Accountability and Achievement, NCLB to ESSA

Thomas S. Dee

Education and Inequality in the 21st Century America

May 19, 2016

Can Federal Education Policy Reduce Educational Inequality?

Sometimes…

Overview – Achievement under Federal Reforms

- The No Child Left Behind (NCLB) Act (2002 to ~2012)
- Stimulus-funded school turnarounds (2010-2013)
- Gap-specific school reforms under NCLB Waivers (~2012-2016)
- Looking ahead to the "Every Student Succeeds Act (ESSA)" (2017-?)

What was NCLB?

- Annual testing of all students in grades 3-8 in math and reading and at least once in grades 10-12
- States developed their own tests and proficiency standards
- Public reporting of performance at the school level, *including results for individual subgroups*
- Schools required to make Adequate Yearly Progress (AYP) – 100% prof. by 2014
- Escalating supports and sanctions ("corrective actions") for failing to make AYP

The Achievement Consequences of NCLB

- NCLB brought test-based school accountability to scale across the U.S. simultaneously

- A National Research Council (2011) report on test-based incentives employed 3 filters for studies of NCLB (and its predecessors), excluding…
  - studies without a comparison group (including a US DoED evaluation)
  - cross-sectional studies that don't account for selection into program
  - regression-discontinuity (RD) studies that examine comparative effects of sanction risk but not the "whole" effect of the policy

- NRC (2011): "Our preferred estimate for these programs is 0.08 standard deviations, reflecting the national results for both the pre-NCLB period by Lee (2008) and the NCLB period by Dee and Jacob (2011)."

The Achievement Consequences of NCLB

- How can we credibly identify the effects of a policy like NCLB that came into effect for every school at the same time?

- Dee and Jacob (2011) leverage the fact that many states had NCLB-like (i.e., consequential) accountability prior to 2002
  - In other states, NCLB implied a *de novo* experience with consequential accountability

- Wong, Cook, and Steiner (2015): similar design with emphasis on heterogeneity by *ex-post* proficiency standards; also public-private school comparisons

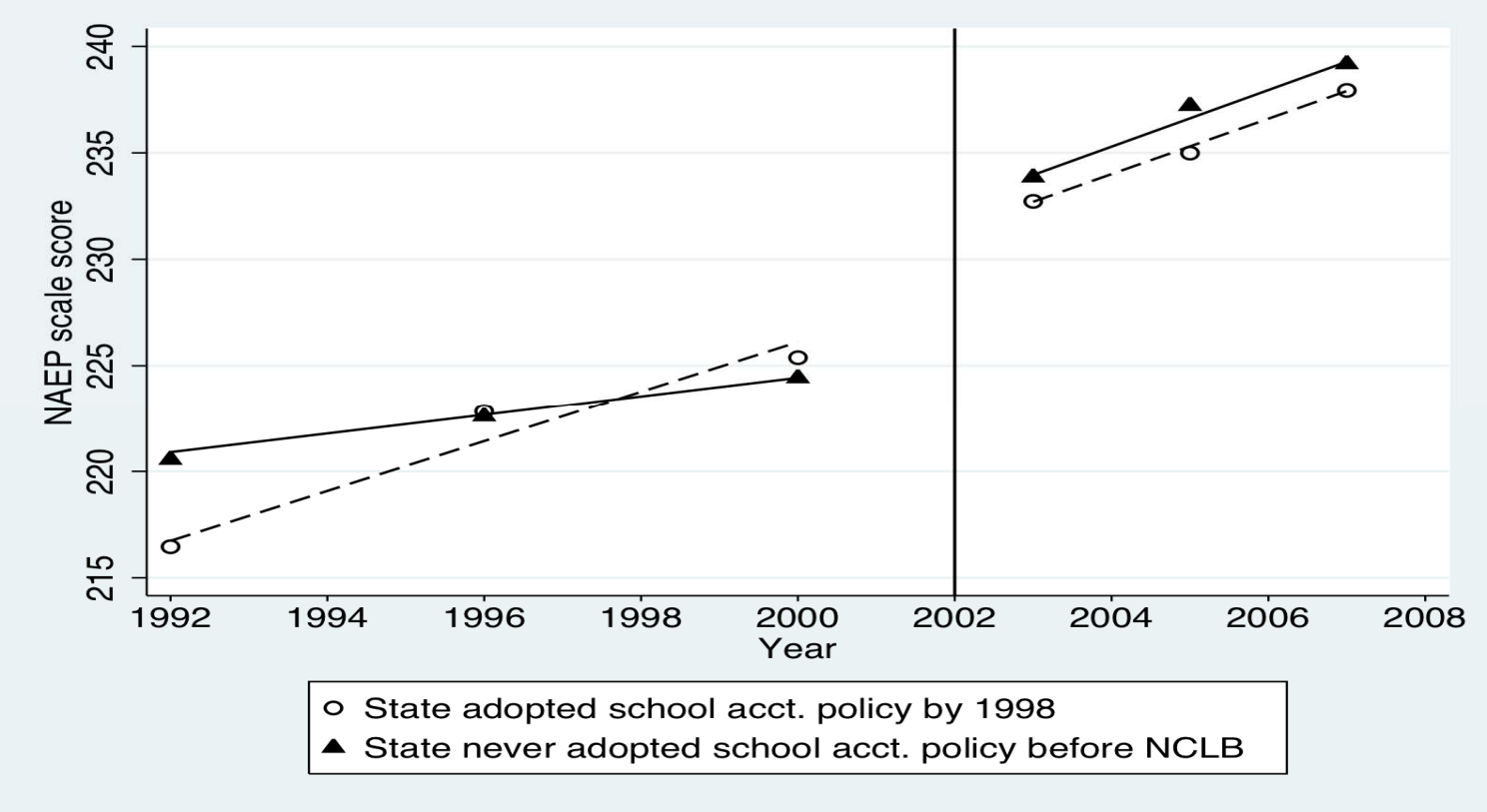
Dee and Jacob (2011)

- Achievement data at the state-year-subject-grade level from main NAEP (1992-2007)

