2022 School Quality Measures Conference

Breakfast: 8:00 — 9:00 am
Rooms 314, 324

WiFi: MIT GUEST

Blueprint Labs

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	**Welcome & Introductions**
9:40 — 10:20 am	Constructing bias-free SQM
10:20 — 10:55 am	Equity and School Quality
10:55 — 11:10 am	Break
11:10 am — 11:30 am	Case Studies: Boston and Chicago
11:30 am — 12:00 pm	Small group discussions
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	Developing and communicating multiple SQM
1:15 — 2:10 pm	Measuring school quality beyond test scores
2:15 — 2:45 pm	Involving the community in developing SQM
2:45 – 3:00 pm	Small group discussions

Blueprint Labs

Welcome & Introduction

Parag Pathak
Class of 1922 Professor of Economics, MIT
Faculty Director, MIT Blueprint Labs

Our mission:

To use data, economics, and analytic tools to uncover the consequences of policy decisions and improve society.

Education

Health Care

Workforce

Our Specialties

Research Design

- What is the impact of an intervention or policy?
- Statistical and econometric methods
- ~~Correlation~~ *Causal impact*

Market Design

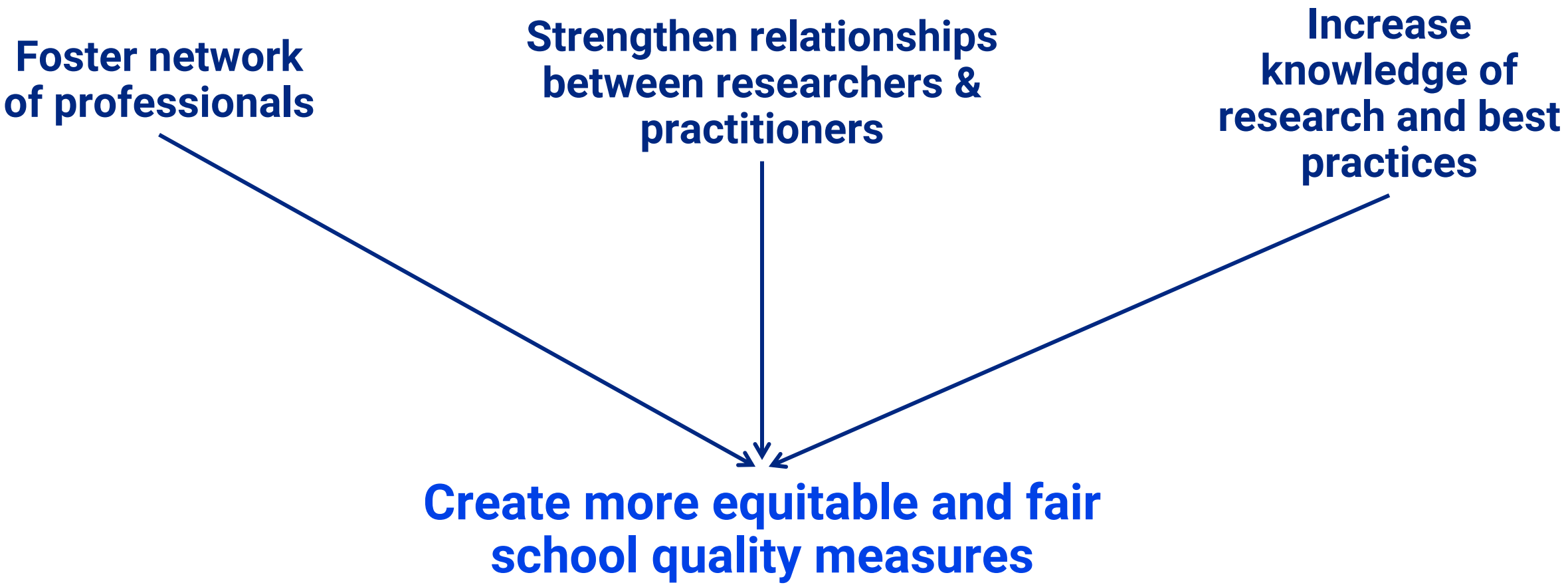
- How can we efficiently and fairly allocate scarce resources?

These tools are useful for answering questions in a variety of fields, including education.

Research and Policy Partnerships

Research	→	Policy	→	Impact
Research design studies on centralized assignment systems	→	Helped design and improve systems in Boston, Chicago, Denver, & NYC	→	More equitable enrollment processes

Today's Goals



Blueprint Labs

Overview & Introductions

Talia Gerstle
Policy and Communications Manager, MIT Blueprint Labs

Hybrid Norms

If on Zoom:

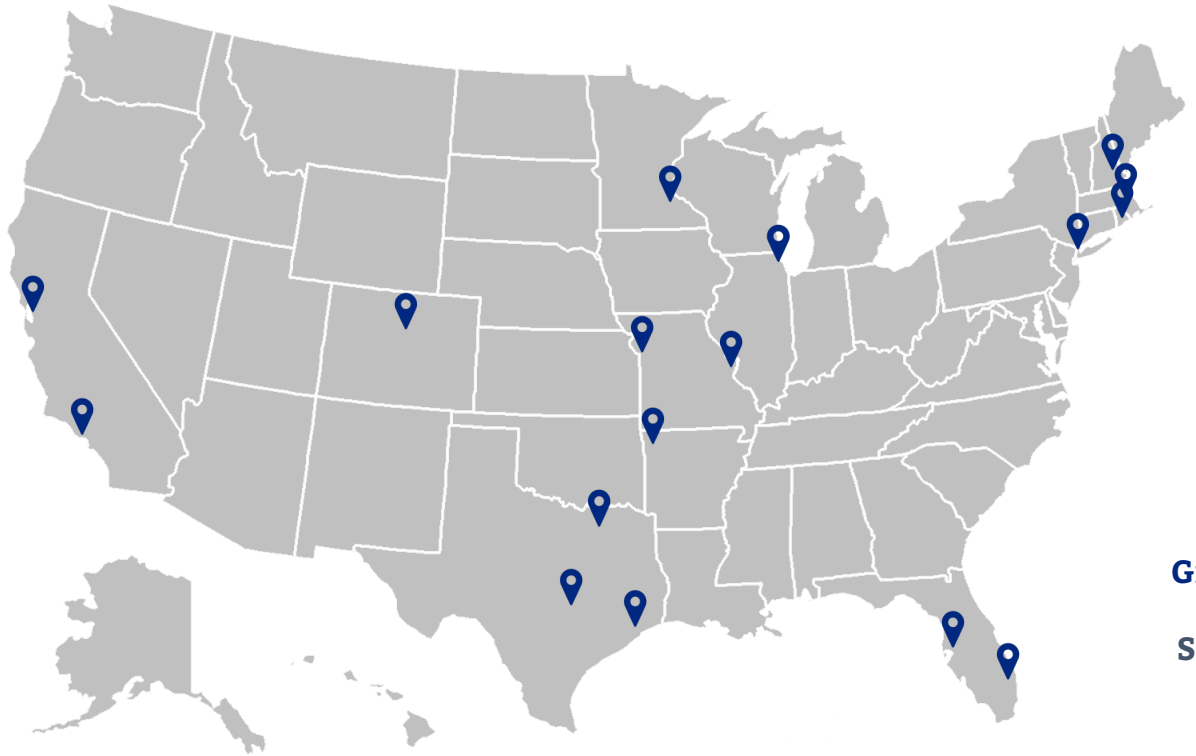
- Mute yourself when not speaking
- Direct message Jennifer Jackson with any issues or questions
- Leave video on when possible

If in person:

- Room 314 down the hall is available if you want to step out
- Bathrooms: back to elevators and to the left

Who is in the room?

Locations



Titles

Senior Data Strategist
Director of Teaching and Learning
Data Analyst
Director, Research, Evaluation, and Assessment
Director of Analytics
Executive Director of State Strategy and Student Opportunity
Executive Director for School Performance
Director of Research and Data
Chief Transformation Officer
Policy Manager
Assistant Director of Campus and District Accountability
Director of Performance Management
Director, School Quality Measurement and Research
Executive Director of Student Services and School Choice
Senior Data Analyst
Director of Research, Assessment, and Accountability
Executive Vice President, Strategy and External Affairs
Director of Research and Evaluation
Associate Director
Executive Director, Research, Data, and Assessment
Director, Research & Data Partnerships
Director of Product Design
Gifted/Talented Education Administrator
Director of Analytics
Chief Executive Officer
Strategy, Learning, and Evaluation Officer
Senior Associate
Executive Director of Accountability and Assessment
Vice President, Data Strategy
Executive Director
Vice President, Product and Design

Introductions

1. Name

2. Organization and role (please introduce others from your team too)

3. Districts:

- Elevator pitch of school quality measure(s)
- One goal regarding accountability

4. Others:

- Quick overview of your org, your work on accountability

Introductions order:

1. Aldine ISD
2. Austin ISD
3. Boston Public Schools
4. Chicago Public Schools
5. Denver Public Schools
6. Garland ISD
7. GreatSchools
8. Highlander Charter
9. Hillsborough County
10. Illinois State Board of Education
11. LAUSD
12. Navigate STL schools
13. NCIEA
14. NYCDOE
15. Oakland Enrolls
16. Oakland USD
17. Palm Beach County Public Schools
18. RIDE
19. St. Paul Public Schools
20. Consultants
21. City Fund
22. Gates Foundation
23. Walton Family Foundation

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	Welcome & Introductions
9:40 — 10:20 am	**Constructing bias-free SQM**
10:20 — 10:55 am	Equity and School Quality
10:55 — 11:10 am	Break
11:10 am — 11:30 am	Case Studies: Boston and Chicago
11:30 am — 12:00 pm	Small group discussions
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	Developing and communicating multiple SQM
1:15 — 2:10 pm	Measuring school quality beyond test scores
2:15 — 2:45 pm	Involving the community in developing SQM
2:45 – 3:00 pm	Small group discussions

Blueprint Labs

Constructing bias-free school quality measures

Peter Hull

Assistant Professor of Economics at Brown University and Research Affiliate at Blueprint Labs

Agenda

1. **Background and Study Motivation**

2. **Credible School Value-Added with Undersubscribed School Lotteries**

3. **Race and the Mismeasure of School Quality**

Two key definitions

School quality: the causal effect of attending a school on later student achievement

- High-quality schools boost achievement for students of a given background and preparation level.

Selection bias: the non-causal variation in school performance measures due to students' background, motivation, or ability

- Schools may appear more or less effective because of the types of students they enroll, rather than because of causal effects.

See an example of selection bias in the Appendix.

Two popular school quality measures

Decreasing selection bias

Measure	Advantages	Drawbacks
Achievement levels	▪ Easy to calculate ▪ Straightforward interpretation ▪ Minimal data requirement	▪ No correction for selection bias, which can lead to misleading conclusions
Student progress / growth	▪ Less selection bias, by adjusting for baseline academic achievement	▪ Some bias remains from unmeasured determinants of student growth ▪ Requires linking student data over time

Our team has developed two new measures that improve on these by leveraging centralized assignment data

Selection bias can have unintended consequences

If ratings have selection bias:

1. Families may be led to favor schools with more advantaged backgrounds (e.g., family income), even though these characteristics don't necessarily predict school quality
2. State/district accountability systems may misidentify schools for interventions (e.g., expansion or closure)

One of Blueprint's core goals is to develop new tools to study and minimize selection bias in school quality ratings.

Agenda

1. Background and Study Motivation

2. **Credible School Value-Added with Undersubscribed School Lotteries**

3. Race and the Mismeasure of School Quality

Study design

Research question:

1. Can we use centralized assignment data to construct reliable estimates of school quality?
2. How much selection bias do such measures remove from more conventional quality measures?

Study sample: Middle and high school students in Denver and New York City

Methodology: Use the quasi-random assignment in centralized systems to construct and validate a new quality measure

Angrist, J., Hull, P., Pathak, P.A., and Walters, C.R. (2021). "Credible School Value-Added with Undersubscribed School Lotteries," *MIT Blueprint Labs Discussion paper 2021.10*.

Risk-Controlled Value-Added Model (RC VAM)

Computing RC VAM

- **Uses centralized assignment** of students to schools, using a matching algorithm that provides data on:
  - Schools that students applied to
  - Priorities that students have at each school (e.g., sibling priority)
  - Tiebreakers (often random)

- **Key idea:** student application patterns and admissions priorities proxy for ambition, family background, and other unmeasured outcome determinants

- **RC VAM** uses linear regression to adjust the end-of-year state test scores of enrolled students at a school by:
  - Baseline student academic achievement
  - Student demographics
  - *Assignment risk,* computed from student application patterns and admission priorities

The result is a highly accurate estimate of the causal impact of attending a school on student test scores!

Using RC VAM to measure school quality requires three strands of data

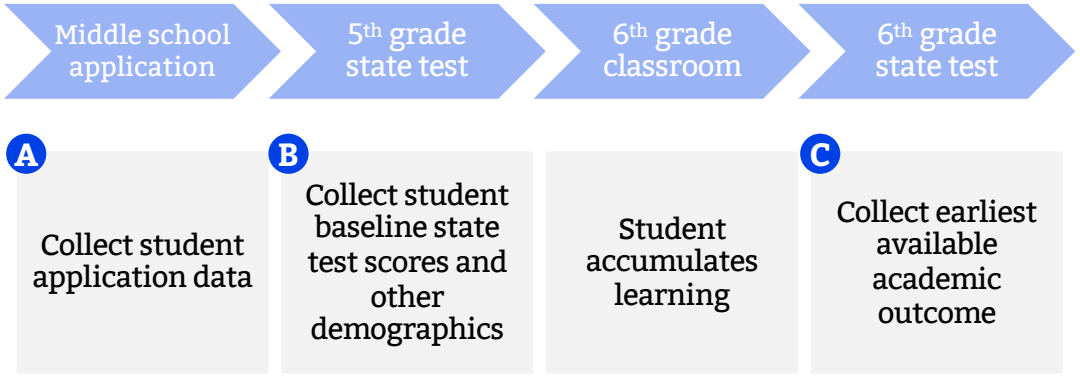
Ingredients to RC VAM

A Data generated by the **centralized student assignment algorithm**:
- Rank-order lists submitted by students
- Priorities of each student at each school
- School tiebreakers
- School offers (algorithm result)

B Baseline **student characteristics**
- Prior year academic achievement
- Demographics

C **Outcomes**, for example
- End-of-year state test scores
- Attendance rate
- Number of suspensions
- HS graduation status
- College-going

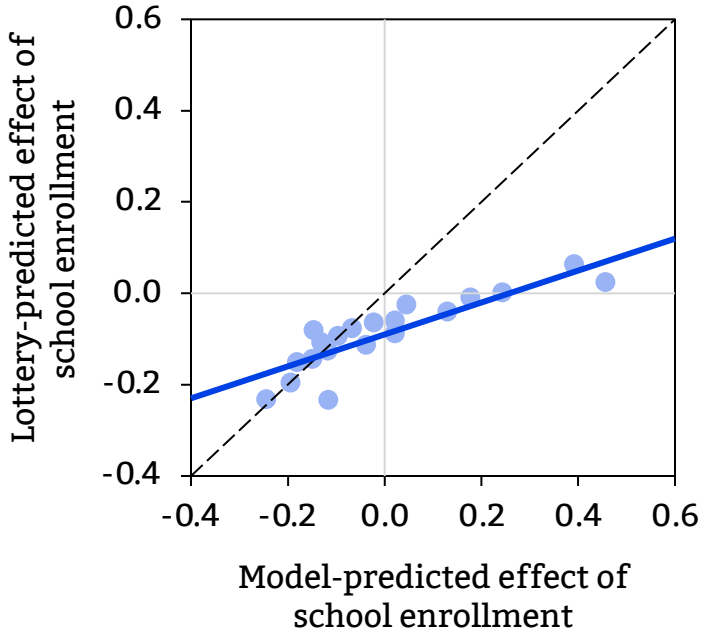
Example timeline for middle school data collection



Validating RC VAM using student assignment lotteries

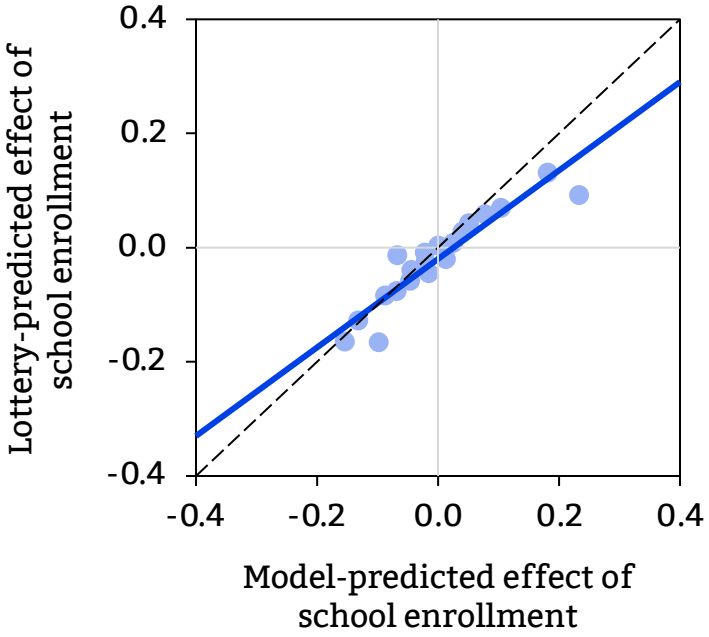
Comparing high school quality measures to test score outcomes in NYC

Uncontrolled VAM



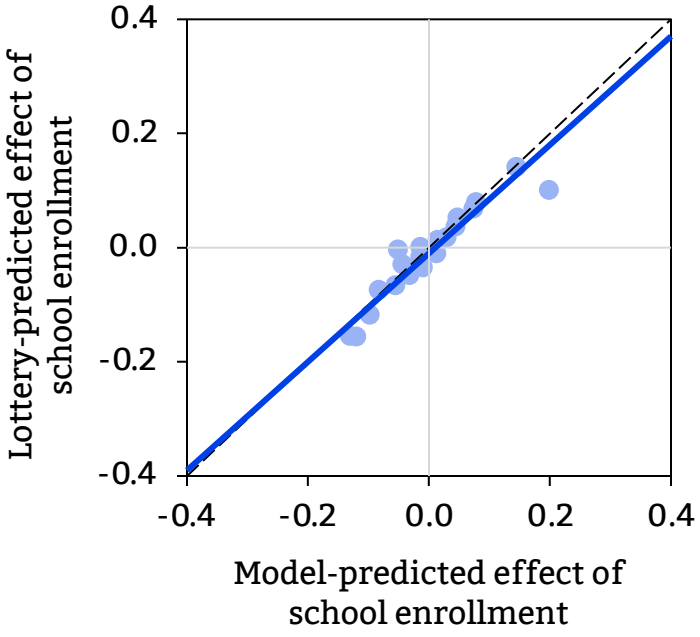
No adjustment (analogous to "levels" quality measure)

Conventional VAM



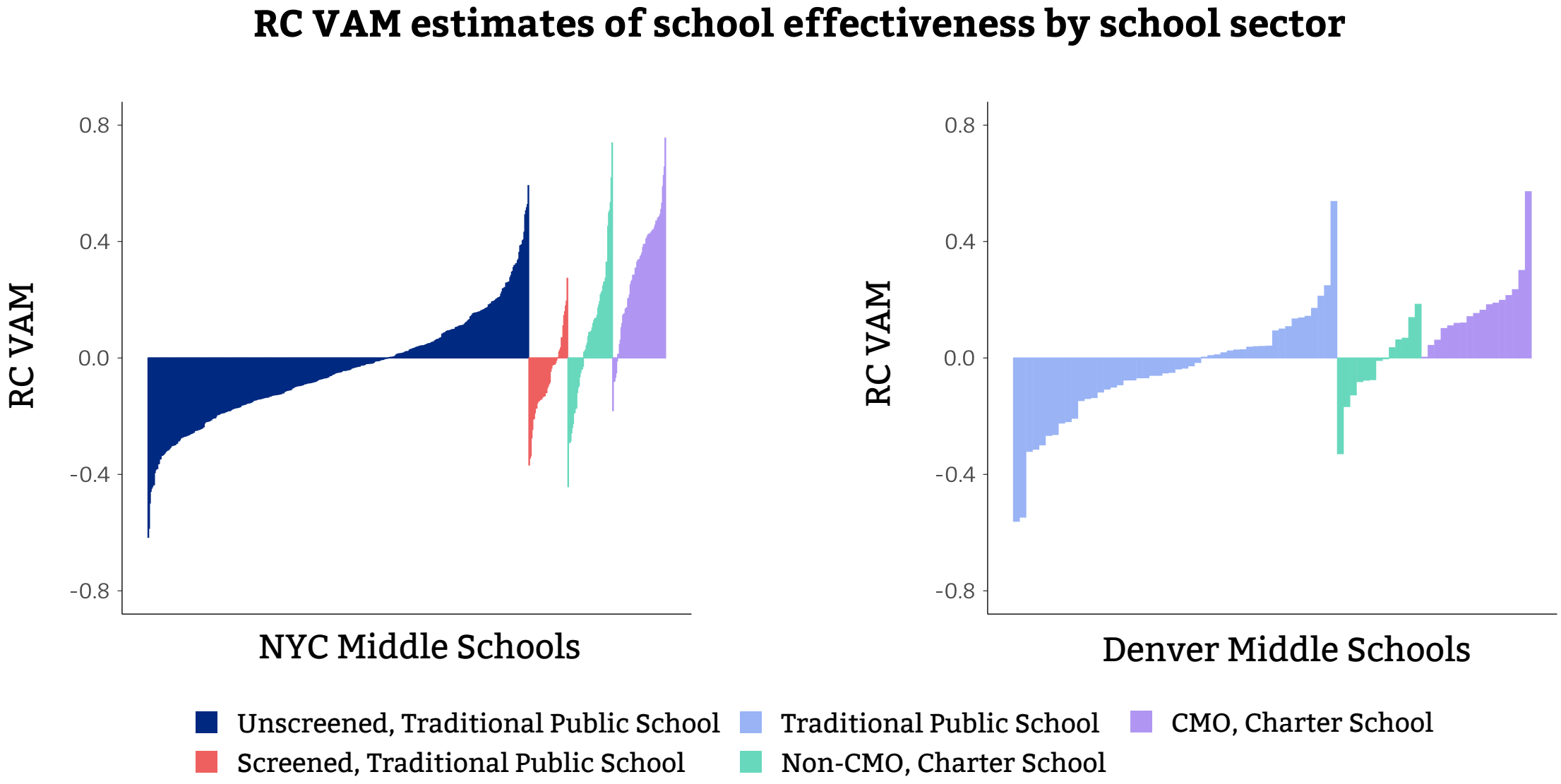
Adjustments for:
- 8th grade state test scores
- 8th grade demographics

RC VAM



Adjustments for:
- 8th grade state test scores
- 8th grade demographics
- **Assignment risk**

Using RC VAM: middle and high school impacts differ by school sector



Agenda

1. Background and Study Motivation

2. Credible School Value-Added with Undersubscribed School Lotteries

3. **Race and the Mismeasure of School Quality**

Motivation for this paper

1. Popular school ratings often exhibit a strong correlation with student racial composition.

Higher-ranked schools tend to have more white/Asian students, fueling concerns over racial segregation.

