REL MID-ATLANTIC
Regional Educational Laboratory at Mathematica

Delaware • District of Columbia • Maryland • New Jersey • Pennsylvania

Measuring educational performance for improvement:
A framework for diagnostic use of data

Brian Gill
Director
REL Mid-Atlantic

Carolyn Phenicie
Council of Chief State School Officers

Donna Johnson
Director of Accountability
DC Office of the State Superintendent of Education

Evan Kramer
Assistant Superintendent of Data, Assessment, and Research
DC Office of the State Superintendent of Education

Gene Pinkard
Director, Education and Society Program
The Aspen Institute

Tonya Wolford
Chief of Evaluation, Research, and Accountability
School District of Philadelphia

Agenda

- Welcome and introduction
- Present a framework for diagnostic use of data to measure educational performance for improvement
- Panel discussion
- Questions and answers

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
- Helping educators and policymakers make sense of data and use it for improvement is a major focus of the Regional Educational Laboratories (RELs)
- In the wake of pandemic-related disruptions, measures that are rich, broad, and actionable may be even more important
- Many districts are drowning in data from learning management systems at the same time that states are flying blind due to assessment disruptions

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

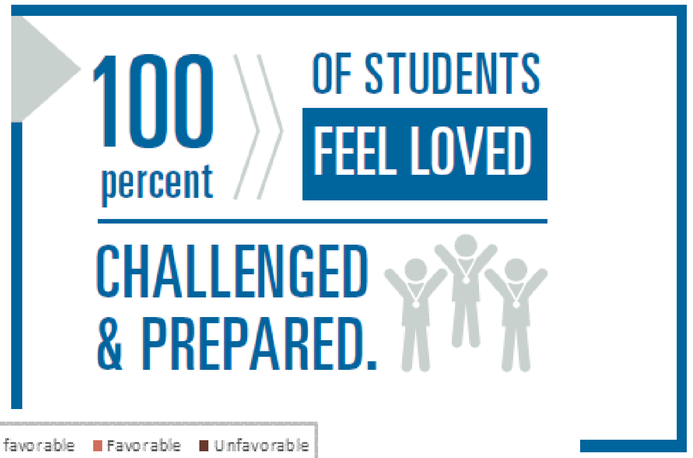
How are the kids doing?

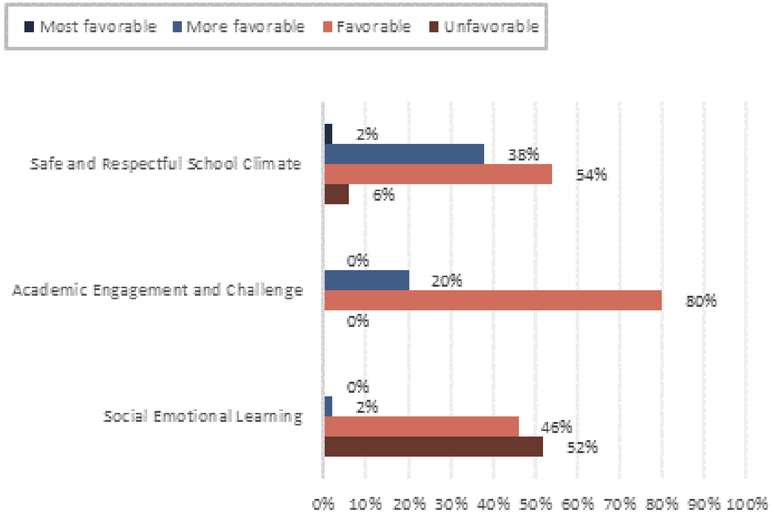
We need broader and richer measures of student outcomes

- 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 PA and 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
- Pennsylvania: Analyzed student and staff survey data to create school climate index (Amos & Xue 2021)
  - Not used for formal accountability: Schools opt into survey participation