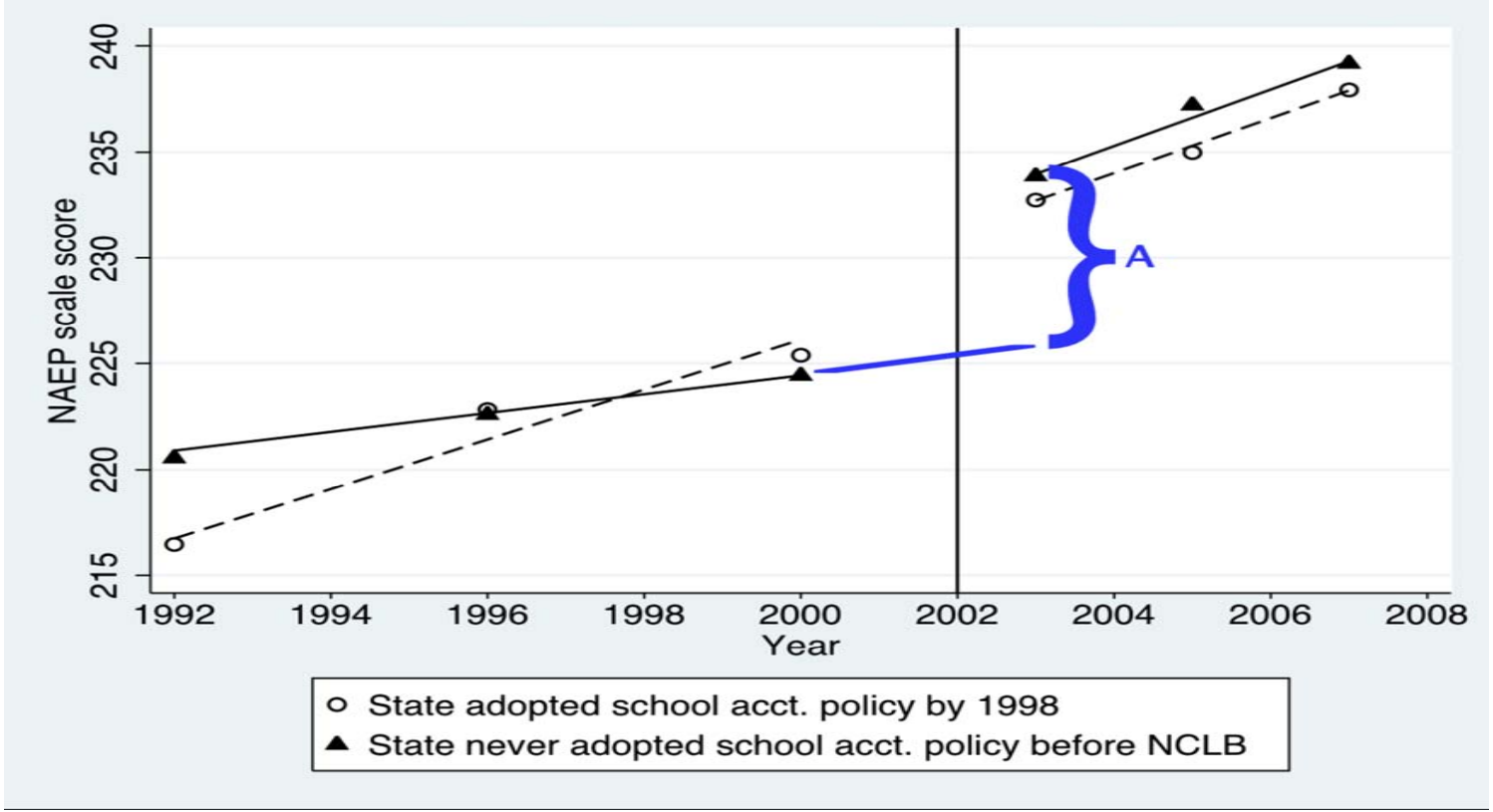
- NAEP is low-stakes test but limited coverage prior to NCLB (37 to 39 states with 2+ pre-NCLB observations)

- "Comparative Interrupted Time-Series (CITS)" design
  - Related to panel-based "difference in difference" (DD) design
  - Accommodates differential pre-trends among "control" states (i.e., those with pre-NCLB accountability) and "treatment" states (i.e., those without pre-NCLB accountability)

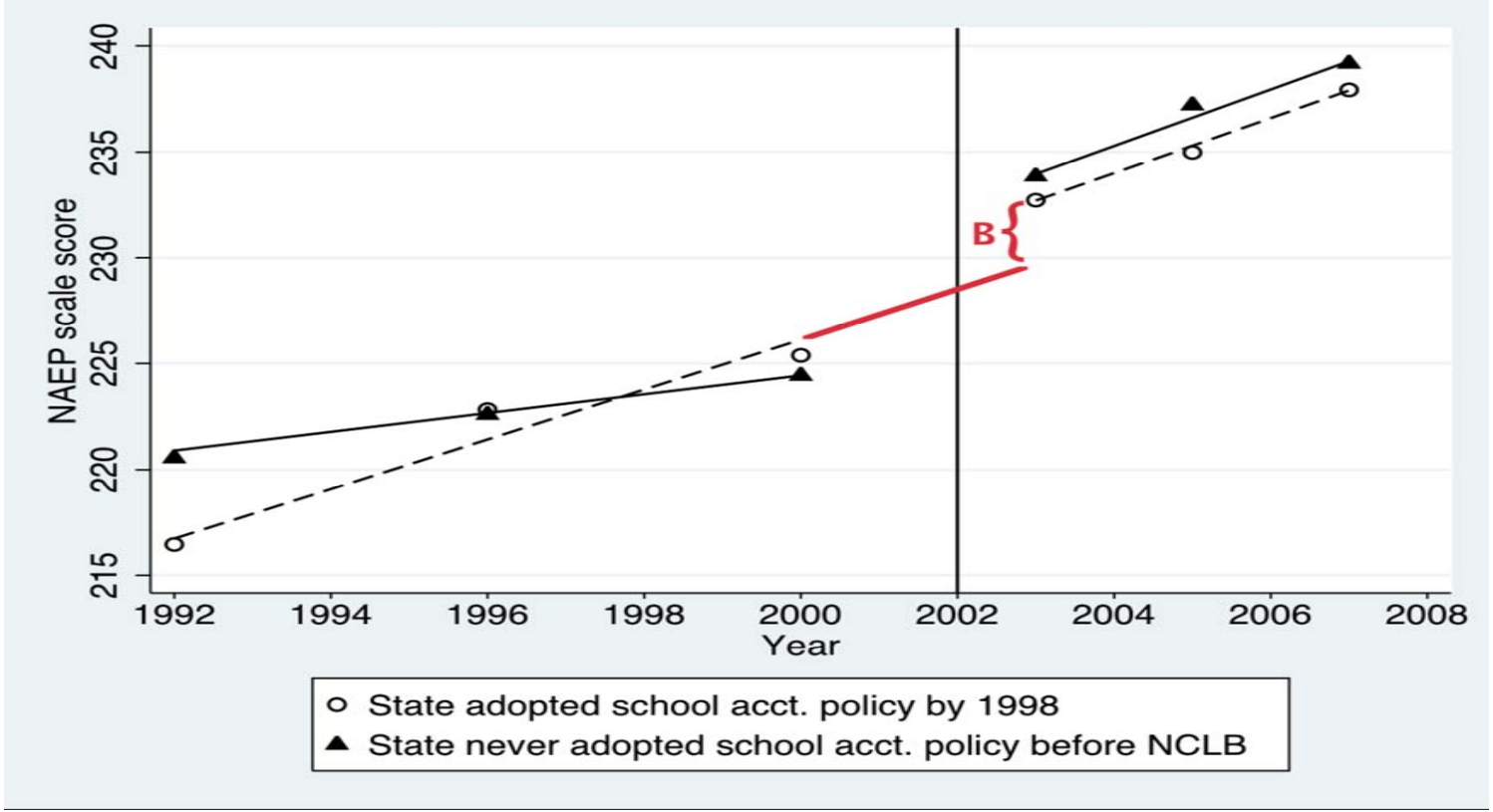
NAEP 4th Grade Math Trends by Pre-NCLB Accountability



