Uncovering Public School Objectives

Hugh Macartney
Duke University and NBER

Robert McMillan
University of Toronto and NBER

Uros Petronijevic
York University

NBER Summer Institute
July 2020

Organizational Behavior: Theory vs. Practice

- Familiar (and standard) approach to understanding behavior of economic agents:
  - posit agents take actions to maximize well-defined objectives, given technology and resource constraints
  - rationalize observed behavior using model (informal or explicit)
- Given almost infinite variety of possible models, important role for empirics: a powerful **disciplining** device
  - some models simply don't accord with the facts
- Today, we will develop that theme in light of some *new facts* we have uncovered

Applying the Standard Approach

- Standard approach is appealing
- Key questions to address before applying it in practice:
  - What are the agents' objectives?
  - What is the underlying technology?
- Answers matter for
  - rationalizing observed behavior
  - predicting counterfactual behavior – if objectives and technology (or environment more generally) were to change
- Often non-trivial to uncover objectives and technology in a given context

Applying Approach in Education Context

- Particular challenges with education applications:
  - Objectives
    - **highly regulated** public schools are not firms and not organized like firms (so not profit-maximizing)
    - objectives likely **complex** and **not fully revealed** to researchers
  - Technology
    - **unknown** form
    - combines observed and **unobserved** inputs
    - reflects **actions** of many agents (e.g, teachers, administrators, parents, students)
- In a nutshell, neither aspect fully understood
- Yet to understand education delivery in practice or to make sensible policy predictions, need a good handle on both

Public School Objectives

- Main **goal** today: *uncover the public school objective*
  - (shed light on technology along the way)
- Our strategy (in essence): leverage *changes* in regulation
- Motivating the strategy: given heavy regulation and complexity of objectives, unlikely that traditional public schools simply maximize
  - quality, or
  - some measure of efficiency (e.g., output per unit cost)
- Thus, policies that increase resources for schools might not translate into better outcomes
  - depends on incentive environment, as has been well-noted
    - e.g., Hoxby (2000)
- This recognition leads the spotlight to be placed, naturally, on *incentives*

Accountability Incentives

- Strengthening incentives an important – perhaps central – theme in education policy over past two decades
- Most far-reaching policy: federal NCLB Act of 2001
  - mandated performance standards based on standardized tests
  - introduced sanctions (more severe with repeated failure)
- NCLB has been studied very widely, and basic incentive properties of proficiency-count schemes are well-understood
  - e.g., Reback (2008), Neal and Schanzenbach (2010)
- Variety of other accountability schemes already in place when NCLB came in, with different incentive features
  - e.g., offering monetary rewards to schools and teachers upon attainment of value-added targets (as under the ABCs in North Carolina – coming up!)

Using Accountability Variation to Make Headway

- Variation plays to economists' strengths: we know how to think about the effects of heightened incentives
- Key observation (from perspective of gaining a sharper insight into complex incentives of public schools):
  - accountability reforms seek to alter objectives
    - new incentives (punishments or rewards) cause agents to place greater weight on academic achievement
    - able to capture change using explicit expression
- Thus, accountability variation gives us a handle on a portion of what schools and teachers are maximizing
  - though, as we will see, unobserved portions still likely matter

Outline

- Data and Policy
- Responses to NCLB
  - Overall Effects
  - Actions and Agency
  - School Heterogeneity
  - Non-Incentivized Targets
- Conclusion
  - The Objective
  - Policy Implications

Data

- In order to establish new facts, take advantage of North Carolina **data** and **policy variation**
- **Data** first: Rich longitudinal administrative education info covering all public school students in North Carolina
  - yearly student-level test scores (1996-97 to 2004-05)
    - comparable across grades and years using developmental scale
  - student, teacher and school characteristics
    - student: parental education, gender, race, disability, limited English proficiency, economic disadvantage
    - teacher: experience, gender, licensure test scores
    - school: accountability program status, grade span
  - teachers matched to students; both followed over time

