not just Title I), while enabling State to use them to identify Priority, Focus, Rewards Schools for ESEA purposes
- NYC Example: All schools graded A, B, C, D, F based on progress- and performance-based AMOs
  - A = Reward Schools
  - C, D = Focus Schools (failing some populations, causing achievement gaps) (≥10%)
  - F = Priority Schools (failing all populations) ( $\geq 5\%$ )

# **NYC Progress Report (page 1)**



Progress Report

**ELEMENTARY** 

#### Progress Report Grade



#### What does this grade mean?

Schools are assigned letter grades based on their overall Progress Report score. Schools that get As and Bs are eligible for rewards. Schools that get Ds and Fs, or 3 Cs in a row, face consequences, including change in school leadership or school closure.

#### How did this school perform?

- This school's overall score for 2007-08 is 87.1
- This score places the School in the 98 percentile of all elementary schools Citywide—i.e., 98 percent of those schools scored lower than this school
- This school met 100% of its improvement target from last year

#### This Progress Report is for:

P.S. 023 Carter C. Woodson

(14K023) Sharon Meade

ENROLLMENT 342

SCHOOL

PRINCIPAL

SCHOOL TYPE ELEMENTARY

PEER INDEX 62.2

Category	Calculated Score		Category Grade
School Environment	<b>11.9</b> out of 15	0 15	Α
Student Performance	14.4 out of 25	0 25	В
Student Progress	56.3 out of 60	Ď 60	Α
Additional Credit	4.5 (15 max)	I	
Overall			

#### How scores translate to grades:

- Schools receive letter grades based on their overall score
- Schools with an overall score between 59.6–100 receive a letter grade of A
- 45% of schools earned an A in 2007-08

#### Elementary Table - Overall Grades

Grade	Score range	City summary
Α	59.6-100	45% of schools
В	45.8-59.5	38% of schools
С	32.6-45.7	13% of schools
D	28.4-32.5	3% of schools
F	9.6-28.4	2% of schools

#### In This Report:

Each school's Progress Report (1) measures student year-to-year progress, (2) compares the school to peer schools and (3) rewards success in moving all children forward, especially children with the greatest needs. The Progress Report measures four areas:

#### School Environment

uses parent, teacher and secondary student surveys and other data to measure necessary conditions for learning: attendance, academic expectations, communication, engagement and safety and respect.

#### Student Performance

measures student skill levels in English Language Arts and Math.

#### Student Progress

measures average student improvement from last year to this year in English Language Arts and Math.

#### Closing the Achievement Gap

gives schools additional credit for exemplary gains among high-need students.

The back page provides specific information about how the school performed in each of these areas.

#### **Quality Review Score**

Score

This school's 2007-08 Quality Review score is: Proficient

To see this school's Quality Review report, find the school's Web site at http://schools.nyc.gov/, click 'Statistics' and scroll down to Quality Review Report.

87.1 out of 100

#### State Accountability Status

Based on its 2006-07 performance, this school is: <u>In Good Standing</u>

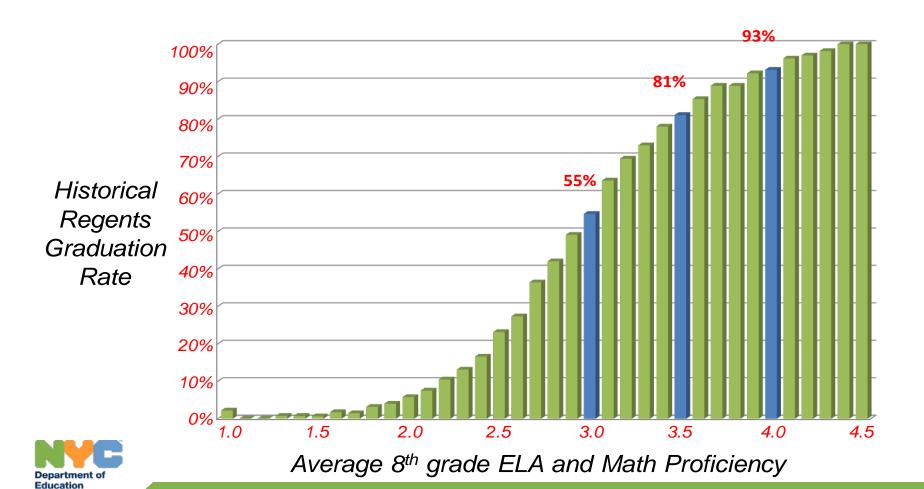
This status is determined by the New York State Department of Education under the No Child Left Behind (NCLB) Act. It is separate from the school's Progress Report Grade.