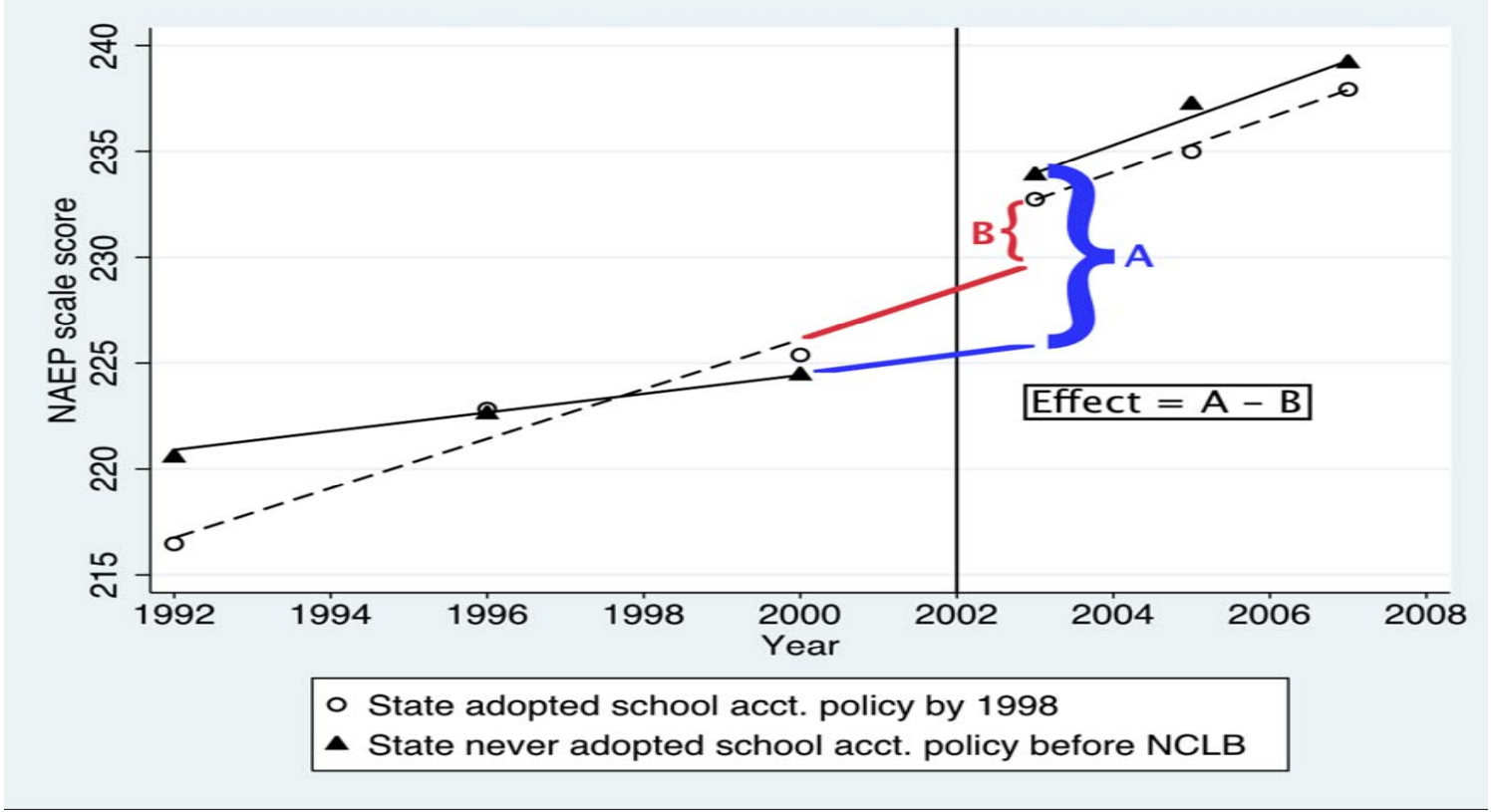
NAEP 4th Grade Math Trends by Pre-NCLB Accountability



NAEP 4th Grade Math Trends by Pre-NCLB Accountability



NAEP 4th Grade Math Trends by Pre-NCLB Accountability



Estimated Effects of NCLB by Grade, Subject, and Subgroup

	Effect Size
Subject	Overall
Grade 4 Math	0.23*
Grade 8 Math	0.10
Grade 4 Reading	0.06

Estimated Effects of NCLB by Grade, Subject, and Subgroup

	Effect Size	
Subject	Overall	White
Grade 4 Math	0.23*	0.16*
Grade 8 Math	0.10	0.05
Grade 4 Reading	0.06	0.15*

Estimated Effects of NCLB by Grade, Subject, and Subgroup

	Effect Size		
Subject	Overall	White	Black
Grade 4 Math	0.23*	0.16*	0.47*
Grade 8 Math	0.10	0.05	0.23
Grade 4 Reading	0.06	0.15*	-0.02

Estimated Effects of NCLB by Grade, Subject, and Subgroup

	Effect Size			
Subject	Overall	White	Black	Hispanic
Grade 4 Math	0.23*	0.16*	0.47*	0.32*
Grade 8 Math	0.10	0.05	0.23	0.22*
Grade 4 Reading	0.06	0.15*	-0.02	0.01

