



# What Problems Are Current Accountability Systems Intended to Address? Have They?

Scott Marion

National Center for the Improvement of Educational Assessment

*Reidy Interactive Lecture Series*

*Portsmouth, NH*

September 21-22, 2023



# Advanced Organizer

---

- Introduction
- Where did we start? Looking back to ESEA
- History of the various reauthorizations
- Theory of action for NCLB and test-based accountability
- Has it worked? It depends on who you ask.
- My ideas

# Disagree without being disagreeable

---

- I will undoubtedly say things that will exasperate some of my colleagues and many of you.
- I'm sure I will disagree with some things I hear from others.
- Accountability discussions invite this sort of disagreement.

# Accountability Systems Are Value Statements

- Accountability systems are policy mechanisms designed to instantiate values.
- Therefore, it is not surprising that accountability discussions invite considerable disagreement.



## “Same Values”/Different Methods

---

- In fact, some of the biggest disagreements occur when different people or groups purport to share the “same values.”
- In many cases, people are using the same word to mean different things.
- But even when the term is being used in similar ways, people can disagree over how to instantiate the values.
- Let’s look at how “equity” may be operationalized in accountability as an example...

# The “Common Methods” Position

---

- Common metrics are used to monitor and evaluate equitable opportunities to learn
- Which outcomes?
  - Achievement and growth on state tests
  - Graduation expectations and rates
  - Postsecondary readiness
  - Other
- Outcomes based on what?
  - State-defined standards
  - State-defined tests and achievement levels
  - Common design and administration to ensure high levels of comparability
- Key Groups:
  - Ed Trust, the Alliance for Excellent, NCLD, ERN, DQC

# The “Personalized and Culturally Relevant” Position

---

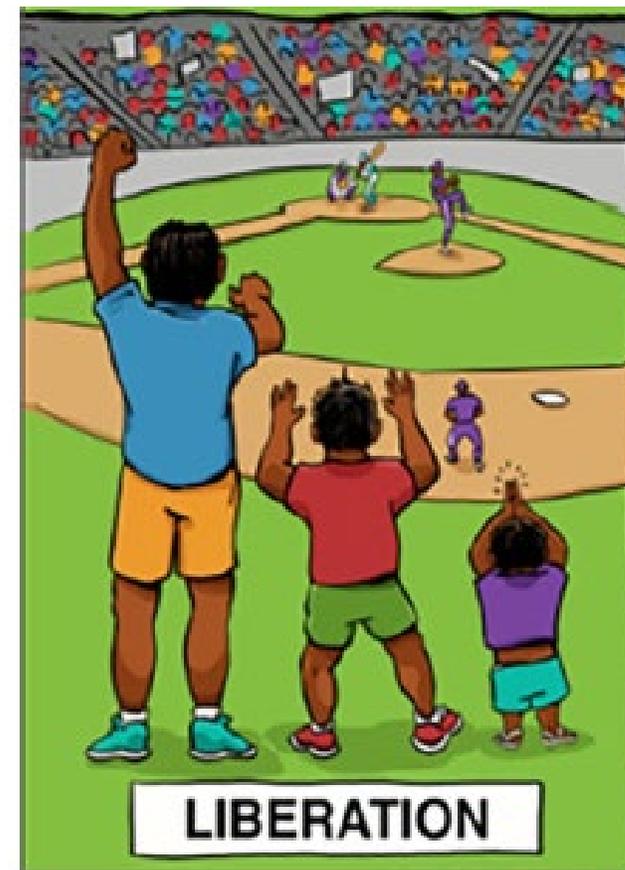
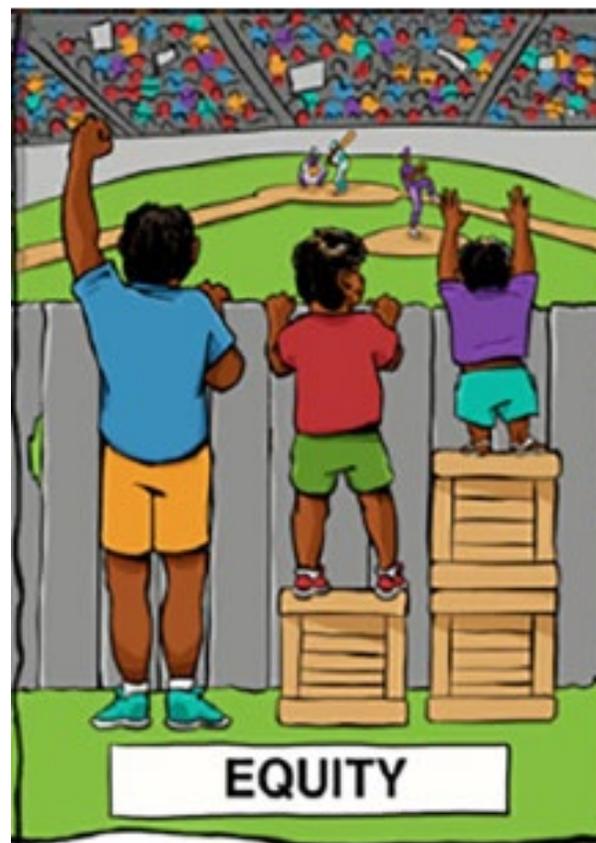
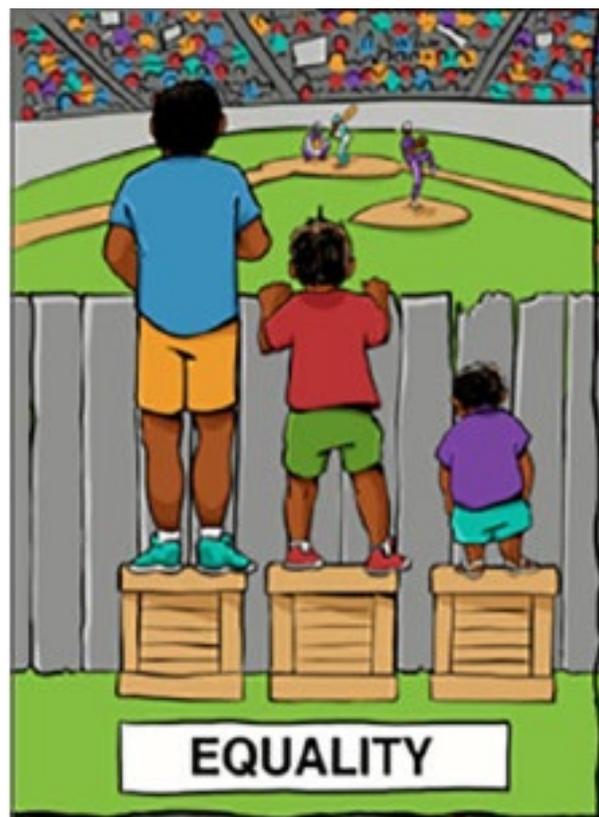
- We must honor and lift up students’ cultural and linguistic heritages and consider questions such as:
  - Whose knowledge?
  - Whose way of knowing?
- These questions lead us to culturally responsive pedagogy and culturally sustainable assessment practices that support student voice and choice in defining:
  - Learning targets
  - Evidence collection
  - Evaluations of quality
- This leads to more flexible assessment and accountability systems
- Key Groups:
  - LPI, KnowledgeWorks, NEPC, NAACP

# Equity

---

- OK, perhaps I picked a challenging example for first thing in the morning.
- My point isn't to provide a deep philosophical dive into conceptions of equity, but since we're here, I strongly recommend that everyone using the term "equity" read this relatively recent article:
- Levenson, M., Geron, T., & Brighthouse, H. (2022). Conceptions of educational equity. *AERA Open*. 8,. 1, 1–12  
<https://doi.org/10.1177/23328584221121344>