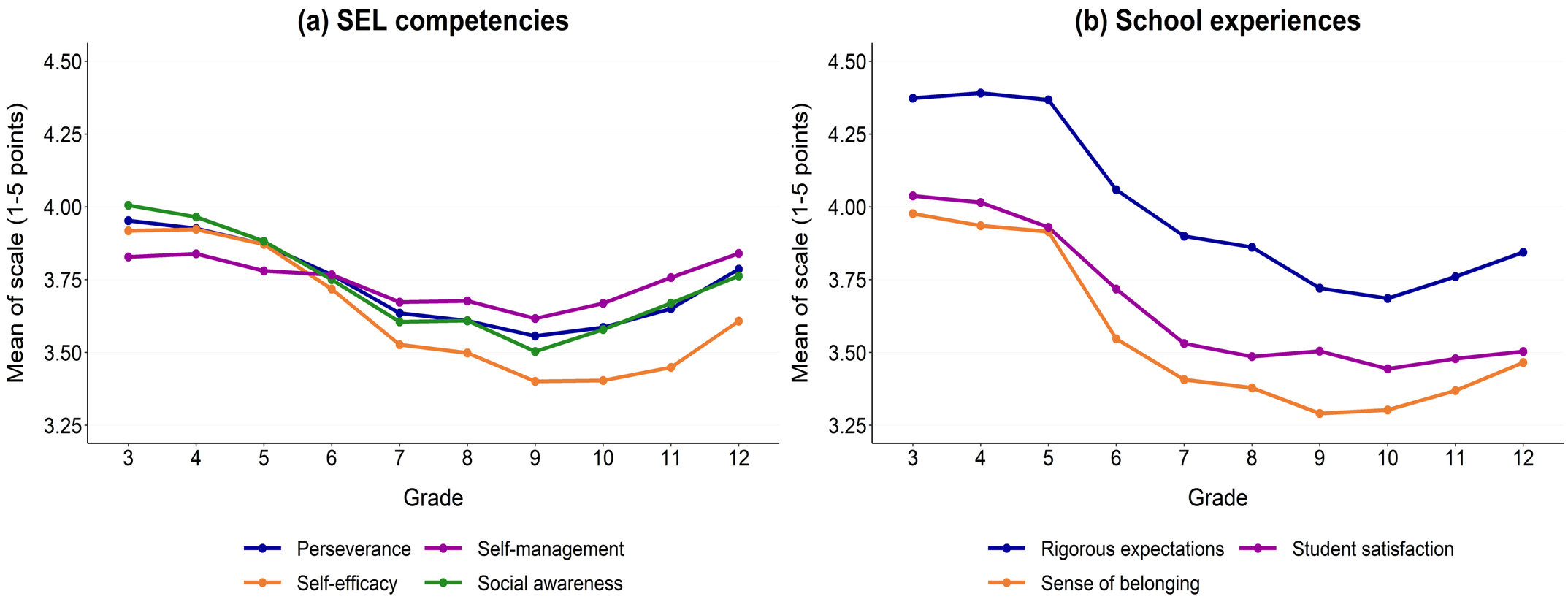




Surveys can produce useful SEL information at the school level

- 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

DCPS: SEL competencies highest in elementary school and lowest in middle school and early high school