Summing up NCLB

- Effects on achievement that are meaningful (roughly a whole year of learning in math) but targeted and not transformational (e.g., largely redundant in many states)

- NCLB contributed to declines in achievement gaps in states with no prior school accountability

- CITS evidence on mediators (Dee, Jacob, and Schwartz 2013):
  - Increased spending from state/local sources ($600 per pupil)
  - Higher teacher salaries but no class size reductions
  - No change in overall instructional time but modest reallocation of time from science and social studies to math and reading

Introduction to School Turnarounds

- June 22, 2009: Arne Duncan calls for a nationwide focus on "turning around" chronically underperforming schools (i.e., the lowest 5 percent)
  - "We want transformation, not tinkering"

- The American Recovery and Reinvestment Act (ARRA) of 2009
  - $3 billion for redesigned School Improvement Grants (SIGs) to support this effort
  - New US DoED guidance targets prioritized SIG eligibility to "*persistently lowest-achieving*" (PLA) schools
  - SIG awards increased to a maximum of $2 million per school annually for 3 years
  - But SIG recipients *required* to implement one of three, highly prescriptive reform models (transformation, turnaround, restart) or to close

- Year 1 and 2 results (2011-12 and 2012-13) from SIG-funded reforms in CA

The Transformation Model

- Teacher and principal effectiveness (e.g., principal replacement, staff evaluations, PD)
- Comprehensive instructional reform (e.g., data use and differentiated instruction)
- Extended learning time, longer school day and year
- Operational flexibility, technical assistance from district, state and/or outside providers
- Socio-emotional & community services (e.g., health, nutrition, social services)

Other Federal Models

- The *turnaround* model is similar to the transformation model but requires replacing at least 50% of the school's prior staff

- The *restart* model requires reopening under the management of a charter school operator, a charter management organization, or an educational management organization.

- "Transformation" is commonly characterized as the "least disruptive" of the federally prescribed models (Hurlburt et al. 2011)
  - Nationwide, 74% of Tier 1/Tier 2 SIG recipients chose transformation
  - 20% chose turnaround, 4% chose restart (n = 33) and 2% (n = 16) chose closure

Theories of Change?

- Chronically underperforming schools serving students in concentrated poverty suffer from multiple, deep-rooted, self-reinforcing problems
  - Weak leadership, ineffective instructional practices, poor working conditions, high turnover
  - Genuinely effective change has to be quick, dramatic, and extensive rather than marginal and targeted (operational authority, supports, and monitoring)

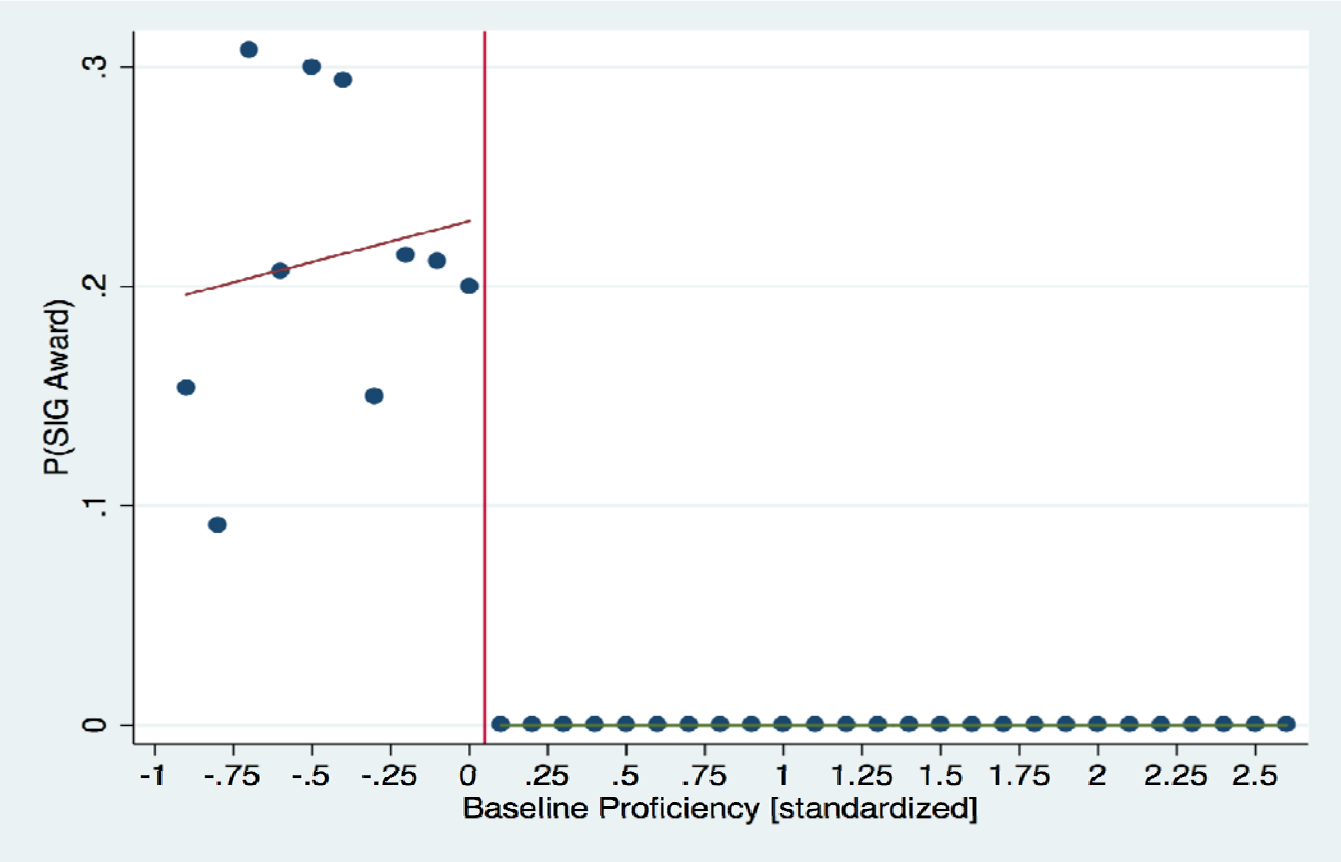
- Unintended consequences of top-down, highly prescriptive reforms?
  - "Counterproductive micromanagement" (Darling-Hammond and Hess 2011). Weak buy-in? Low-quality implementation? Actively disruptive?
  - Are these concerns attenuated by new school leadership and state/federal monitoring?