# From: Levenson, Geron, Brighthouse (2022)



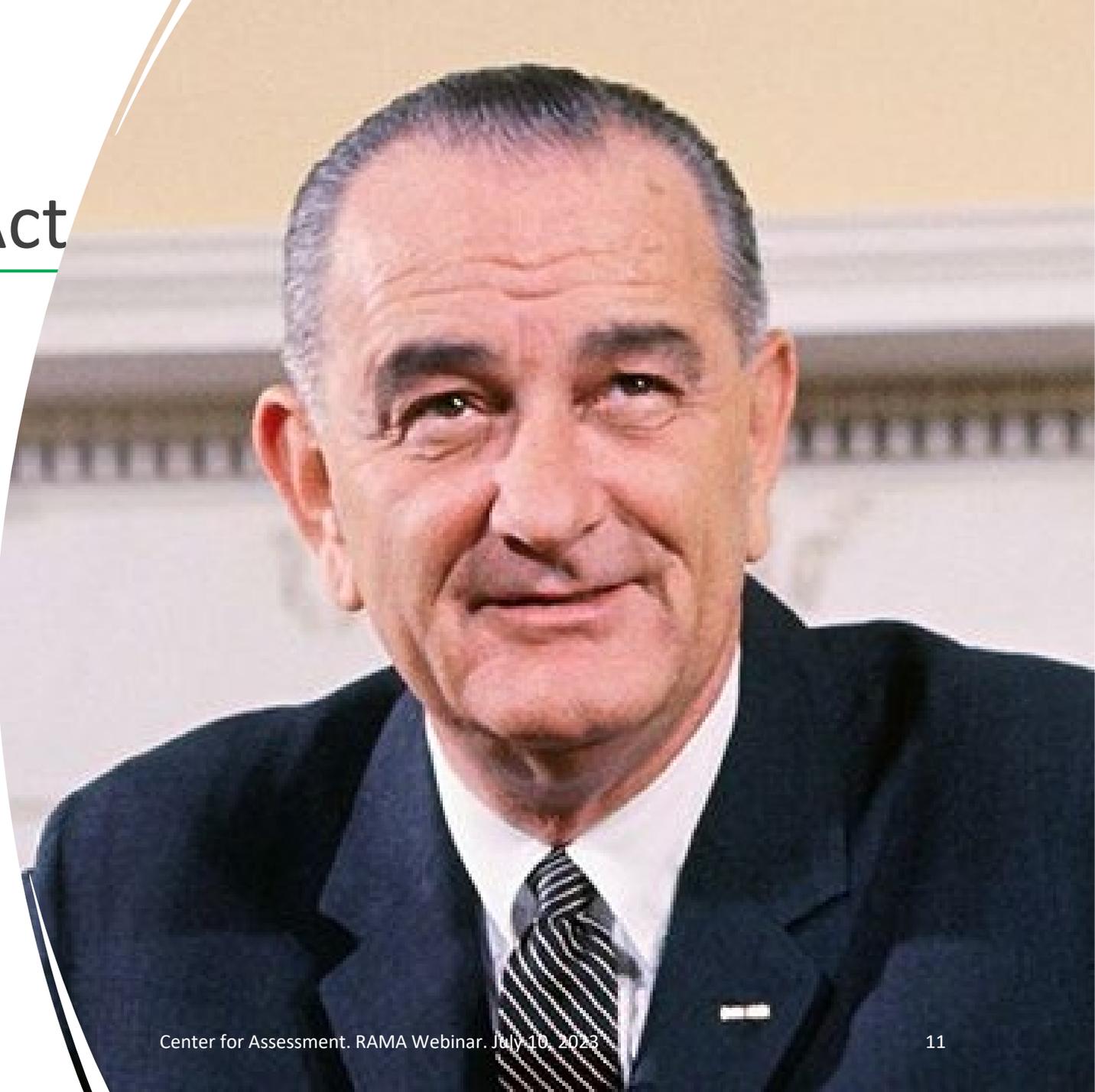
# Where did we start?

---



# Elementary and Secondary Education Act

- Originally enacted in 1965 as part of Lyndon Johnson's "Great Society" programs designed to address income inequality in the United States
- It provides approximately 7% of state education budgets targeted to low-income, special education, and English learner students.
- The most recent reauthorization is called The Every Student Succeeds Act.

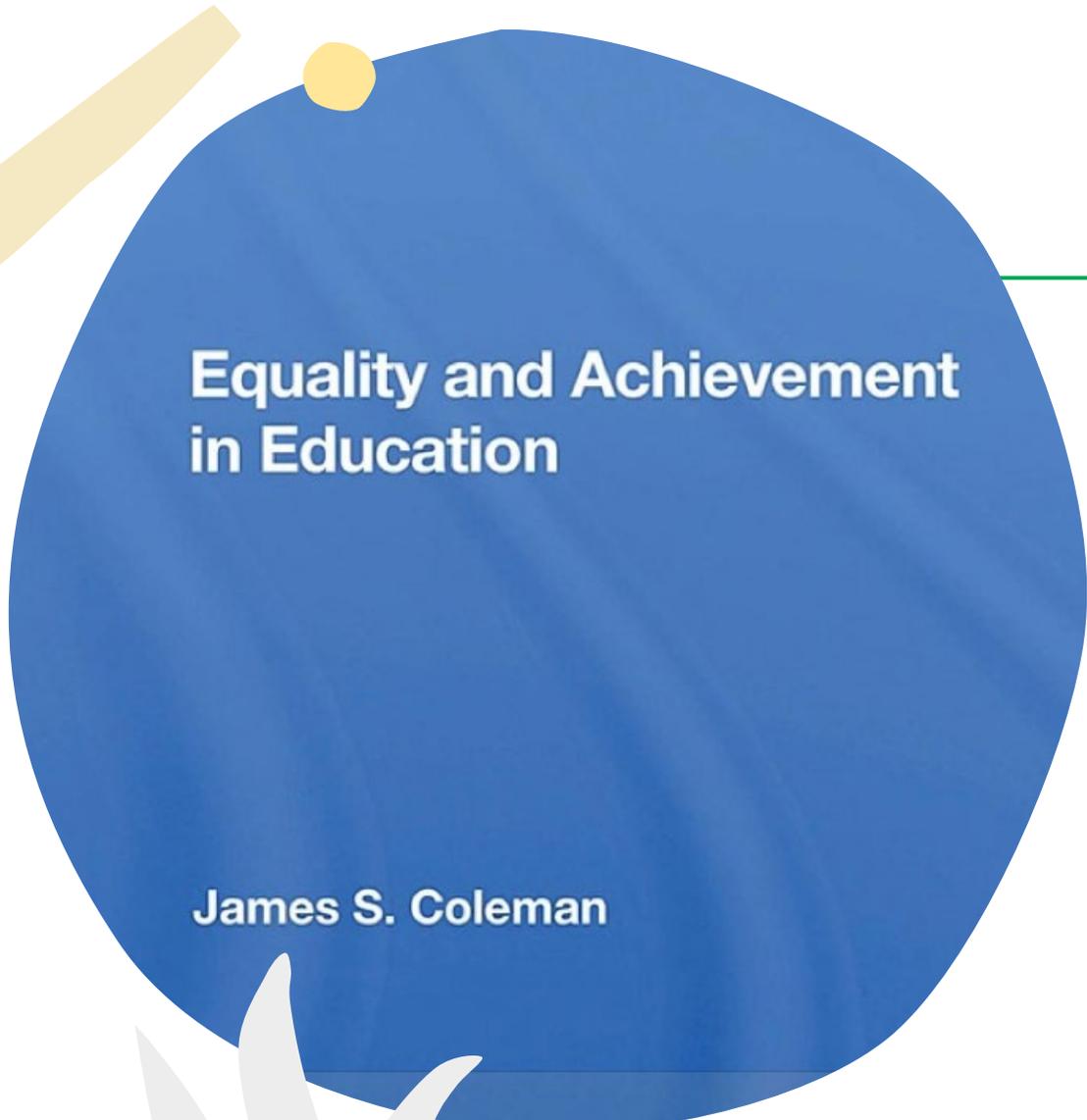


# What Problems Was ESEA Designed to Address?

---

*SEC. 201. In recognition of the special educational needs of children of **low-income families and the impact that concentrations of low-income families have on the ability of local educational agencies to support adequate educational programs**, Congress hereby declares it to be the policy of the United States to provide financial assistance (as set forth in this title) to local educational agencies serving areas **with concentrations of children from low-income families to expand and improve their educational programs by various means (including preschool programs) which contribute particularly to meeting the special educational needs of educationally deprived children.***

PUBLIC LAW 89-10-APR. 11, 1965, p. 27 (emphasis added)



## Equality and Achievement in Education

James S. Coleman



# Implied and Explicit Goals

---

- Equity was limited by the social, cultural, and political context of 1965
  - In other words, student groups defined by race, ethnicity, special education, or language learner status were absent from the law. Only focused on income.
- Multiple conceptions of equity were later framed by Coleman (1966 & 1967), ranging from “equal” inputs to equal outcomes
  - The original landmark study was justifiably criticized for several reasons, but his conceptualization of equality of educational opportunity was important.

