School Turnarounds: Evidence from the 2009 Stimulus

THOMAS S. DEE
STANFORD GSE & NBER

PACE Seminar, December 13, 2013

Introduction

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 added to 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

THIS STUDY

- "Regression discontinuity" (RD) evidence on the early impact of SIG-funded reforms in California
  - 2nd-year results (AY 2011-12) presented for the first time today

The Broader Context – Why SIGs Matter

- An expensive federal initiative to make dramatic changes within the most struggling schools
- A novel addition to prior whole-school reform efforts (e.g., CSRs, SFA, DI, SDP, Title I School-wide programs)
- A leading example of similarly prescriptive, highly controversial federal reforms (e.g., Race to the Top, “Priority Schools” in NCLB waiver process)
- Part of a broader debate about the capacity of schools alone to be meaningful agents of social equality (e.g., “No Excuses” vs. “Broader, Bolder” initiatives)
- All combined with a research design that has some promise of a strong causal warrant (i.e., leveraging sharp, discontinuous assignment to SIG eligibility based on lowest-achieving criterion)

Federal guidance on SIG Eligibility

- States identify persistently lowest-achieving (PLA) schools → highest priority for SIG funding
- Two "tiers" of schools eligible for PLA status
  - Tier 1 candidates: Title 1 schools in improvement, corrective action, or restructuring
  - Tier 2 candidates: "secondary" schools eligible for Title I support
- Lowest 5 percent in baseline math/ELA achievement among otherwise eligible schools in Tier 1 & 2 pool → eligible for PLA status
- Lowest achievement growth → eligible for PLA status
- Other little-used mechanisms for PLA status: graduation-rate criteria & "newly eligible" status
- Lower-priority "Tier 3"schools are eligible for SIGs, no prescriptive reforms required (no Tier 3 awards made in CA)

SIG Eligibility in California

- 3,652 schools (out of ~9,000) were in the Tier 1/Tier 2 pool
- "Lowest Achieving" assignment rule: 3-year (2007-2009) math/ELA AYP proficiency rate below thresholds specific to school levels (~19% qualify)
  - Elementary: ≤ 29.97%, Middle ≤ 22.44%, High ≤ 37.31%
- "Lack of Progress" assignment rule: sum of API growth over five years (2005-2009) < 50 (~40% qualify)
- Other PLA eligibility requirements: (1) Baseline API < 800 and (2) n-size requirement for AYP calculations
  - These are candidate RDs but underpowered
- 5% of original 3,652 schools (i.e., n = 183) identified as PLA, eligible to apply for a 2010-11 SIG
  - N = 92 Cohort 1 SIG awards made

Federally Prescribed School Reforms

- The widely used *transformation* model has several key features
- (1) Teacher and principal effectiveness
  - Replacing the principal
  - Staff evaluations based in part on student performance and used in personnel decisions
  - Embedded professional development
- (2) Comprehensive instructional reform: aligned vertically and to state standards, continuous use of data to inform & differentiate instruction
- (3) Extended learning time, longer school day and year
- (4) Operational flexibility, technical assistance from district, state and/or outside providers
- (5) Socio-emotional & community-oriented services (e.g., health, nutrition, social services)

Federally Prescribed School Reforms

- 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
- Nationwide, 74% of Tier 1/Tier 2 SIG recipients chose transformation; 20% chose turnaround (Hurlburt et al. 2011)
  - 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

IMPLICIT ASSUMPTIONS ABOUT UNDERLYING "MARKET FAILURES"?

- Imperfect information: staff cannot easily identify effective practices and have underpowered incentives because of imperfect monitoring
- Public goods: productivity-enhancing norms and supports around instructional practice, staff collaboration, shared organizational purpose (social K) are underprovided collective goods

UNINTENDED CONSEQUENCES OF TOP-DOWN, HIGHLY PRESCRIPTIVE REFORMS?

- "Counterproductive micromanagement" (Darling-Hammond and Hess 2011). Weak buy-in? Low-quality implementation? Actively disruptive?
- Or are these concerns attenuated by new leadership and some prescriptive changes that are easily monitored (e.g., extended learning time, staff performance evaluations)