Evidence from California

- 3,652 schools (out of 9K) were in the "Tier 1" and "Tier 2" pool
  - Tier 1: Title I schools in improvement, corrective action, or restructuring
  - Tier 2: "secondary" schools eligible for Title I support

- Lowest-achieving 5% (i.e., n = 183) identified as "persistently lowest achieving" (PLA) and eligible to apply for a 2010-11 SIG
  - N = 92 Cohort 1 SIG awards made

- Key assignment rule: 3-year (2007-2009) math/ELA AYP proficiency rate below thresholds specific to school levels
  - Elementary: ≤ 29.97%, Middle ≤ 22.44%, High ≤ 37.31%



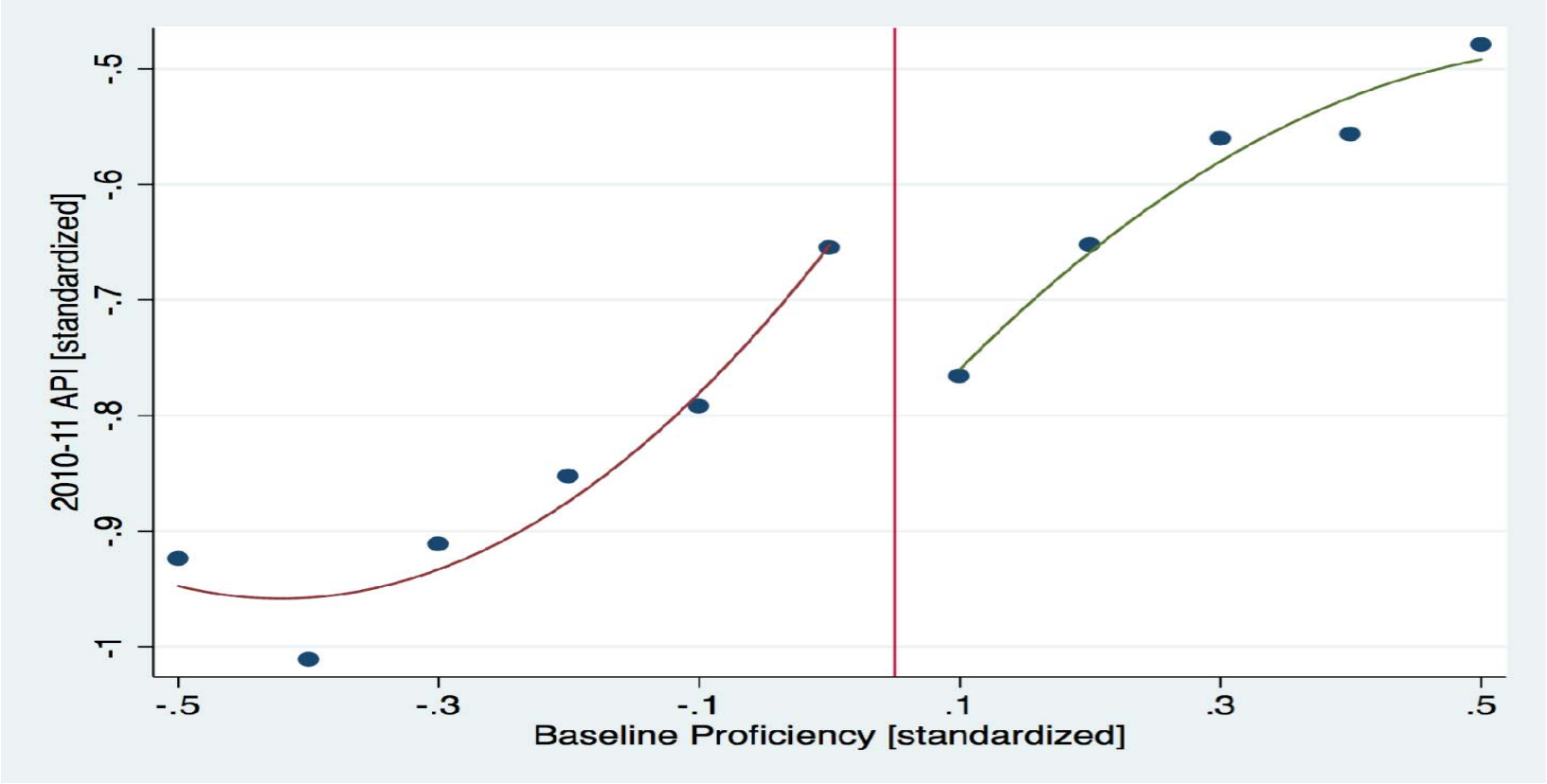
First Stage: SIG Awards and Baseline Achievement

Evidence from California

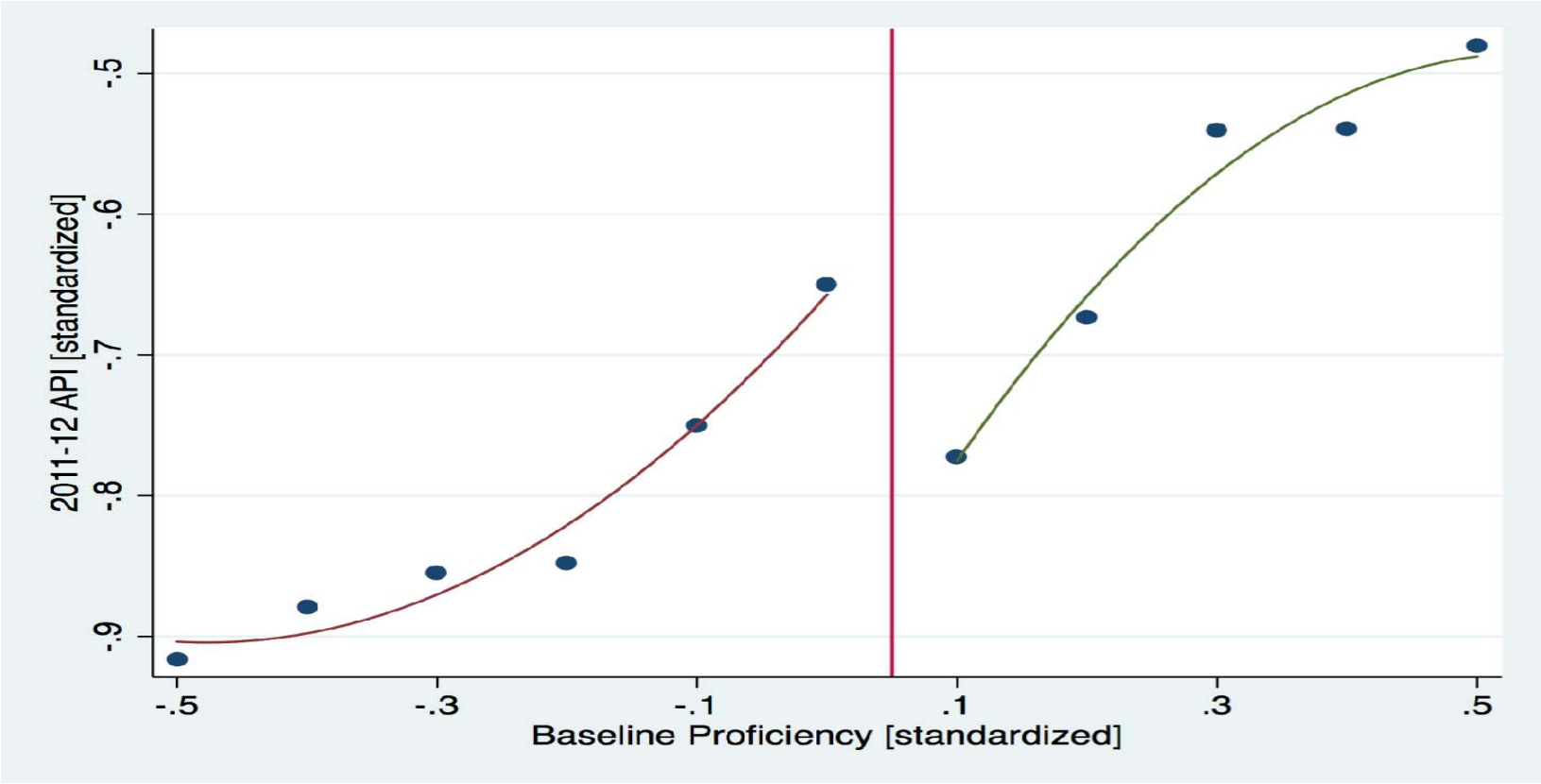
- Students served in PLA schools: 74% Hispanic, 13% black, 85% FRL eligible, 42% English-language learners