# A Nation at Risk

The Imperative for Educational Reform



A Report to the Nation and the Secretary of Education

## Introducing Standards and Testing

- ESEA went along for almost 30 years without state requirements for common content standards or assessments.
- The advent of standards-based reform (e.g., AAAS, 1986, NCTM, 1989, and Smith & O'Day, 1991) and major national policy documents such as *A Nation at Risk* (1983) and *Goals 2000* (1989) led to new requirements for state standards and assessments in the Improving America's Schools Act (IASA), the 1994 iteration of ESEA.



# Key Accountability Provisions of IASA, NCLB, and ESSA

**IASA** required states to assess all students (not just Title I schools).

IASA required states to establish “adequate yearly progress (AYP)” standards for students in schools and districts receiving Title I funds and to report assessment results at the state, district, and school levels disaggregated by gender, race, limited English proficiency status, migrant status, disability, and economic status.

IASA required AYP to be based on the performance of all students with annual school targets established by each state.

IASA required the participation of all students on assessments but did not specify a minimum participation threshold or explicitly mandate the participation of certain subgroups – e.g., students with disabilities.

**NCLB** required states use assessments to hold all schools accountable.

NCLB built on the reporting requirements mandated under IASA by requiring states to develop and report AYP statewide measurable objectives for improved achievement by all students and for specific subgroups .

NCLB required all students to meet proficiency 2014, including all subgroups.

NCLB required that 95 percent of all students and all students in each subgroup participate in state assessments. Any subgroup or the all students group not meeting the 95% requirement would cause a school not to meet AYP.

Further, NCLB required that all students, and each student group, meet or exceed AYP targets for that given year. If not, the entire school would be considered as not meeting AYP.

**ESSA** maintained the requirement that states use assessments to hold schools accountable but allowed for additional indicators and state discretion in how various indicators were combined to make judgements about schools.

ESSA eliminated AYP and the 100 percent proficiency requirement, and it included three new subgroups: homeless status, status as a child in foster care, and status as a child with a parent serving in the military.

The ESSA accountability requirements included academic achievement results based on state test scores, an additional academic indicator (generally student longitudinal growth), progress in English learners demonstrating proficiency in academic English, graduation rate for high schools, and one additional indicator of “school quality or student success.” The two indicators based on state test scores (academic achievement and student growth) must be weighted significantly more than other indicators in school ratings.

ESSA maintained the 95 percent participation requirement but instead of leading to failure of AYP, states were required to include the number of non-participants below the 95% threshold in the denominator of the academic achievement calculation, i.e., percent proficient.

# Competing Theories of Action in the Shift to NCLB

## Standards-Based Reform

- Clear and rigorous content standards
- Well-articulated performance standards (later named achievement standards)
- Assessments to reflect ambitious learning goals
- School delivery standards
  - Capacity-building, to enable critical instructional changes and other supports needed to make it possible for all students to reach high standards, was the linchpin in the SBR theory of action.

## Incentives or Test-Based Accountability

- “We’ll regulate less, if schools and school districts will produce better results” (Alexander, 1986, p. 202)
- Incentives theory of change:
  - Ambitious learning goals
  - Tests as the primary basis for school accountability
  - Rewards and sanctions to serve as incentives for improvement
- Guess which became the implied theory of action for NCLB?

# Has it worked?

---



# Has it Worked?

---

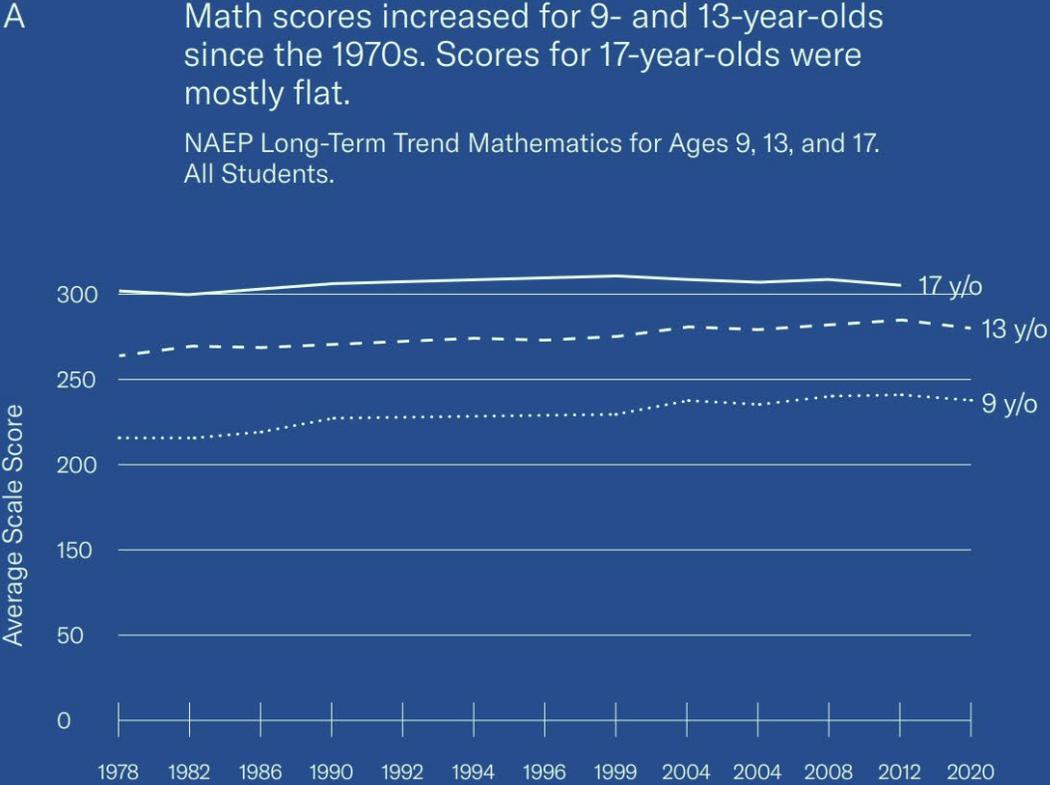
- Well, that depends on who you ask and what you mean by “worked.”
- The original intent of the ESEA – equal access to high-quality education – should be the lens through which we evaluate the theory and outcomes of test-based accountability.
- Using later reauthorizations (NCLB, ESSA), we can look at these outcomes:
  - Improved overall achievement
  - Improved relative achievement of student groups (e.g., closing gaps)
  - Improved graduation rates overall and for student groups
- We must also consider unintended negative consequences, including opportunity costs, in any evaluation of whether “it worked.”