Evaluating SIG-funded School Reforms in California

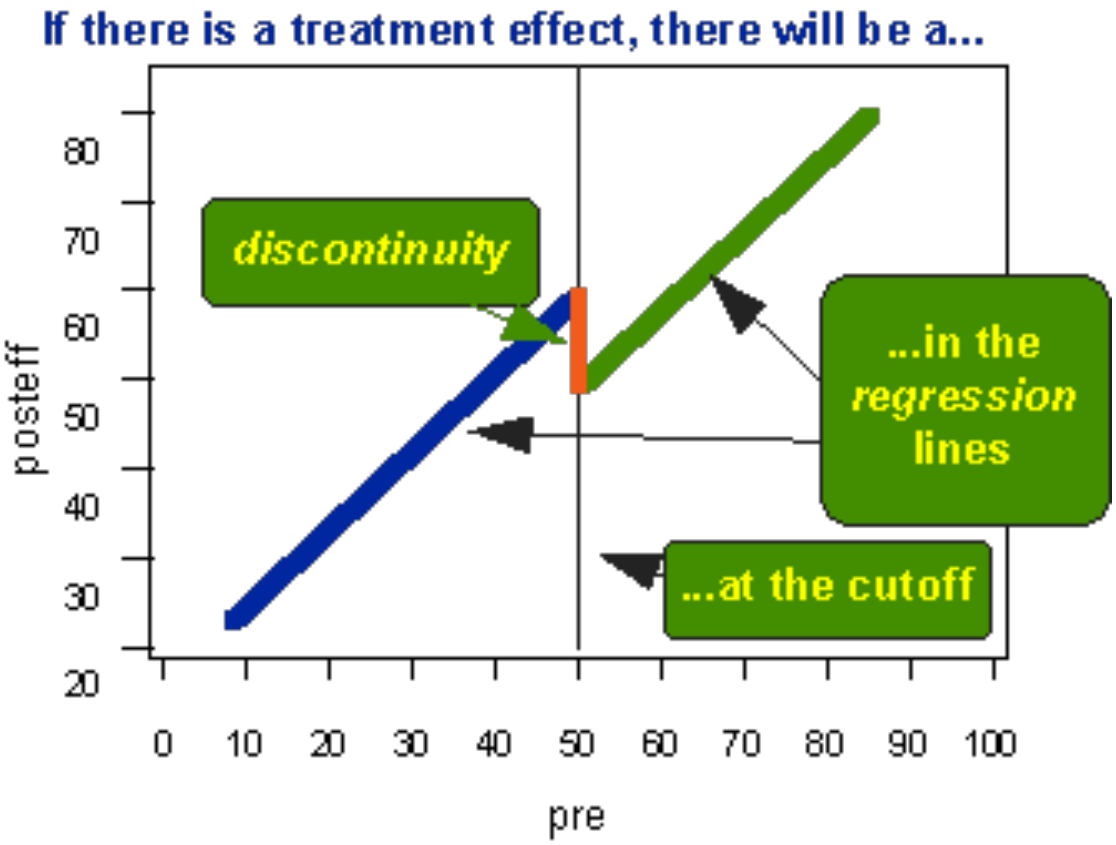
- A mix of encouraging and cautionary anecdotal evidence…

Grants make big difference at some East Bay schools, while others that got money continue to struggle

By Theresa Harrington Contra Costa Times Contra Costa Times

- Descriptive evidence is useful but doesn't provide convincingly causal evidence on the effects of these reforms
- It is possible to implement a "regression discontinuity" (RD) design that does have a strong causal warrant
- RD designs have long been understood as a program evaluation technique (Campbell and Thistlewaite 1960)
  - › New and expansive interest among applied policy researchers over the last 10 years
- RD designs support causal inference by leveraging discontinuous rules for assigning subjects to treatments…

A Quick Primer on RD Designs



-Students with “pre” scores < 50 assigned a treatment (blue line)

-Students with scores at 50 or higher receive no treatment (green line)

-Do *post-treatment* outcomes “jump” at the T/C threshold?

Analytical sample and covariates

N=3,652 SCHOOLS IN THE TIER 1 AND TIER 2 POOLS

- Eliminate n=588 non-standard schools (e.g., continuation schools, juvenile court schools)
  - › Most are missing API scores and SIG-ineligible
- Eliminate 38 special-education schools, 120 charter schools, 3 closed schools, 156 schools without available baseline data

ANALYTICAL SAMPLE OF 2,747 SCHOOLS (TABLE 1)

- 6.1% are PLA schools (n=167), 3% (n=81) received SIG awards
- 47 transformations, 27 turnarounds, 7 restarts

SCHOOL-COVARIATES FOR BOTH AY 2009-10-AY 2011-12 (TABLE 1)

- Students (% race-ethnicity, FRL, EL, disability status)
- Teachers (experience, graduate degree, race-ethnicity)
- Schools (urbanicity, level, enrollment, pupil-teacher ratio)