- School-level "Academic Performance Index" (API) as outcome
  - Weighted measure based on statewide testing (e.g., CSTs, CMAs, CAHSEE)
  - The "cornerstone of the state's accountability system"

- Results…



2010-11 API Scores (0.5 bandwidth)



2011-12 API Scores (0.5 bandwidth)

Summing up: effect sizes and cost effectiveness?

- Year 1 gains concentrated in turnaround schools; by year 2, gains in both transformation and turnaround schools

- Estimated effect of SIG-funded reforms: 34 scale-point API increase
  - 5.2% of mean, baseline API among SIG-eligible schools (650)
  - 23% of average gap between lowest-achieving schools (650) and state goal (800)

- First-year SIG results: 0.3 gain w/r/t school-level SD
  - ~0.09 w/r/t student-level SD at a cost of $1,500 per pupil
  - A benchmark from Project STAR's class-size reductions: 0.2 SD gain for a 47% expenditure increase (approximately $5,000 per pupil)

Changes within SIG schools?

- Nationwide federal survey (NCEE 2014) of low-performing schools (both SIG and non-SIG)
- Few differences in school-level operational authority and school supports
- But SIG schools were significantly more likely to report:
  - Replacing the principal
  - Increasing learning time
  - New parental/community engagement strategies
  - Professional development on Common Core State Standards and turnarounds
  - Use of multiple performance measures for teacher evaluations (but still uncommon)

Implementation Challenges across States?

Education Week's blogs > Politics K-12

Politics K-12

Your education road map

« Santorum Suspends Campaign; Will Romney Start Talking K-12? | Main | Is Hawaii One Step Closer to Losing Race to Top Grant? »

GAO Criticizes Ed. Dept. On School Improvement Grant Program

By Alyson Klein on April 11, 2012 7:34 PM

- States report staff recruitment & retention challenges
- In 23 of 44 states, grants renewed *without* meeting annual goals
- Sustainability of extended learning time in doubt when grant expires

School Reforms under NCLB Waivers

- Congressional inaction on reauthorizing NCLB ceded authority to the executive branch
- In 2011, the US DoED began offering waivers to states that implemented the priorities of the Obama administration
- Waivers: College and career-ready standards, flexibility to define achievement goals but concentrated school improvement efforts in 15% of most troubled schools
- Priority Schools: lowest-achieving 5% to adopt federal turnaround principles
- Focus Schools: 10% with largest achievement gaps or lowest subgroup performance

The Effects of Focus School Reforms in 3 States

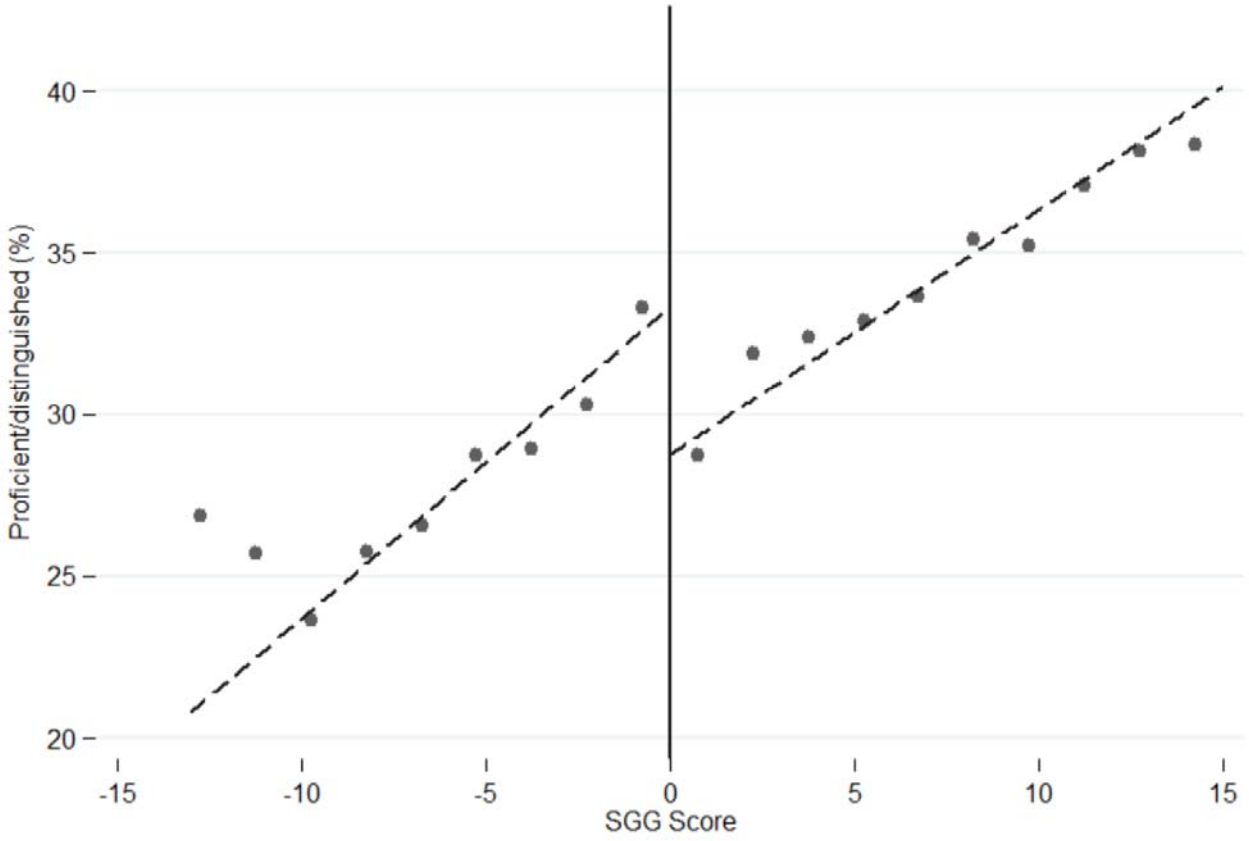
- Kentucky identified Focus Schools using a “super subgroup”. AY 2013-14 was first full year of implementation (Bonilla and Dee 2016).
- North Carolina identified Focus Schools using gap between each school’s highest/lowest subgroups. First full year of implementation in AY 2013-14 (Dee and Doss 2016)
- Louisiana (Dee and Dizon-Ross 2016) identified Focus Schools as those with lowest School Performance Scores (i.e., an “F” letter grade). Priority Schools are those in state-run Recovery School District (RSD)
- All 3 studies use regression-discontinuity (RD) designs to identify the causal effects of Focus School reforms