Policy Backdrop

- NCLB system came into effect in 2002-03 on top of pre-existing statewide 'ABCs' accountability scheme
- ABCs introduced in 1996-97 school year:
  - VA target for each student, based on prior performance
  - targets aggregated to school level
  - for schools exceeding aggregate target, all teachers and the school principal given monetary reward
- Relatively uniform incentives under ABCs: give all students similar attention
- *Dynamic* consideration (Macartney 2016): substantially out-performing target today (mechanically) results in
  - harder subsequent targets
  - higher probability of future failure
  - financial loss

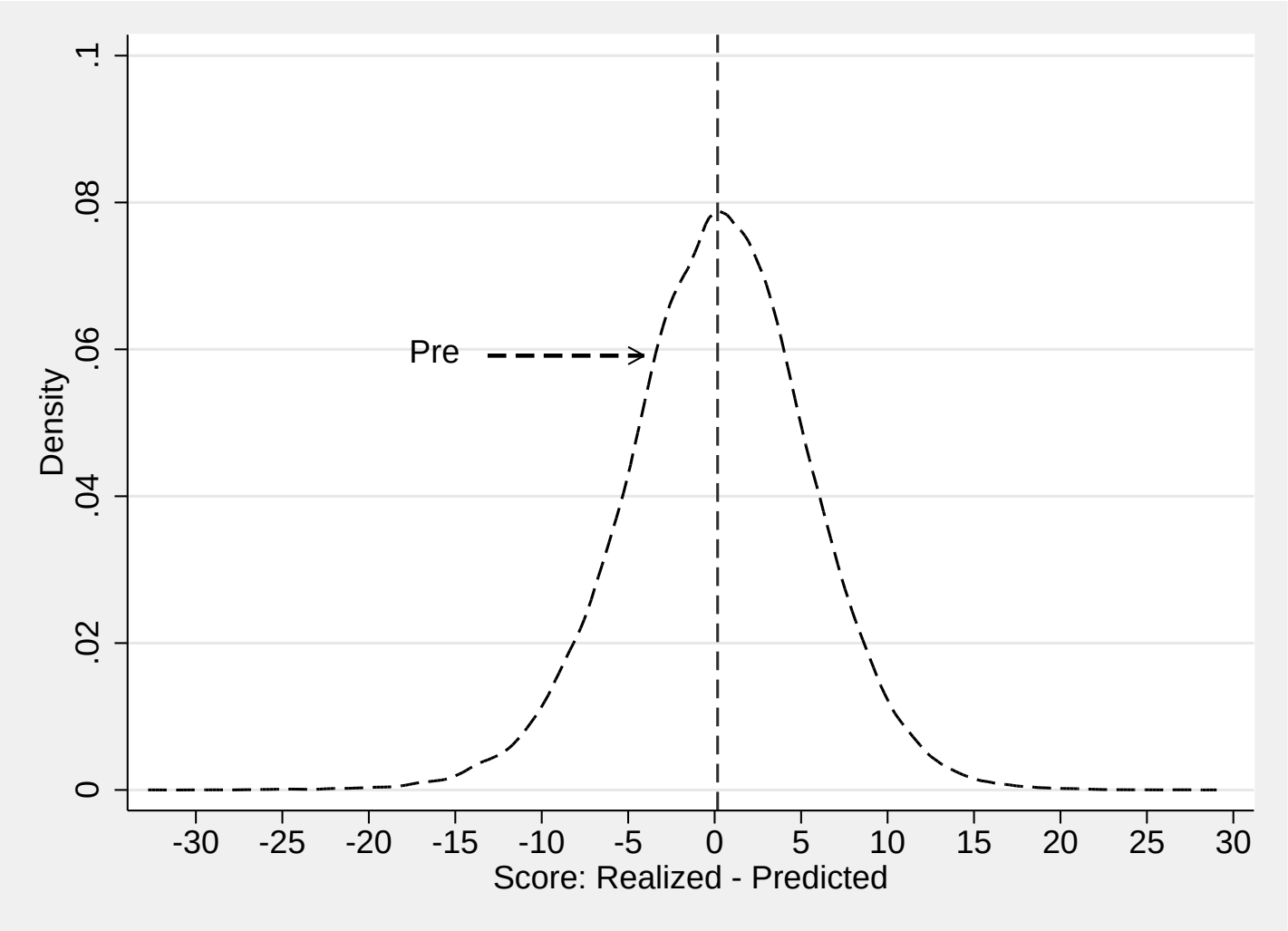
Policy Backdrop (2)