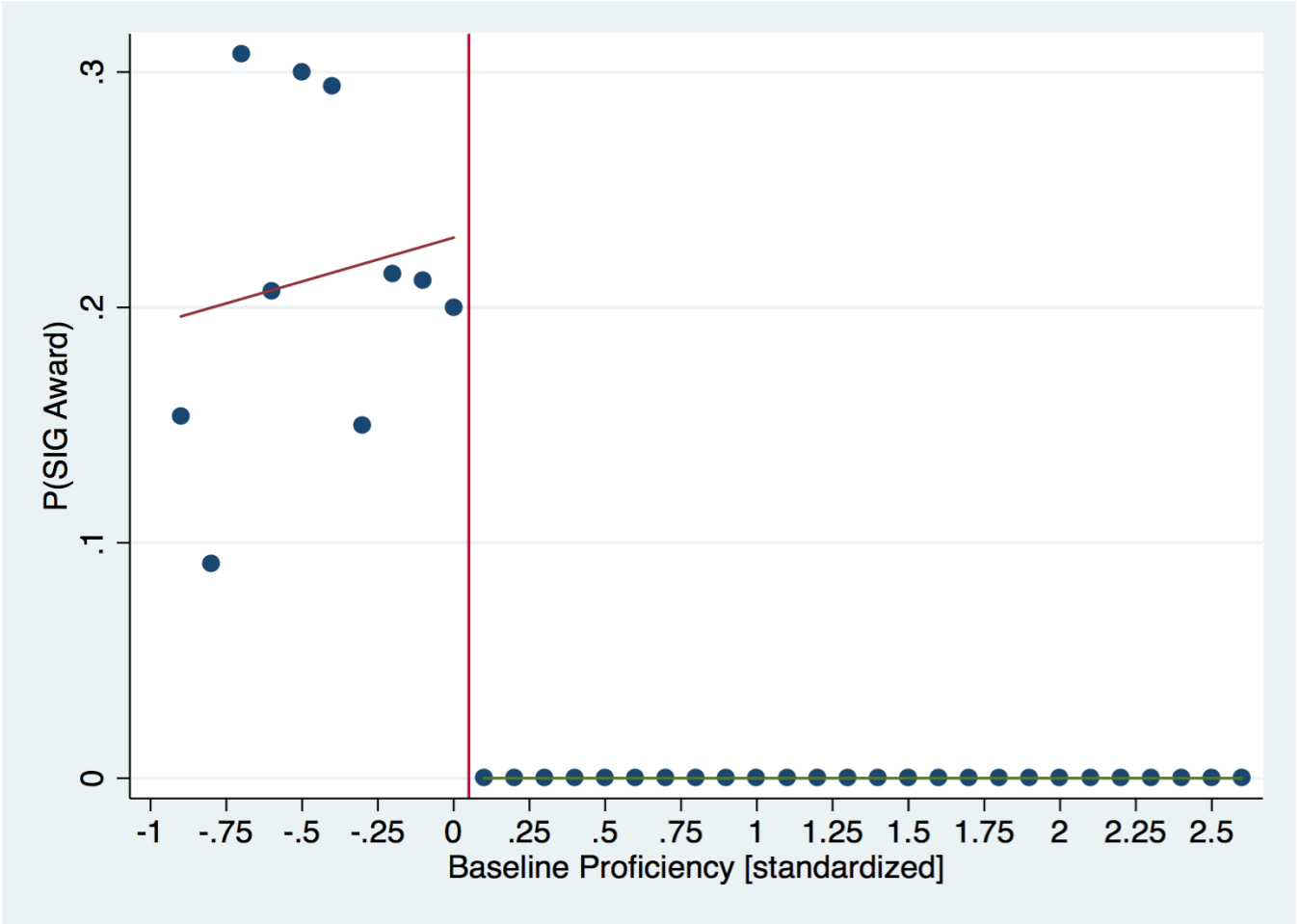
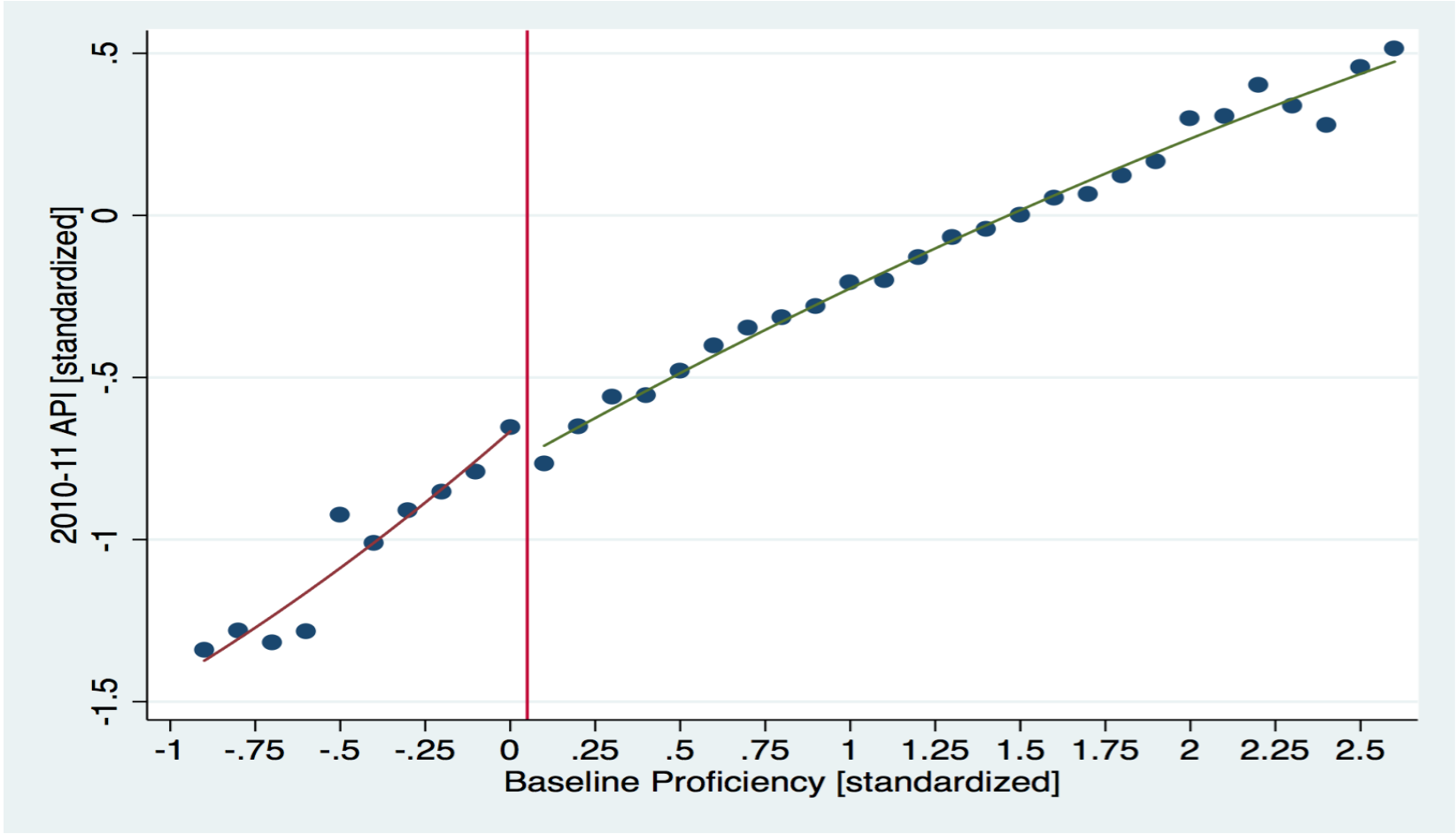