Focus School Implementation in KY

Focus School students using computer aided instruction. © Herald-Leader

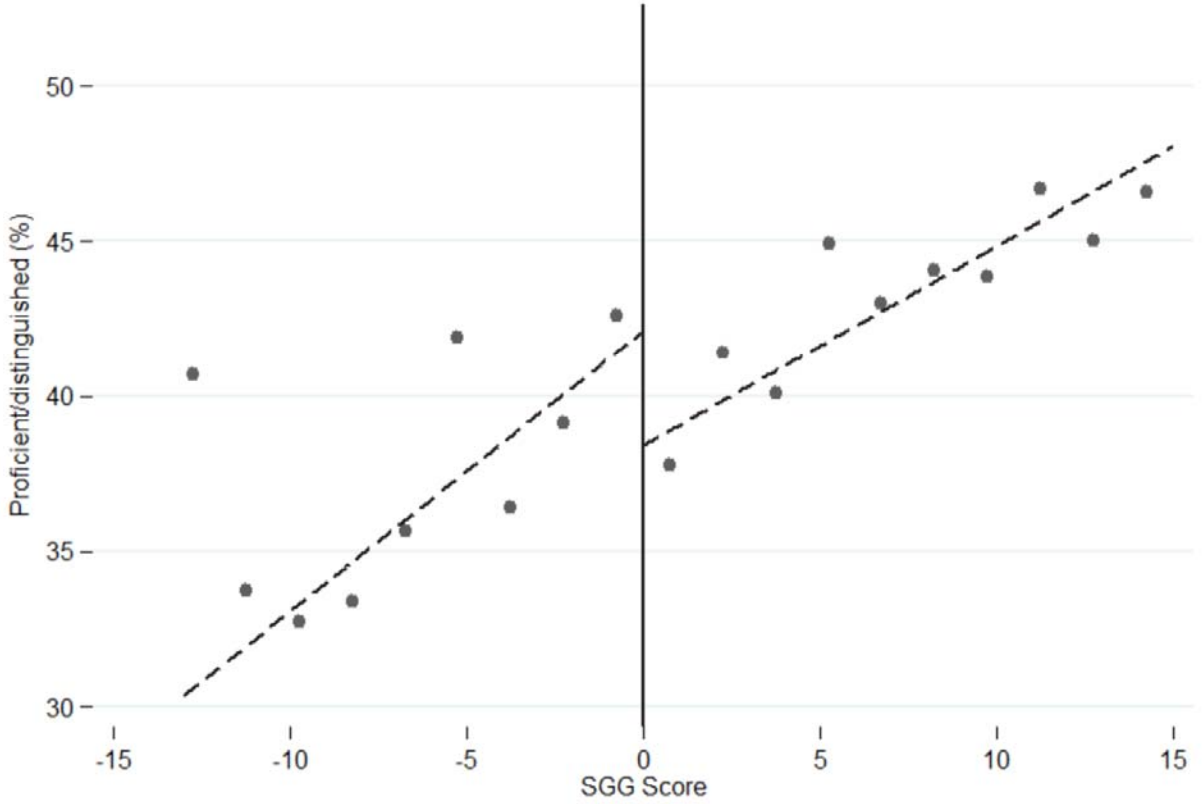
- Flexibility to implement Title I School-wide programs
- Teacher teams (PLCs) to collaborate & to use data
- Extended learning time using CAI
- Technical Assistance (planning & implementation)
- Comprehensive School Improvement Plans (CSIP)
- State Monitoring

Figure 4(a) 2014 Math Proficiency Rate, Gap Group, Fuzzy RD



- 7.5 percentage point increase in math proficiency
- ~25 percent gain relative to schools to right of threshold