# It Worked!

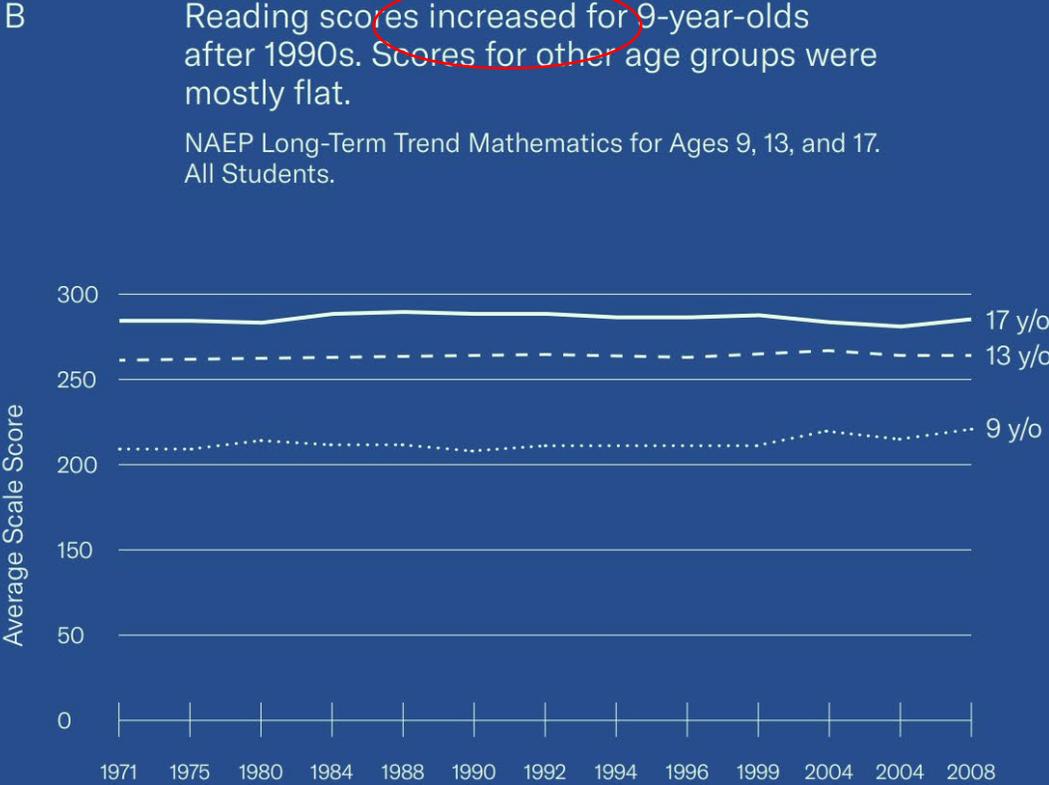
---

- It's fascinating that multiple groups of smart people can look at the same set of evidence and come to wildly different conclusions:
- *First, it's clear from multiple studies and analyses that student achievement in the United States **improved dramatically** from the mid to late 1990s until the early 2010s—especially in math, especially at the elementary and middle school levels, and especially for the most marginalized student groups ([Petrilli, 2023](#)), [emphasis added]*

### Figure 1 Long-Term NAEP Trends in Math and Reading by Age



Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) Long-Term Trend Mathematics Assessments.

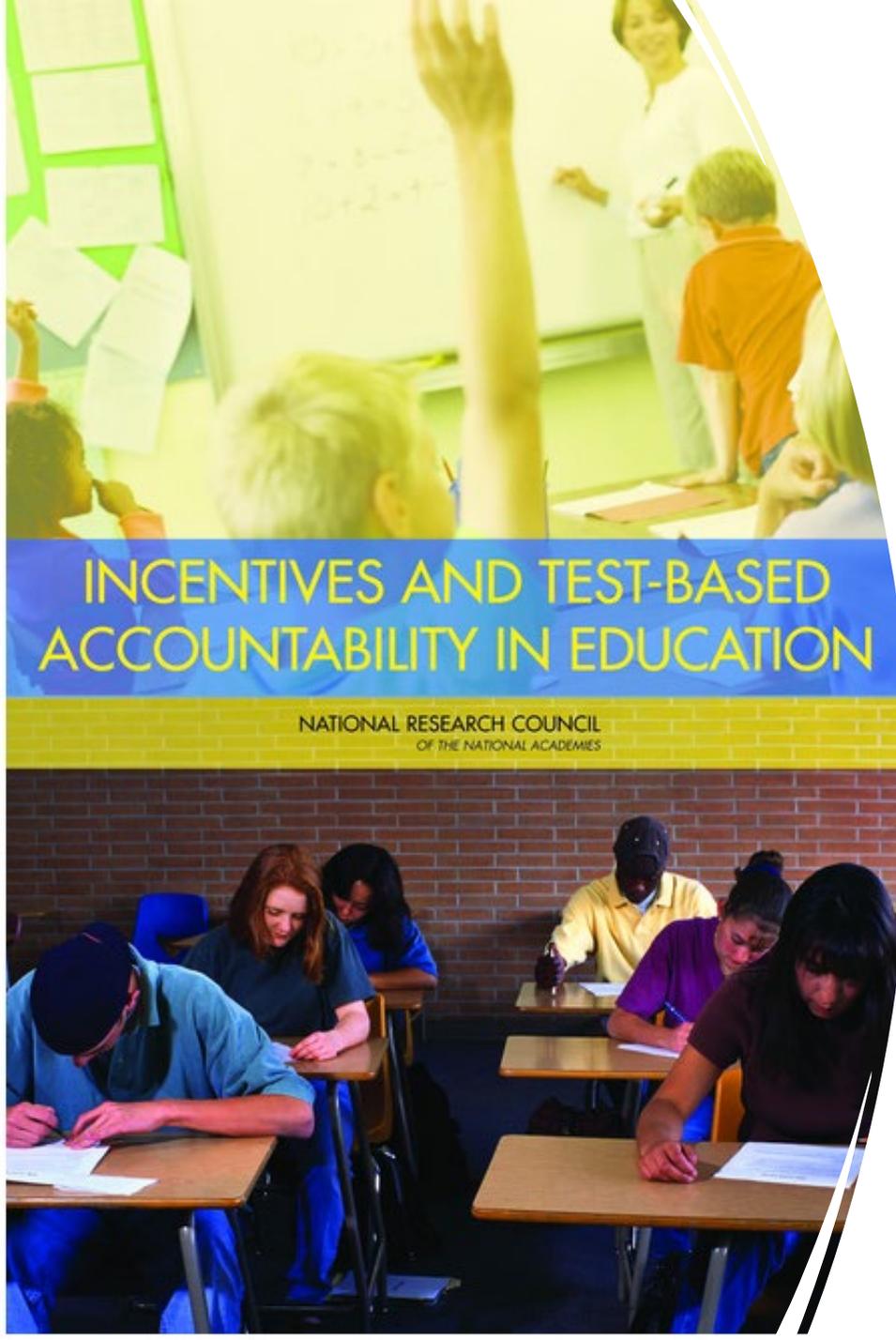


Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) Long-Term Trend Mathematics Assessments.

# Dee and Jacob (2011)

---

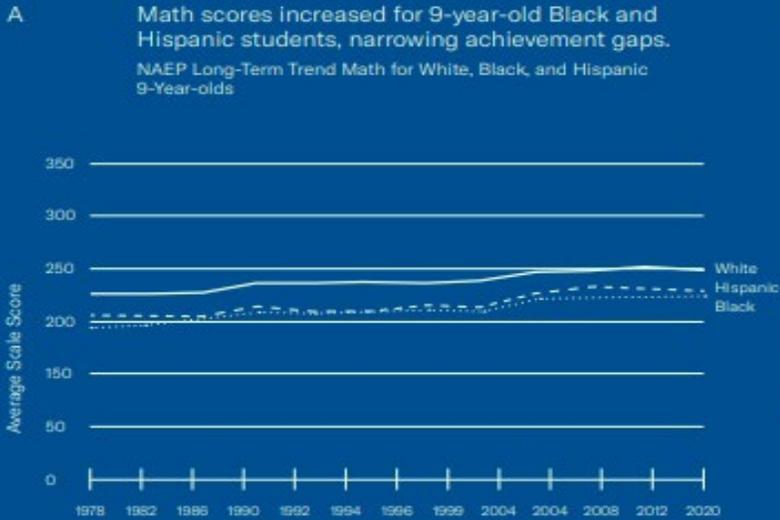
- One of the most cited studies in support of test-based accountability
- They quantified the effect of NCLB on a state's NAEP achievement scores using regression analyses that took account of whether a state had had consequential accountability in place prior to NCLB and, if so, for how long.
- According to the model's logic, states with accountability in place the longest (the earliest being 1991-1992) would be expected to experience the least effect on NAEP achievement from NCLB. Dee and Jacob's results are complex, with many robustness checks.
- **Their main finding** was that for states without prior school accountability, NCLB produced a gain in 4<sup>th</sup> grade math by 2007 of **.23 standard deviations**. The .10 standard deviation effect in 8<sup>th</sup> grade math was not statistically different from zero, and there was **no discernible effect of NCLB on 4<sup>th</sup> or 8th grade reading**.