# **Guiding Principles**

#### AMOs should motivate and diagnose, not simply sort.

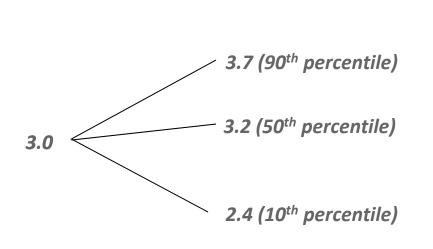
- **Proficiency and growth**: Evaluate districts and schools based on <u>learning levels</u> students reach and <u>gains</u> students make
- Measures of what educators add, not what students start with: Focus on outcomes that are not correlated with socioeconomic status, special populations or demographic characteristics
- All students and students most in need: E.g., AMOs based on median outcome of all students + median outcome of each or all "ESEA" populations
  - **Diagnostic transparency**: Ensure educators can re-create and verify results, so they know how they're being assessed and why they did or didn't do well—which schools and kids are not succeeding: no regressions
- Multiple measures: Measure desired outcomes and actions that generate them.
- Realistic measures based on <u>demonstrated</u>, <u>comparable</u> experience: E.g., base AMOs on how districts and schools performed compared to recent performance of <u>all</u> districts and schools statewide and <u>peer</u> districts and schools that serve similar student populations; modify periodically to reflect recent gains
- **Known targets**: Give districts and schools known targets to aim for each year; in theory, everyone can get an 'A' if all make significant progress (criterion referenced)

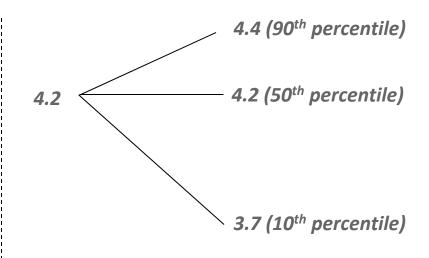
# Longitudinal Progress to and Beyond Proficiency Leads to in High School And After



Joel I. Klein, Chancellor

# Growth Percentiles Measure Progress on the Same Scale Even if Students Start from Different Places





- Going from a 3.0 to a 3.2 may be a 50<sup>th</sup> percentile outcome, while staying at 4.2 for two years may also be a 50<sup>th</sup> percentile outcome.
- Starting point can be <u>previous year's scale score or proficiency rating on the same test</u> (e.g., grades 3-8); or a <u>demographic profile</u> (e.g., for grades K-2); or a <u>prior score on a different test</u> (e.g., grade 9-12 exit or end –of-course exams anchored to scores on 8<sup>th</sup> grade test)
- Growth percentiles can be adjusted, e.g., for special education status; overage for grade

# **Guiding Principles**

#### AMOs should motivate and diagnose, not simply sort.

- **Proficiency and growth**: Evaluate districts and schools based on <u>learning levels</u> students reach and <u>gains</u> students make
- Measures of what educators add, not what students start with: Focus on outcomes that are not correlated with socioeconomic status, special populations or demographic characteristics
- All students and students most in need: E.g., AMOs based on median outcome of all students + median outcome of each or all "ESEA" populations
- **Diagnostic transparency**: Ensure educators can re-create and verify results, so they know how they're being assessed and why they did or didn't do well—which schools and kids are not succeeding: <u>no regressions</u>
- Multiple measures: Measure desired <u>outcomes</u> and <u>actions</u> that generate them.
- Realistic measures based on <u>demonstrated</u>, <u>comparable</u> experience: E.g., base AMOs on how districts and schools performed compared to recent performance of <u>all districts and schools statewide</u> and <u>peer districts and schools</u> that serve similar student populations; modify periodically to reflect recent gains
- **Known targets**: Give districts and schools known targets to aim for each year; in theory, everyone can get an 'A' if all make significant progress (criterion referenced)

# **Closing the Achievement Gap**

- Student <u>Progress</u> and <u>Performance</u> give 50% weight to <u>all students</u> and 50% weight to <u>lowest performing one-third</u> in the school
- Points also awarded for exemplary gains with ESEA populations:
  - We award schools "additional credit" for closing the achievement gap with high need populations
  - Elementary/Middle/K-8 Schools earn additional credit through exemplary gains on State tests, with their high need populations
  - High Schools earn additional credit through exemplary gains based on credit accumulation and/or Regents results with their high need populations