- NCLB a proficiency-based scheme
  - fixed proficiency target set for all students in each grade
  - (also targets for demographic subgroups)
  - penalties for failing schools
- Strong incentives under NCLB (and proficiency schemes in general) to focus on students predicted to score near target
  - e.g., Booher-Jennings (2005), Burgess, Propper, Slater and Wilson (2005), Reback (2008), Neal and Schanzenbach (2010)
- So expect boost in scores close to proficiency threshold
- Related questions (looking ahead):
  - How should two types of scheme interact in theory?
  - How do they interact in practice?
- First, some basic theory, and some basic corresponding facts

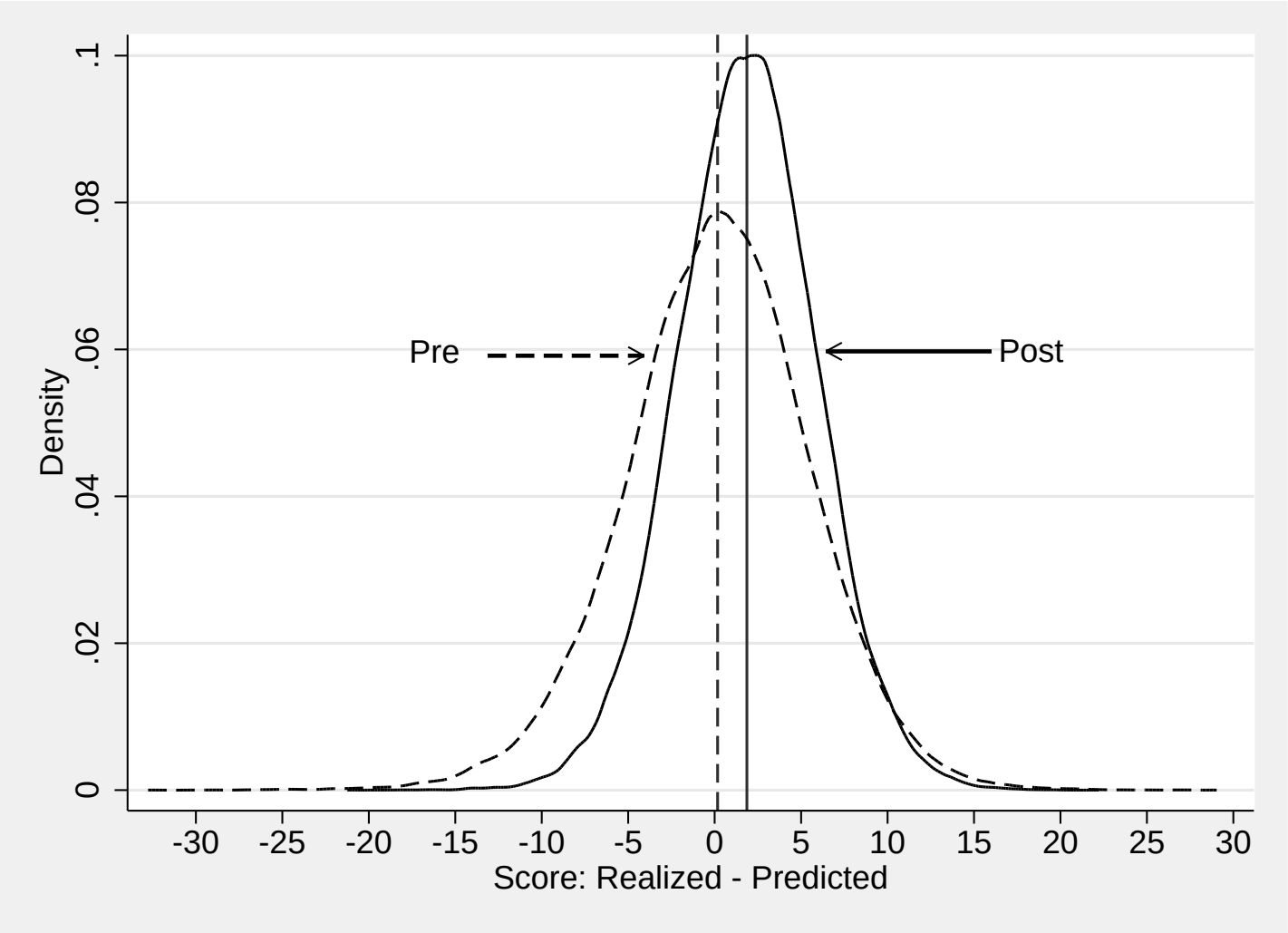
Test Score Effects of NCLB

- **Prediction**: General shift to stricter performance standards (backed by incentives) should lead schools and teachers to place more emphasis on incentivized output
- As documented widely, they do
  - e.g., Carnoy and Loeb (2002), Figlio and Winicki (2005), Hanushek and Raymond (2005), Lavy (2009), Dee and Jacob (2011), Figlio and Loeb (2011), Muralidharan and Sundararaman (2011), Imberman and Lovenheim (2015), Deming, Cohodes, Jennings and Jencks (2016)
- In our setting, able to measure response to NCLB, comparing pre- and post-reform