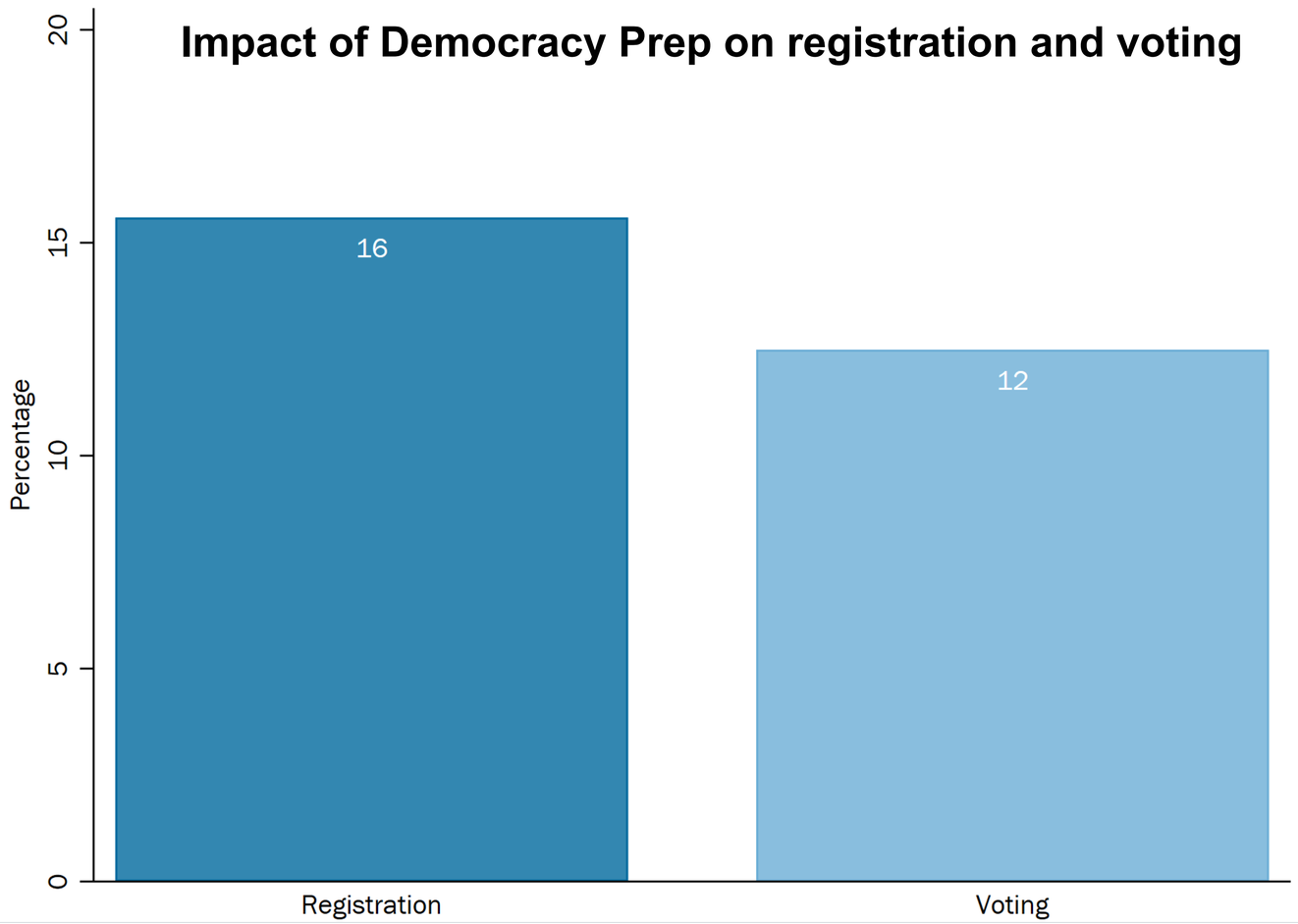
Kautz et al 2021

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

2: Impacts on student outcomes

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

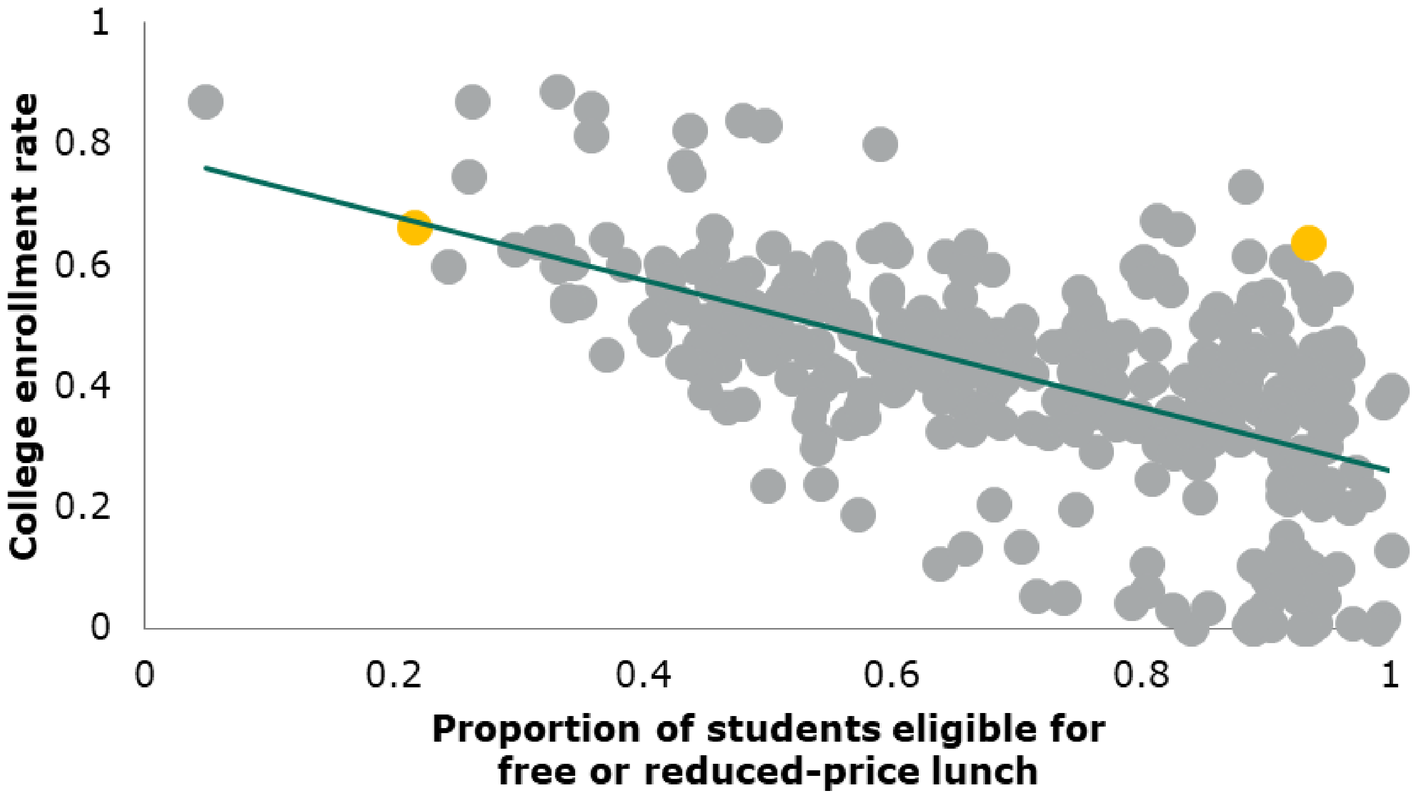
Impacts must be distinguished from raw outcomes

- NCLB critics recognized raw proficiency results conflated school performance/student characteristics
  - Critiques prompted development of value-added and student growth measures