#### **5 ESEA Populations**

- 1. English Language Learners
- 2. Special Education Students
- 3. Hispanic Students who are in the Lowest Third Citywide
- 4. Black Students who are in the Lowest Third Citywide
- 5. Other Students who are in the Lowest Third Citywide



# **Guiding Principles**

#### AMOs should motivate and diagnose, not simply sort.

- **Proficiency and growth**: Evaluate districts and schools based on <u>learning levels</u> students reach <u>and gains</u> students make
- Measures of what educators add, not what students start with: Focus on outcomes that are not correlated with socioeconomic status, special populations or demographic characteristics
- All students and students most in need: E.g., AMOs based on median outcome of all students + median outcome of each or all "ESEA" populations
- **Diagnostic transparency**: Ensure educators can re-create and verify results, so they know how they're being assessed and why they did or didn't do well—which schools and kids are not succeeding: <u>no regressions</u>
- Multiple measures: Measure desired <u>outcomes</u> and <u>actions</u> that generate them.
- Realistic measures based on <u>demonstrated</u>, <u>comparable</u> experience: E.g., base AMOs on how districts and schools performed compared to recent performance of <u>all districts and schools statewide</u> and <u>peer districts and schools</u> that serve similar student populations; modify periodically to reflect recent gains
- **Known targets**: Give districts and schools known targets to aim for each year; in theory, everyone can get an 'A' if all make significant progress (criterion referenced)

### Multiple Measures: All Schools, Elementary/Middle Schools

# All schools – metrics giving parents, teachers, students a stake in evaluation and measuring how well schools are organized to accelerate learning

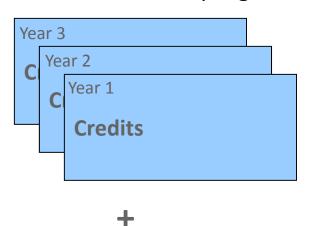
- Attendance
- Annual School Survey of Parents, Teachers, Secondary Students
  - > Academic Expectations
  - Communication
  - Engagement of Students in Learning
  - Safety and Respect

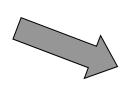
# MS/K-8 – metrics to expand data sources and subject areas beyond State tests in ELA and Math and to promote high school readiness

- English core course (Common Core) passing rate
- Math core (Common Core) course passing rate
- Science core course passing rate
- Social studies core course passing rate
- High school credit earned in 8<sup>th</sup> grade

# Multiple Measures: High School

We measure graduation rates as well as the key performance indicators that track progress toward graduation

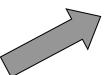








- Regents (how fast students complete them; how many pass)
  - ✓ English✓ Math☐ Science☐ U.S. History☐ Global Studies



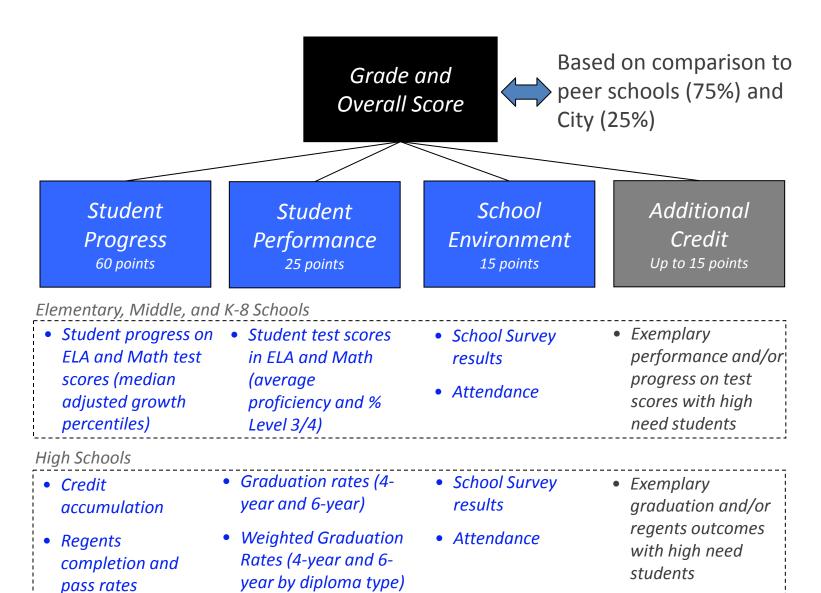
- Associates Degree
- Advanced Regents
   Diploma with Honors
- Advanced Regents
   Diploma
- Regents Diploma
- Local Diploma
- GED