2. This correlation could be due to either school quality or selection bias.

Do higher-quality schools enroll a larger share of white/Asian students?

3. If the explanation is selection bias, there may be a "free lunch."

We can remove the "racial imbalance" from existing ratings without reducing their ability to predict school quality.

Study design

Research questions:

1. Does the racial imbalance of levels and growth ratings reflect school quality or selection bias?
2. Can the correlation between ratings and race be addressed, without making the ratings less effective guides to true school quality?

Study sample: 6th grade students in Denver and New York City

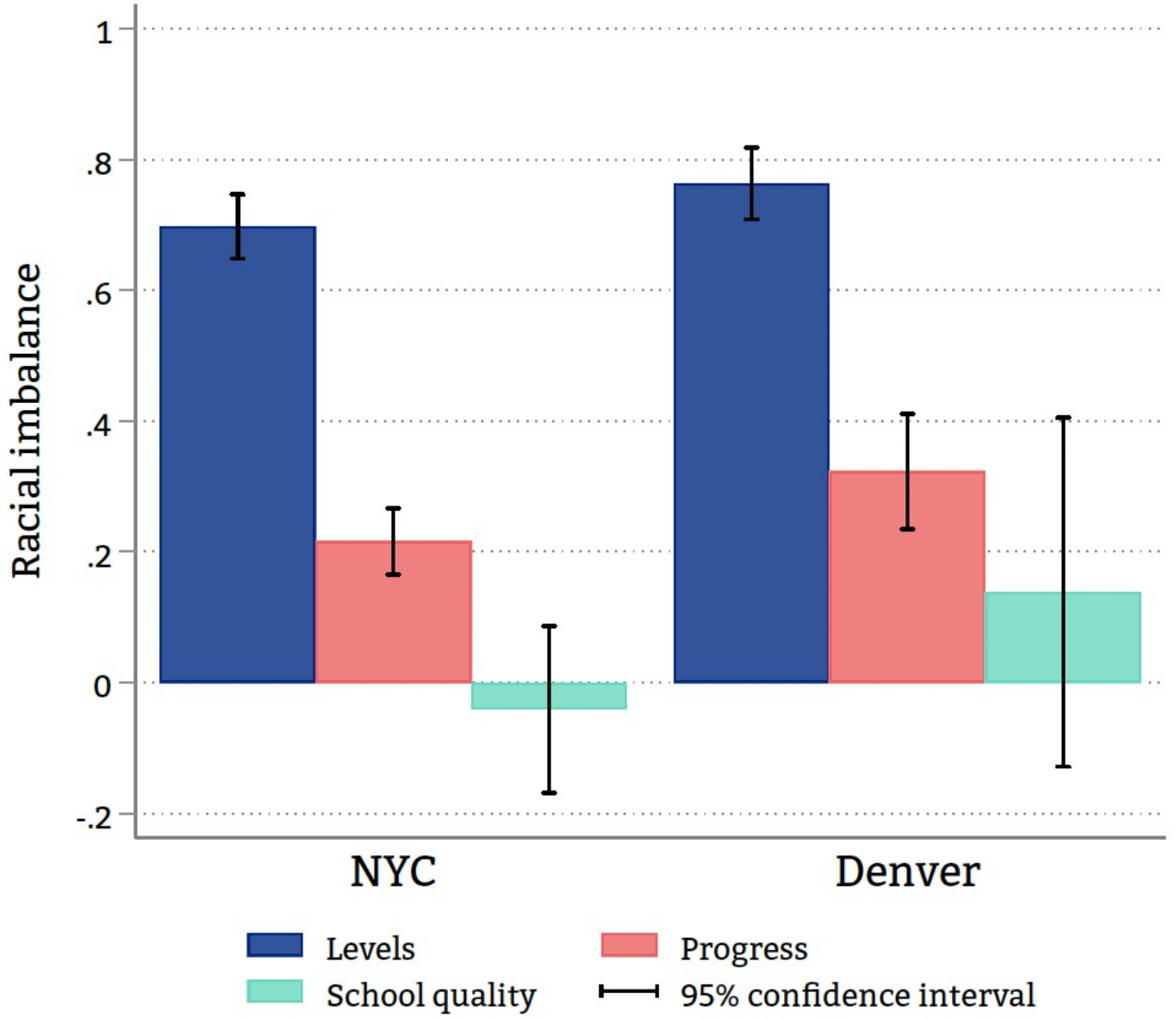
Methodology: Use quasi-random centralized assignment to some schools to measure school quality and correlate it with racial shares.

Angrist, J., Hull, P., Pathak, P.A., and Walters, C.R. (2022): "Race and the Mismeasure of School Quality," *MIT Blueprint Labs Discussion Paper #2022.01*

Finding #1: Levels ratings are strongly correlated with race while progress ratings are much less so. Causal school quality is uncorrelated with race.

Graph interpretation:
"Racial imbalance" is the relationship between the rating and the share of enrolled white students (specifically, the linear regression coefficient).

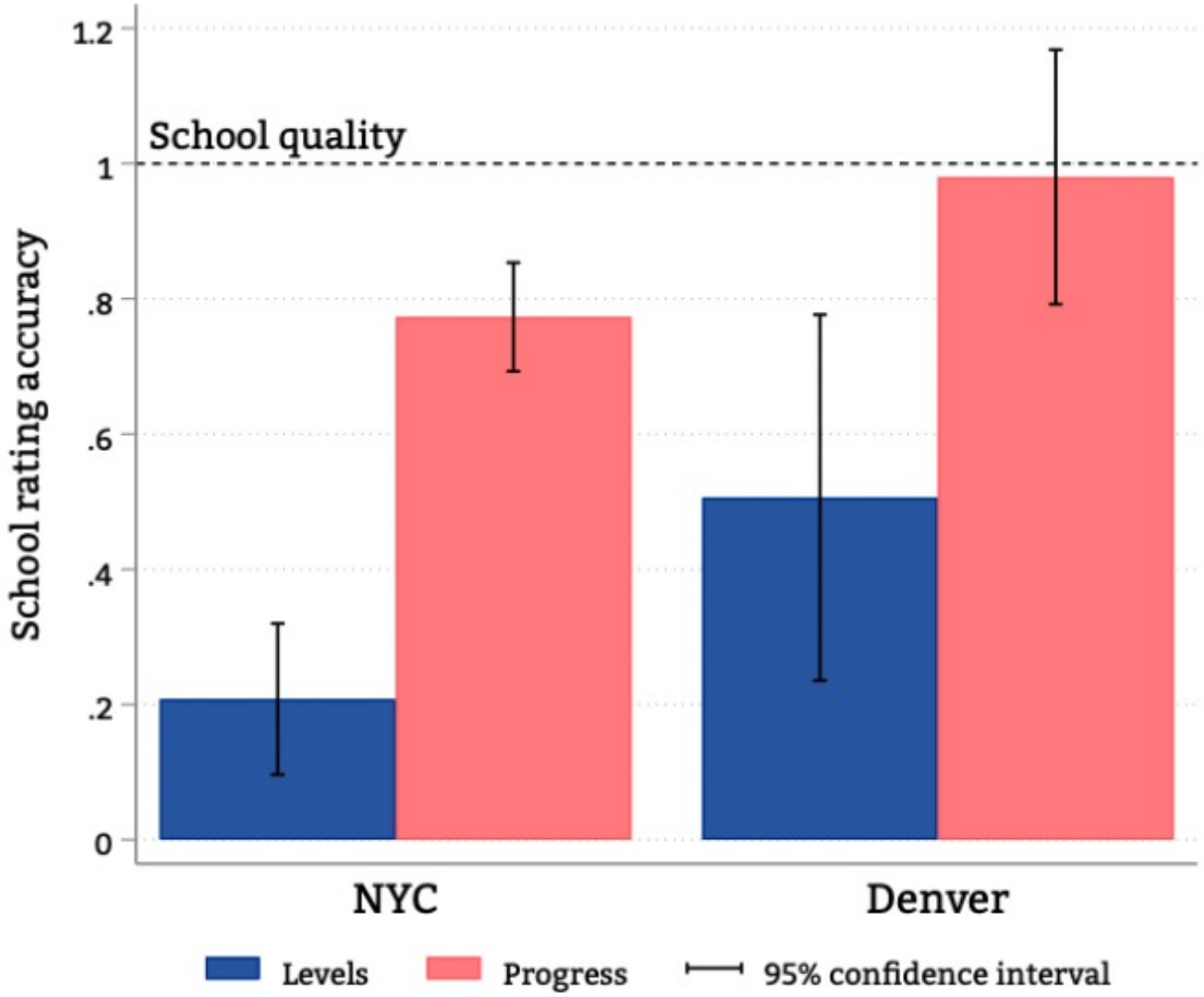
E.g. the dark blue bar in NYC indicates that moving from a school with zero white students to all white students is associated with a 0.7 standard-deviation increase in the levels rating.



Finding #2: Progress ratings predict school quality much more accurately than levels ratings.

Graph interpretation:
"Rating accuracy" is the relationship between school quality and the rating (specifically, the linear regression coefficient). A coefficient of one means a one-for-one prediction of quality.

E.g. the dark blue bar in NYC indicates that moving to a school with a one standard deviation higher levels rating is associated with a 0.2 standard deviation increase in school quality.



Together, these findings suggest the correlation between ratings and race in New York City and Denver can be removed from a highly predictive measure of school quality.

Finding #3: A new "race balanced" progress measure eliminates the correlation with race and performs as well or better than existing measures.

Race-balanced progress uses a simple regression adjustment to remove the existing progress rating's correlation with race.

- We find it predicts school quality as well or better than the unadjusted measure.
- It is simple to compute, and could easily be adopted by school districts.
  - Given by the residuals of an OLS regression of progress on race.
  - Does not require centralized assignment data.

	Predictive accuracy (R^2)	Racial imbalance
Levels	.043	.697
Progress	.597	.216
Race-balanced progress	**.632**	**.000**

Next Steps

1. **Generalizability**
   a. Do these findings hold in other cities? We'd be very excited to partner with other districts/states in order to find out.

2. **Parent/family/district decision-making**
   a. How would families respond to these new school quality measures?
   b. How can these measures be best used in school finder websites or district/state decisions?

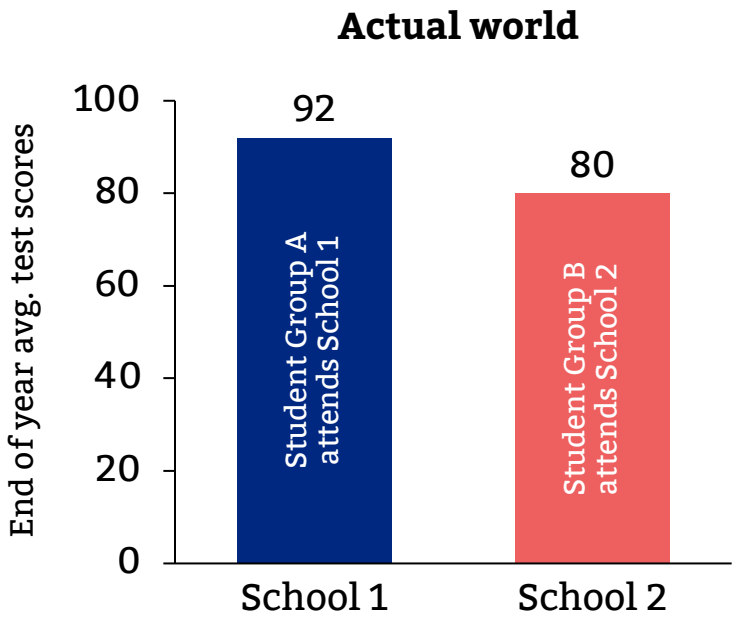
3. **Implementation**
   a. How can these findings be most useful for school ratings providers, districts, and states?

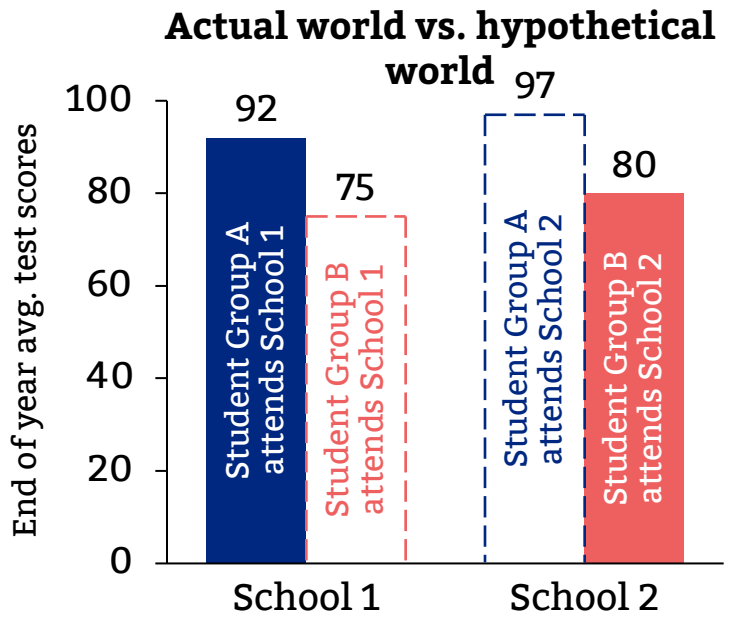
Thank you!

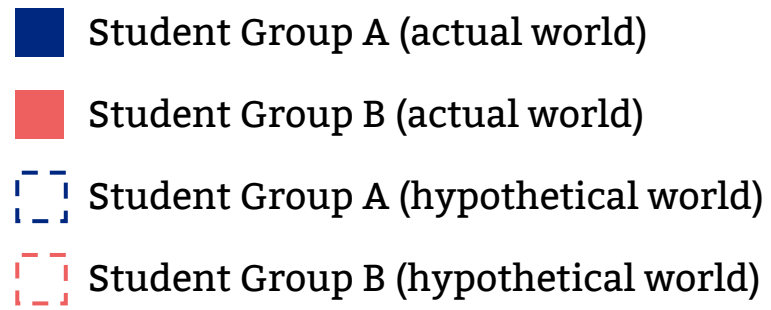
Questions/comments? ***peter_hull@brown.edu***

Blueprint Labs

Illustrative example of selection bias







School 2 is the higher quality school. Test score averages in the actual world imply School 1 is more effective than School 2. If hypothetically School 2 had been assigned Student Group A, Student Group A would have **higher** test scores at School 2 than they do attending School 1 in the actual world. Differences between Student Group A and B are obscuring measurement of school effectiveness using test score averages.

Selection bias definition

To learn more, check out materials on the Blueprint website

For more on **RC VAM**, see our working paper and results website

For more on **Race and Mismeasure of School Quality**, see our working paper and policy brief

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	Welcome & Introductions
9:40 — 10:20 am	Constructing bias-free SQM
10:20 — 10:55 am	**Equity and School Quality**
10:55 — 11:10 am	Break
11:10 am — 11:30 am	Case Studies: Boston and Chicago
11:30 am — 12:00 pm	Small group discussions
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	Developing and communicating multiple SQM
1:15 — 2:10 pm	Measuring school quality beyond test scores
2:15 — 2:45 pm	Involving the community in developing SQM
2:45 – 3:00 pm	Small group discussions

Blueprint Labs

Equity and School Quality

Jon Deane
CEO, GreatSchools

Orville Jackson
VP, Data Strategy, GreatSchools

Great! SCHOOLS.org

MIT Blueprint Labs' School Quality Conference

April 29, 2022

Equity and School Quality

Jon Deane
Chief Executive Officer

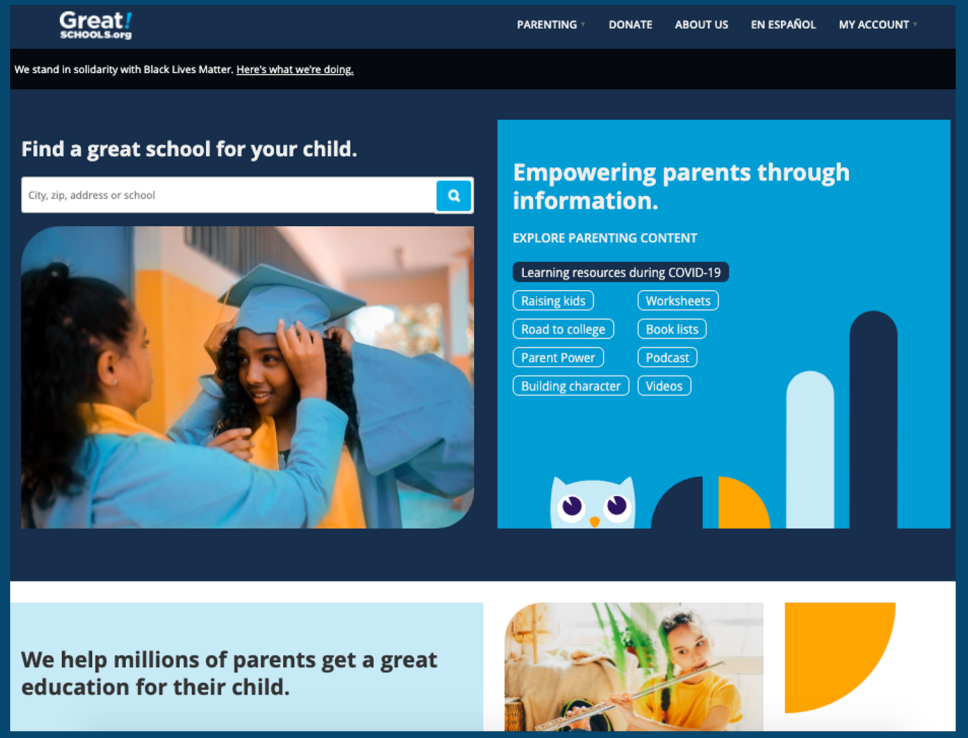
Orville Jackson, Ph.D.
Vice President, Data Strategy

Great! SCHOOLS.org

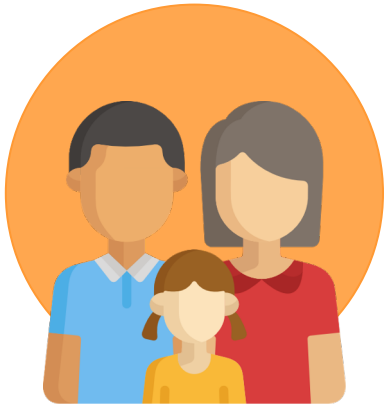
Agenda

- GreatSchools' mission
- Context for the evolution of our Equity rating
- Overview of recent changes
- Exploring implications of new research
- Next steps
- Q&A

Our mission is to provide high-quality information that supports **parents** pursuing a great education for their child, **schools** striving for excellence, and **communities** working to diminish inequities in education.



Great! SCHOOLS.org



Parents & Caregivers

- **49 million** unique site visitors per year
- **~40%** from low-income households
- **750,000** weekly newsletter subscribers

Educators

- **~17%** of our audience
- **30,000** subscribers to our educator newsletter
- Ability to **add info** to school profiles

Researchers

- Share our national and state data sets
- Examples: MIT, Columbia, Tulane, University of California

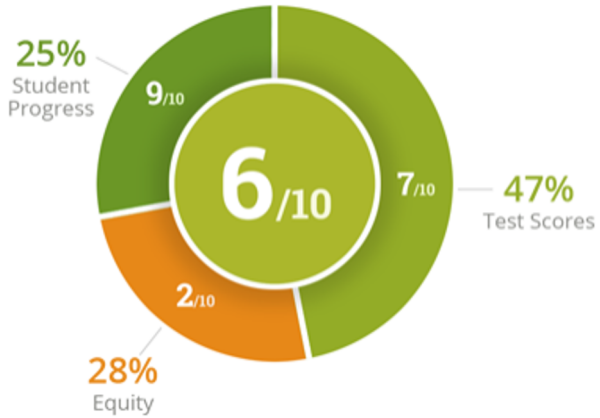
Icons by Freepik

Evolution of GreatSchools' ratings

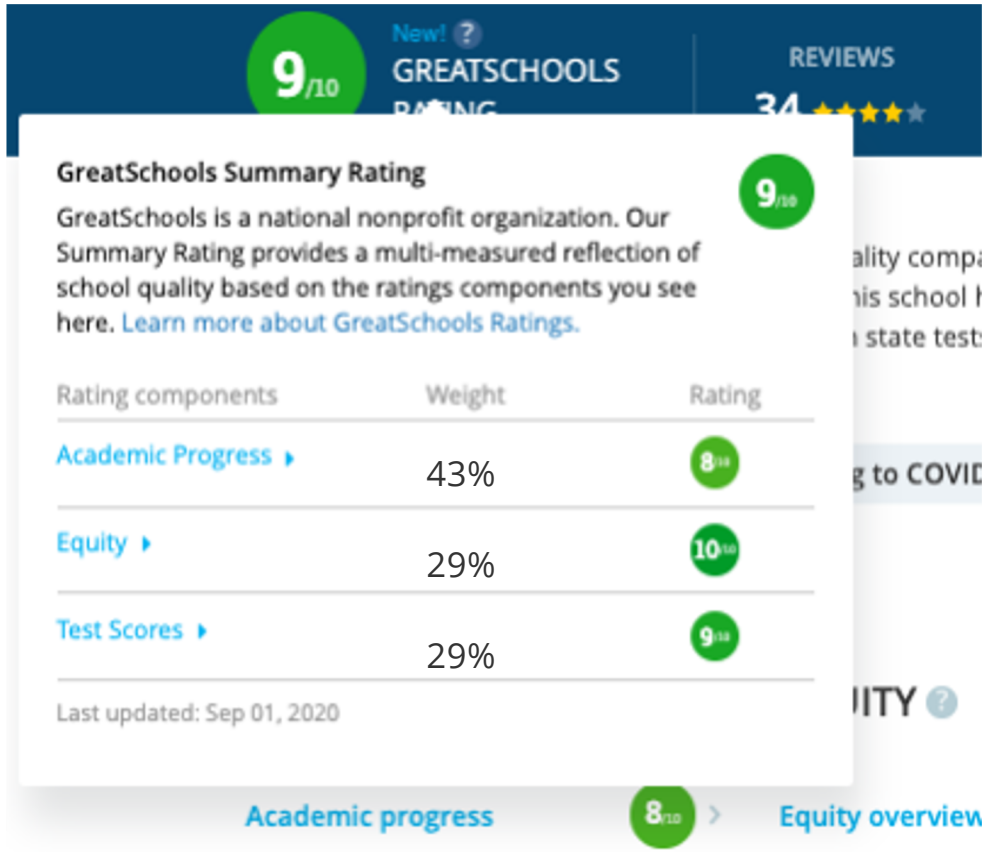
Expanding data beyond test scores and focusing on equity

9
out of 10

Original Rating
based mainly
on test scores.