- Expanding outcomes included in accountability systems risks re-creating 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

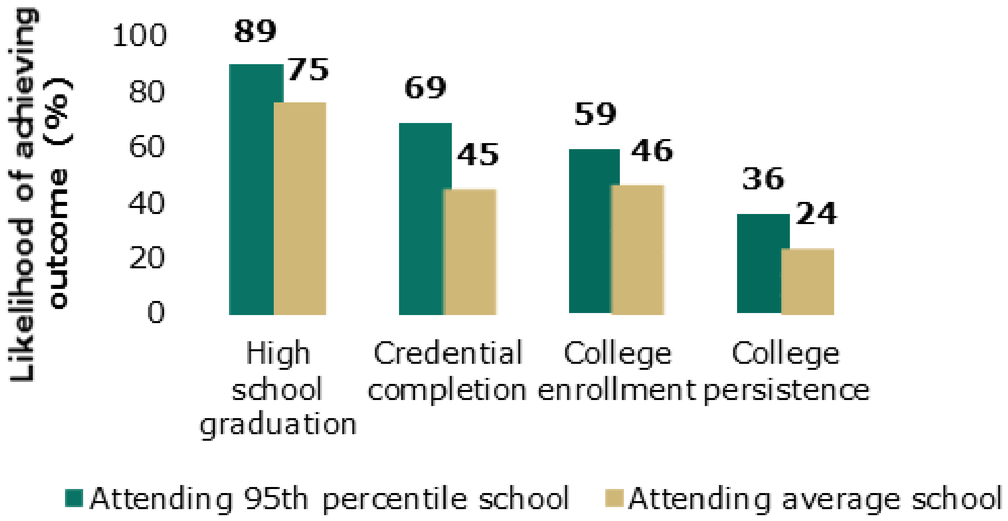
- Use statistical methods similar to those in value-added/growth models (Deutsch et al. 2020; Gross et al 2021)
- Account for differences in advantages and disadvantages of students served
  - Poverty, prior achievement, IEP and ELL status, anything relevant and measured prior to high-school entry
- Aims to 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