# Multiple Measures: College Readiness (HS)

#### **Metrics**

- 1. College Prep Course Index: Percentage of students in the graduation cohort who have:
  - taken/scored 65+ on Algebra II Regents exam,
  - taken an Advanced Placement (AP) course/scoring 3+ on AP exam,
  - taken an International Baccalaureate (IB) course/scoring 4+ on IB exam,
  - received college credit through a dual enrollment program (College Now, Early College, etc.), or
  - taken/passed another approved college ready course/assessment.
- 2. College Readiness Index: Percentage of students in the graduation cohort who have passed out of remedial requirements set by City College of NY, by the time their cohort is scheduled to graduate. Calculated using both SAT and Regents exam scores.
- 3. College Enrollment Rate: Percentage of students in the graduation cohort who enroll in a two- or four-year postsecondary institution in the fall after graduating, according to data from the National Student Clearinghouse.

# How Multiple Measures Roll Up Into Overall Score



### Multiple Measures: Quality Review

- An 2.5-day on-site review of how effectively schools use data to improve student achievement. The Quality Review:
  - Provides schools with feedback on what is working well and areas in need of improvement
  - Informs school goals and plans for aligning resources to meet student needs
- A narrative report on every school is published on each school's Web site

#### Possible scores:

- Well-Developed
- Proficient
- Underdeveloped with Proficient Features
- Underdeveloped

#### Key events of the Quality Review:

- Case study
- Class visits
- Conversations with multiple constituencies
- Observation of collaborative activity

# **Multiple Measures: Quality Review Criteria**

Quality Statement 1: Instructional and organizational coherence					
1.1: Rigorous and engaging curriculum	1.2: Differentiated instruction aligned to beliefs	1.3: Aligned resource-use	1.4: Positive learning environment		
Quality Statement 2: Gather and analyze data					
2.1: School-level assessment data analysis	2.2: Alignment of assessments to curriculum	2.3: Data tools for decision-making	2.4: Data reflection and communication		
Quality Statement 3: Plan and set goals					
3.1: School-level theory of action and goals	3.2: Teacher team and classroom-level goals	3.3: Assessments used to make adjustments and provide feedback	3.4: Clear expectations and family engagement		
Quality Statement 4: Align capacity-building					
4.1: Instructional focus and differentiated support for teachers	4.2: Teacher teams engaged in collaborative inquiry	4.3: Teacher leadership development and instructional capacity	4.4: Support for meeting child/youth development needs		
Quality Statement 5: Monitor and revise					
5.1: Evaluate instructional and resource decisions	5.2: Evaluate assessment and data systems	5.3: Evaluate planning and goal-setting systems	5.4: Evaluate adult capacity-building systems		

# **Guiding Principles**

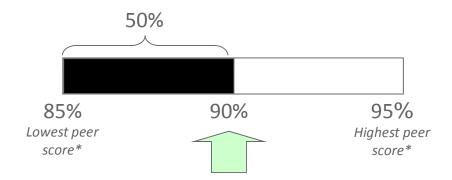
#### AMOs should motivate and diagnose, not simply sort.

- **Proficiency and growth**: Evaluate districts and schools based on <u>learning levels</u> students reach <u>and gains</u> students make
- Measures of what educators add, not what students start with: Focus on outcomes that are not correlated with socioeconomic status, special populations or demographic characteristics
- All students and students most in need: E.g., AMOs based on median outcome of all students + median outcome of each or all "ESEA" populations
- **Diagnostic transparency**: Ensure educators can re-create and verify results, so they know how they're being assessed and why they did or didn't do well—which schools and kids are not succeeding: <u>no regressions</u>
- Multiple measures: Measure desired <u>outcomes</u> and <u>actions</u> that generate them.
- Realistic measures based on <u>demonstrated</u>, <u>comparable</u> experience: E.g., base
   AMOs on how districts and schools performed compared to recent performance of
   <u>all districts and schools statewide</u> and <u>peer districts and schools</u> that serve similar
   student populations; modify periodically to reflect recent gains
- **Known targets**: Give districts and schools known targets to aim for each year; in theory, everyone can get an 'A' if all make significant progress (criterion referenced)