Figure 4(c) 2014 Reading Proficiency Rate, Gap Group, Fuzzy RD



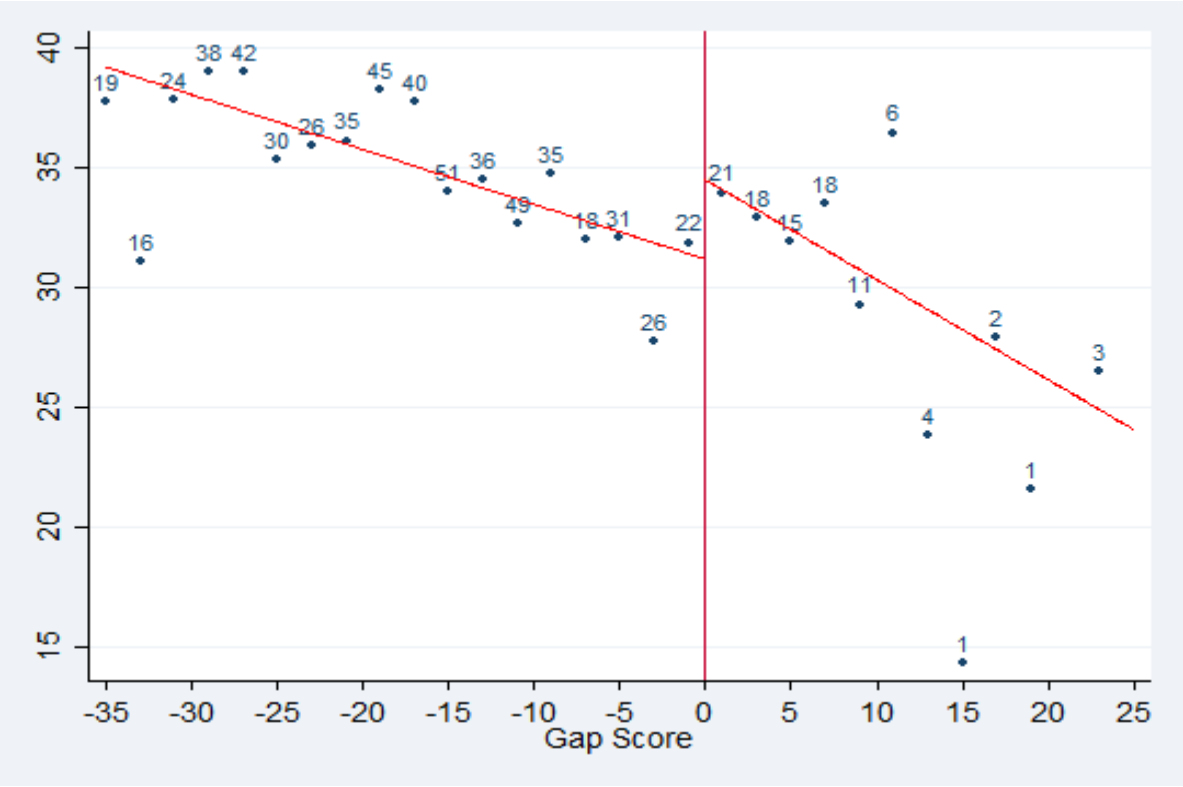
- 4.5 percentage point increase in reading proficiency
- ~12 percent gain relative to schools to right of threshold
- Teachers also report higher-quality PD and instructional supports

Focus School Implementation in NC

School needs assessment and revised improvement plans

School plans emphasize PLCs and teacher PD, extended learning time, and reforms to discipline policies

Many plans noted the *specific subgroup* for which they received Focus status



2013-14 Reading Proficiency, Black Students

- 3.7 percentage point increase in reading proficiency
- ~10 percent gain relative to schools to left of threshold
- Gains faded in 2014-15?
- Evidence of reductions for Hispanic students. Possible “triage” behavior?

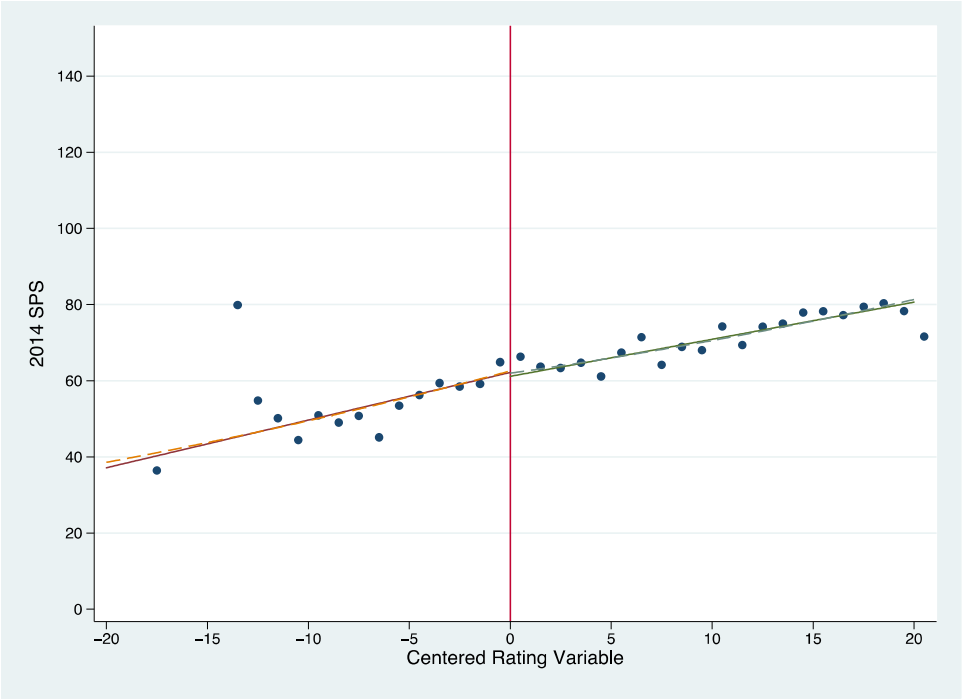
Focus School Implementation in LA

Data review and needs assessment?

Technical assistance focusing on Common Core implementation, teacher evaluation systems, and school choice?

Plus accountability pressure from “F” label?

US DoED monitoring found implementation “not meeting expectations”



2014 School Performance Scores

- No evidence of effects on school performance in 2014 or in 2015
- No evidence of effects on subject-specific schools or by school level

Reflections on an Era of Federal Activism

- The US system (100,000 public schools in 14,000 districts) is federalist with strong state and local control
- The federal role in influencing school policies and instructional practices is limited
- Despite this, federal policy has demonstrated a non-trivial capacity to catalyze reductions in the inequality of student outcomes
- However, federal policies appear effective in some states and districts and not others?
  - Why? Historical accident? Or are there traits of states and districts that systematically moderate the effects of federal policies?

Looking ahead to ESSA?

- Continued testing: math/ELA in grades 3-8; once in high school
  - Required subgroup reporting: EL, racial minorities, students in poverty
- States with autonomy to define more flexible accountability systems that also include non-test measures
  - Also, "challenging" academic standards but not necessarily CCSS
- Mandated identification and intervention in 15% of lowest-performing schools (i.e., Priority and Focus School reforms)
- In other 85% of schools, ineffectual "report card" accountability?
- Research agenda needs renewed attention to state/local implementation?

Thank you!

tdee@stanford.edu