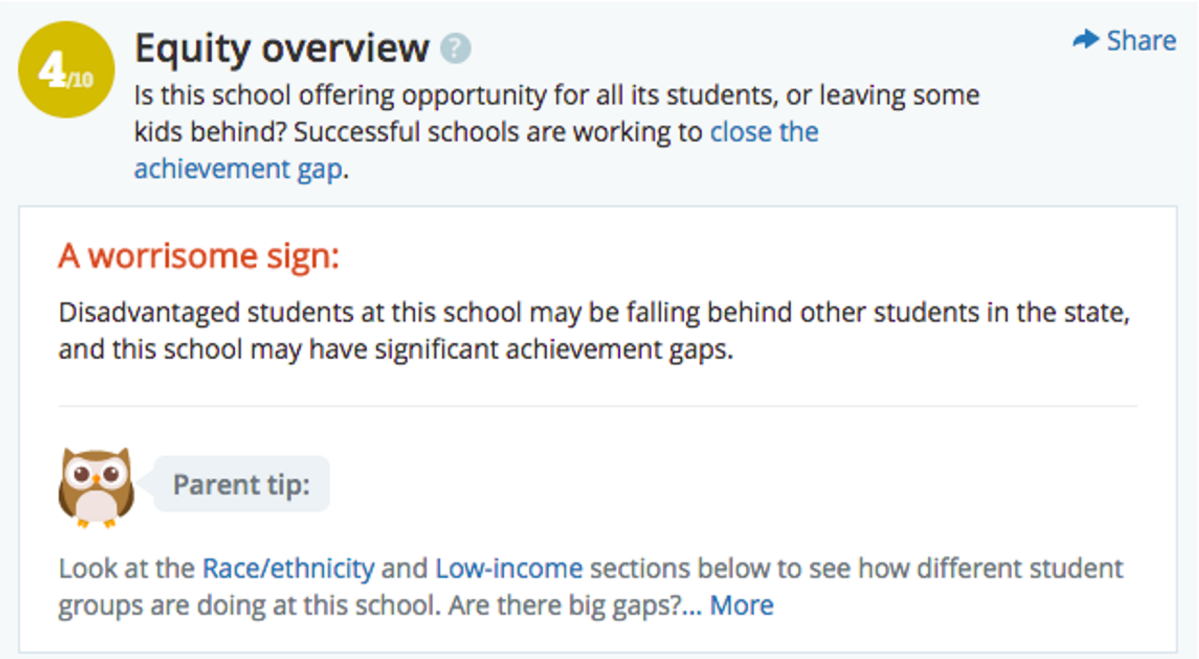


2017
Summary Rating update
Included test scores, equity, advanced coursework, and discipline/attendance flags.



2020 update now puts a greater weight on equity and student growth.

Initial Equity rating goals

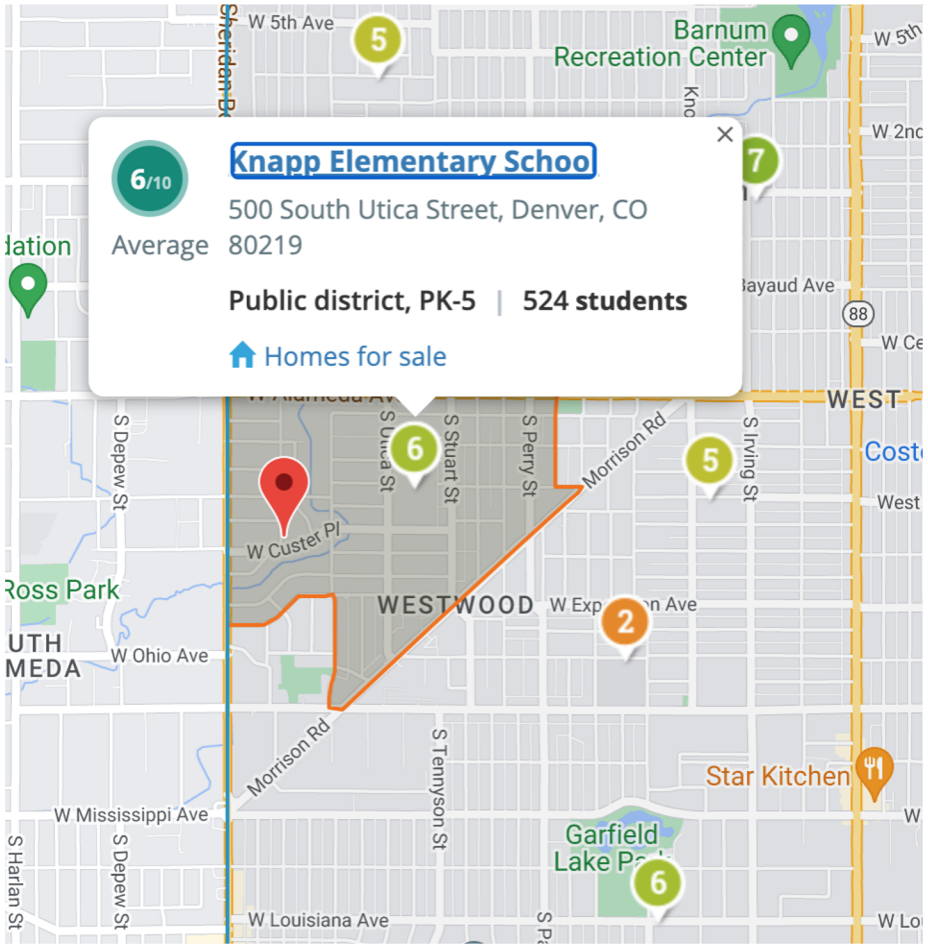


GreatSchools.org, 2017

- Reflect a school's success in serving historically disadvantaged students
- Measure whether students at a school achieve the same level of academic success

Equity challenges

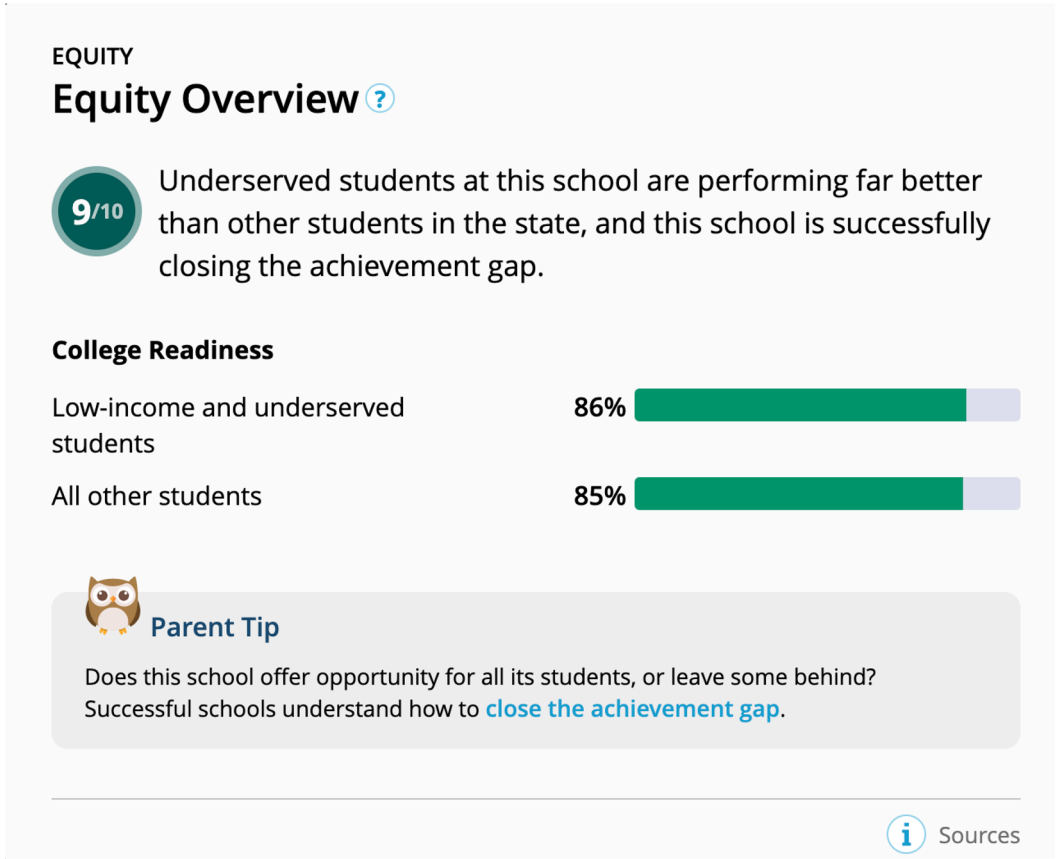
Moving beyond proficiency and highlighting strengths



- The initial Equity rating's focus on proficiency levels overlooked high-need schools with low test scores that are doing well with regard to equity in other domains.
- Schools with very few underserved students were unable to receive Equity ratings.
- As a result, we incorporated additional measures into our equity methodology, including student growth and college readiness.

2020 Update: Improving our equity lens

- **Research Advisory Committee**
  - Facilitated regular convenings to gather input on our approach
- **Learned from parents and partnered with advocacy organizations**
  - Engaged GreatSchools' Parent Council and advocacy organizations in discussions about equity
- **Pre-launch pilot**
  - Gradual rollout of the ratings in Michigan and California to examine the impact of the ratings methodology



2020 Ratings methodology

Emphasizing growth

Rating	From		To
Growth	• • •	↑	• • • •
Equity	• •	↑	• • •
Growth Proxy	•	↑	• • • •
Test	• • • •	↓	• • •
College Readiness	• • • •	↓	• • •

2020 update emphasized growth above all, as well as growth proxy. We also emphasized equity's importance in the overall Summary Rating, weighing it equally with Test/College Readiness.

Blueprint research: Moving forward

- Blueprint Labs' research has exciting implications to inform the next step in ratings evolution.
  - Opportunity to present a more accurate picture of school impact that controls for non-school factors
- What are we doing currently?
  - Actively modeling and sharing with researchers and others to work through complex questions regarding research.
    - *For example, how might this work apply to other districts, regions, states? Can it be used with other measures?*
  - Exploring how we might best use this information to inform parents through user testing and research with real parents and audiences to understand how they understand and act on this data

Beyond Ratings: Additional quality indicators

1) **Data reflecting equitable practices and resources**
- Climate surveys
- School programs and curriculum
  - Data partnerships
- Career pathways

1) **Highlighting achievement**
- Data-driven awards / badges
  - College Success Award
  - Thrive Award

3) User-generated content
- Parent insights
  - Reviews
  - Surveys
- School leader input
  - Direct profile edits

Opportunities

- Encourage school leaders in your district to **claim their profile**
- Partner with us to **pilot new data** types in your district or state
- **Share your input.** What else do you think parents need to know?

Contact our team at

research@greatschools.org

Thank you

Great! SCHOOLS.org

Q&A

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	Welcome & Introductions
9:40 — 10:20 am	Constructing bias-free SQM
10:20 — 10:55 am	Equity and School Quality
10:55 — 11:10 am	**Break**
11:10 am — 11:30 am	Case Studies: Boston and Chicago
11:30 am — 12:00 pm	Small group discussions
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	Developing and communicating multiple SQM
1:15 — 2:10 pm	Measuring school quality beyond test scores
2:15 — 2:45 pm	Involving the community in developing SQM
2:45 – 3:00 pm	Small group discussions

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	Welcome & Introductions
9:40 — 10:20 am	Constructing bias-free SQM
10:20 — 10:55 am	Equity and School Quality
10:55 — 11:10 am	Break
11:10 am — 11:30 am	**Case Studies: Boston and Chicago**
11:30 am — 12:00 pm	Small group discussions
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	Developing and communicating multiple SQM
1:15 — 2:10 pm	Measuring school quality beyond test scores
2:15 — 2:45 pm	Involving the community in developing SQM
2:45 – 3:00 pm	Small group discussions

Blueprint Labs

Case Study: Boston Public Schools

Jacob Stern
Director of Performance Management, Boston Public Schools

Agenda

- Introduction
- Background
- Our School Quality Measures
- Contact Information

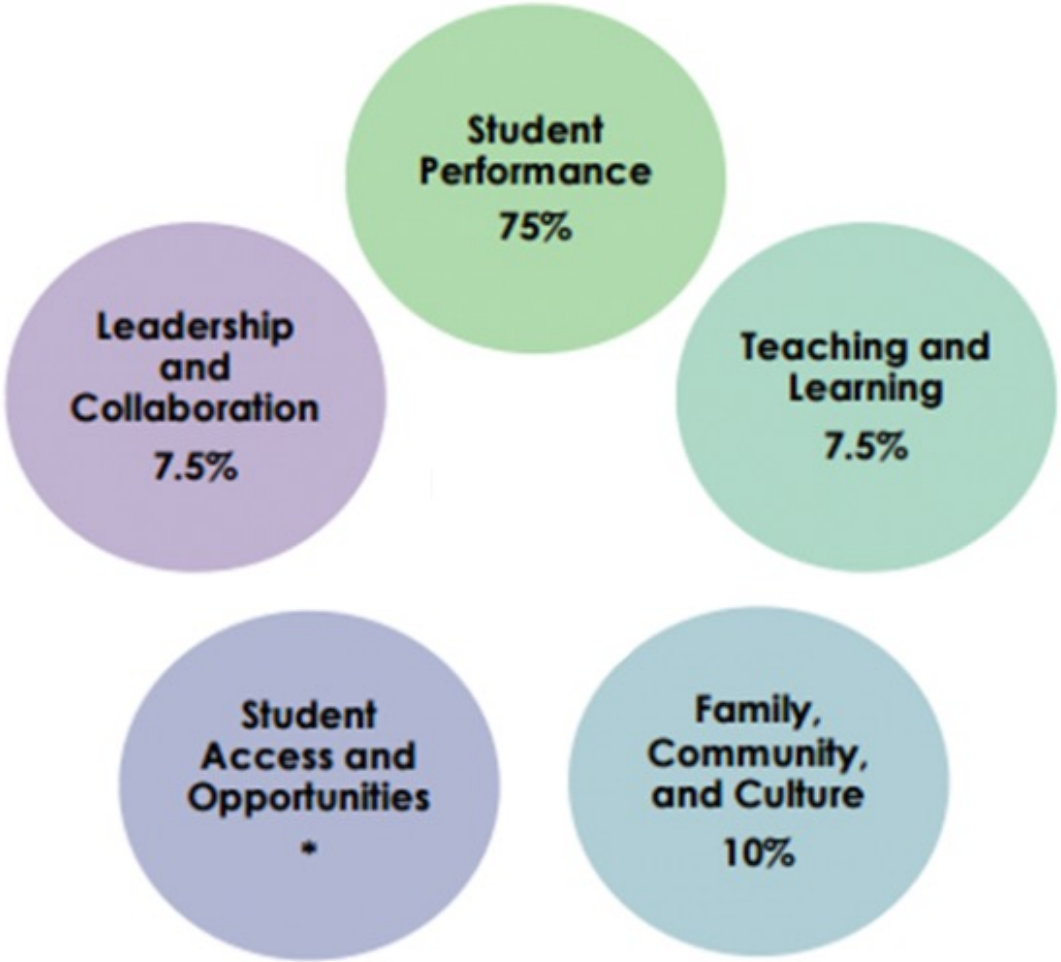
District Background

Race/Ethnicity	Percentage
Asian	9%
Black	29%
Hispanic	43%
White	15%
Other or Multi-Racial	4%

Other Demographics	
Students enrolled in SY21-22	46,169
Schools in SY21-22	113
Low Income	71%
Students with a disability (with an IEP)	22%
Students classified as English Language Learners	30%

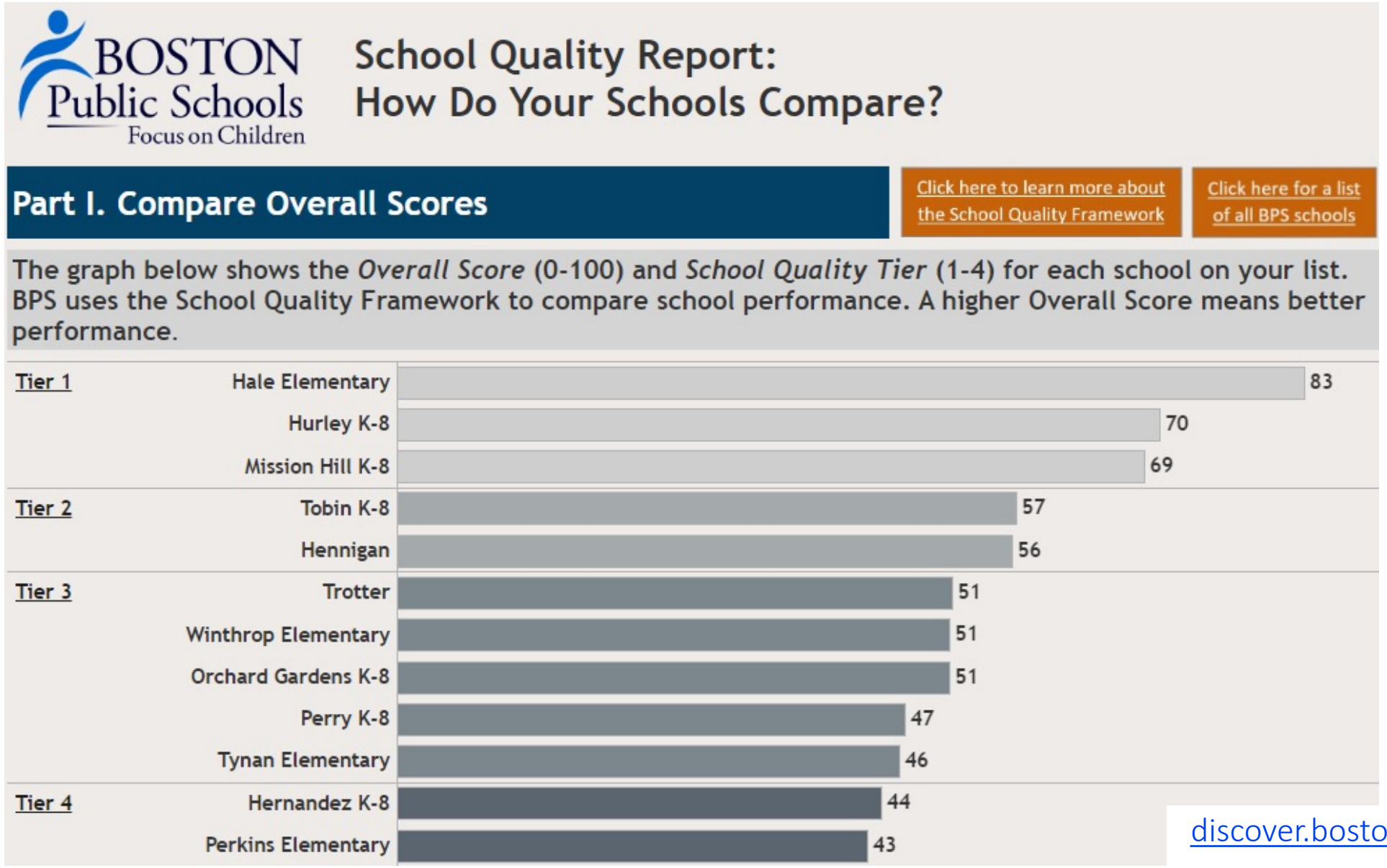
Our School Quality Measures (1/2)

- There are 30 metrics used to **calculate an overall score** in five domains
- Each school is placed in an assignment tier based on its score
- Gives student **growth** more weight
- Rewards schools for making progress w/ low performing students
- Heard (students, teachers, parents)
- Explicitly measures achievement gaps
- **Criterion referenced**, allowing all schools the opportunity to reach Tier 1
- **Overlap with state accountability** framework in data and utilization
- Access and Opportunities presents measurement challenges