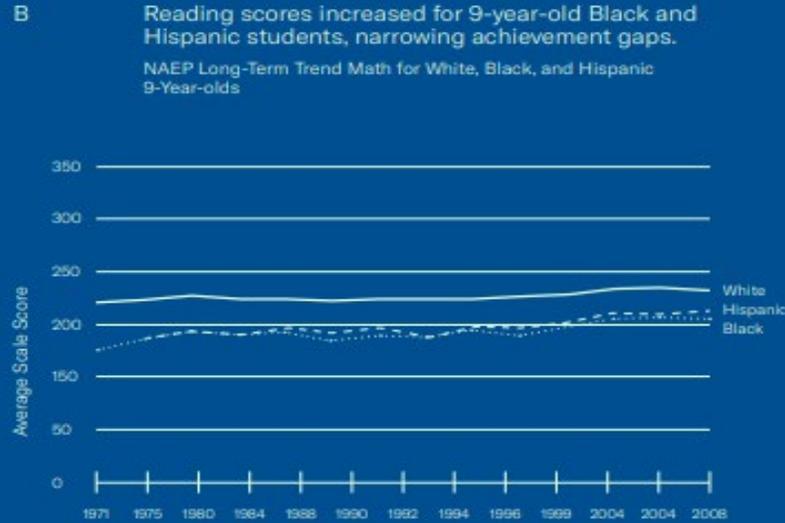
# NRC Committee on Incentives and Test-Based Accountability in Public Education (NRC, 2011)

- The NRC Committee concluded that the **average effect of consequential, test-based accountability policies on student achievement is .08 standard deviations**. Combined effects across grades (most often NAEP grades 4 and 8) and subject areas (reading and math).
- An alternative summary of the same literature might be to say that NCLB and other accountability policies **had a positive effect on 4<sup>th</sup> grade mathematics achievement and, sometimes, possibly improved 8<sup>th</sup> grade mathematics**.
- There is **no consistent evidence that accountability policies improve reading achievement**. Further, policymakers must consider the opportunity costs associated with these minimal score improvements.

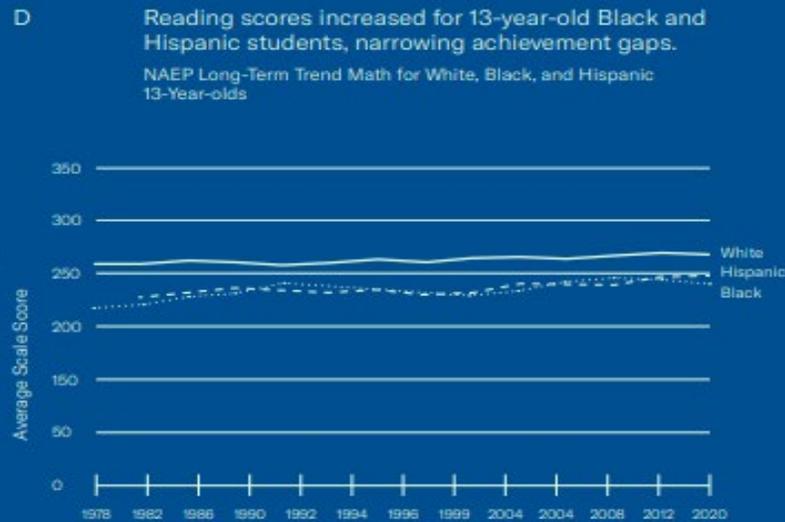
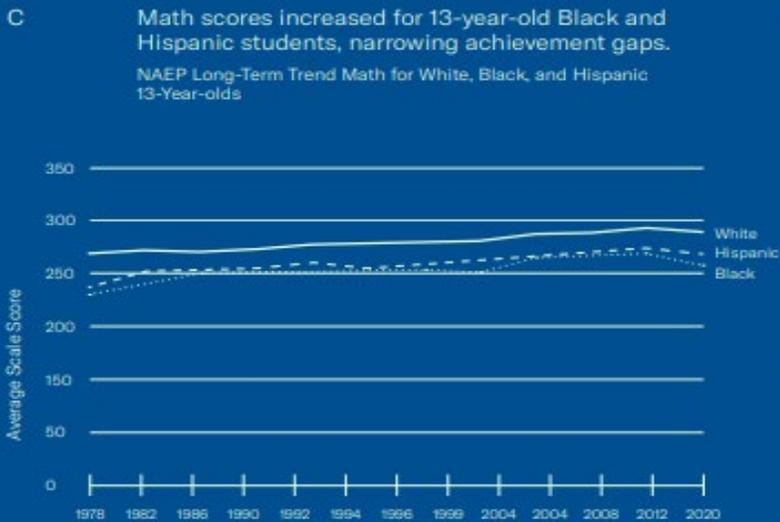
Figure 3 Long-Term NAEP Trends in Math and Reading by Race/Ethnicity



Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) Long-Term Trend Mathematics Assessments.



Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) Long-Term Trend Mathematics Assessments.



## Performance of Student Groups

- Based on NAEP Long-term Trend data
- Some conclude that these charts show the positive effects of NCLB and consequential accountability
- Others question the very slow pace of change
- A little bit of “NAEPery”

# It Didn't Work

---

- Or, more accurately, it didn't work enough to justify the negative consequences given the modest achievement gains.
- Let's take a look at the achievement issues before turning to some unintended consequences...

# Did we meet the achievement goals?

Of course, Bob Linn provided fair warning back in 2003 when he coined the concept of “ambitious but reasonable expectations” and noted that NCLB had the first part but not the second.

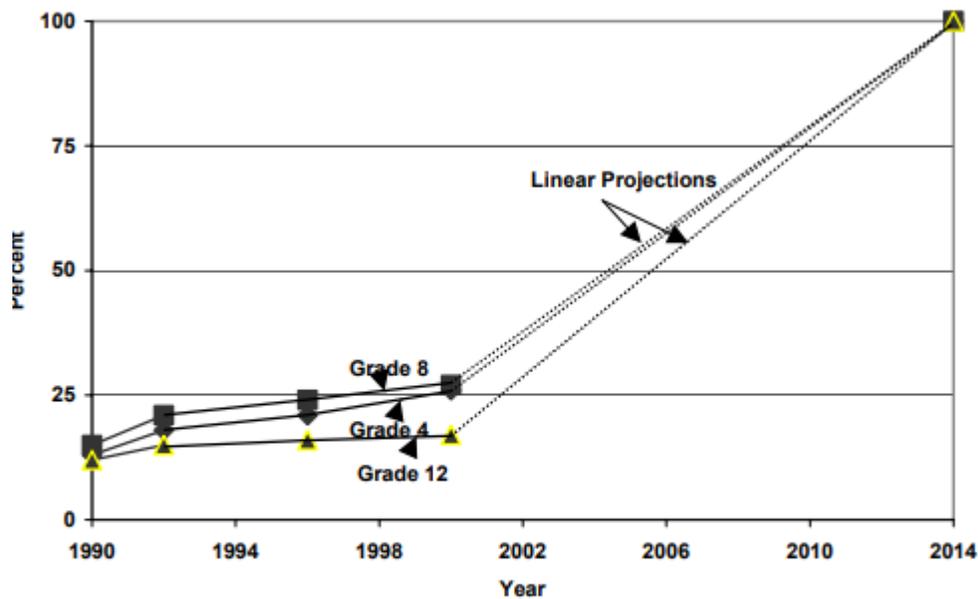


Figure 1. NAEP Mathematics trends with projections to 2014: percent proficient or above.