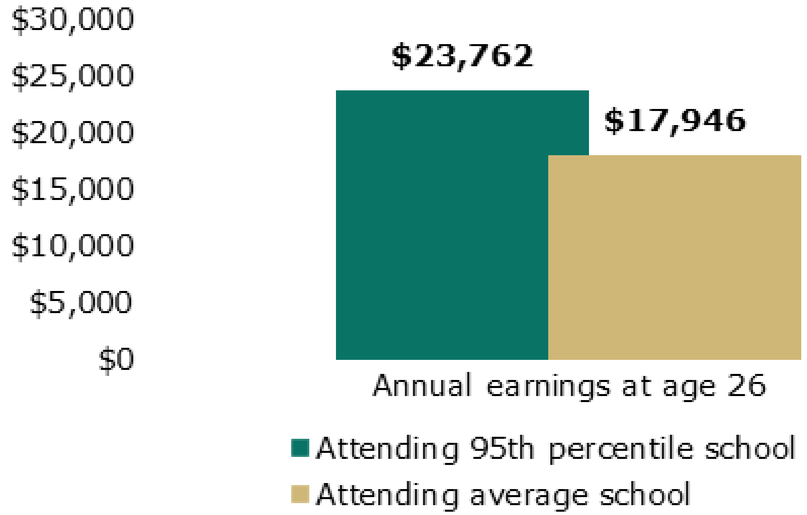
High schools can have the same student outcomes with very different promotion power



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

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

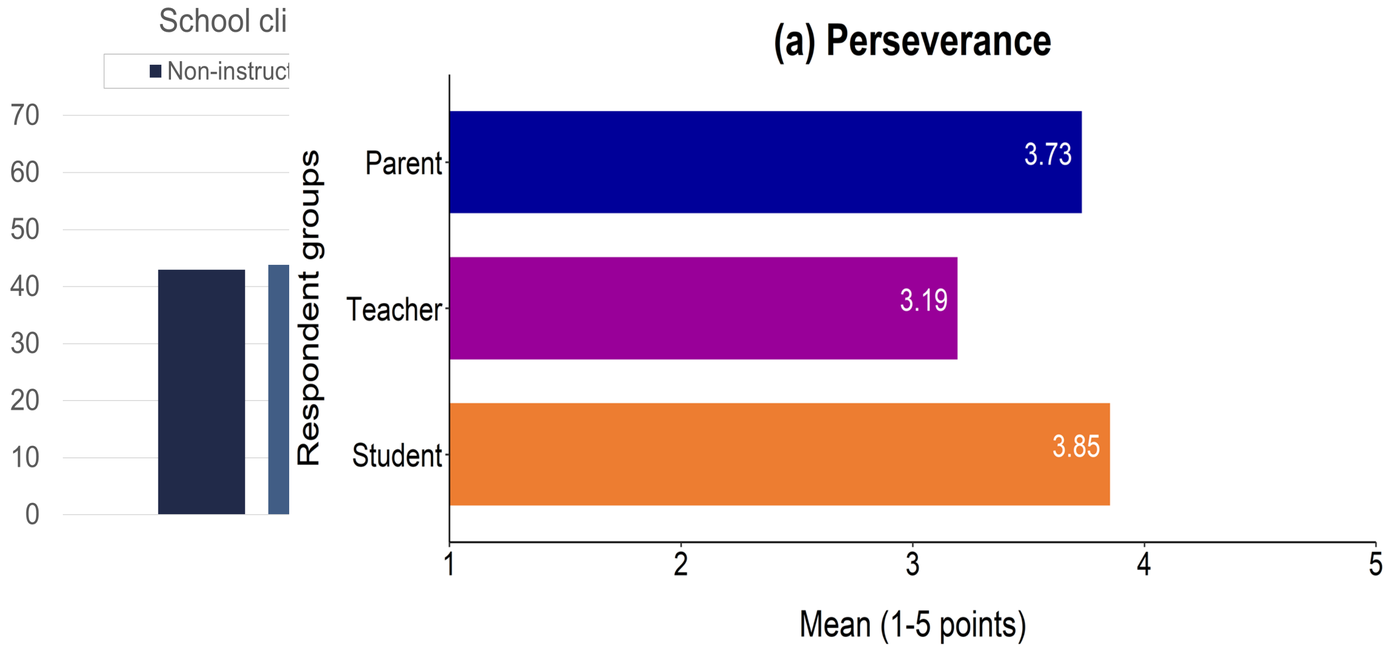
Process measures might compensate for weaknesses of outcome and impact measures