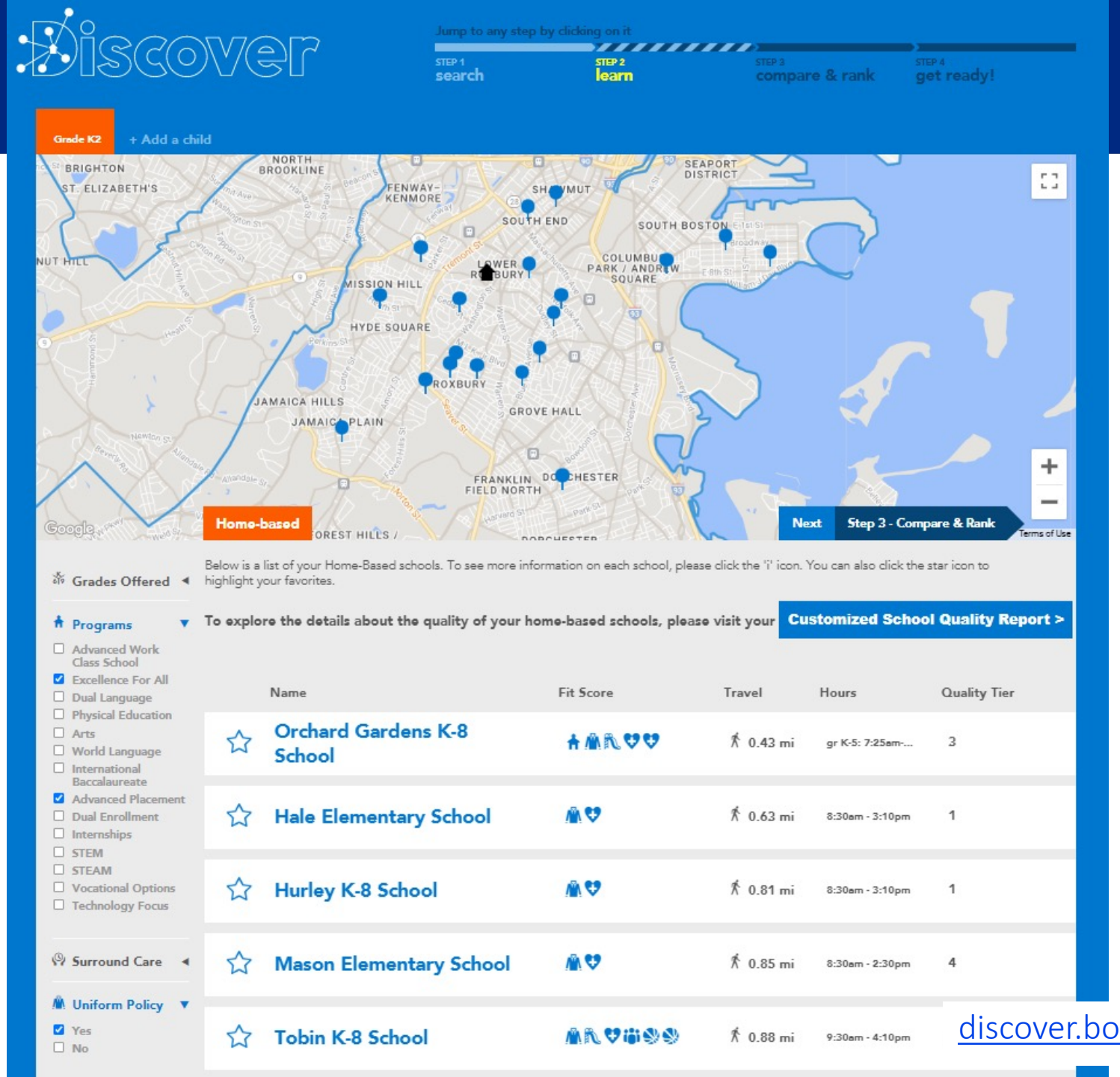
https://www.bostonpublicschools.org/schoolquality

Our School Quality Measures (2/2)



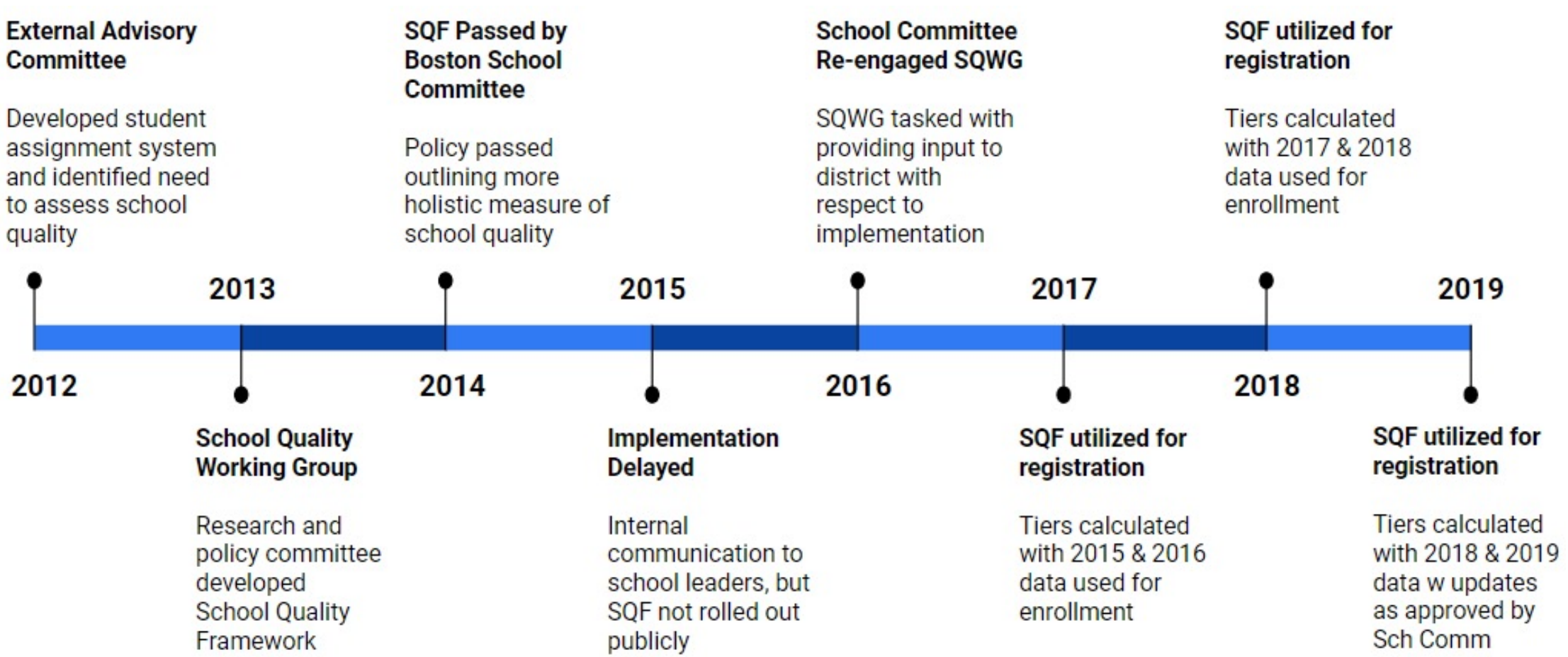
discover.bostonpublicschools.org

Quality Measures & Student Assignment



discover.bostonpublicschools.org

Implementation Context



Contact Information

- **Jake Stern, Director of Performance Management – Boston Public Schools**

  jstern2@bostonpublicschools.org

Blueprint Labs

Case Study: Chicago Public Schools

Jennifer Kish
Senior Data Analyst, Chicago Public Schools

Introduction

Chicago is currently in the process of redesigning our school performance framework, shifting from our former School Quality Rating Policy (SQRP) to a cutting edge approach to accountability that is co-created and responsive to community feedback.

Accountability Redesign
Meaningful Measurement

Chicago Public Schools

District Demographics SY21-22

Race/Ethnicity	Percentage
Hispanic	46.6%
Black	36.0%
White	10.8%
Asian	4.4%
Other or Multi-Racial	2.2%

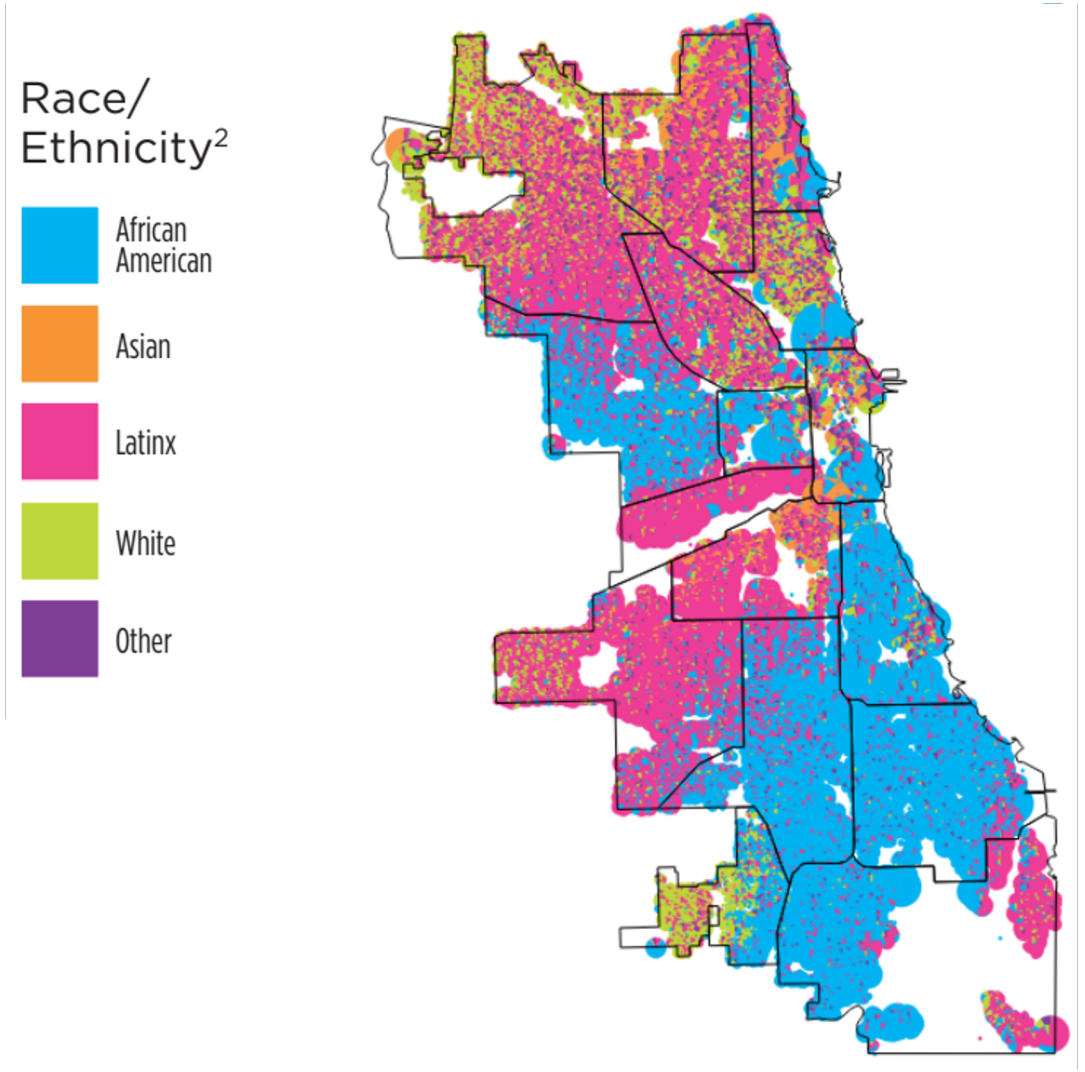
Other Demographics	
Students enrolled in SY21-22	330,411
Schools in SY21-22	636
Economically disadvantaged students	69.8%
Students with a disability (with an IEP)	21.0%
Students classified as English Language Learners	21.0%

Chicago Public Schools

Our School Quality Measures

The greatest strength of the SQRP summative rating system was comparability between schools - any given rating reflects the same expectations across ES, HS, charter vs. non-charter, etc.

The level rating system's greatest strength was also its greatest weakness.



Chicago Public Schools

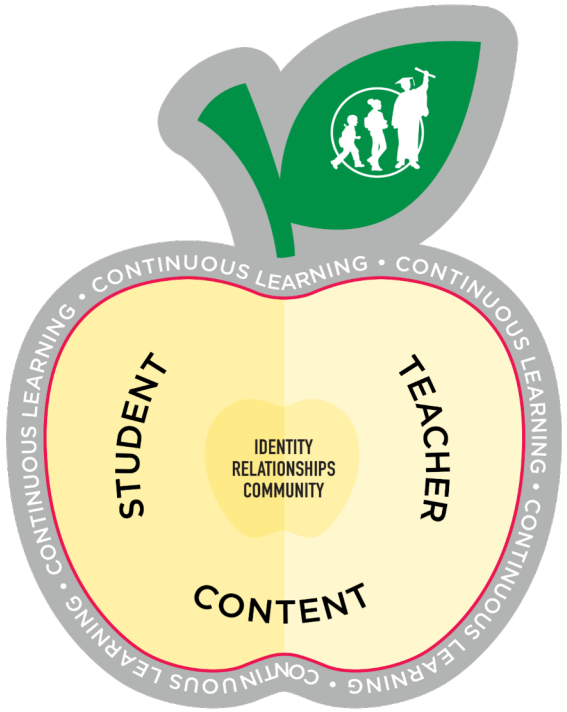
Accountability Redesign

Chicago Board of Education charged us with creating a new system in June 2019. To do that:

- Convened a 31-member **Accountability Redesign Advisory Group (AG)**, which will recommend a new policy to CPS
- Hosted **8 meetings** of the AG, with 4 more scheduled for this summer
- Assembled a **Stakeholder Engagement Design Team**, which advises CPS and our partners on the who, what, and how of our stakeholder engagement efforts
- Conducted 5 town halls, 20 focus groups, and a community-wide survey, resulting in the district hearing and incorporating feedback from **over 10,000 stakeholders to date**
- Developed base **principles** and a **framework** for the system at large
- Started **proposing potential components of the new system** to the Advisory Group

Accountability Redesign

Components of our system that went up for a Board vote in a resolution this week include:

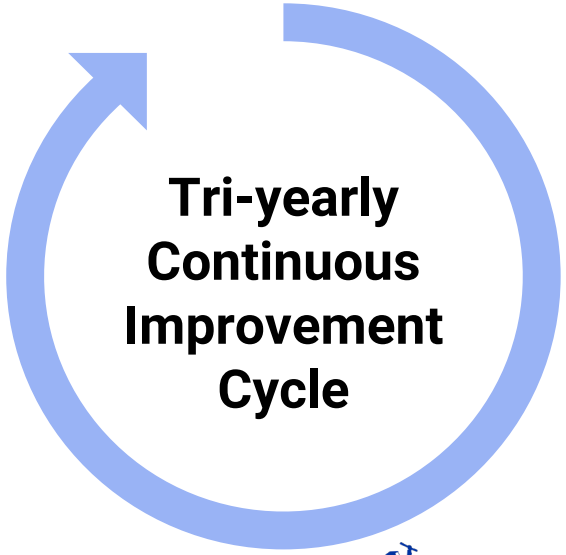


CPS Instructional Core Vision

Eliminate barriers to success and ensure opportunities for all students		
Build safe, supportive environments for students to learn and grow	Promote a wide range of life skills	Prepare students for academic success

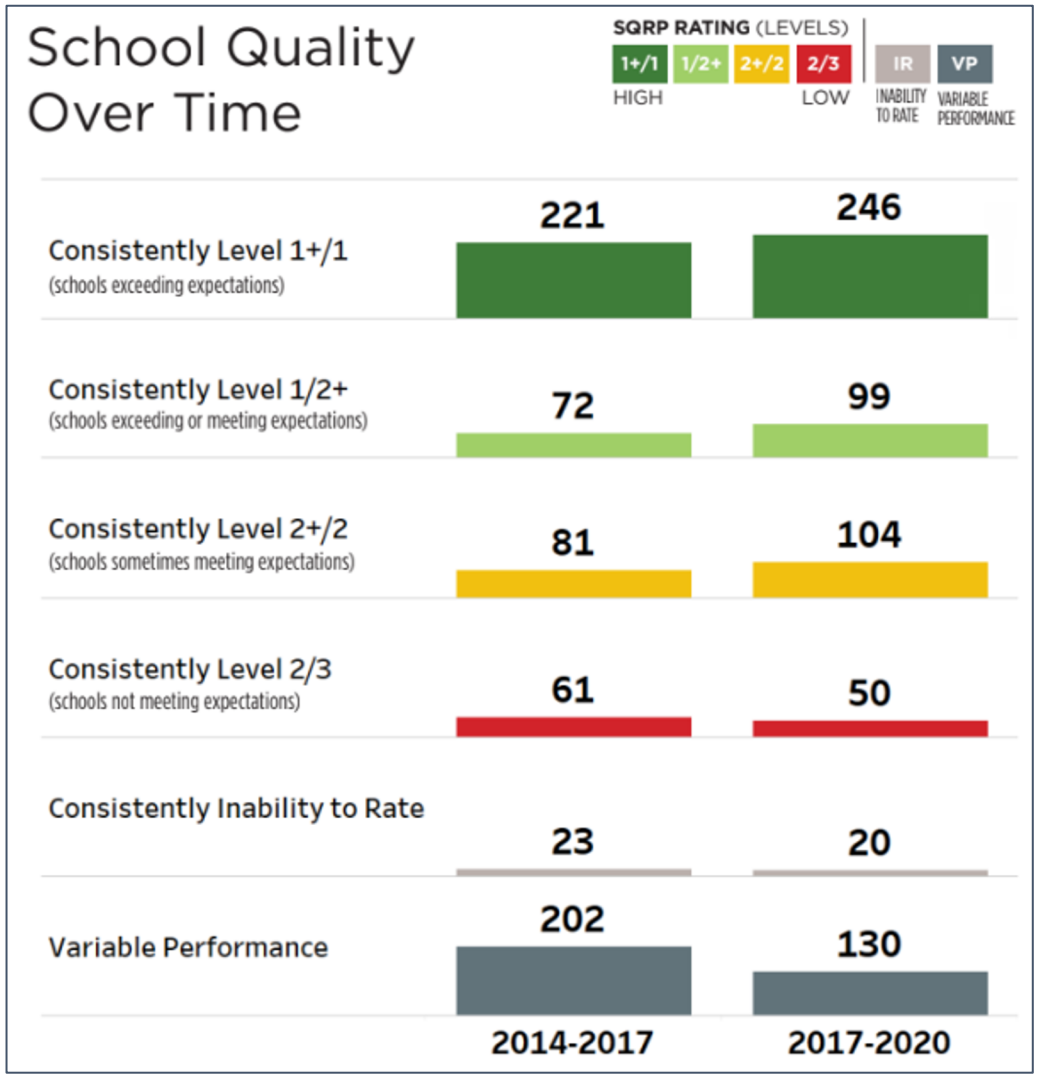
Accountability Redesign Foundational Principles

No single, summative ratings



Chicago Public Schools

Our School Quality Measures



Source: 2020-21 Annual Regional Analysis, CPS

SQRP defines school quality using a single, summative rating for each school.