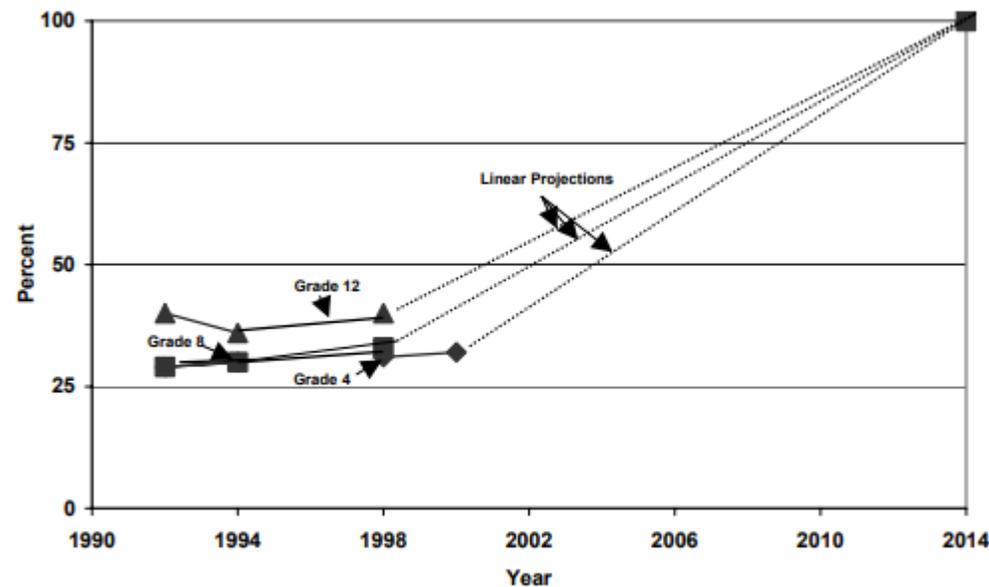
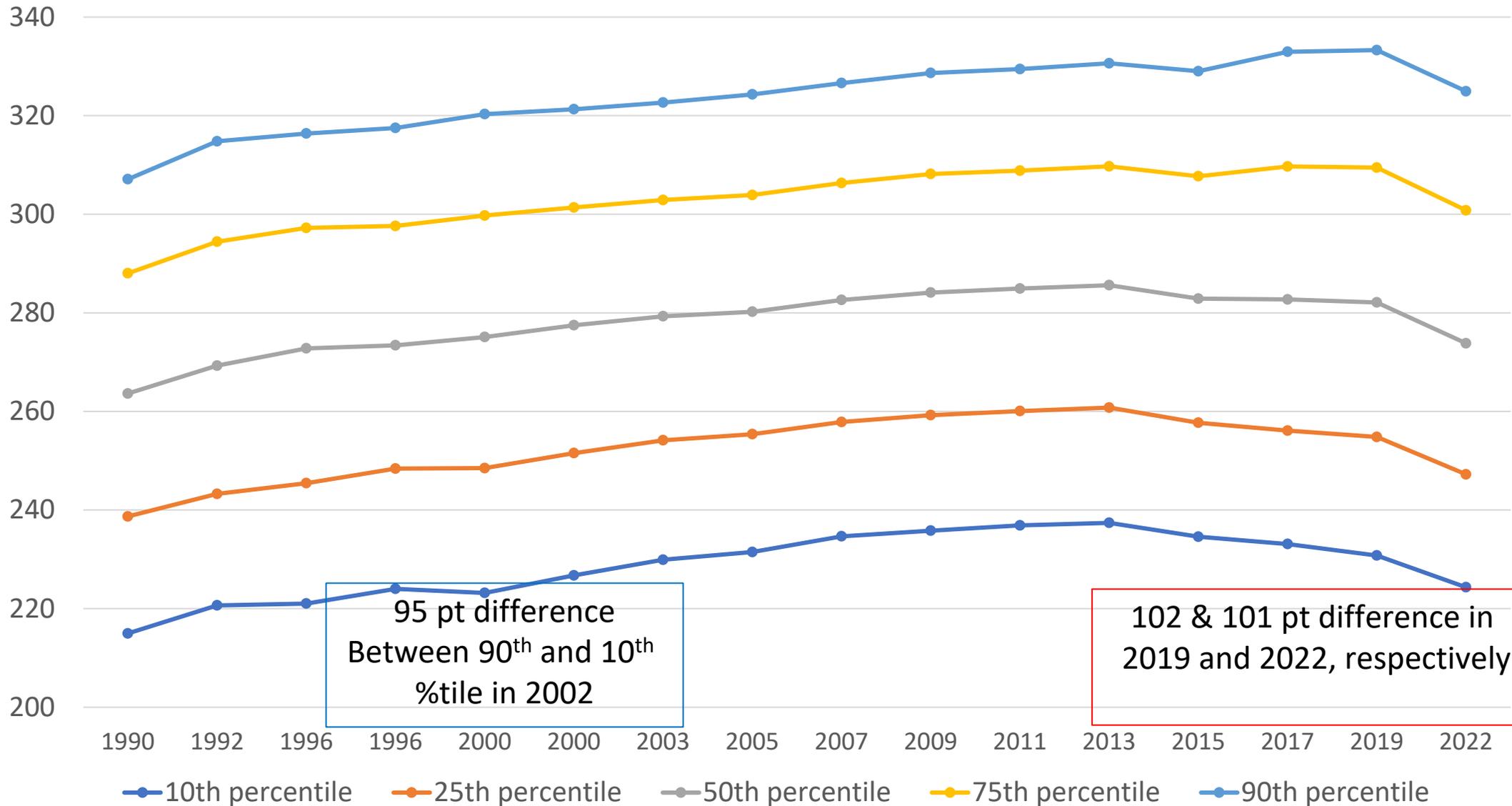


Figure 2. NAEP Reading trends with projections to 2014: percent proficient or above.

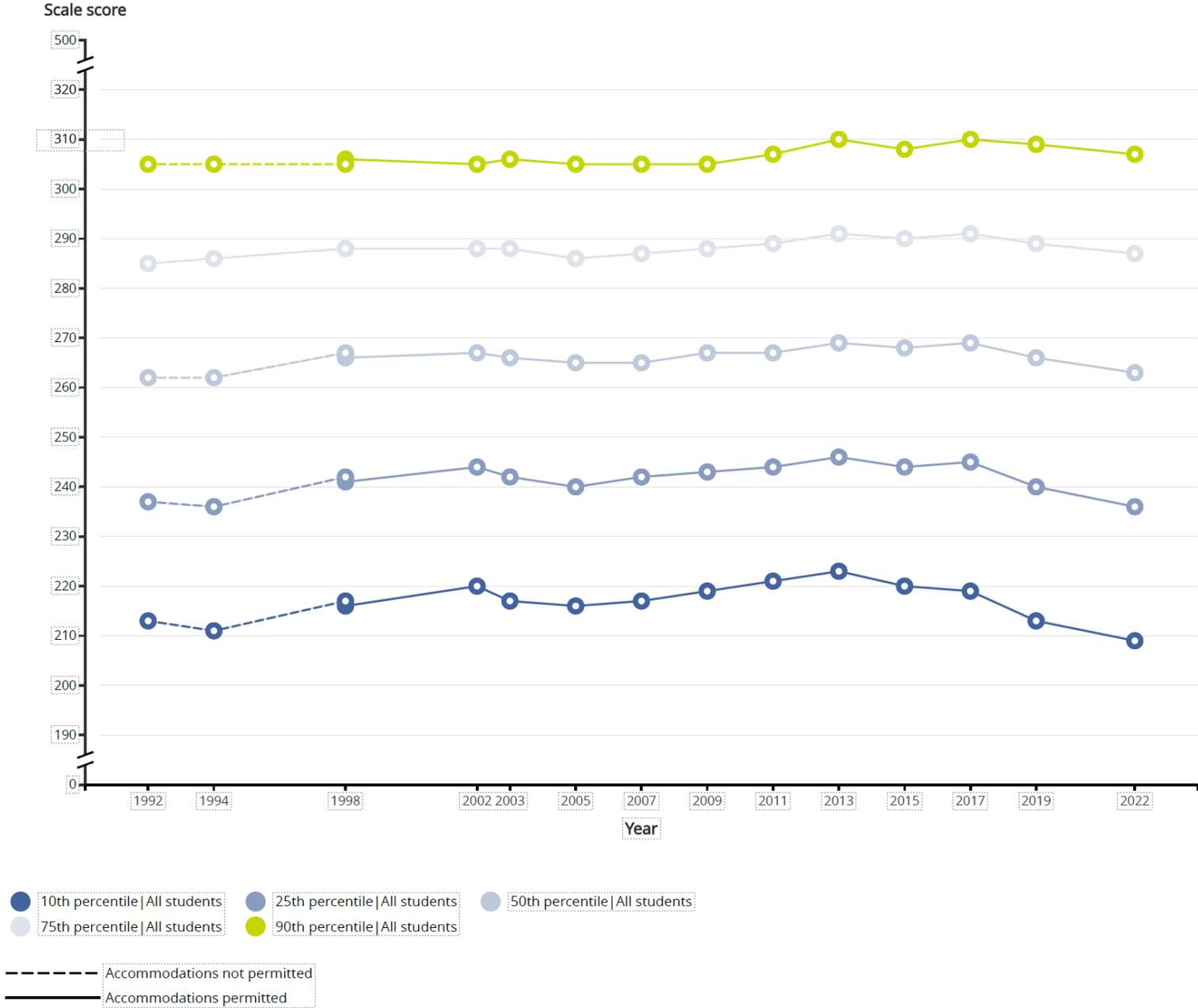
Linn, R. L. (2003). Accountability: Responsibility and Reasonable Expectations. *Educational Researcher*, 32(7), 3–13.