Overall Test Score Response to NCLB



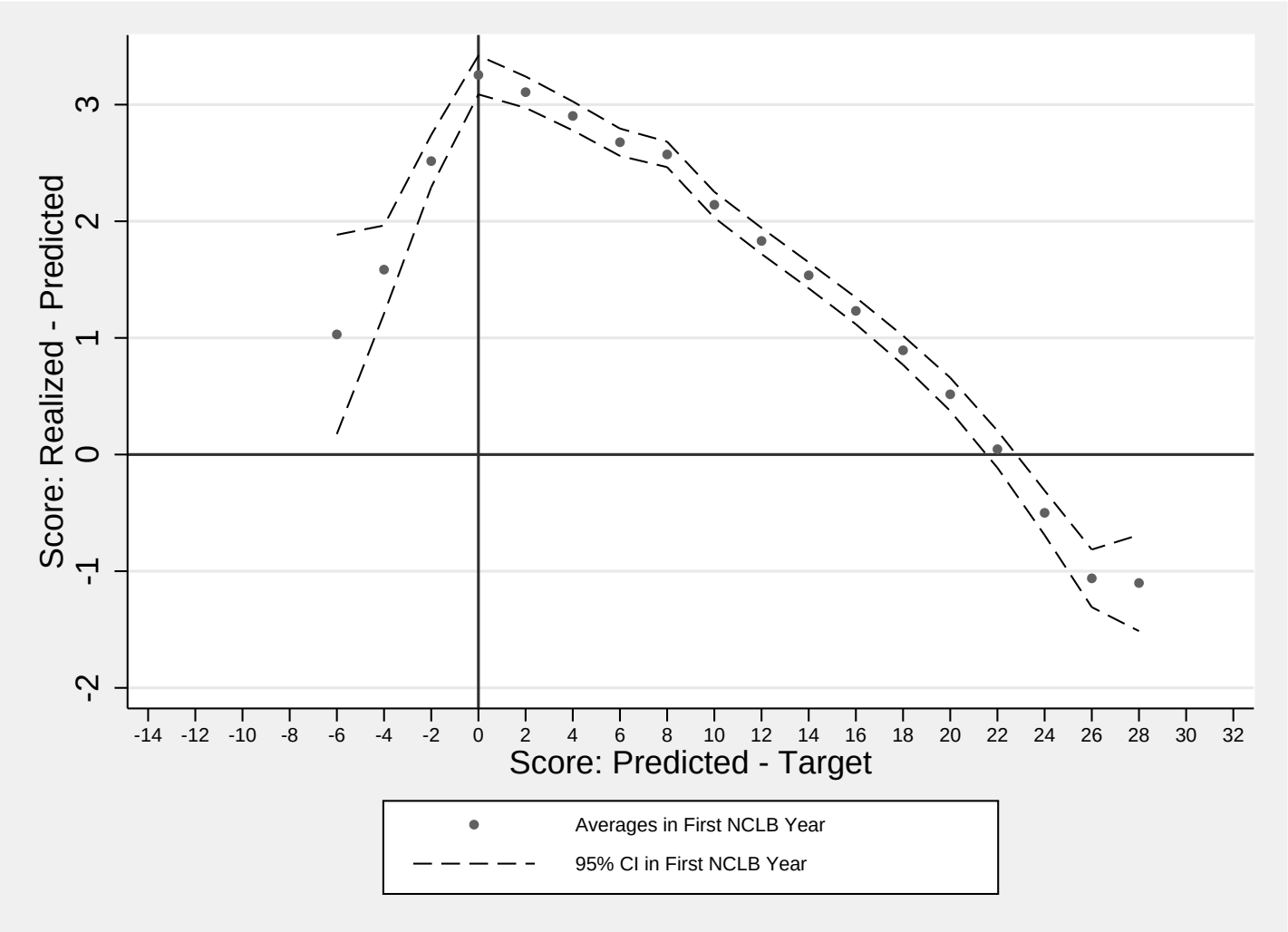
Overall Test Score Response to NCLB



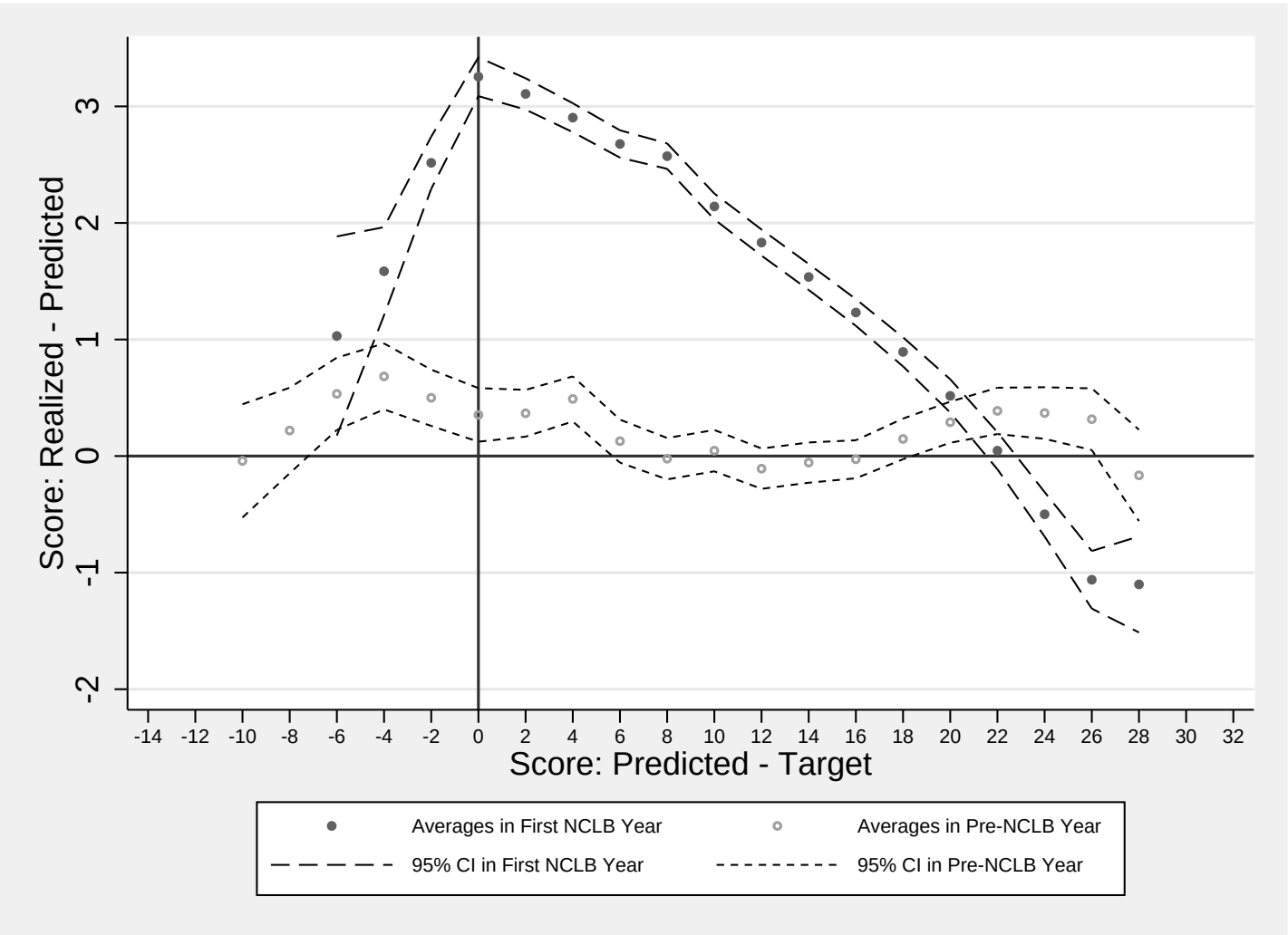
Test Score Response w.r.t. Target

- **Prediction**: NCLB and other proficiency schemes should lead educators to focus on students non-uniformly (as noted), yielding *non-uniform* effects