- 5 levels: **1+**, **1**, **2+**, **2**, and **3**
- 3 Illinois accountability status ratings, **Good Standing**, **Provisional Support**, and **Intensive Support.**

Chicago Public Schools

Our School Quality Measures

Elementary School Metrics

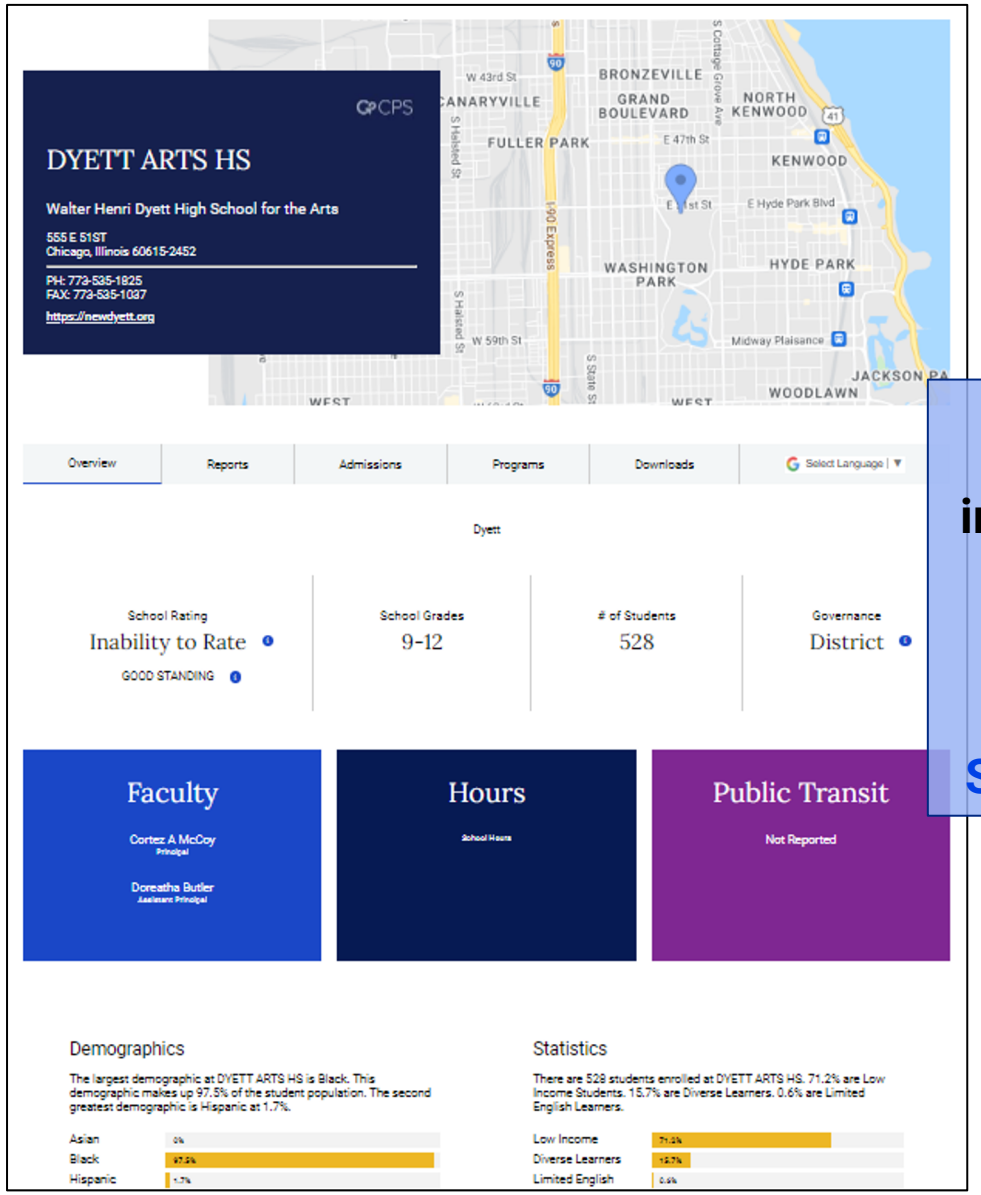
Metric	Weight
Student Growth on NWEA MAP	25%
Student Attendance	20%
Growth of Priority Groups on NWEA MAP	10%
Percentage of Students Making National Average Growth on NWEA	10%
5Essentials Survey	10%
Student Attainment on NWEA MAP (Grades 3-8)	10%
Student Attainment on NWEA MAP (Grade 2)	5%
ELL Language Development Growth on ACCESS	5%
Data Quality	5%

High School Metrics

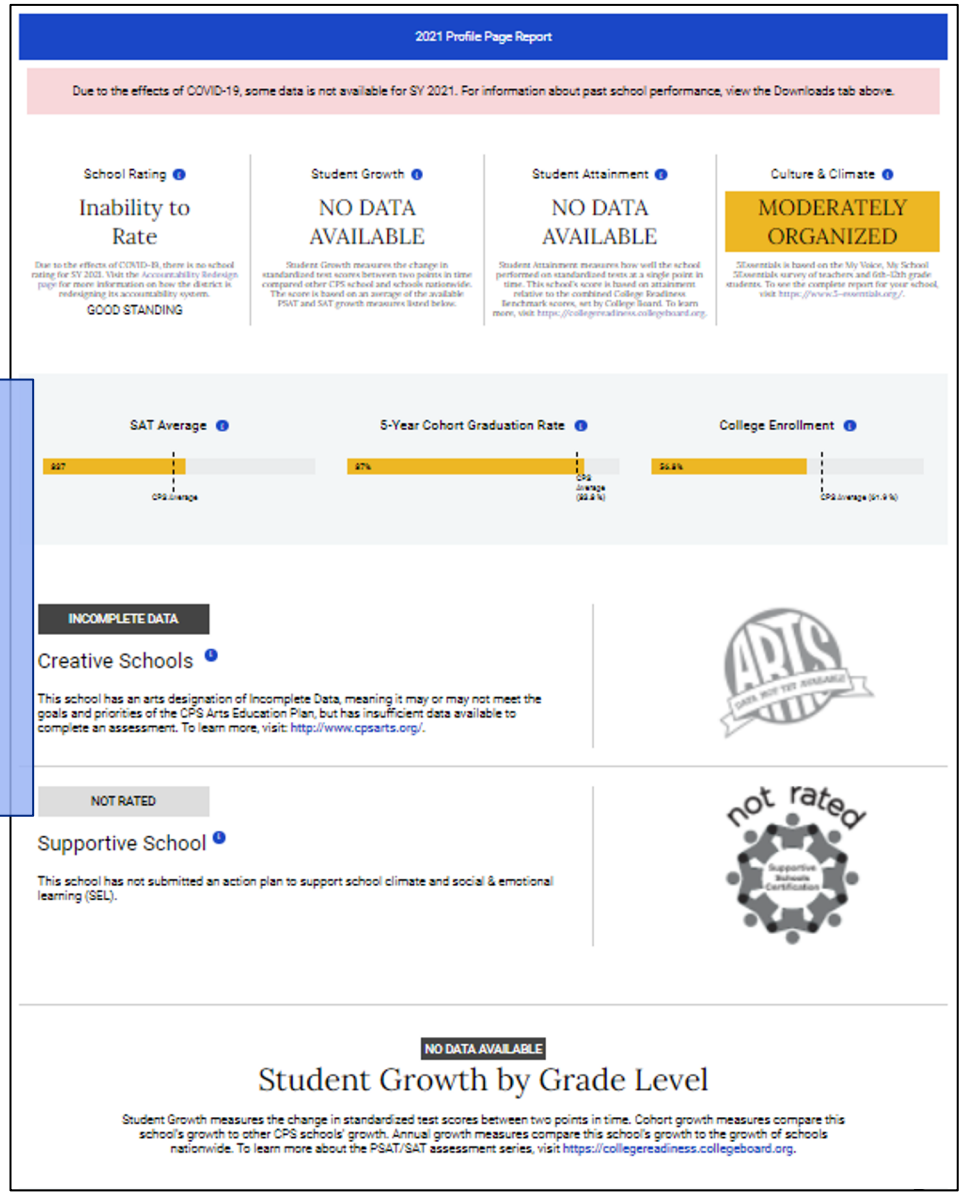
Metric	Weight
Student Growth on PSAT/SAT	10%
Growth of Priority Groups on SAT	Up to 5%
Student Attainment on SAT	10%
Student Attendance	12.5%
Freshman On-Track Rate	12.5%
4-Year Cohort Graduation Rate	12.5%
Early College / Career Credentials	6.25%
1-Year Dropout Rate	6.25%
College Enrollment	6.25%
College Persistence	6.25%
5Essentials Survey	6.25%
Data Quality	6.25%

Chicago Public Schools

Our School Quality Measures



School quality information is shared publicly on the web via CPS' School Profile Pages



Preview of Accountability Redesign

No single, summative ratings

Conditions
instructional, environmental/cultural, leadership and other dynamics that influence student and family experience in a school

Outcomes
student performance on certain indicators

Resources
supports (budgetary and otherwise) available to school communities

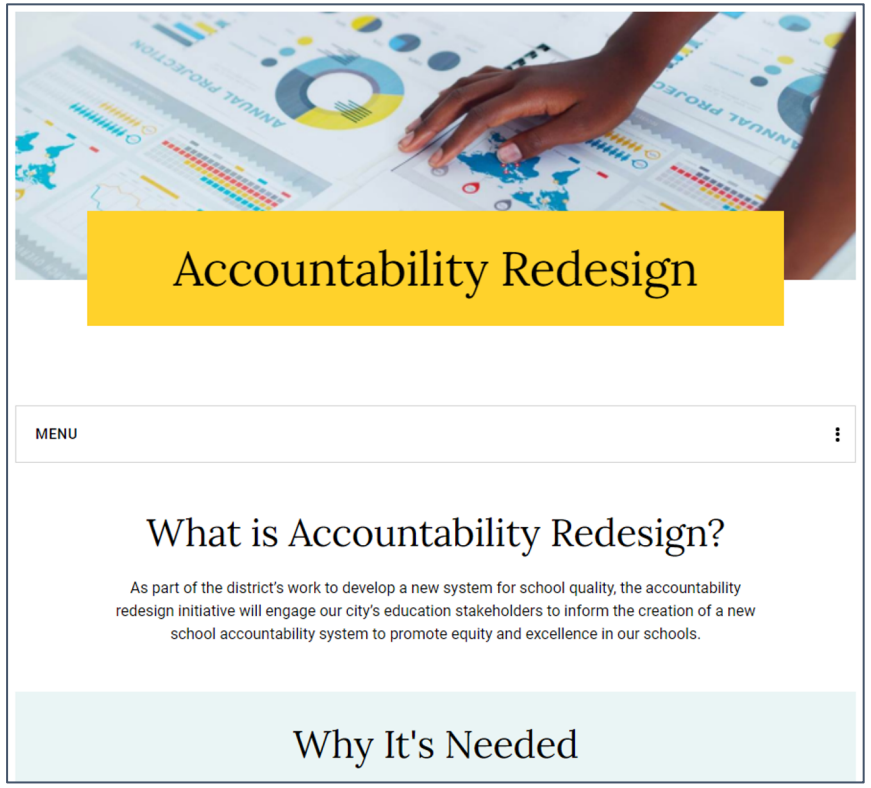
Chicago Public Schools

Contact Information

Jenn Kish, Senior Data Analyst
jlkish@cps.edu

Jeff Broom, Director of Performance Data & Policy

jcbroom@cps.edu



More information can be found on our website:
https://www.cps.edu/strategic-initiatives/accountability-redesign/

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	Welcome & Introductions
9:40 — 10:20 am	Constructing bias-free SQM
10:20 — 10:55 am	Equity and School Quality
10:55 — 11:10 am	Break
11:10 am — 11:30 am	Case Studies: Boston and Chicago
11:30 am — 12:00 pm	**Small group discussions**
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	Developing and communicating multiple SQM
1:15 — 2:10 pm	Measuring school quality beyond test scores
2:15 — 2:45 pm	Involving the community in developing SQM
2:45 – 3:00 pm	Small group discussions

Discussion Questions

1. **There is sometimes a trade-off between the rigor and communicability of a measure. How do you navigate this trade-off?**
2. **If two measures of school quality seem contradictory, how might we guide families to understand them?**
3. **How do we consider equity in terms of communicating measures to different audiences and populations?**

Small group discussions will **end at 11:50**.

Blueprint Labs

Lunch

Room 324, 314, or outside.

We'll reconvene at 12:45pm.

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	Welcome & Introductions
9:40 — 10:20 am	Constructing bias-free SQM
10:20 — 10:55 am	Equity and School Quality
10:55 — 11:10 am	Break
11:10 am — 11:30 am	Case Studies: Boston and Chicago
11:30 am — 12:00 pm	Small group discussions
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	**Developing and communicating multiple SQM**
1:15 — 2:10 pm	Measuring school quality beyond test scores
2:15 — 2:45 pm	Involving the community in developing SQM
2:45 – 3:00 pm	Small group discussions

Blueprint Labs

Developing and communicating multiple measures of school quality

Eric Ashton
Executive Director for School Performance, New York City Department of Education

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	Welcome & Introductions
9:40 — 10:20 am	Constructing bias-free SQM
10:20 — 10:55 am	Equity and School Quality
10:55 — 11:10 am	Break
11:10 am — 11:30 am	Case Studies: Boston and Chicago
11:30 am — 12:00 pm	Small group discussions
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	Developing and communicating multiple SQM
1:15 — 2:10 pm	**Measuring school quality beyond test scores**
2:15 — 2:45 pm	Involving the community in developing SQM
2:45 – 3:00 pm	Small group discussions

Blueprint Labs

Measuring school quality beyond test scores

Brian Gill
Senior Fellow, Mathematica

Marty West
Professor and Academic Dean, Harvard Graduate School of Education

Using Student Surveys to Track Social-Emotional Learning: Evidence from the CORE Districts

Martin R. West, Harvard University

MIT School Quality Measures Conference
April 29, 2022

Mathematica

Beyond test scores: A framework for measuring school performance

Brian Gill, Ph.D., J.D.
Senior Fellow, Mathematica
Director, Mid-Atlantic Regional Educational Laboratory

April 2022

How can we make data *useful* to educators and policymakers for improvement?

Policymakers and educators are implored to be "data-driven" and "evidence-based"

But data often aren't actionable

- **In the wake of pandemic-related disruptions, measures that are rich, broad, and actionable may be even more important**

Beware the descent into Taylorism: Using measures for performance management can go wrong in various ways

- **Incompleteness of measures**
- **Distorting effects of high stakes (Campbell's Law)**
- **Overinterpretation of measures that aren't reliable (fooled by randomness)**
- **Misattribution (misinterpreting cause of an outcome)**
- **Disconnect between outcome measures and processes producing them ("black box" problem)**

How do we avoid these pitfalls and ensure that data are diagnostic and actionable?

The framework: Measurement to manage performance requires distinct kinds of data for distinct diagnostic purposes

1. **Student outcomes: How are the kids doing (in the short and long term)?**

   Diagnostic for identifying needs of students (and collectively, needs of schools)

2. **Impact on outcomes: How much does the school contribute to how kids are doing?**

   Diagnostic for identifying underperforming (and overperforming) schools

3. **Processes related to outcomes: What is happening in the school (or classroom)?**

   Diagnostic for identifying possible processes for intervention

   Potentially diagnostic for identifying underperforming (and overperforming) schools

1: Student outcomes

How are the kids doing?

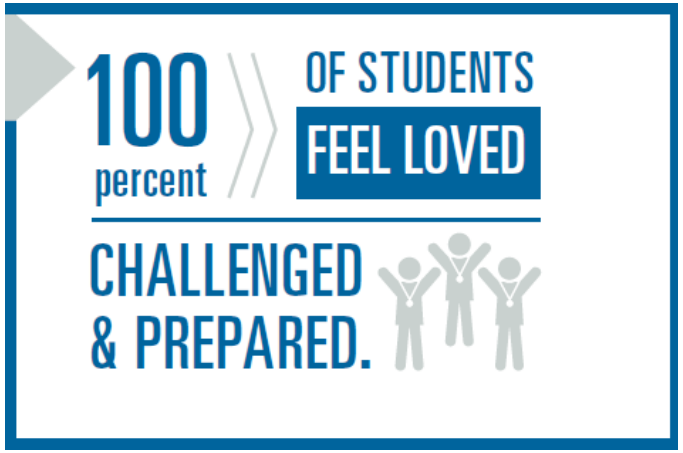
We need broader and richer measures of student outcomes

- **Strongest critique of high-stakes testing is that tests don't measure everything we want kids to learn**
- **ESSA has expanded measures beyond test scores to include graduation, chronic absenteeism, college readiness, and other measures of success**
- **Other possibilities include:**
  - Postsecondary enrollment, persistence, degree completion
  - Workforce participation in adulthood
  - Social-emotional learning
  - Citizenship, including registration and voting

REL Mid-Atlantic assessed social-emotional learning in DC

DC Public Schools: Analyzed student survey data to create index of whether students are "loved, challenged, and prepared" (Kautz et al. 2021)

Not used for formal accountability, but districtwide measure is publicly reported



Surveys can produce useful SEL information at the school level

- **It is possible to get good response rates from students (and staff)**
- **Schoolwide SEL measures show good psychometric characteristics**
- **SEL varies systematically at the school level**
  - There is more variation within schools than between schools, but between-school variation is meaningful
- **Some caution in interpretation may be needed, e.g. for grade-level differences**

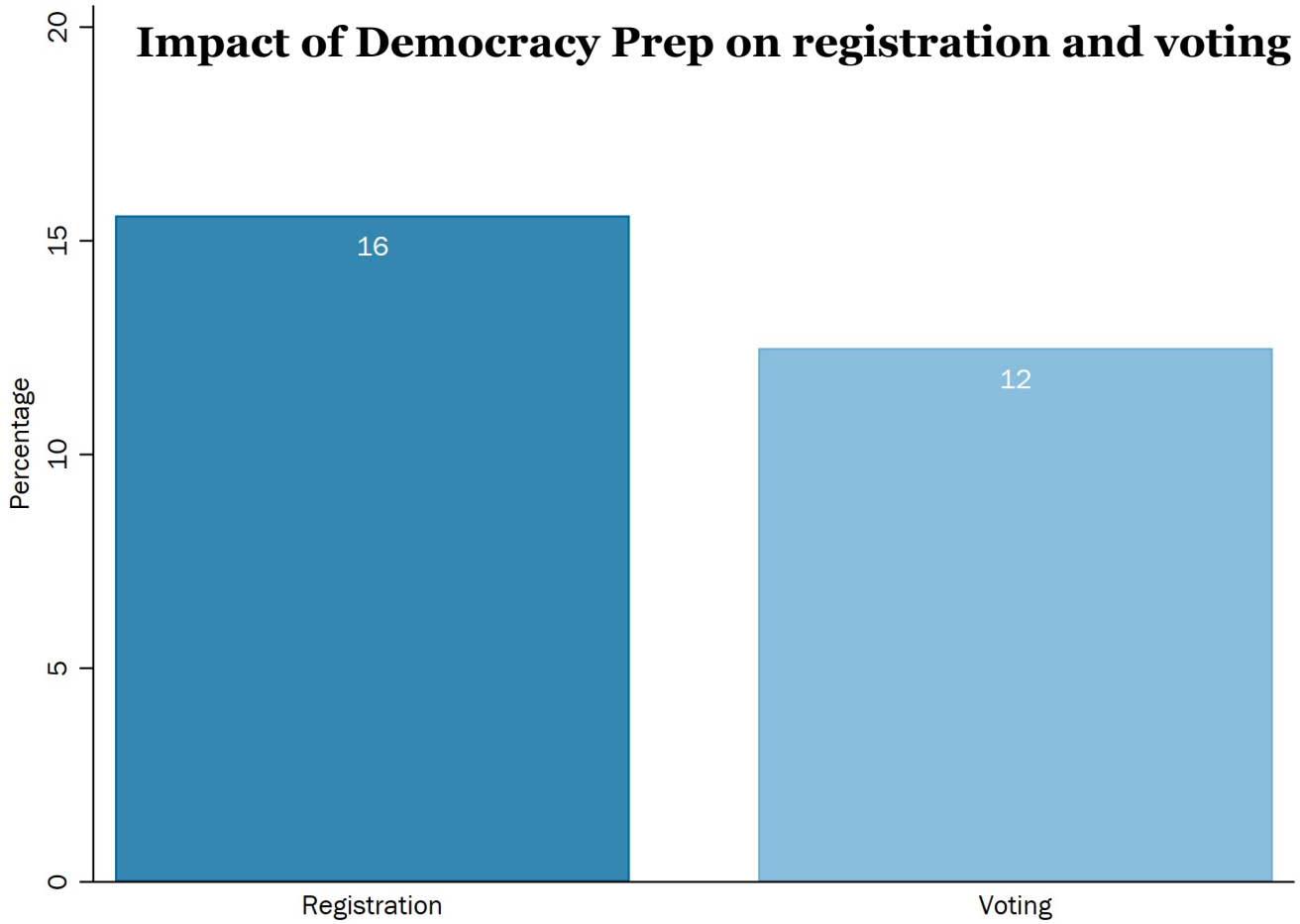
Civics: Preparation for citizenship is foundation of public education

Horace Mann, 1855: A well-functioning democracy requires an educated citizenry

> "Education must be universal . . . The qualification of voters is as important as the qualification of governors, and even comes first, in the natural order . . . The theory of our government is … that every person… shall become fit to be a voter. Education must bring the practice as nearly as possible to the theory."

Civic purpose of education remains relevant today

Schools can have large impacts on registration and voting



Gill et al 2020

Mathematica

2: Impacts on student outcomes

How much does the school contribute to how kids are doing?