Figure 1 – Assignment to SIG “Treatment”

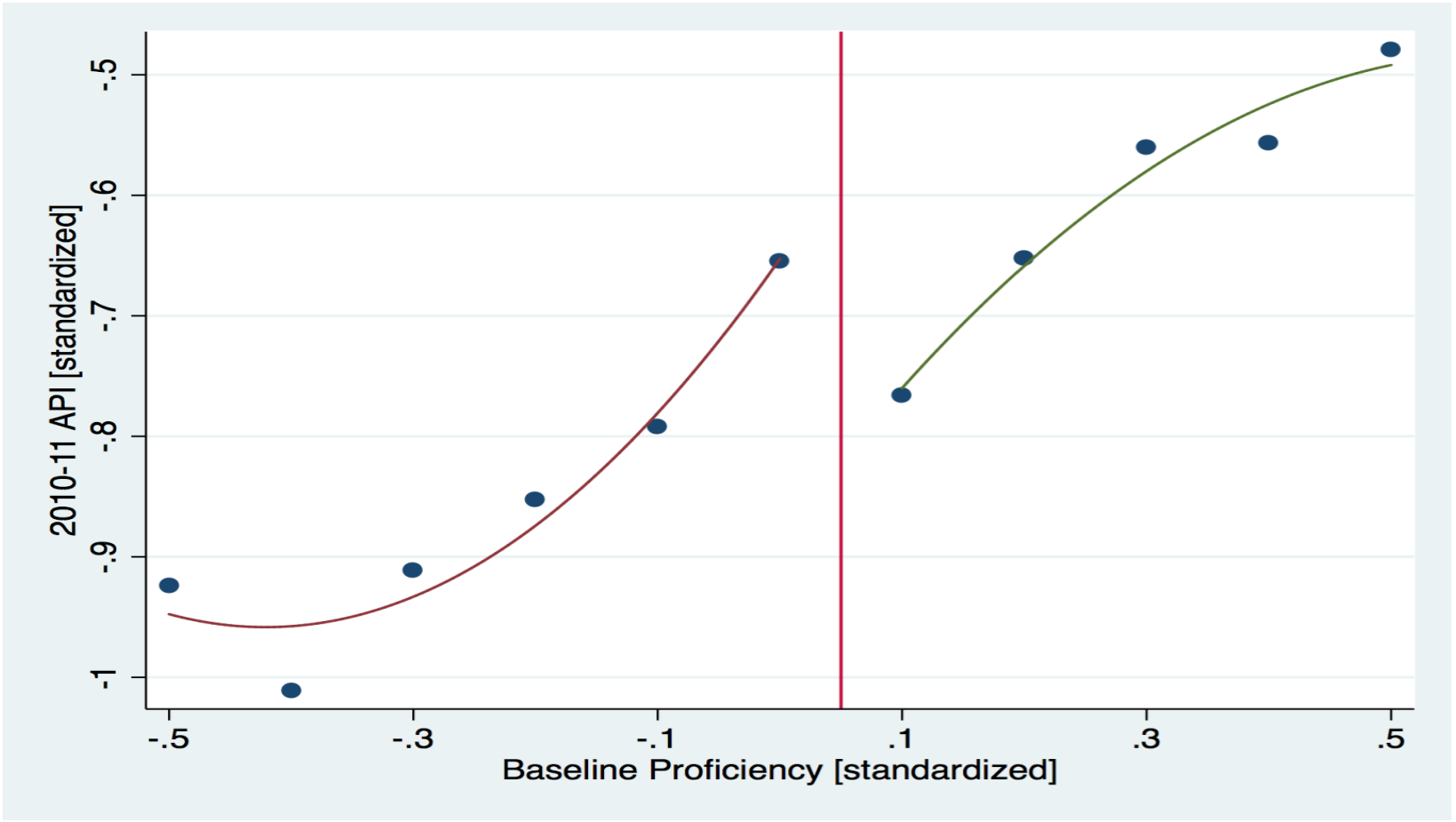
Academic Performance Index (API)

- School-level performance measure based on statewide testing (e.g., CSTs, CMAs, CAHSEE); standardized using school-level mean and SD
- The "cornerstone of the state's accountability system" used to identify schools of distinction, target interventions, and in AYP calculations
- The weighting applied to test results in different subjects varies by grade level
  - For elementary and middle-school students, math and ELA are heavily weighted
  - For high-school students, more balanced weighting of math, ELA, social studies, and science
- Some controversy over growing use of CMAs; implications for construct and internal validity?
- A common performance measure across schools makes it possible to harness power by using schools at all levels
  - Also, math and ELA results based on school-grade-year CST data