- 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)

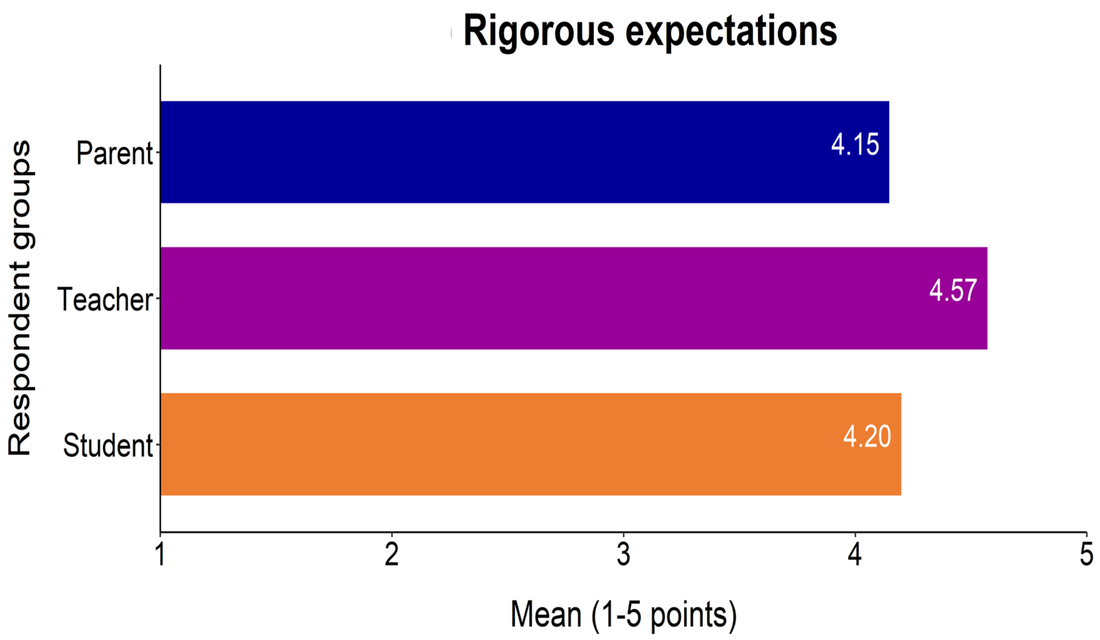
- Some impacts may be impossible to measure
  - We can probably produce acceptably valid and reliable estimates of the value-added of teachers and schools, but…
  - Nobody knows how to measure the impact of an individual principal on student outcomes (Chiang, McCullough, Lipscomb, Gill 2016)

School climate: Student and staff ratings are correlated but also differ systematically