Impacts must be distinguished from raw outcomes

- **NCLB critics recognized raw proficiency results conflated school performance/student characteristics**
  - Angrist et al (2022) show how raw outcome measures unfairly penalize schools serving black and brown students
  - NCLB critiques prompted development of value-added and student growth measures
- **Don't recreate NCLB's flaw in failing to distinguish school's contribution**
  - Why not apply statistical techniques to identify school impacts on SEL, graduation, college enrollment?
- **Accountability arguments about relative weight of outcomes vs impacts (status vs growth) miss the point: Outcomes and impacts are diagnostic for *different purposes***
  - Low-status/high-growth school could end up with same overall rating as high-status/low-growth school, but they need very different interventions

Promotion power measures separate schools' contributions from other factors for non-test outcomes

Use statistical methods similar to those in value-added/growth models (Deutsch et al. 2020; Gross et al 2021)

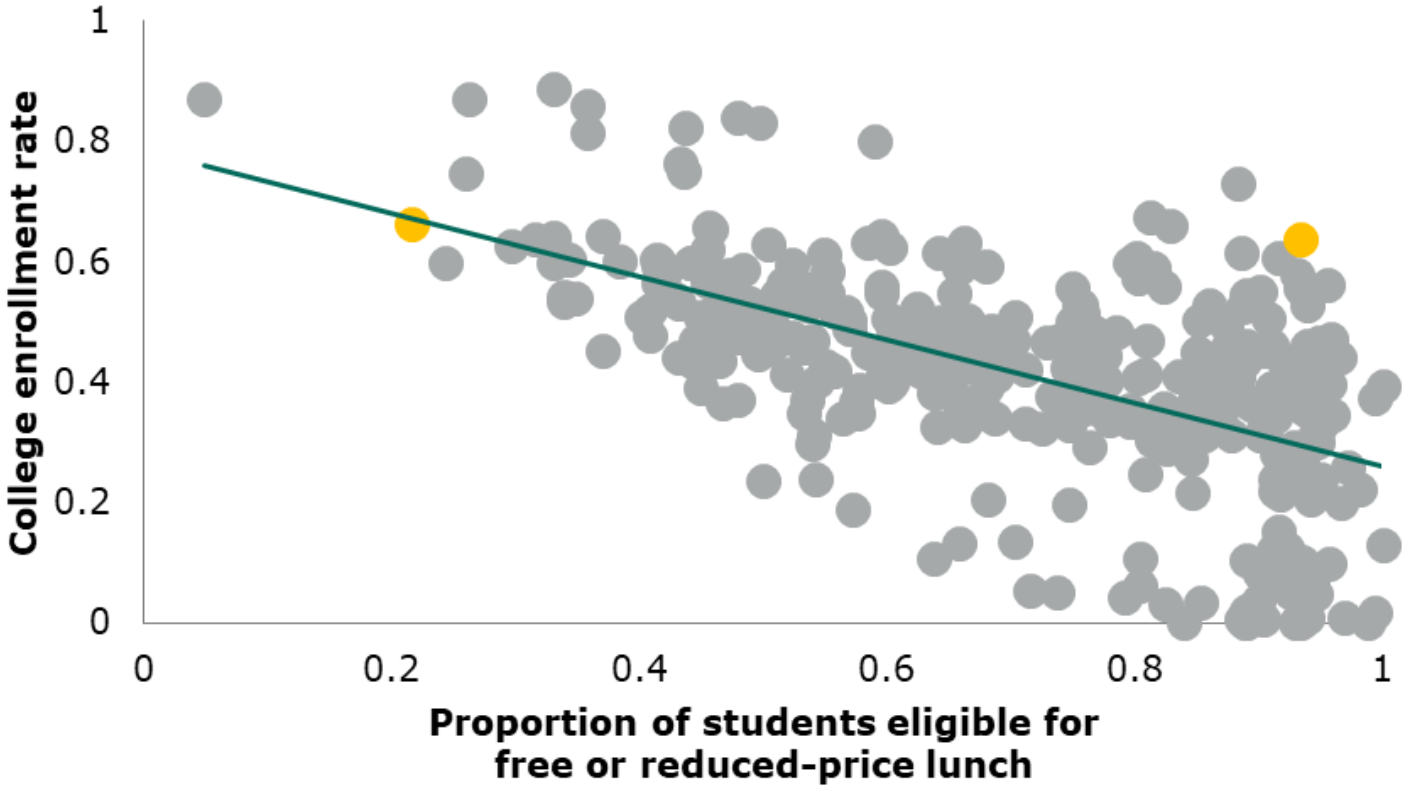
Account for differences in advantages and disadvantages of students served

- Poverty, prior achievement, IEP and ELL status, anything relevant and measured prior to high-school entry

Aims: Be fair to schools and provide better diagnostic info to districts and states

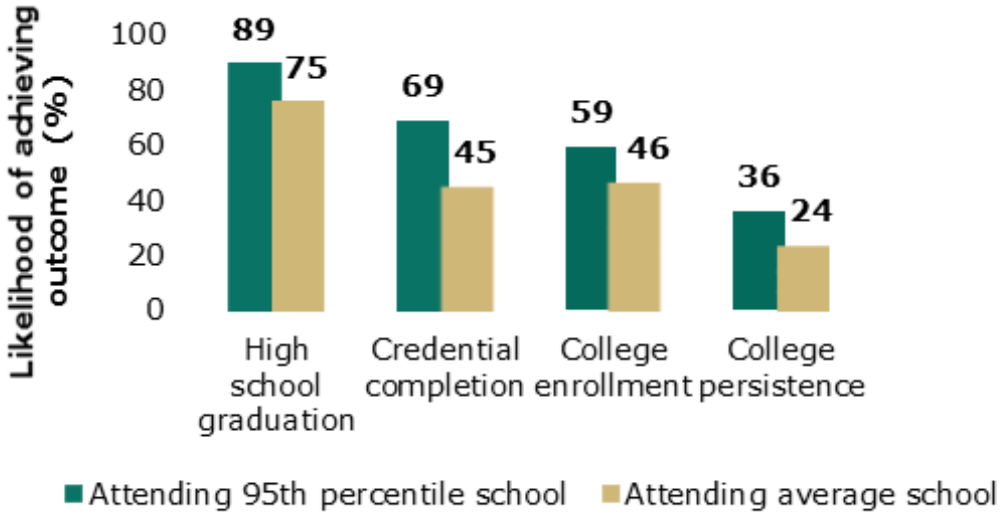
Under development in partnerships with state education agencies in Louisiana and DC

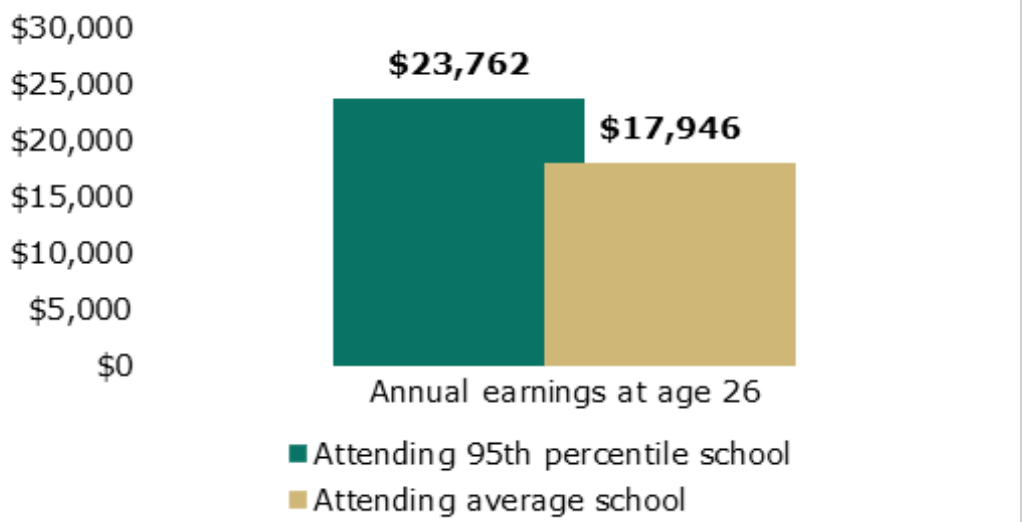
High schools can have the same student outcomes with very different promotion power



Deutsch et al 2020

Attending a high school with high promotion power can substantially improve a student's long-term success





Deutsch et al 2020

Mathematica

3: Educational processes

What is happening in the school?

Process measures can help identify areas for possible intervention

Even if an impact measure provides a valid and reliable measure of performance, it is a black box: doesn't tell us how or why

Process measures and mechanisms might include;

- Observations of instructional practice
- Climate surveys
- Student participation indicators from learning management system
- School inspections
- Exclusionary discipline
- Class size, teacher qualifications & experience

Process measures might compensate for weaknesses of outcome and impact measures

- **Outcome/impact measures, even if valid & reliable, are thin/incomplete**
  - But we can recognize good schools and good teaching from observation
  - Evidence indicates that even students can recognize good teaching (Raudenbush & Jean 2015; Chaplin et al 2014)

- **Some impacts may be impossible to measure**
  - We can probably produce acceptably valid and reliable estimates of the value-added of teachers and schools, but…
  - Nobody knows how to measure the impact of an individual principal on student outcomes (Chiang, McCullough, Lipscomb, Gill 2016)

School climate changes substantially with a new principal

Kozakowski, Gill, Shiferaw 2021

Mathematica

Conclusion:
Use the right measures for the right purposes!

For additional information, please contact:

Brian Gill
BGill@mathematica.org

Motivation

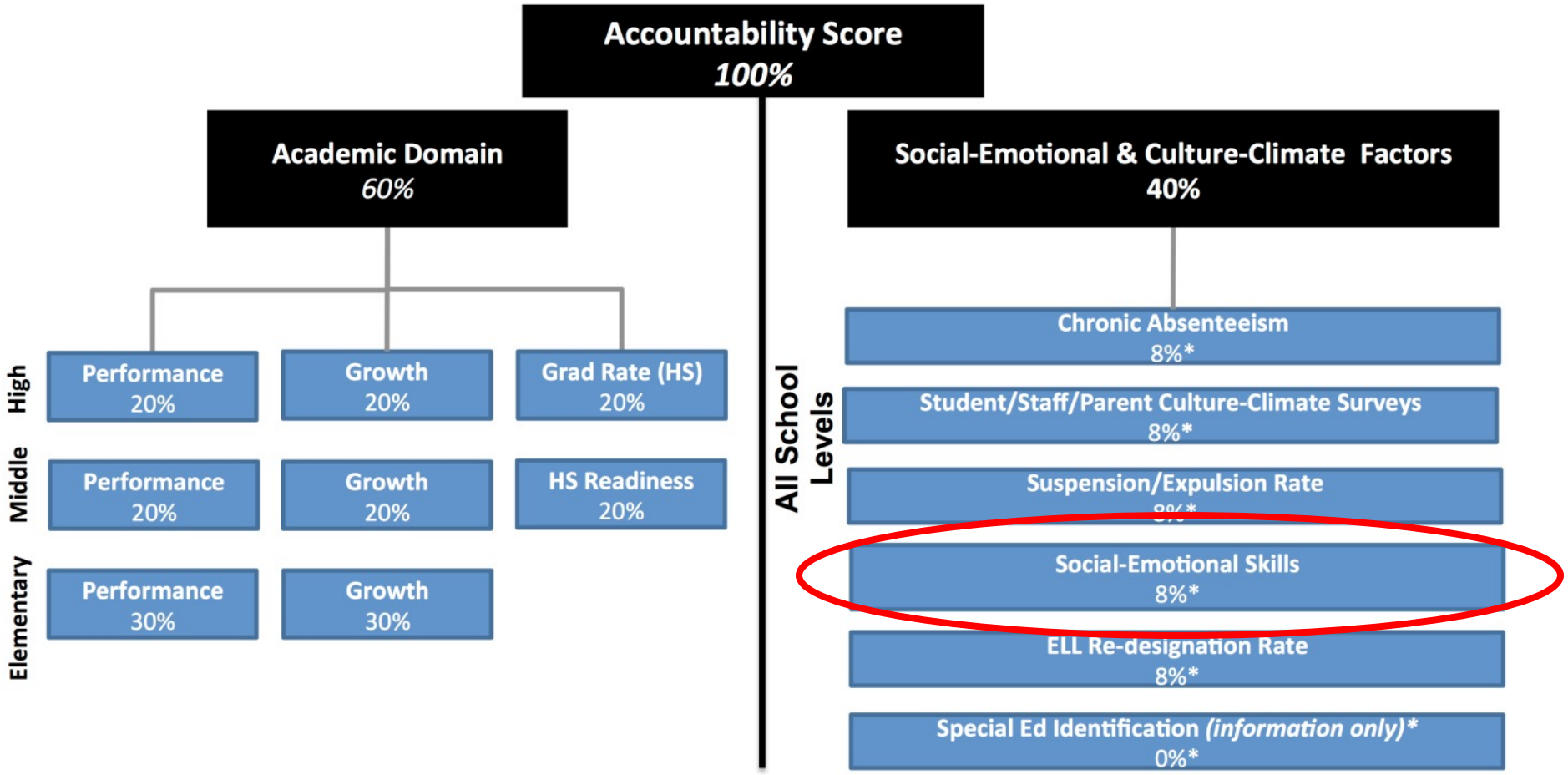
- Growing evidence of the importance of students' non-cognitive or social-emotional skills for success in school and life
- Growing interest among educators and policymakers in measuring and supporting students' social-emotional development
- Growing use of self-report surveys to track social-emotional skills
- Little evidence on the properties of these surveys when administered at scale or about typical patterns in social-emotional development
- CORE Districts provide an opportunity to learn about these issues

CORE Districts

- A consortium of nine California school districts serving over one million students

- Six of these districts used a federal waiver to develop a new accountability system from 2013-2015: Fresno, Los Angeles, Long Beach, Oakland, Santa Ana, and San Francisco

School Quality Improvement Index (SQII)



Development and Validation of SEL Measures

- Constructs selected based on 3 criteria:
  - **Measurable** (in ~10 minutes using free instruments)
  - **Meaningful**ly predictive of academic and life outcomes
  - **Malleable** through school-based interventions

- 2013-14: pilot test of SEL measures on ~32,000 students
  - Focus on comparing alternative versions of survey scales and the use of anchoring vignettes to adjust students' self-reports
- 2014-15: field test of final SEL measures on ~390,000 students
  - Focus on reliability and validity when administered at scale
- 2015-16: SEL data reported publicly and count as 8% of SQII ratings

SEL Constructs

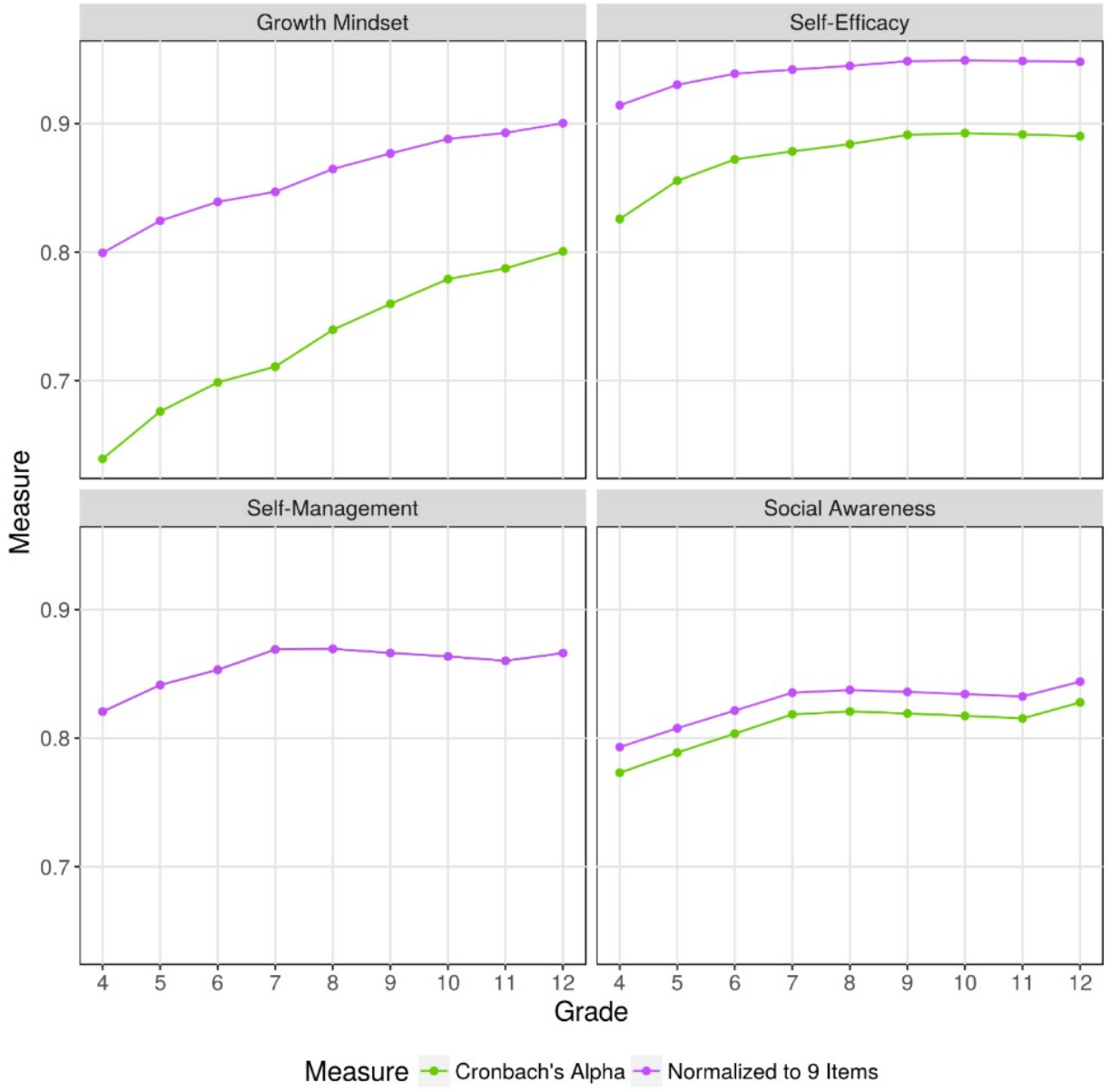
Skill	Definition	Sample item
Self-management	The ability to regulate one's emotions, thoughts, and behaviors effectively in different situations. This includes managing stress, delaying gratification, motivating oneself, and selecting and working toward personal and academic goals.	Please answer how often you did the following during the last 30 days. During the past 30 days… I remained calm even when criticized or otherwise provoked.
Growth mindset	The belief that one's abilities grow with effort. Students with a growth mindset see effort as necessary for success, embrace challenges, learn from criticism, and persist in the face of setbacks.	Please indicate how true each of the following statements is for you: My intelligence is something that I can't change very much.
Self-efficacy	The belief in one's ability to succeed in achieving an outcome or reaching a goal. Self-efficacy reflects confidence in the ability to exert control over one's motivation, behavior, and environment.	How confident are you about the following at school? I can do well on my tests even when they are difficult.
Social awareness	The ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources and supports.	During the past 30 days… How carefully did you listen to other people's points of view?

Insights from Prior Research

- Potential concerns about SEL surveys
  - Social-desirability bias
  - Reference bias
  - Stereotype threat
  - Differences in item interpretation across subgroups
  - Unmotivated responding

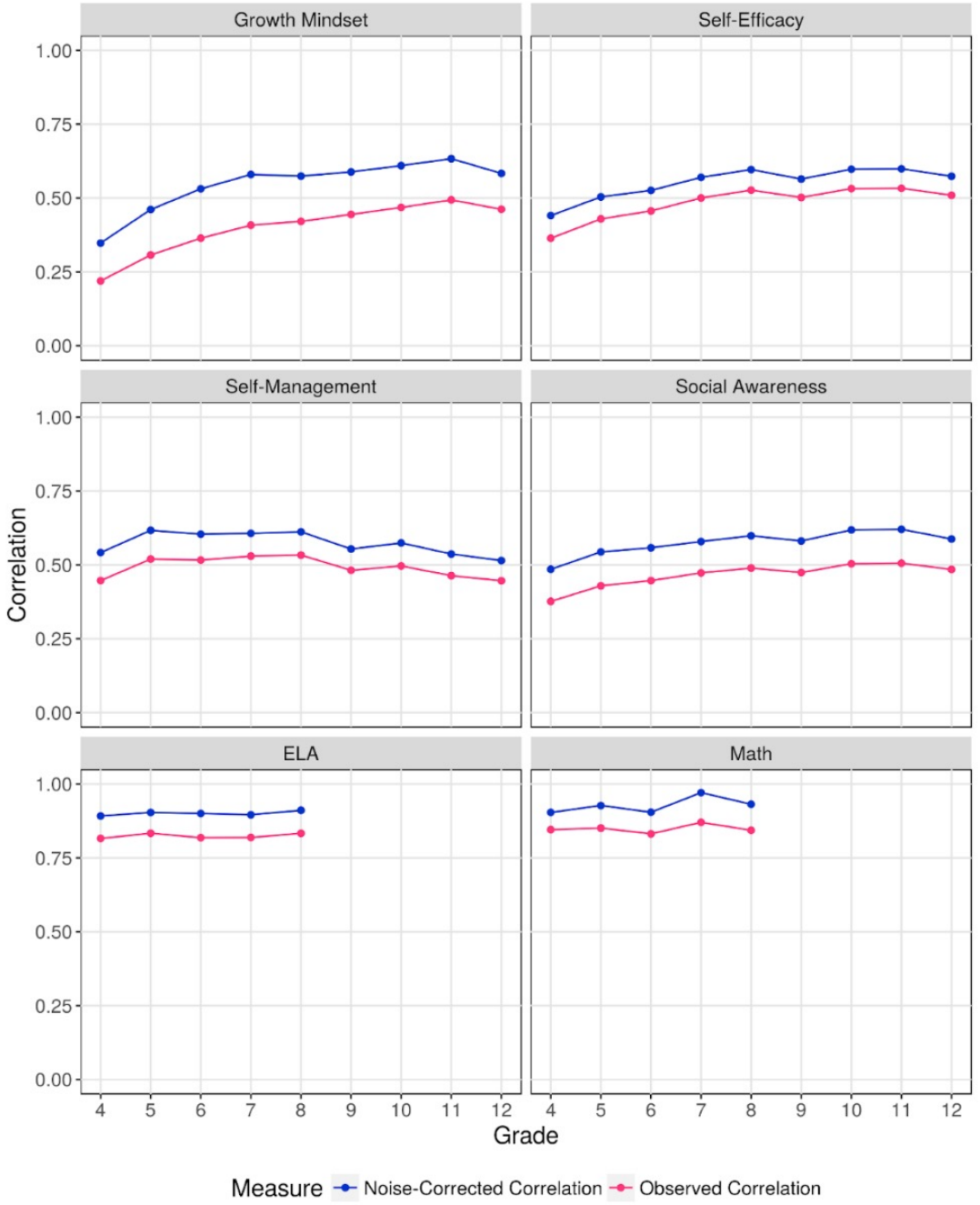
- One takeaway from developmental psychology: no expectation of linear or consistent "growth" in social-emotional constructs
  - Metacognitive ability improves with age, but social-emotional arousal peaks during adolescence → specific social-emotional skills may develop differently depending on the degree to which they require cognitive control relative to emotional regulation

Internal consistency of survey scales is moderate-to-high (except Growth Mindset in early grades) and increases with grade