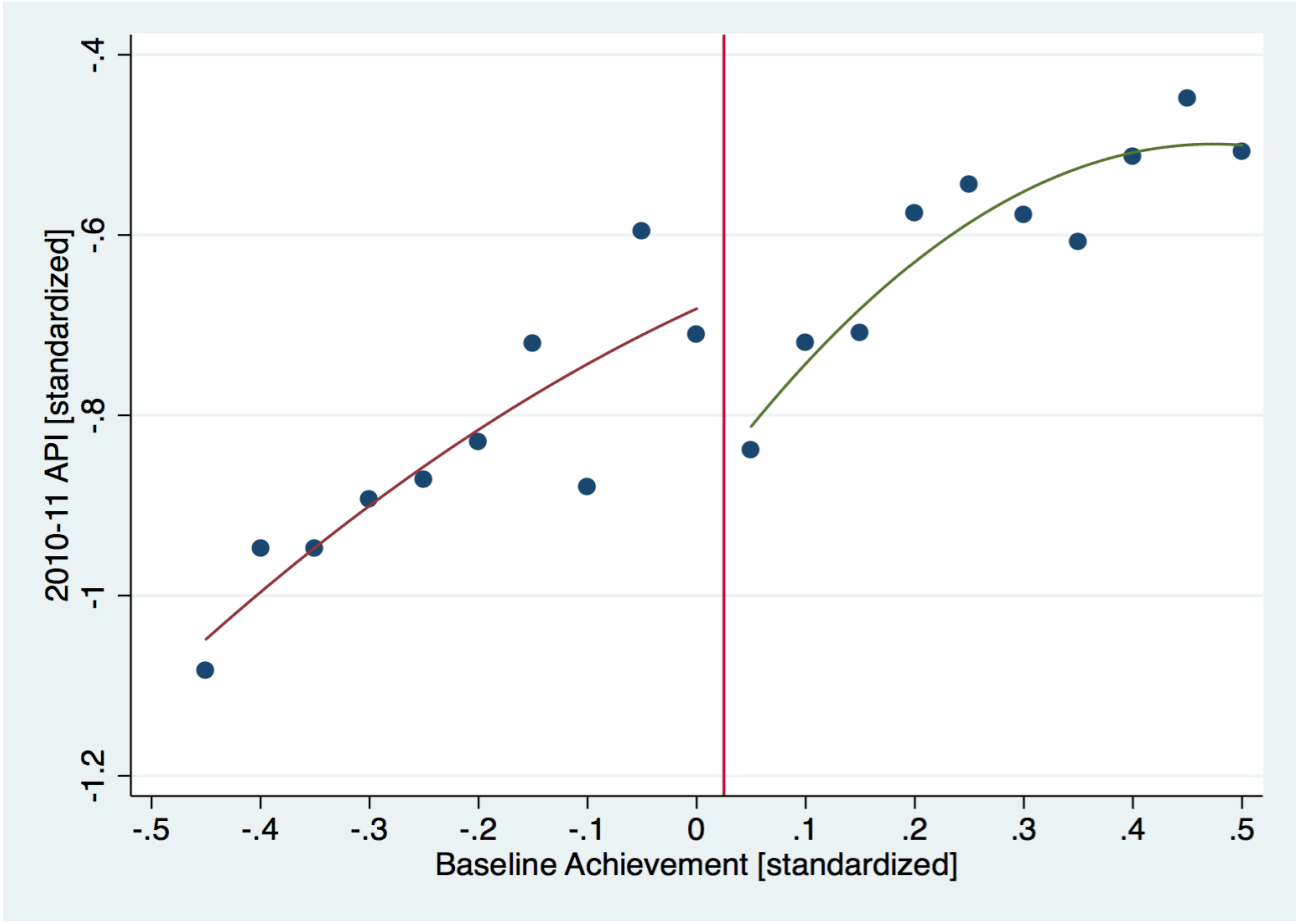
Results



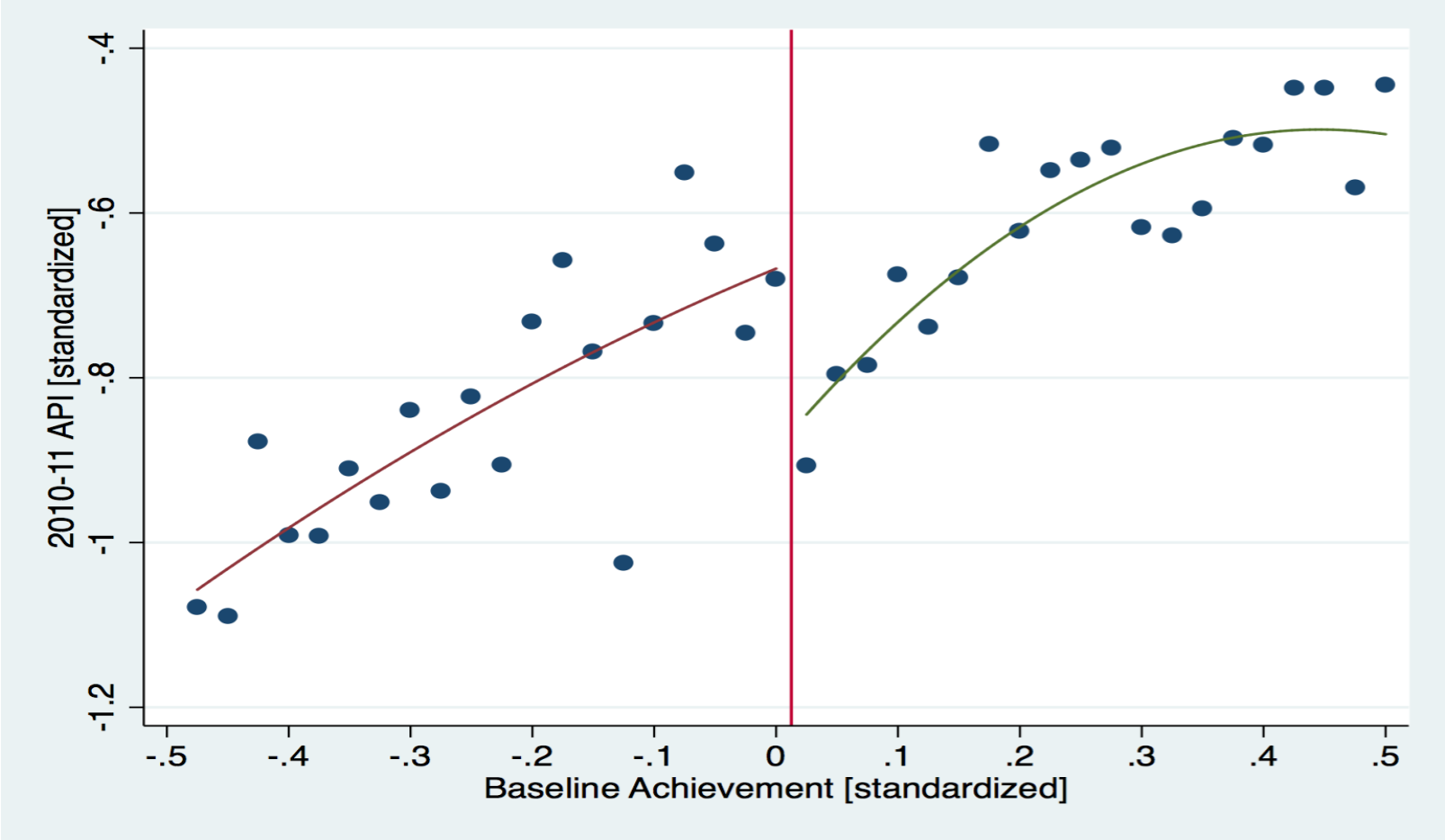
2010-11 API Scores around SIG-eligibility threshold