Pennsvlvania

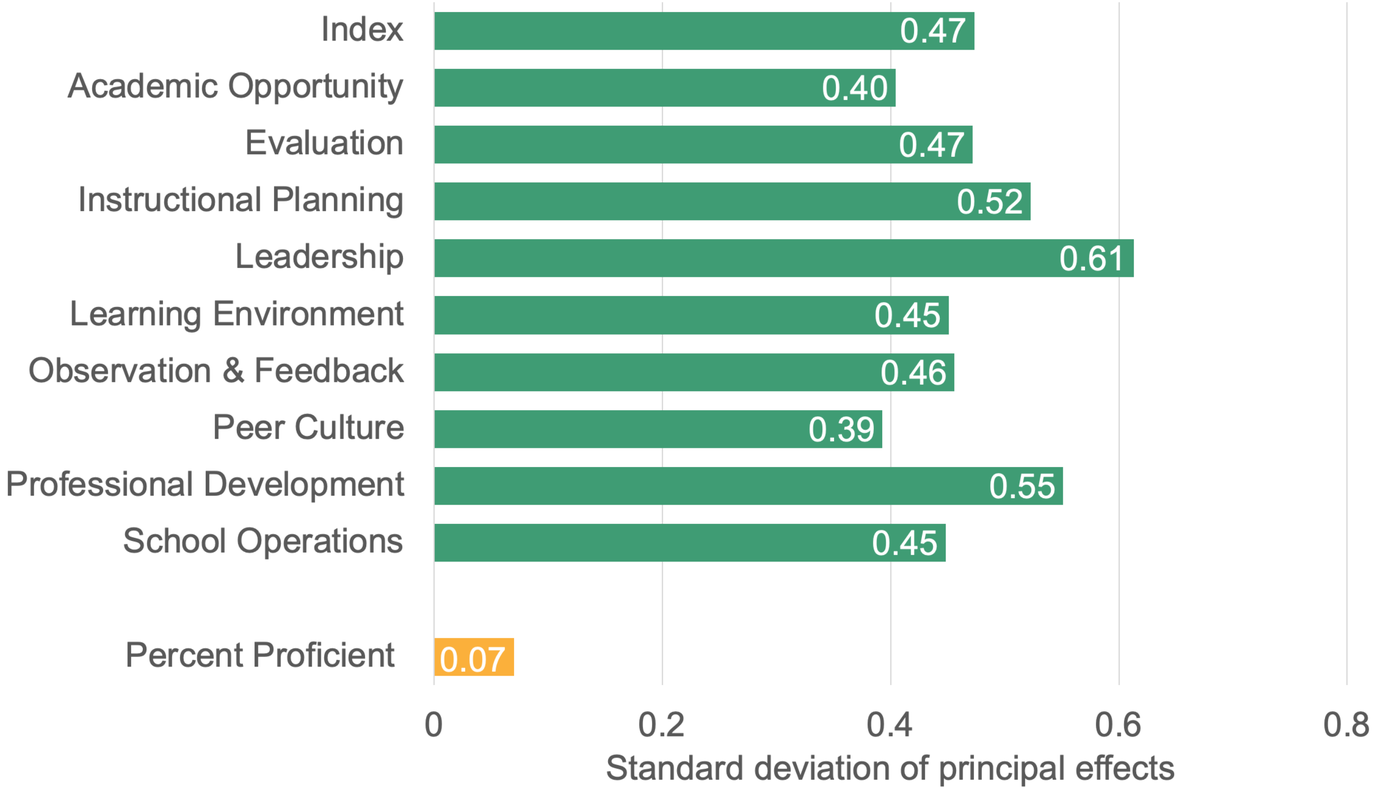


DCPS



Kautz et al 2021

School climate changes substantially with a new principal



Kozakowski, Gill, Shiferaw 2021

Some possible implications for policy and practice

1: Broaden discussion of performance beyond student outcomes and formal consequences

- Accountability can be created without high stakes (Gill, Lerner, and Meosky 2016)
  - Information alone can motivate
  - Campbell's Law is real, and some measures are more corruptible than others
  - Lower stakes may lower the temperature
  - Transparency can create accountability
- Outcomes are not the only measures that matter

2: Recognize the limits of data and measurement

- Good measures can identify:
  - Student and school needs (how are the students doing?)
  - Underperforming schools (what is the school contributing?)
  - Broken processes (what is happening in the school?)
- But measurement alone can’t tell an educator or administrator what to do
- Actions to improve must be informed by expert knowledge from the field

Panel discussion and Q&A

Carolyn Phenicie
Council of Chief State School Officers

Brian Gill
Project Director
REL Mid-Atlantic

Donna Johnson
Director of Accountability
DC Office of the State Superintendent of Education

Tonya Wolford
Chief of Evaluation, Research, and Accountability
School District of Philadelphia

Evan Kramer
Assistant Superintendent of Data, Assessment, and Research
DC Office of the State Superintendent of Education

Gene Pinkard
Director, Education and Society Program
The Aspen Institute

Stakeholder feedback survey

- Please complete a short, 2-minute, survey to provide feedback on the webinar
- The survey is voluntary, and the results are anonymous and only used to improve future webinars
- The survey will pop up automatically at the conclusion of the webinar. Please stay logged into WebEx to complete it.