- In our setting, able to measure response to NCLB at different points in predicted performance distribution
  - (abstracting from sub-groups, as we do here, still leaves potent variation)
- Also able to conduct a pre-reform placebo test

Response w.r.t. NCLB Target



Response w.r.t. NCLB Target



Performance Response Findings

1. Clear evidence of overall performance response to NCLB
2. Response is non-uniform (as basic theory would predict)

- **Question**: What explains these facts?
  - in particular, what is being done, and by whom?
- Shedding light on *actions* and *agents* will be informative as to the objectives (and the technology)

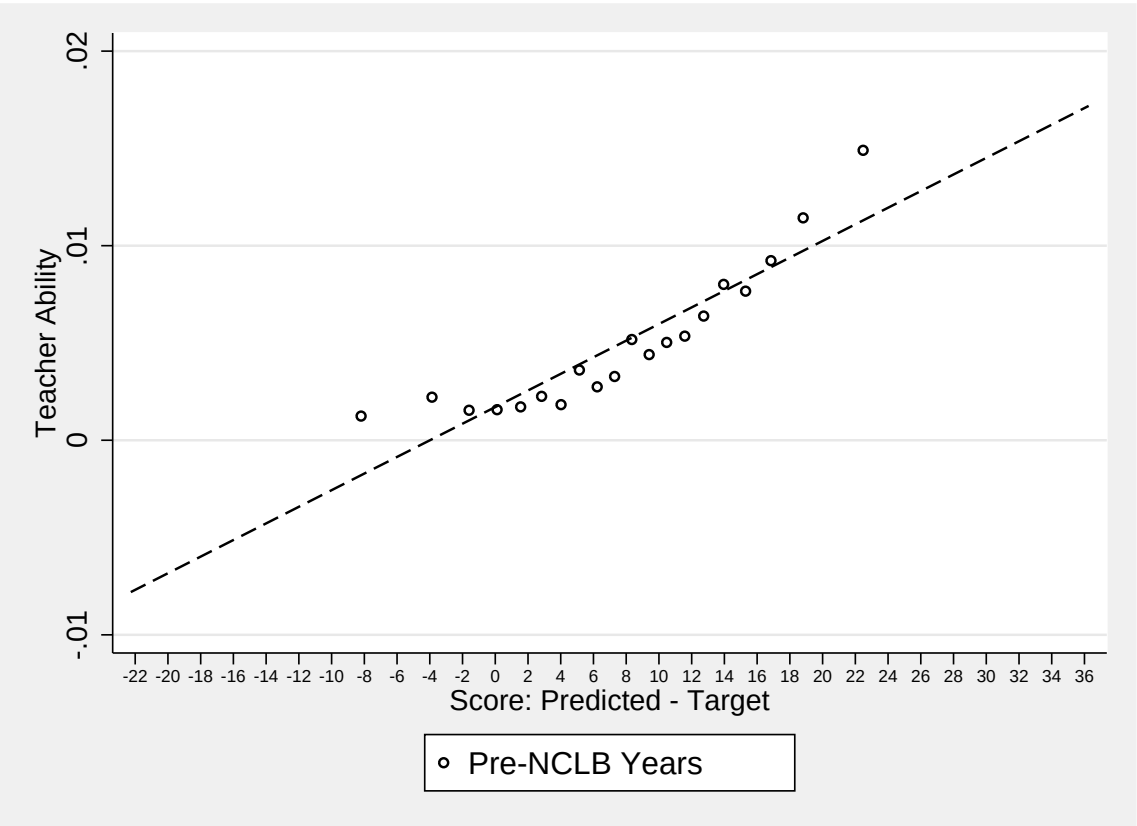
School Responses?