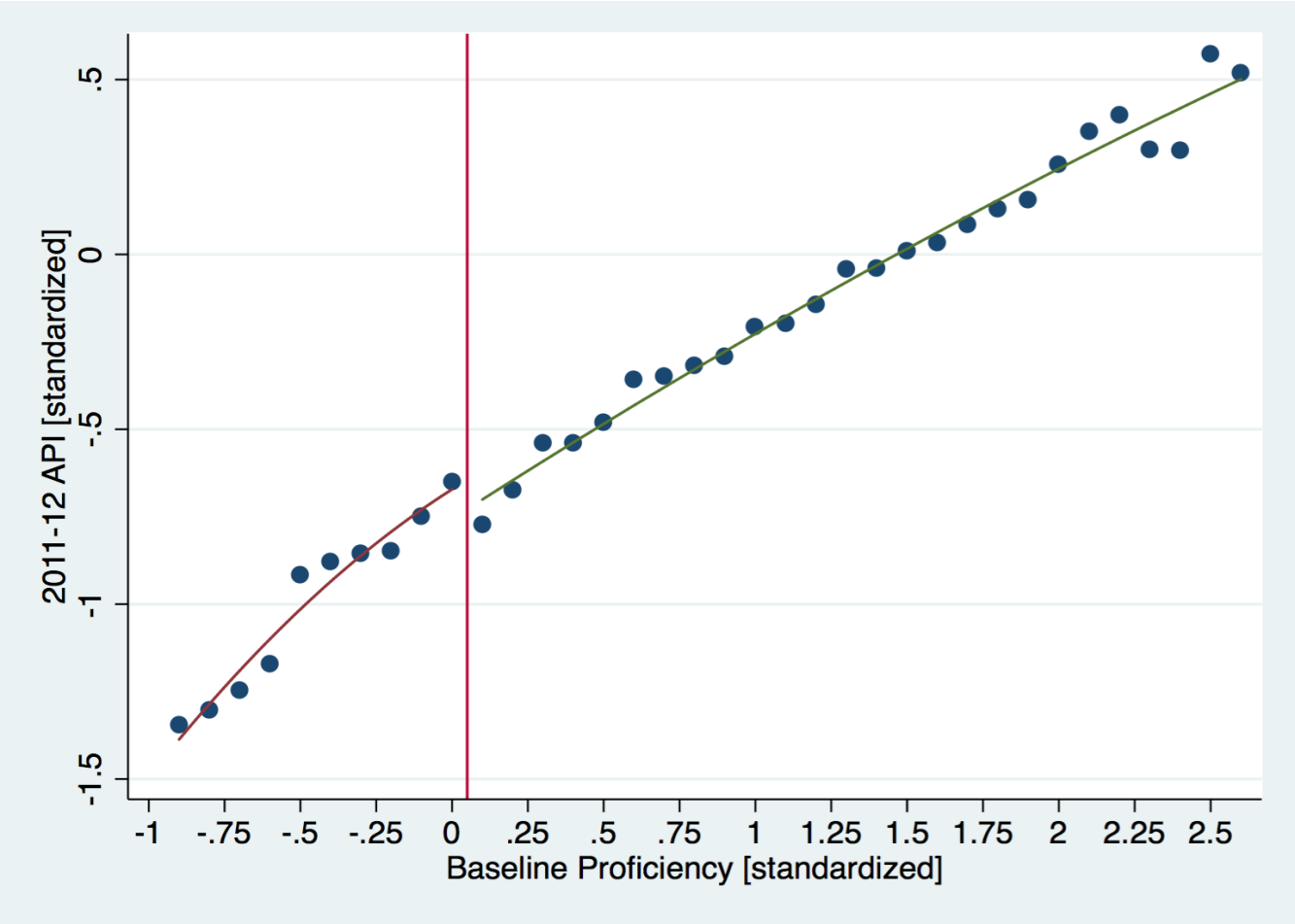
2010-11 API Scores (0.5 bandwidth)