# **Example: Attendance (Peer Horizon)**

School A has an attendance rate of 90%

- The attendance rates for schools in <u>School A's peer</u> group ranged from 85% to 95%
- School A scored exactly in the middle between the lowest and highest score in its peer group
- Therefore, School A's
   <u>Peer Horizon score</u> would be 50%



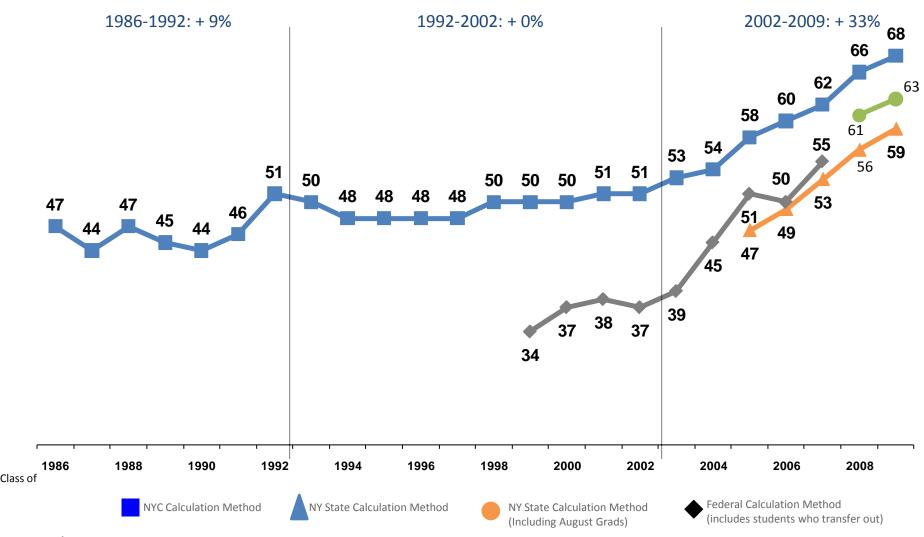
School A's attendance is **50%** of the distance
between the lowest and
highest scores <u>in its peer</u>
group



Note: Minimums and maximums are established using +/- 2 standard deviations from the mean

# After Remaining Flat for 16 Years, NYC's Graduation Rate Has Increased by 33% Since 2002

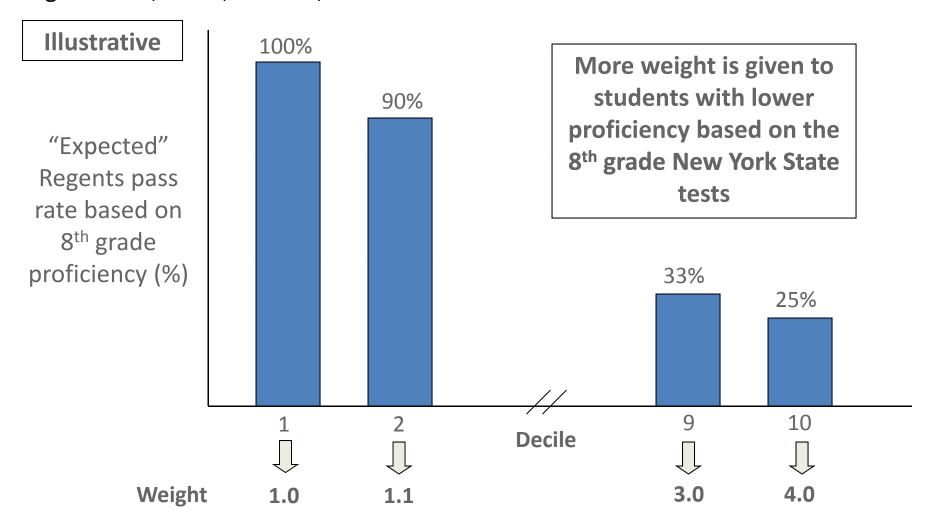
Percent of Students in a Cohort Graduating from High School in 4 Years



# **Appendix**

# **High School: Weighted Regents Pass Rates**

The Weighted Regents Pass Rates measure progress made since the 8<sup>th</sup> grade ELA, Math, Science, and Social Studies tests