<https://doi.org/10.3102/0013189X032007003>

# Main NAEP Mathematics Grade 8 Scores by Percentiles



# Main NAEP 8<sup>th</sup> Grade Reading Trends by Percentile since 1992



# Key Unintended Consequences

---

- Lorrie Shepard, Chris Saldaña, and I (in press) just wrote a chapter, “Standards-Based Reform and School Accountability,\*” where we conclude that despite the evidence just discussed, the negative consequences far outweigh any of these gains.
  - Curriculum Narrowing and Test-Score Inflation
  - Stress and Deprofessionalization of Teaching
  - Deficit Views of Students and Dead-End Placements
  - Proliferation of unproductive testing

\*In Cohen-Vogel, L., Scott, J., & Youngs, P. (eds) *Handbook of Education Policy Research, Volume 2*. Washington, DC: American Educational Research Association.



## It wasn't all bad

- Most notably, NCLB and the Individuals with Disabilities Education Act (IDEA) caused a dramatic increase in participation in the general education curriculum and statewide assessments for students with identified disabilities.
  - The same eventually became true for English learners.
- And graduation rates improved considerably overall and for all student groups.
- These are really good things.

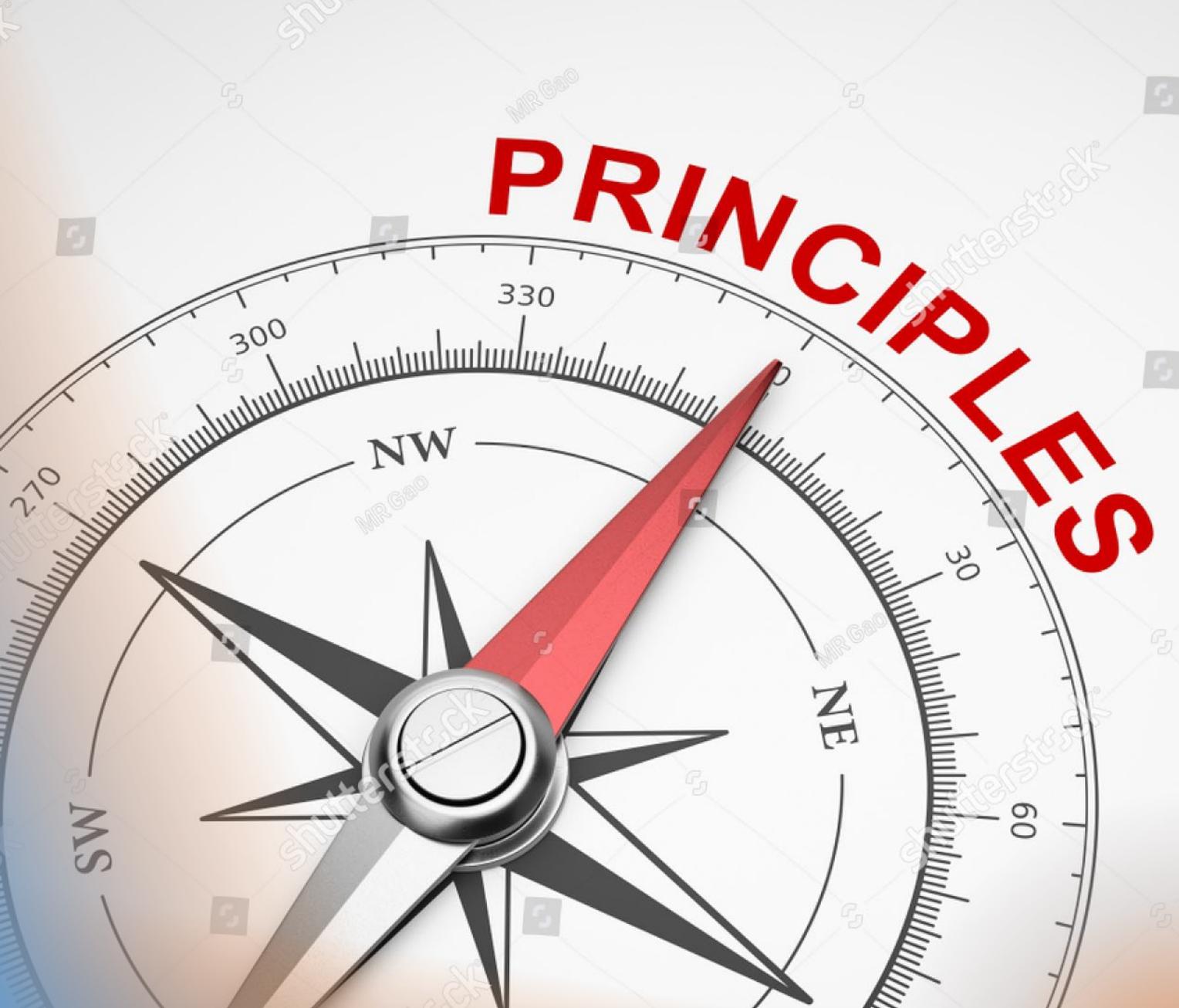
# Can we do better?

---

# My Ideas

---

- Remember the part about cringing...well, in the few minutes I have left, I want to share some of my ideas for improving accountability.
- First, given the almost \$80 billion federal investment in Title I (and that's just one program), there needs to be some form of accountability.
  - We'd expect that of any other federal program



# PRINCIPLES

## Guiding Principles

---

- Clear goals
  - Research-based
  - Coherence
  - Reciprocal
  - Flexible
  - Useful
- 
- Based on my recent blog:  
<https://www.nciea.org/blog/school-accountability-is-broken/>

# Clear Goals

---

Maximizing equity and social justice

Maximizing school system functioning to support meaningful opportunities to learn for all students and adults



# Drive

The Surprising  
Truth About What  
Motivates Us

'Provocative and  
fascinating'

Marion, RILS, September 21, 2023

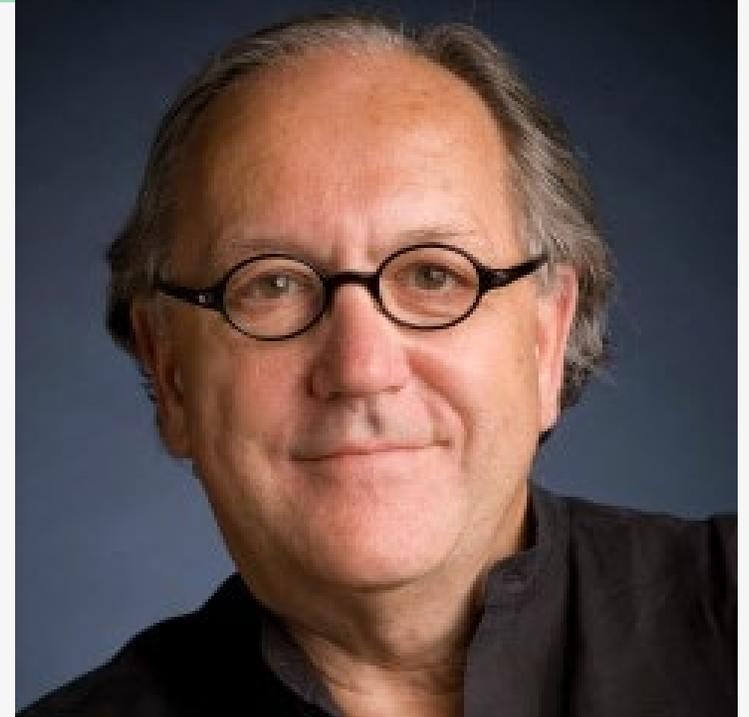
## Research-based

---

- Daniel Pink argues that people are motivated to improve when they have a sense of:
  - **autonomy** (the ability to direct our own lives),
  - **mastery** (the urge to get better and better at something that matters), and
  - **purpose** (the yearning to do what we do in the service of something larger than ourselves).
- Our current top-down accountability systems run counter to all three of Pink's elements of motivation.