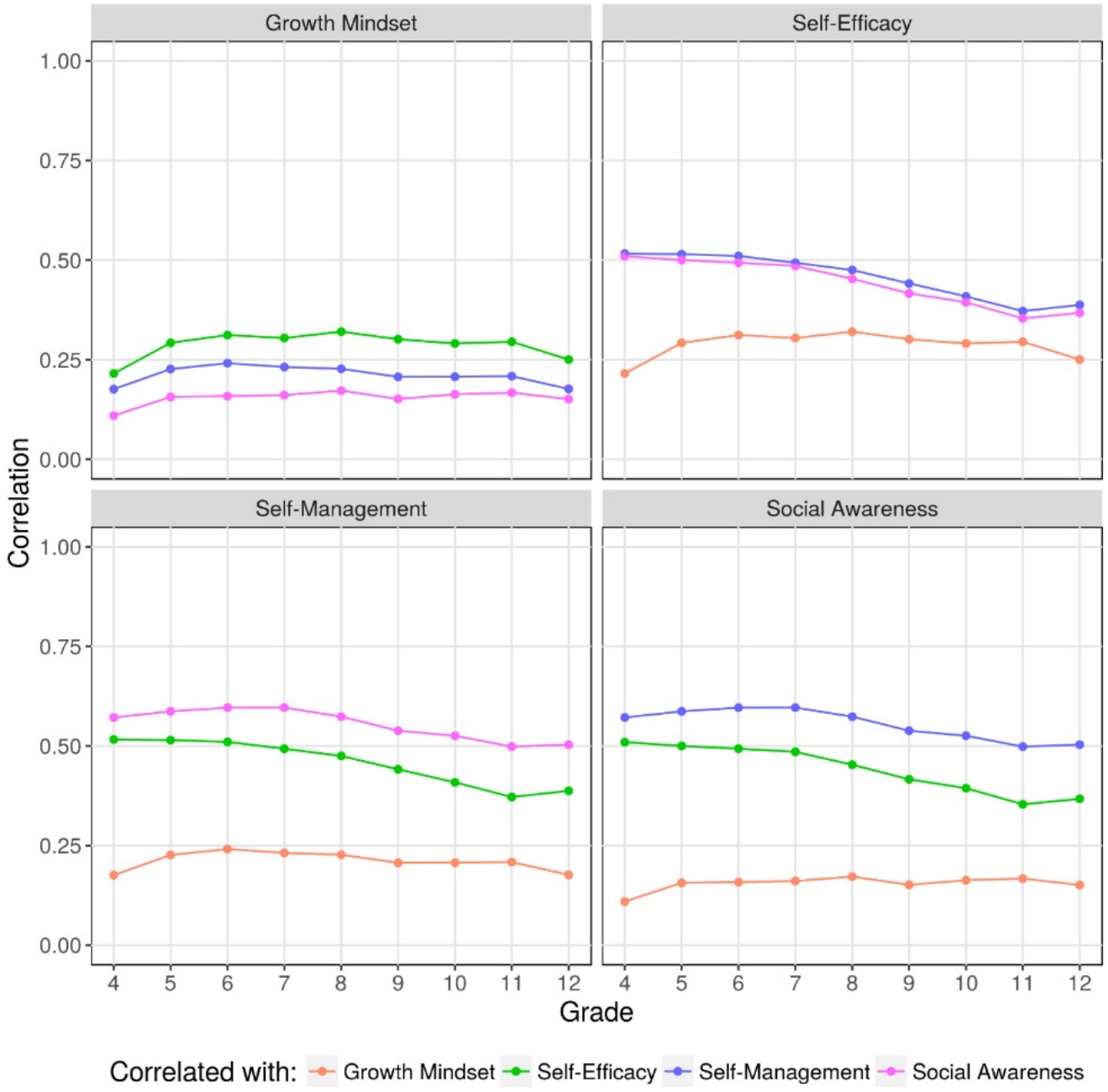
SEL measures for individual students vary more from year to year than test scores, even when corrected for measurement error

- Noise-corrected correlations range from ~0.5-0.6
- Pattern is consistent with SEL constructs being more malleable (or with impacts on SEL being transitory)

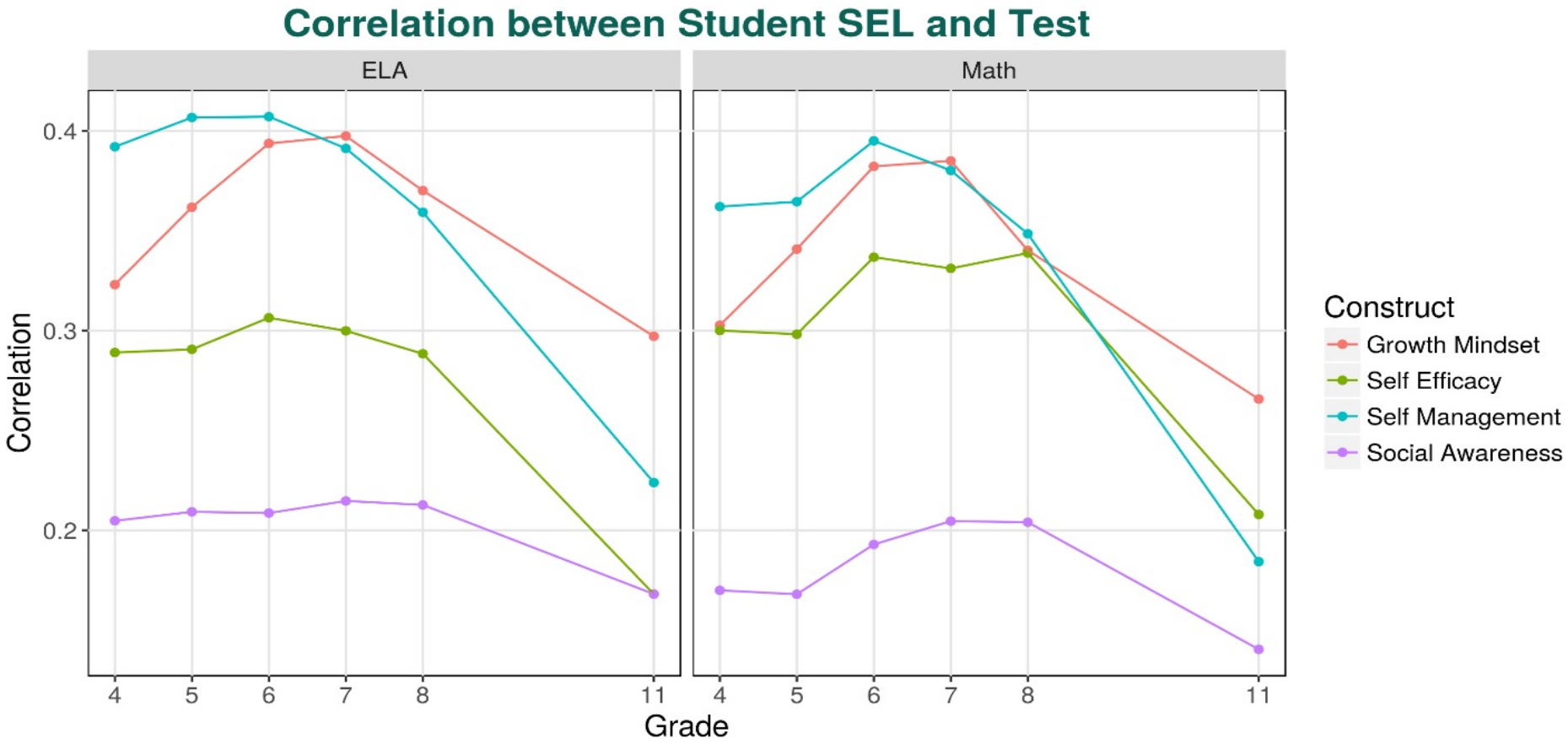


Within students, SEL constructs are positively correlated (to varying degrees)

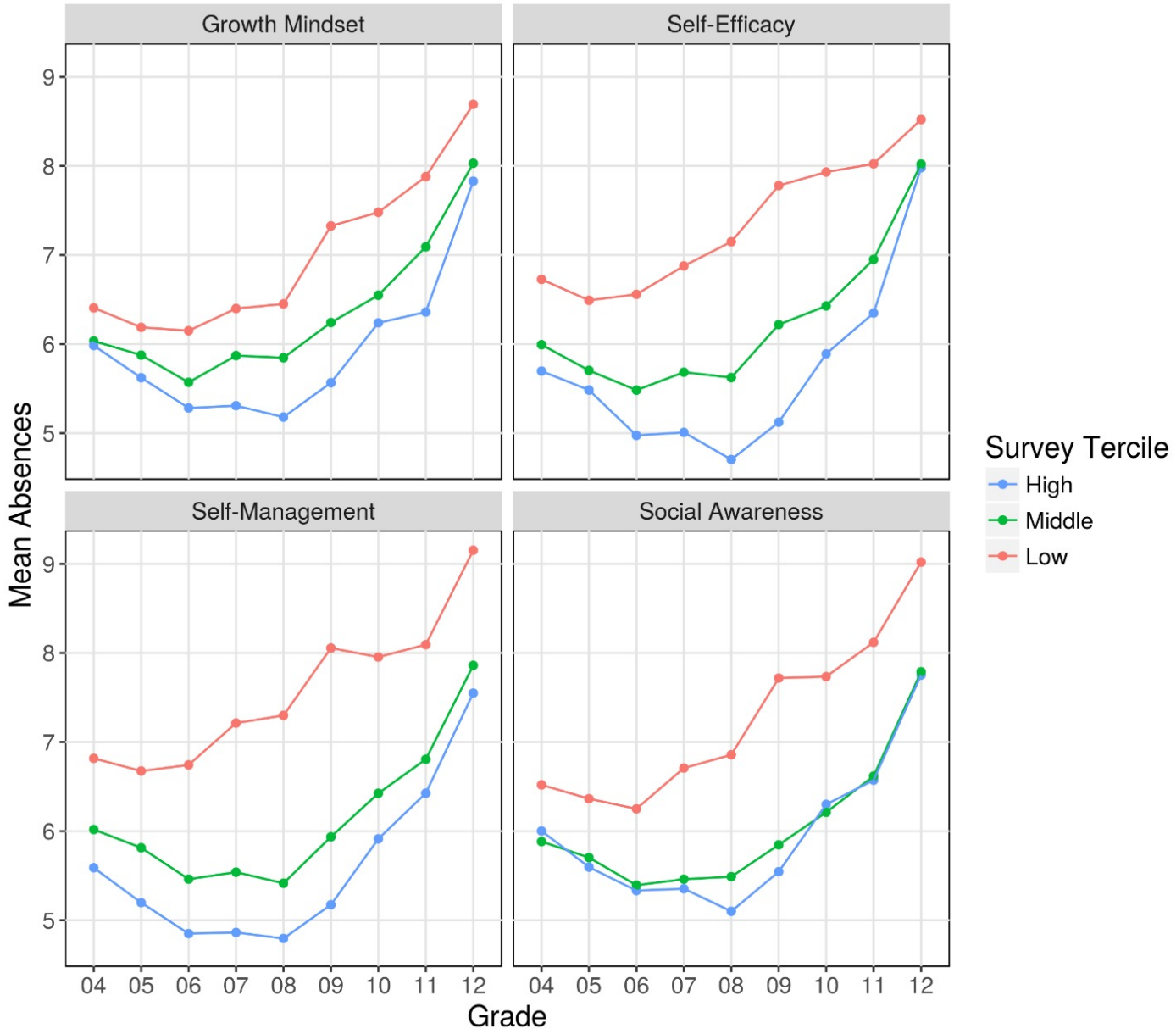
- Social Awareness and Self-Management are correlated at ~0.5-0.6
- Correlations with other constructs are lowest for Growth Mindset



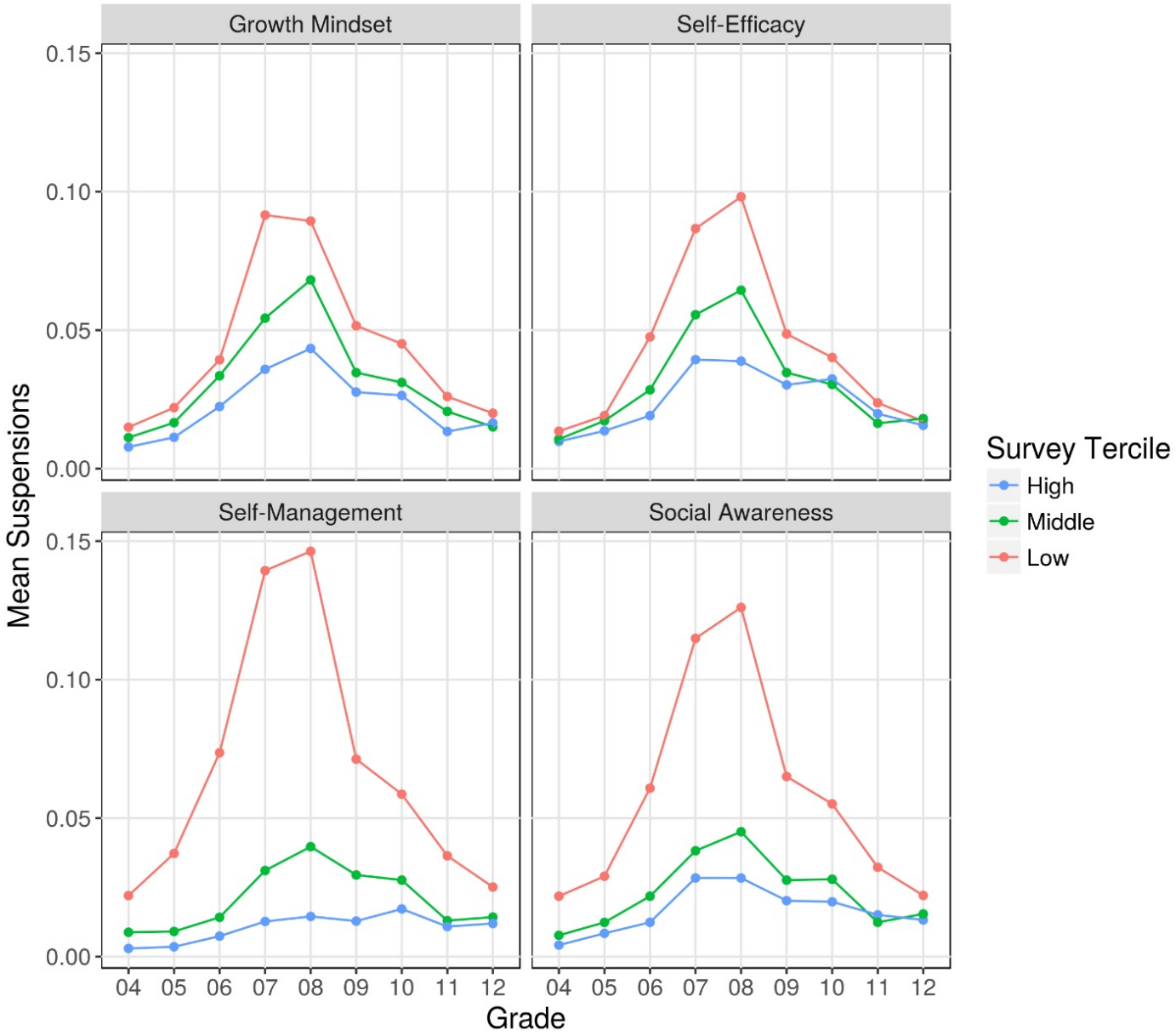
Self Management and Growth Mindset are most predictive of ELA and math test performance; all correlations drop markedly in high school



Students with lower self-reports have more absences



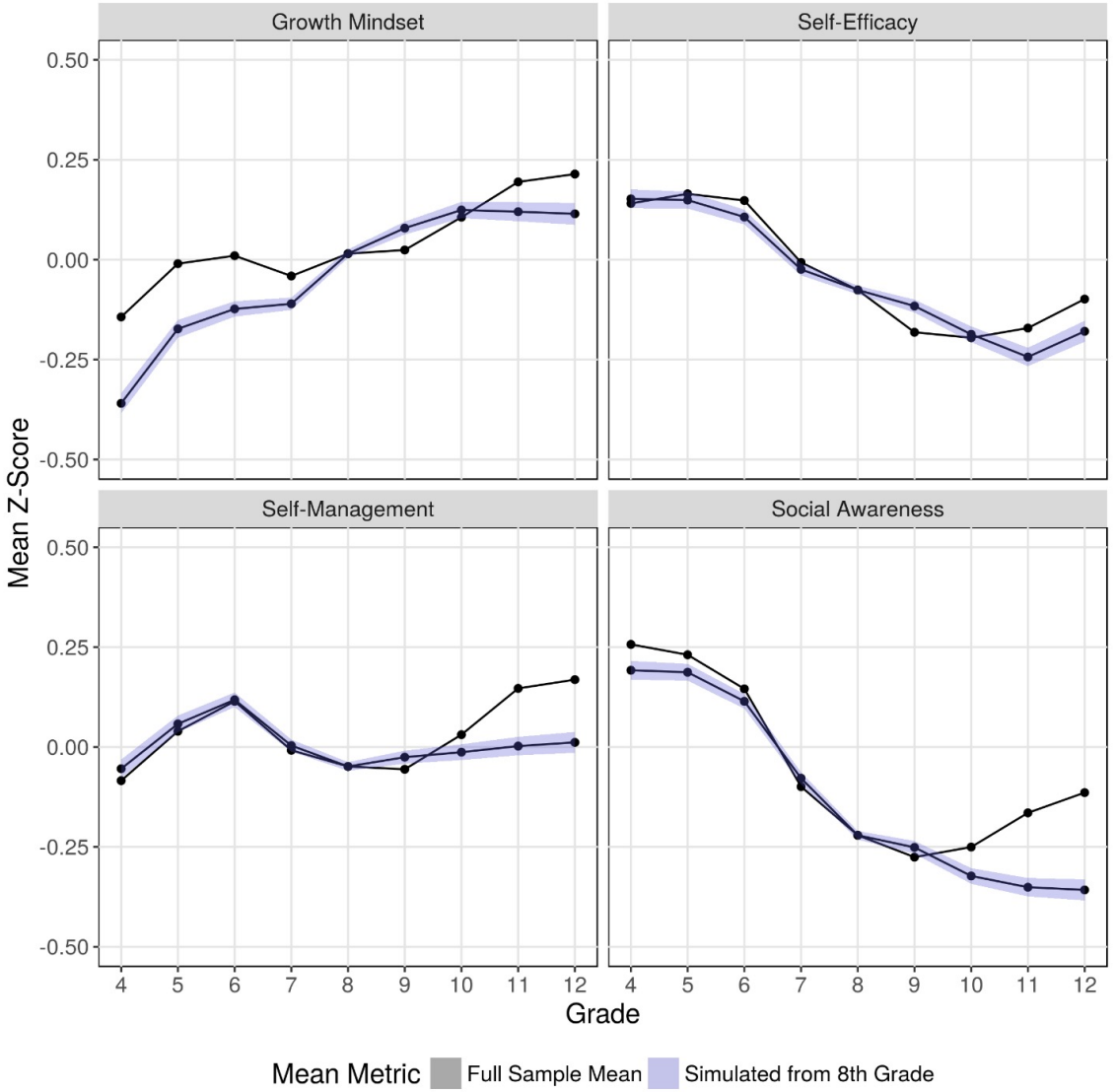
Students with lower self-reports are suspended more frequently



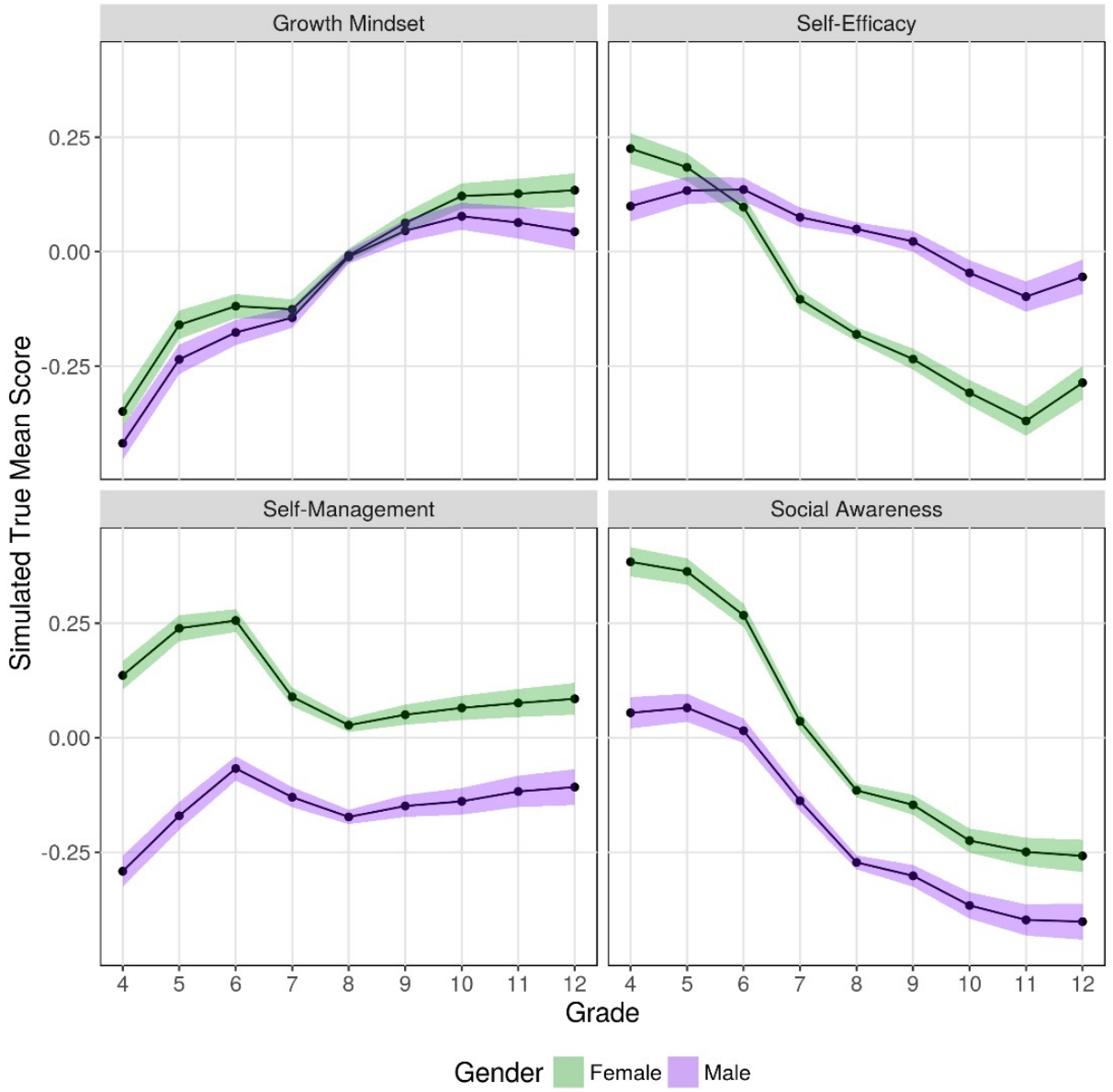
SEL constructs do not "grow" linearly or monotonically across grades: three constructs decline in middle school

- Positive (negative) selection out of (into) the sample in middle school with respect to growth mindset