Paperwork Burden Statement: *According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless such collection displays a valid OMB control number. The valid OMB control number for this information collection is* ***1880-0542****. The time required to complete this information collection is estimated to average* ***10 minutes*** *per response, including the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection.* ***If you have any comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to:*** *U.S. Department of Education, Washington, D.C. 20202-4700.* ***If you have comments or concerns regarding the status of your individual submission of this form, write directly to: Felicia Sanders****, Institute of Education Sciences, U.S. Department of Education, 550 12th Street, SW, Room 4117, Washington, D.C. 20202; Felicia.Sanders@ed.gov.*

Citations

- Amos, Lauren, and Yange Xue. "Development of Pennsylvania Department of Education School Climate Index Summary." Washington, DC: Regional Educational Laboratory Mid-Atlantic, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, June 2021.
- Chaplin, Duncan, Brian Gill, Allison Thompkins, and Hannah Miller. "Professional Practice, Student Surveys, and Value-Added: Multiple Measures of Teacher Effectiveness in the Pittsburgh Public Schools." Washington, DC: Regional Educational Laboratory Mid Atlantic, Institute of Education Sciences, U.S. Department of Education, July 2014.
- Chiang, Hanley, Moira McCullough, Stephen Lipscomb, and Brian Gill. "Can Student Test Scores Provide Useful Measures of Principals' Performance?" NCEE Report No. 2016-002. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, September 2016.
- Deutsch, Jonah, Matthew Johnson, and Brian Gill. "The Promotion Power Impacts of Louisiana High Schools." Chicago: Mathematica, March 2020.
- Gill, Brian, Brandon Coffee-Borden, and Kristin Hallgren. "A Conceptual Framework for Data-Driven Decisionmaking." Cambridge, MA: Mathematica Policy Research, June 2014.
- Gill, Brian, Emilyn Ruble Whitesell, Sean P. Corcoran, Charles Tilley, Mariel Finucane, and Liz Potamites. "Can Charter Schools Boost Civic Participation? The Impact of Democracy Prep Public Schools on Voting Behavior." *American Political Science Review,* 2020. doi:10.1017/S000305542000057X.
- Gill, Brian, Jennifer S. Lerner, and Paul Meosky. "Re-Imagining Accountability in K-12 Education: A Behavioral Science Perspective." *Behavioral Science & Policy,* vol. 2, no. 1, 2016, pp. 57-70.
- Gross, Max, Menbere Shiferaw, Jonah Deutsch, and Brian Gill. "Using Promotion Power for College- and Career-Ready SAT Scores, Graduation, and College Enrollment to Identify the Effectiveness of Public High Schools in the District of Columbia." REL 2021-098. Washington, DC: Regional Educational Laboratory Mid-Atlantic, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, September 2021.
- Kautz, Tim, Kathleen Feeney, Hanley Chiang, Sarah Lauffer, Maria Bartlett, and Charles Tilley. "Using a Survey of Social and Emotional Learning and School Climate to Inform Decisionmaking." REL 2021-114. Washington, DC: Regional Educational Laboratory Mid-Atlantic, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, August 2021.
- Kozakowski, Whitney, Brian Gill, and Menbere Shiferaw. "Exploring the Potential Role of Staff Surveys in School Leader Evaluation." REL 2021-117. Washington, DC: Regional Educational Laboratory Mid-Atlantic, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, August 2021.
- Raudenbush, Stephen, and Marshall Jean. "To What Extent Do Student Perceptions of Classroom Quality Predict Teacher Value Added?" In *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project,* edited by T. Kane, K. Kerr, and R. Pianta (pp. 170–202). San Francisco: Jossey-Bass, September 2015.

Thank You & Contact Information

- Brian Gill: BGill@mathematica-mpr.com
- Carolyn Phenicie: Carolyn.Phenicie@ccsso.org
- Donna Johnson: DonnaR.Johnson@dc.gov
- Evan Kramer: Evan.kramer@dc.gov
- Gene Pickard: Eugene.Pinkard@aspeninstitute.org
- Tonya Wolford: Twolford@philasd.org

Disclaimer

This work was funded by the U.S. Department of Education's Institute of Education Sciences (IES) under contract ED-IES-17-C-0006, with REL Mid-Atlantic, administered by Mathematica. The content of the presentation does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

REL MID-ATLANTIC
Regional Educational Laboratory
at Mathematica

https://ies.ed.gov/ncee/edlabs/regions/midatlantic/