## Elmore's Wise Words (2004)

*It is difficult to imagine an organizational form that is any less adapted to the demands of consistent, high-level engagement of students and teachers around content in the ways described above. Schools, in their modal form, are designed to buffer teachers from virtually any interference in the academic core (p. 288)*



Elmore, R. F. (2004). Moving forward: Refining accountability systems. In Fuhrman, S. H. & Elmore, R. F. *Redesigning accountability systems for education* (pp.276-296). New York, NY: Teachers College Press

## Elmore (2004, p. 288)

---

*It is absolutely essential to understand that when policies lay down stakes on incoherent organizations, the stakes themselves do not cause the organizations to become more coherent and effective.... If the schools had the assets in advance of the stakes, they presumably would not need the stakes to mobilize them. In this context, stakes make no sense as policy instruments unless they are joined in some systematic way with assistance that is designed to create the organizational assets that are required to respond to the stakes. In the absence of this kind of assistance, most schools and systems will respond within the constraints of their existing assets, which are, by definition, inadequate to respond to the task (p.288).*

# Coherence

---

- Many are rightfully pushing for more personalized educational approaches based on advances in human learning and development and a growing understanding of the importance of culturally responsive pedagogy.
- If we agree—and we might not—that we need to move away from a factory model of education, then why would we expect all the adults in the system to be subject to the same assessment and accountability requirements?





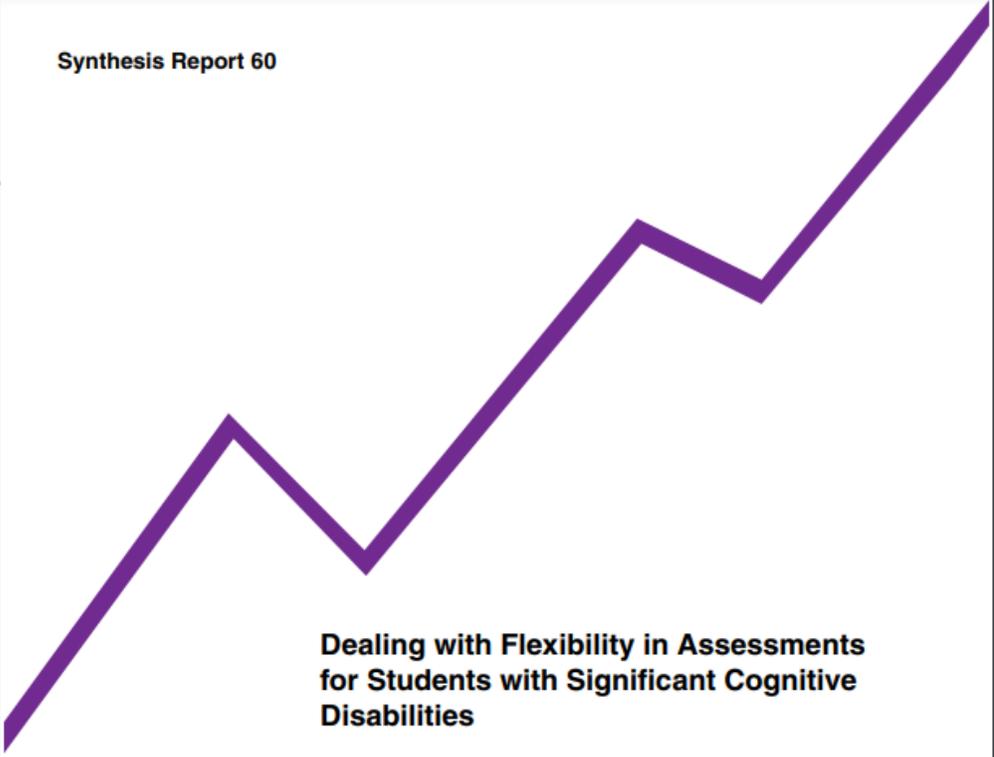
## Reciprocal

- Drawing (again) on Elmore's concept of reciprocal accountability, **I never understood why schools have been the sole (essentially) focus of accountability.**
- I'm drawing on my roles in a state DOE and my nine years as a school board member.
- State and especially districts have a lot to do with the degree to which schools can provide rich opportunities to learn.
  - If they don't, then what are they doing?
- Many of our case studies address reciprocity.

# Flexible

- If our goal is to learn about how best to improve schools' ability to enhance student learning and development, requiring all states, districts, and schools to do essentially the same things seems to run counter to a learning mindset.
- I'm not saying there's nothing in common or standardized, but we need to loosen the reins on comparability a bit.

Synthesis Report 60



**Dealing with Flexibility in Assessments  
for Students with Significant Cognitive  
Disabilities**

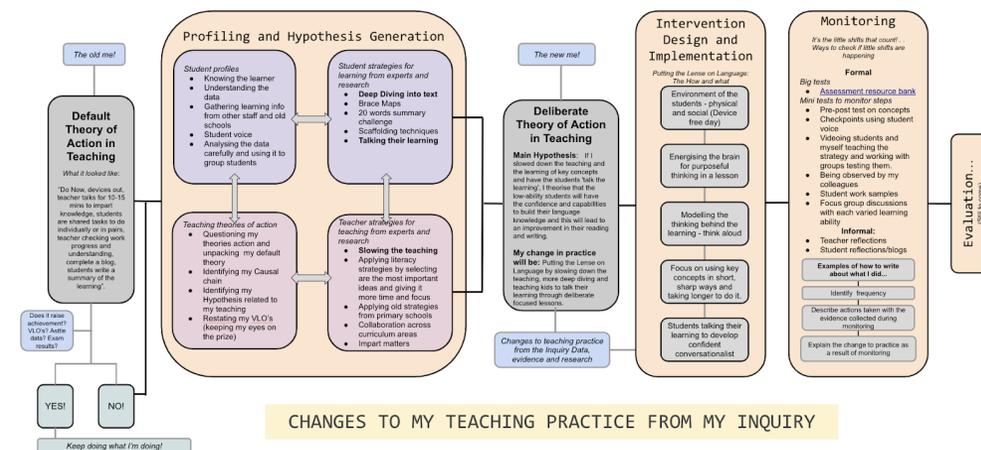


*In collaboration with:*

Council of Chief State School Officers (CCSSO)  
National Association of State Directors of Special Education (NASDSE)

# Useful

- It's no secret that most of us at the Center are a bit obsessed with theories of action.
- Why? Because theories of action are critical for clarifying design ideas, but just as importantly, they serve as frameworks for evaluation and continuous improvement.
- Why would we keep doing something if we don't have any evidence that it is working?



# What's next?

---

- Carla and Laura will lead you through an exercise to clarify the problems you might be trying to address.
- Chris<sup>2</sup>, Brian, Juan, and Laura will share insights from their recent paper on what you can do now to improve accountability systems.
- After lunch, we will engage with five sets of terrific guests about the work they are doing to improve accountability.
- Juan and Chris B. will help us think about evaluation and continuous improvement.