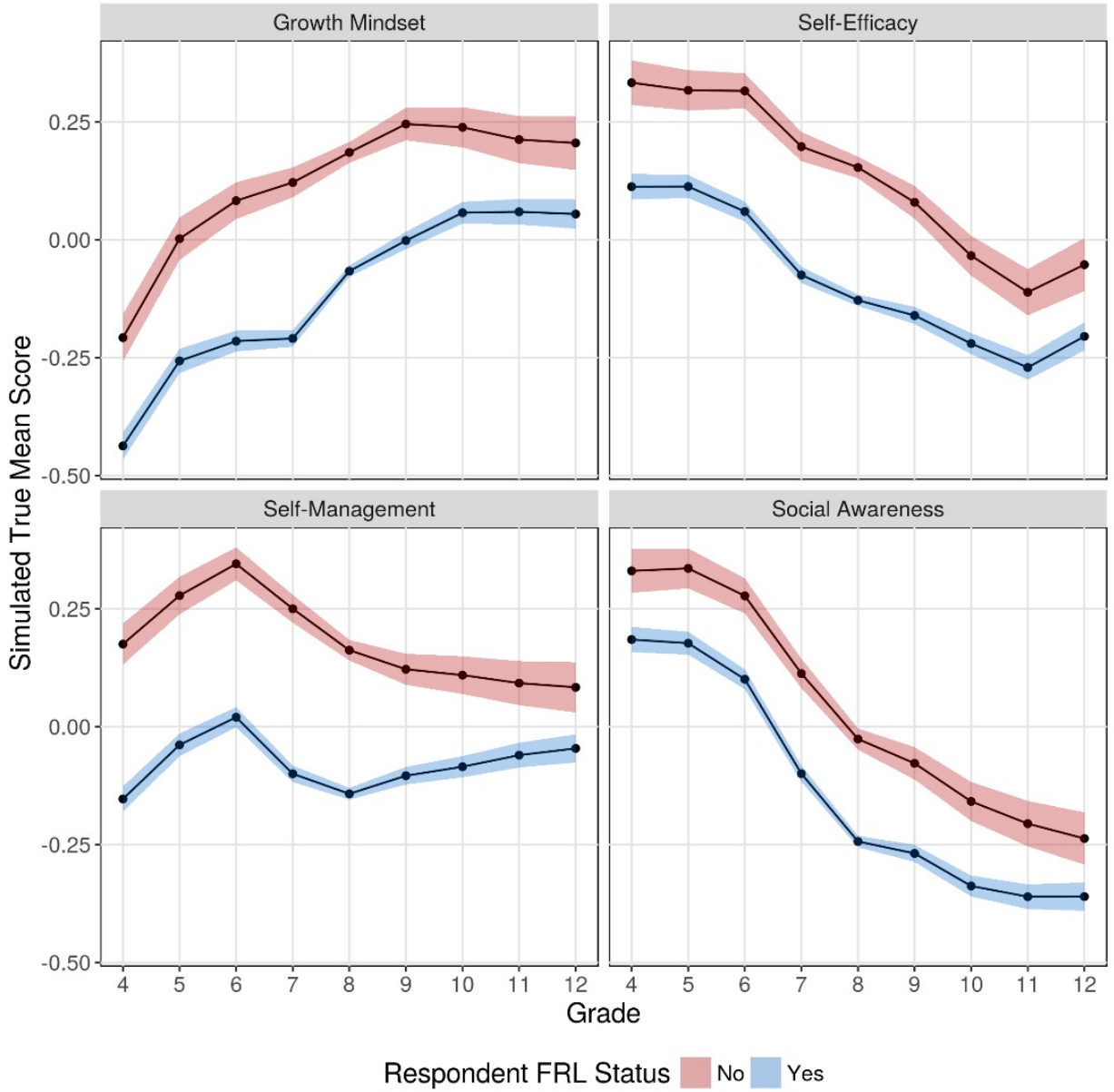
- Negative selection out of the sample on all four constructs in grades 10-12 (due to high school dropout)



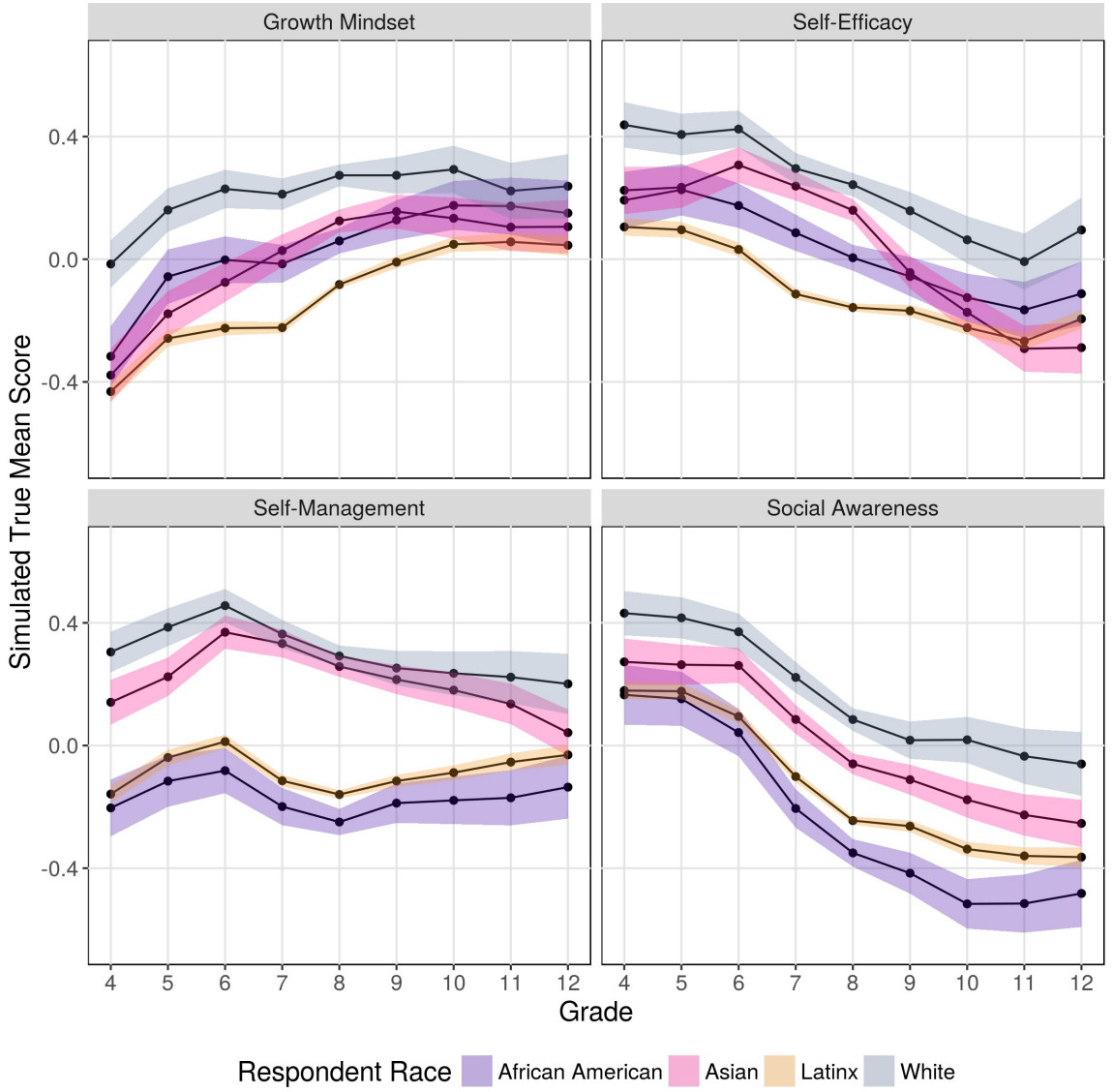
A consistent gender gap favors girls in Self-Management and Social Awareness, but girls suffer a large decline in Self-Efficacy



Economically disadvantaged students report lower levels of each construct; these gaps narrow in high school.



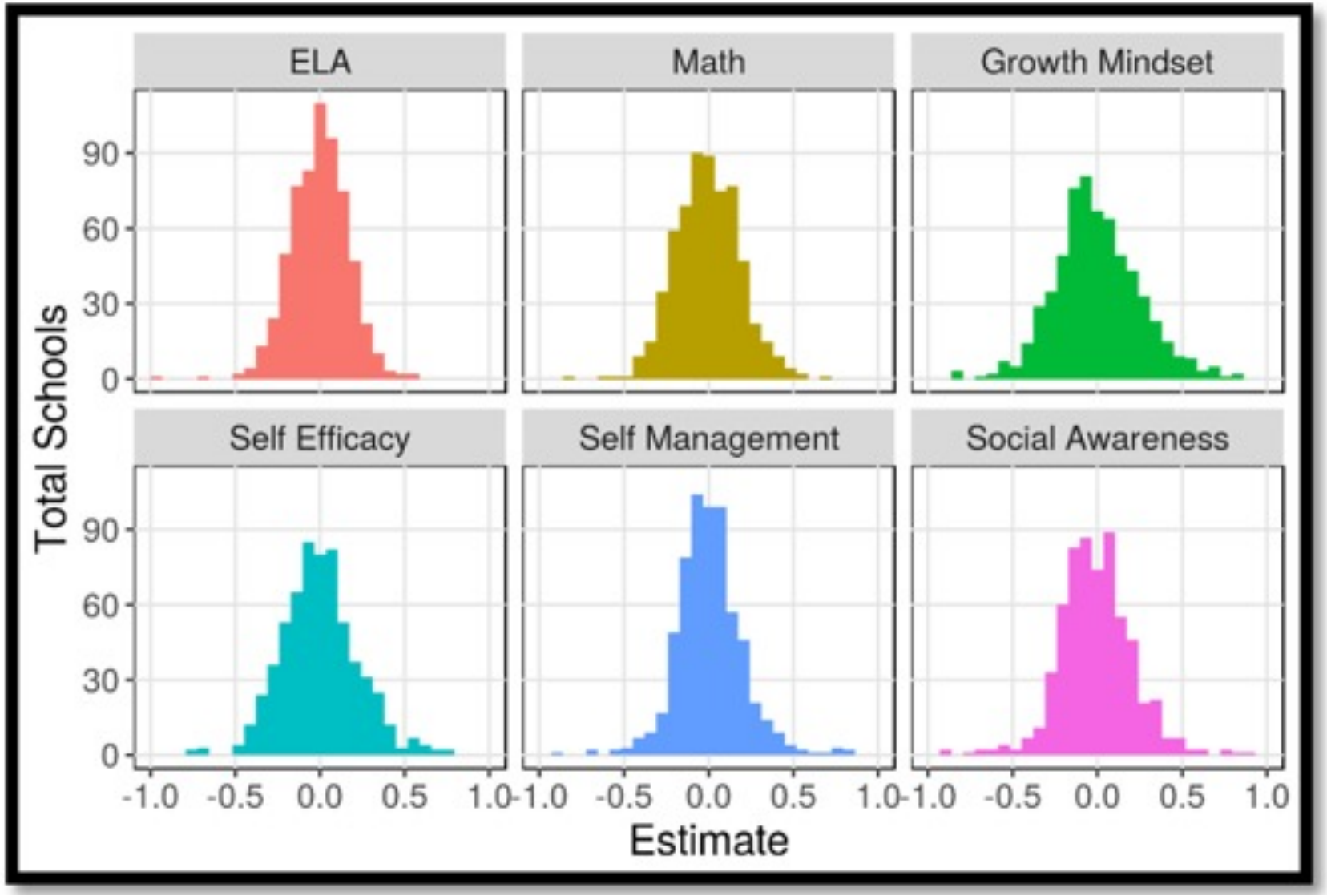
White students report higher levels than do African American and Latinx students; Asian American students report more rapid gains in Growth Mindset in middle school but a drop in Self-Efficacy high school



Implications

- Survey-based measures of SEL can maintain reliability and validity in the context of public reporting/performance management
- Survey timing, administration conditions, question wording, reporting metrics, and stakes all matter for inferences
- Need to interpret SEL survey data in light of normative trends in students' responses over time → e.g., use comparison groups to study the effects of interventions on SEL outcomes
- Aggregate data can be used to set priorities for SEL development and to target interventions and supports
- Differences by student race and economic disadvantage highlight need to use demographic adjustments and value-added methods to make inferences about school impacts (vs. student outcomes)

Within CORE, schools' estimated value added to SEL constructs varies by a similar amount as value added to test scores (but is also less stable across years than test-score value added)



Jackson et al. (2021) provide evidence of the medium-run benefits of attending a Chicago high school with strong value added to SEL

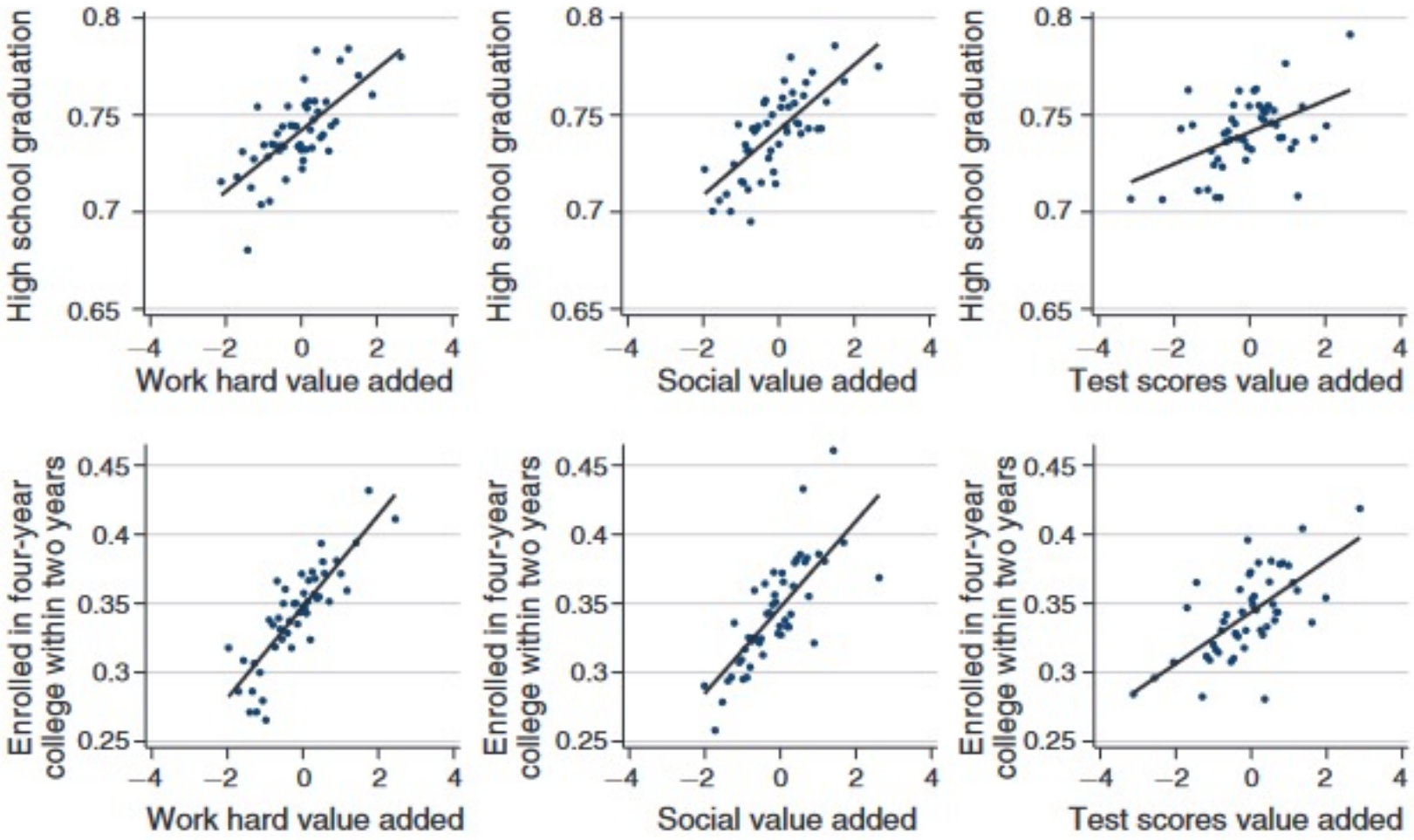


FIGURE 1. BINNED SCATTERPLOT OF HIGH SCHOOL GRADUATION AND FOUR-YEAR COLLEGE GOING BY VALUE ADDED

Implications

- Survey-based measures of SEL can maintain reliability and validity in the context of public reporting/low-stakes performance management
- Survey timing, administration conditions, question wording, reporting metrics, and stakes all matter for inferences
- Need to interpret SEL survey data in light of normative trends in students' responses over time → e.g., use comparison groups to study the effects of interventions on SEL outcomes
- Aggregate data can be used to set priorities for SEL development and to target interventions and supports
- Differences by student race and economic disadvantage highlight need to use demographic adjustments and value-added methods to make inferences about school impacts (vs. student outcomes)
- Recent evidence indicates that surveys also may be able to provide new insight on how schools affect SEL and long-run outcomes

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	Welcome & Introductions
9:40 — 10:20 am	Constructing bias-free SQM
10:20 — 10:55 am	Equity and School Quality
10:55 — 11:10 am	Break
11:10 am — 11:30 am	Case Studies: Boston and Chicago
11:30 am — 12:00 pm	Small group discussions
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	Developing and communicating multiple SQM
1:15 — 2:10 pm	Measuring school quality beyond test scores
2:15 — 2:45 pm	**Involving the community in developing SQM**
2:45 – 3:00 pm	Small group discussions

Blueprint Labs

Involving the community in developing school quality measures

Ericka Burns
Policy Manager, Denver Public Schools

Nivan Khosravi
Principal, Maxwell Elementary, Denver Public Schools

DENVER PUBLIC SCHOOLS

“Reimagine the SPF” Committee & Process

Ericka Burns, PhD
April 2022

Topics

- Background & Challenge of the Denver School Performance Framework (SPF)
- Reimagine SPF Committee process
- Committee recommendations
- Reflection & Recommendation for further discussion

Context: Our District

Denver Public Schools includes the entire city of Denver
~87,000 students
~200 schools

- ~145 district-run and innovation schools
- ~60 charter schools

FRL: 65%
EL/ML: 35%

Governance model:

- Publicly elected Board of Education
- Superintendent appointed by BoE

Background: Understanding the challenge

The DPS School Performance Framework (SPF)

What did the Denver SPF Measure?	How was the SPF used?
90-95% of framework State test scores ● Performance & Growth on state tests grades 3-11 ● Performance & Growth on language learning scores grades 3-12 ● Differences in student group performance and growth on state tests: students of color, students living in poverty ~5% of the rating was based on other measures: attendance and family satisfaction survey	● School Choice and marketing ● Teacher evaluations ● School leader evaluations ● Teacher performance payment ● School leader performance payment ● Facility allocation ● Charter school contracts ● Innovation/autonomy status

Background: Understanding the challenge

Community feedback collected over years of BoE meetings, public feedback, town halls, and superintendent listening tours

- The Single Story:
  - Principals/teachers did not feel the framework fully or accurately represented their schools
  - Families did not value a rating that was solely based on state test scores
- Confusing rating: families and educators noted the complicated calculations
- Polarizing debate:
  - Value for rigorous goals & accountability
  - Value for robust definition of quality schools

Reimagine SPF Committee

In Spring 2019, the Board & Superintendent charged staff with designing a community led, district supported process to *Reimagine the School Performance Framework (SPF)*

Guiding Questions:
What do we value in schools?
How do we measure this?
How are schools held accountable to this goal?

Committee Process

Publicly posted application:

- posted on webpage & social media
- targeted outreach to community groups
- 165 applicants

Selection Panel: reviewed applications with the goal of finding a representative committee of teachers, school leaders, parents/family members, and community members.

Committee Meetings: led by external facilitators

Meetings: 10 monthly meetings

Composition: representative of district regions, governance types, and aligned to demographics of district

- 8 teachers:
- 8 school leaders
- 8 parents/family members:
- 2 district/central office administrators
- 4 community members/advocacy group members

Committee info and process posted on the Reimagine SPF webpage

Committee Recommendations

Full committee recommendation and overview were written and presented by committee members

Overview:

1) **State SPF:** Adopt the state performance framework to capture and track school performance and to meet state accountability requirements.

---

1) **Dashboard**: Create a School *Dashboard* to inform all stakeholders on school performance and growth.

1) **Focus on Improvement:** Leverage a collaborative continuous improvement cycle to assess the ongoing performance of schools across our three value domains: Academics, Whole Child, and Culture/Climate.

Reflection & Recommendations for Further Discussions

- Learnings & reflection on the process:
  - Clear guardrails - what is in/out of the committee's scope
  - Broader community feedback loops
  - Board/District roles and relationships

- Challenges: Context and work production:
  - COVID pandemic
  - Changing leadership

- Further discussion:
  - Defining "accountability" with a public dashboard and school choice model
  - Charter contracting outside accountability frameworks
  - Designing improvement cycles
  - Shifting from *accountability* to *responsibility*

Agenda

Time	Session
8:00 — 9:00 am	Breakfast
9:00 — 9:40 am	Welcome & Introductions
9:40 — 10:20 am	Constructing bias-free SQM
10:20 — 10:55 am	Equity and School Quality
10:55 — 11:10 am	Break
11:10 am — 11:30 am	Case Studies: Boston and Chicago
11:30 am — 12:00 pm	Small group discussions
12:00 — 12:45 pm	Lunch
12:45 – 1:15 pm	Developing and communicating multiple SQM
1:15 — 2:10 pm	Measuring school quality beyond test scores
2:15 — 2:45 pm	Involving the community in developing SQM
2:45 – 3:00 pm	**Small group discussions**

Small group discussion

1. **How do you "measure the measure"**? Said another way, how do you judge the success of a school quality metric?
2. Given that all schools are different, **what are the tradeoffs of using one measure to assess all schools?**
    a. Likewise, **what are the tradeoffs of using multiple measures?**
3. **In your context, what are your goals for school quality measures?**
    a. Is the goal to help families identify the "best" school? Or better understand the features of a school that might make it the best fit for them? Or is it to make decisions about funding, closures, etc.?

Small group discussions will **end at 3:00**.

Next steps

- Post-conference survey
- Recordings

Blueprint Labs

Thank you!

Please join us for a reception, on the 6th floor

Refreshments and snacks to be served