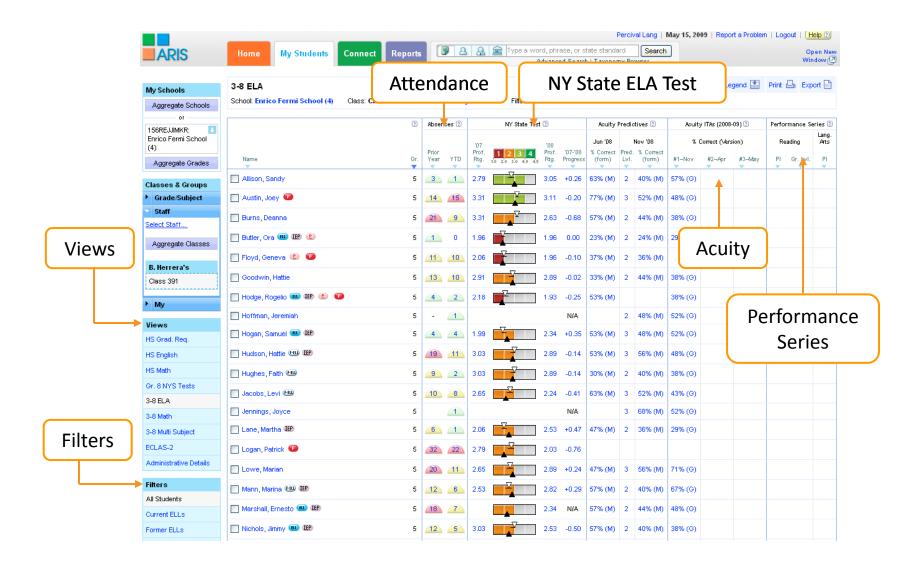


### Peer groups

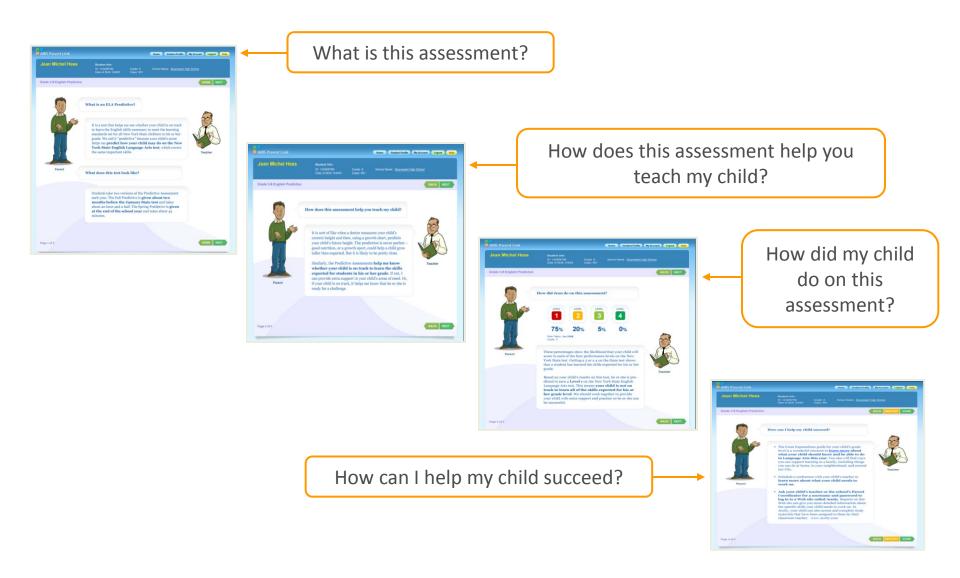
- A peer group is a group of schools with similar student populations that serve approximately the same grade levels
  - > For elementary and K-8 schools, peer groups are determined based on a comparison of student demographics across schools
  - For middle schools and high schools, peer groups are determined based on a comparison of student performance on ELA and Math test scores
- Peer groups consist of up to 40 schools serving approximately the same grade levels (i.e., elementary schools have only other elementary schools in their peer group; same goes for middle schools, K-8 schools, and high schools)
- Each school has a unique peer group (so each school can be in the middle of its peer group)