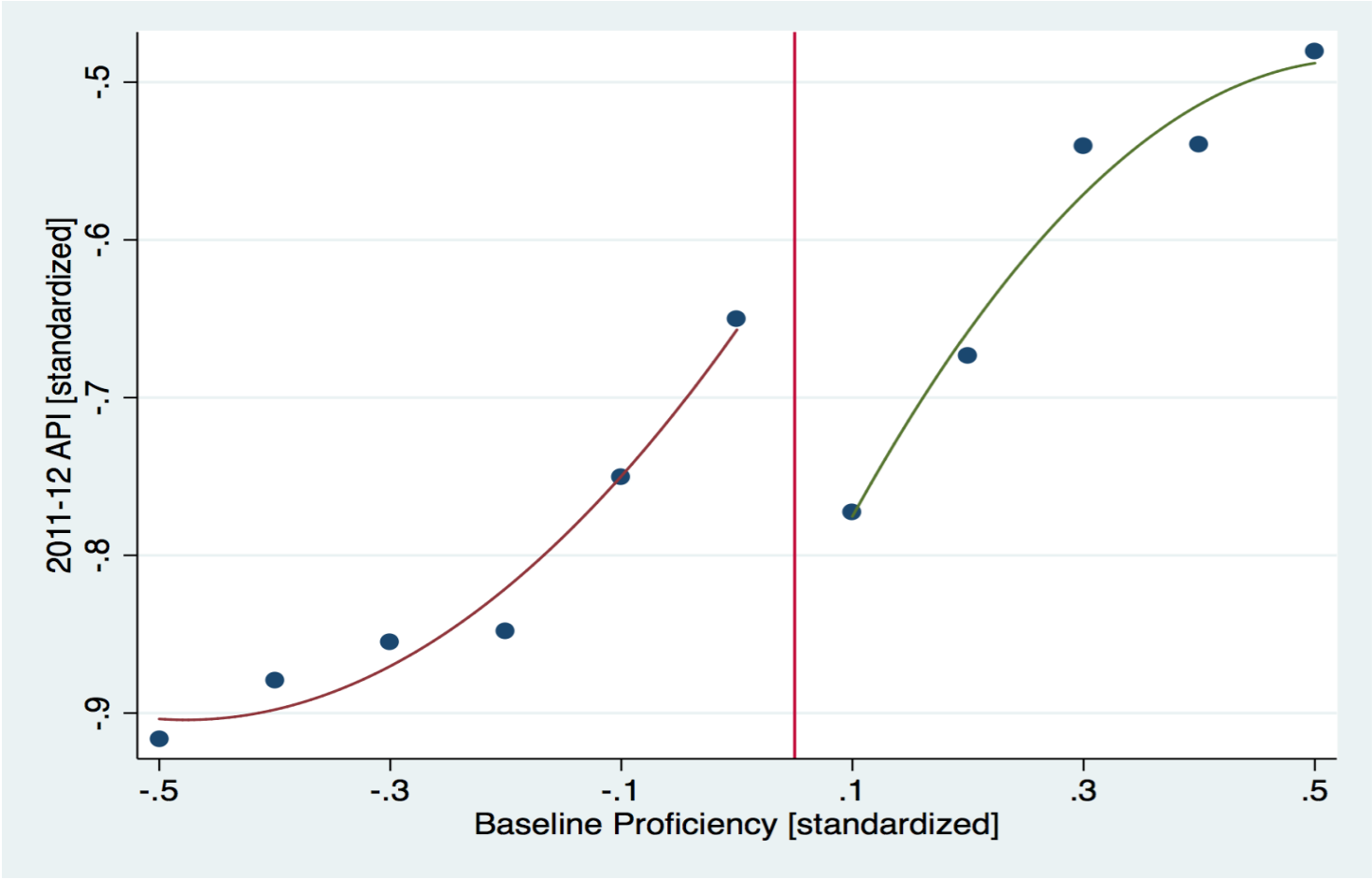
2010-11 API Scores (0.5 bandwidth, 0.05 bin width)



2010-11 API Scores (0.5 bandwidth, 0.025 bin width)



2011-12 API Scores around SIG Eligibility Threshold



2011-12 API Scores (0.5 bandwidth)

Robustness Checks?

Overall results

- API scores "jump" 0.07 SD at SIG-*eligibility* threshold (0.08 SD by 2012)
- Estimated effect of SIG *award* is 0.30 SD in 2011; 0.36 SD in 2012
- Gains on both math and ELA CST scores but math gains larger

Could Schools Manipulate Eligibility Status?

- Pre-determined nature of assignment variables suggest not
- Density test (McCrary 2008) cannot reject smoothness of distribution at threshold

Misleading reliance on Functional Form?

- Importance of graphical evidence
- Use of alternative functional forms
- Use of "local linear regressions" with increasingly restrictive bandwidths
- Balance of *baseline* (AY 2009-10) covariates around discontinuity
- Estimated effects of "placebo" RDs

Robustness Checks?

NON-RANDOM SORTING OF STUDENTS TO/FROM SIG-ELIGIBLE SCHOOLS?

- Bias of uncertain direction?
- Note highly compressed timing of SIG award to CA, LEA applications and awards
- Balance of *post-treatment* covariates around discontinuity

DO SIG-FUNDED SCHOOLS DIFFERENTIALLY USE CMAS?

- Estimated RD effects on % with disability in 2010-11 and 2011-12 are nulls

Any Evidence on Treatment Mediators?

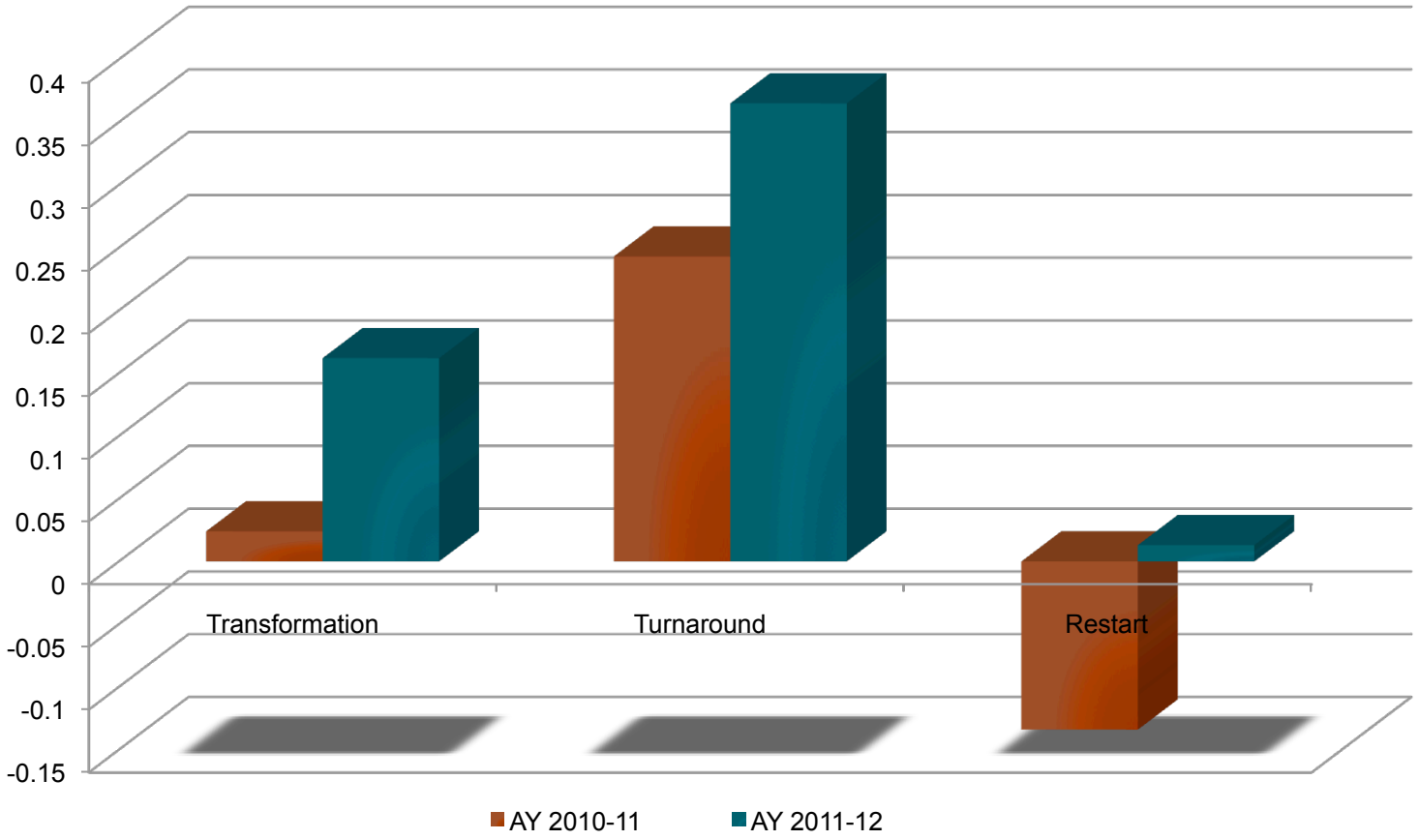
RD estimates of effects of SIG Eligibility on school Staffing?

- Probable leadership change but difficult to establish with measurement error in available data
- New staff: average teacher experience falls by ~5 to 6 years
- More staff: Pupil-teacher ratios fall by ~7 in year 1 (but not year 2?)

Any evidence on the comparative efficacy of the different reform models (e.g., transformation vs. turnaround)?

- "Difference in differences" models where API growth is dependent variable
  - Compare pre/post of SIG schools to contemporaneous pre/post of "control" schools (e.g., all lowest achieving schools, all PLA schools)
- Year-1 gains concentrated in turnaround schools
- Year-2 gains in both turnaround and transformation schools

“Diff in Diff” Estimates: API gains by SIG Model



Summing up: effect size and cost effectiveness?

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

- A cost-effectiveness benchmark from Project STAR's class-size reductions
  - 0.2 student-level SD gain for 47% expenditure increase (approximately $5,000 per pupil)

- First-year SIG results: 0.3 gain w/r/t school-level SD
  - ~0.09 w/r/t student-level SD; cost of $1,500 per pupil

- More cost-effective but not dramatically so?

Discussion

- (Surprising?) evidence on the efficacy of SIG-funded reforms in CA

- Conventional caveats about generalizability
  - Unclear relevance for other states where SIGs were differentially implemented (GAO 2011)
  - Unclear relevance for the *median* school in CA because the RD estimates are "local"

- A more critical external-validity concern?
  - What about SIG-eligible schools that couldn't craft a winning SIG application or didn't even apply?
  - The RD estimates are still causal because they leverage "intent-to-treat" (SIG *eligibility*).
  - But the causal estimates are defined for treatment "compliers"
    - Analogy to prescription-drug trial with imperfect & non-random compliance?

- How to support improvement in low-performing schools that **could not or would not** take up SIG eligibility?
  - Not an academic question for states with NCLB waivers!