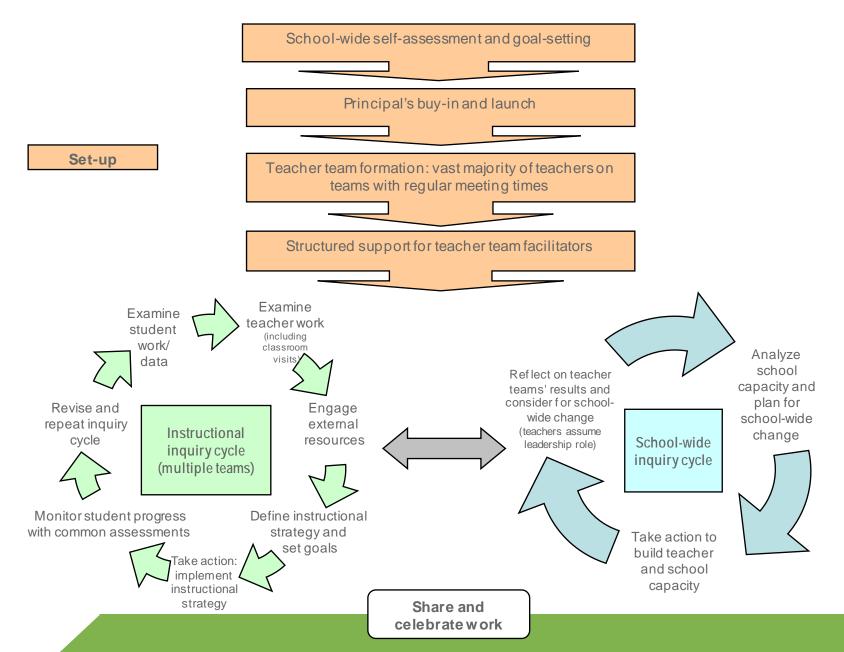
# Enable: ARIS 'My Students' 3-8 English Language Arts View



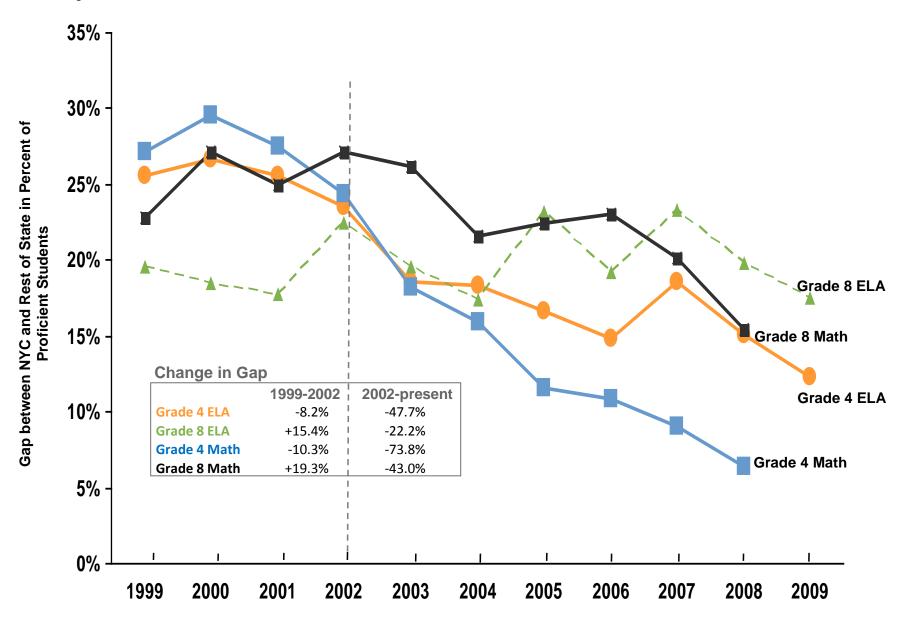
# **Enable: ARIS Parent Link Customized Parent Walkthroughs**



# **Enable: Inquiry Teams**

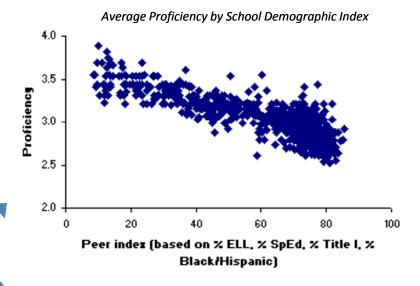


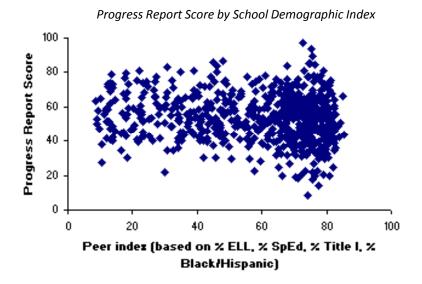
#### Gap in Percent of Proficient Students Between NYC and Rest of State



# **Progress Report Objectives**

- •<u>Broad Scale</u>: integrate diverse sources of observation and data about schools and outcomes
- Outcomes not Inputs: evaluate schools based on measurable results
- Progress and Performance: measure what schools contribute to students, not what students bring to schools
- Peer and City Comparisons:
   compare similar schools on growth
   measures; hold all schools to Citywide
   standard





### **Progress Report**



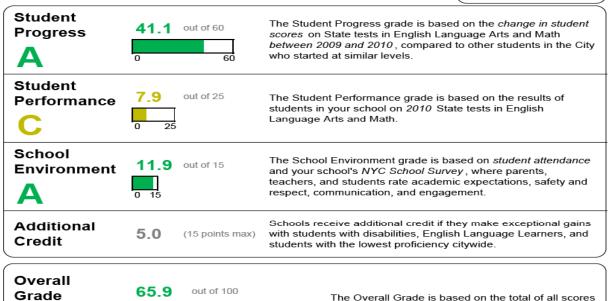
#### 2009-10 Progress Report Overview

ELEMENTARY and MIDDLE SCHOOLS

The NYC School Progress Report informs families about the school's strengths and weaknesses, emphasizing how far students have come in the past year. Progress Report grades are made up of 3 sections: student progress, student performance, and school environment. To view your school's Progress Report in detail, visit ARIS Parent Link at arisparentlink.org.

#### Overall Progress Report Grade





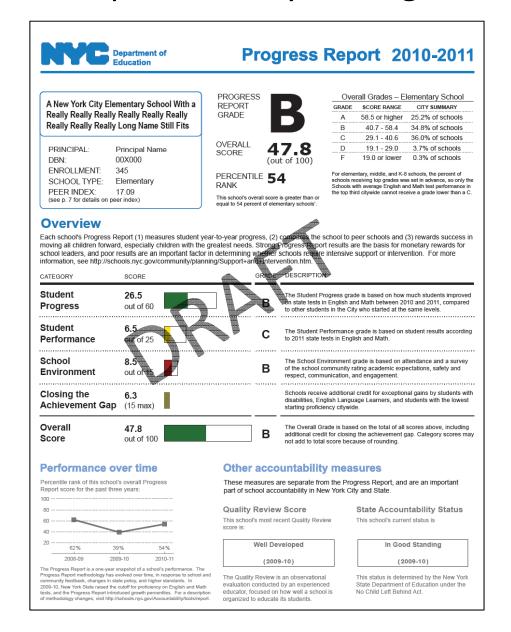


Note: In light of changes in State tests and Progress Report methodology, schools cannot drop more than two letter grades from last year to this year. Further, schools with top performance on English and Math tests cannot receive a grade lower than C.

100

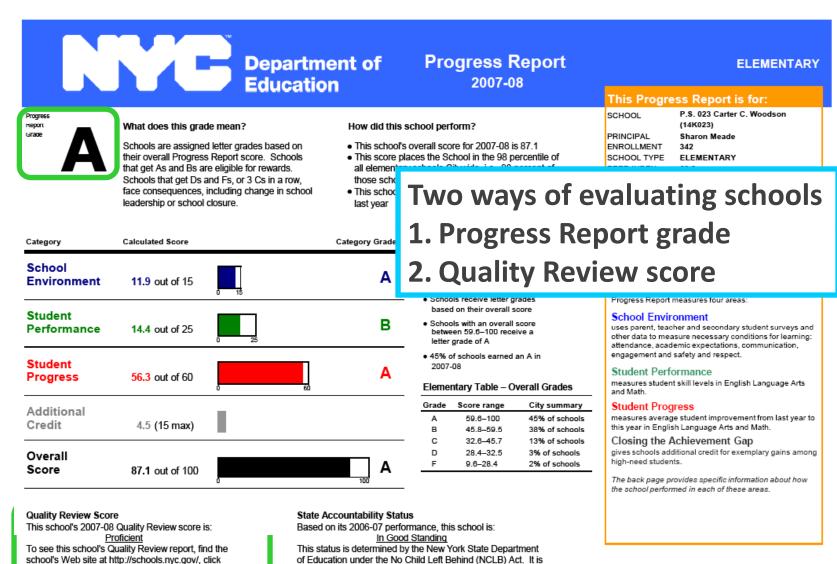
above, including additional credit.

### New templates clarify scoring and metrics



# **NYC Progress Report (page 1)**

'Statistics' and scroll down to Quality Review Report.



separate from the school's Progress Report Grade.