- **Prediction**: For school-level scheme such as NCLB, expect schools to pull various levers at their disposal in service of meeting the accountability threshold
- While information about actions is incomplete, administrative data can provide useful insights
- Consider observable actions that schools *could* take in order to satisfy the NCLB target (for example):
  - assign the most marginal students to the best teachers
  - assign more marginal students to smaller classes
  - coordinate teachers to direct greater effort to marginal students

Pre-Reform Student-Teacher Matching

- Let's take possible changes in student-teacher matching first
- To establish matching baseline, plot student incentive strength vs. teacher ability
- What are these?
  - **incentive strength** $\equiv$ predicted score $-$ target
    - essentially, expected student performance w.r.t. target
  - **teacher ability** captured by pre-NCLB teacher VA
    - standard Empirical Bayes estimator of VA, controlling for cubic polynomials of student prior scores and characteristics (see Kane and Staiger 2008, and Chetty, Friedman and Rockoff 2014)

Teacher Ability and Student Incentive Strength (Pre-Reform)

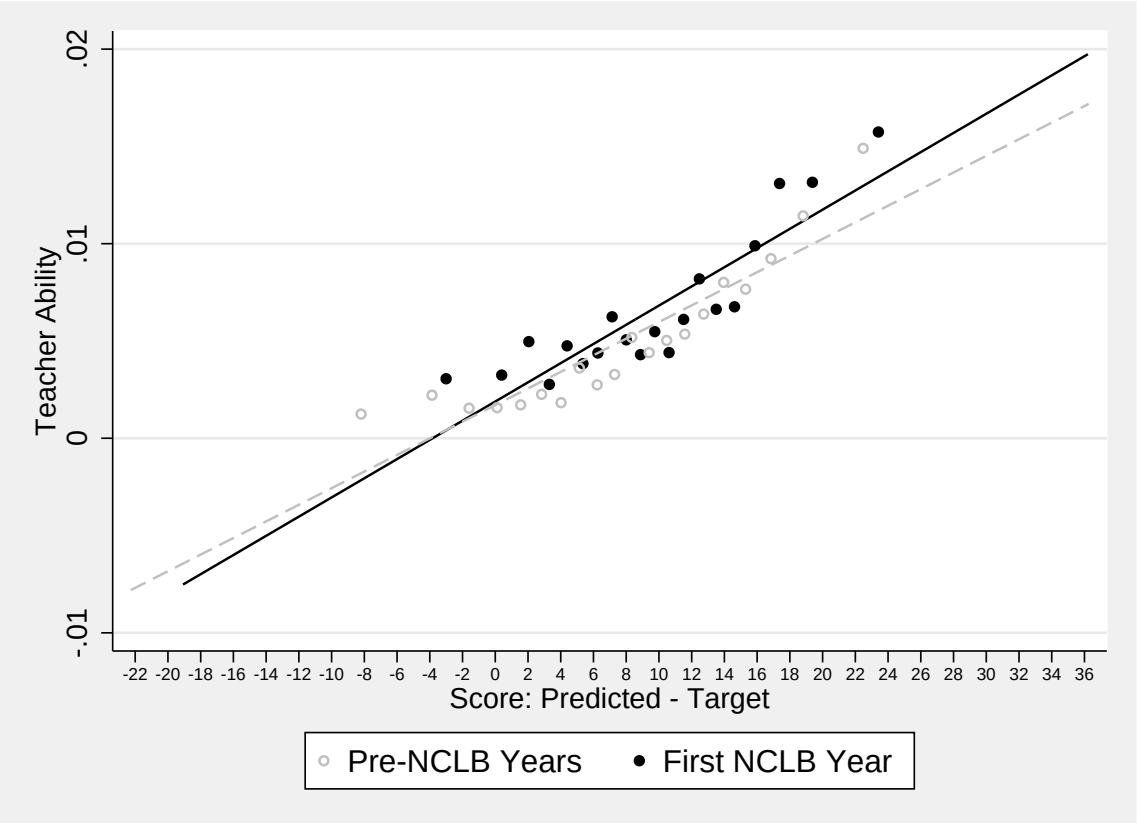


- Positive assortative matching pre-reform
  - controlling for school-year-grade FEs

Post-Reform Matching

- **Prediction**: NCLB should lead higher-VA teachers to be matched with more marginal students
- As North Carolina passing threshold set low in performance distribution, expect *less* positive assortative matching post-reform
  - (graphically, profile should *flatten*)

Teacher Ability and Student Incentive Strength (Post-Reform)

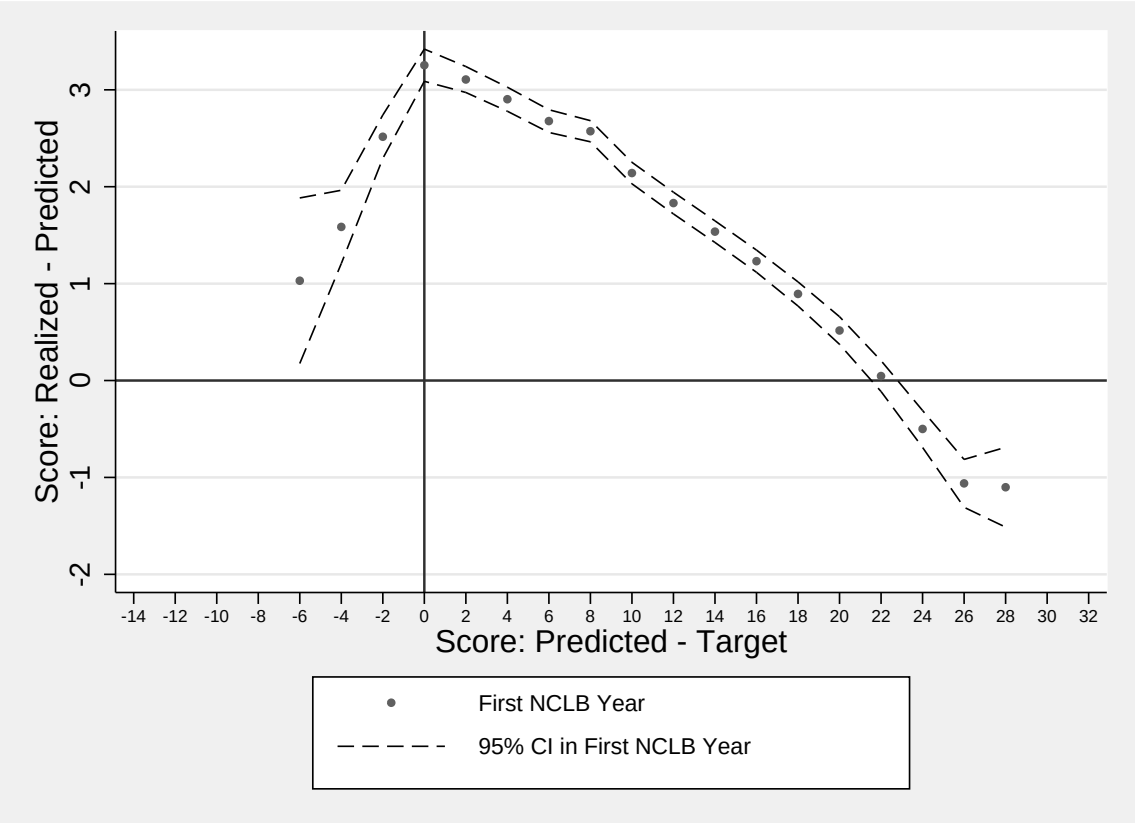


- Not what we see
  - assortative matching little-changed

Interpreting the Matching Evidence

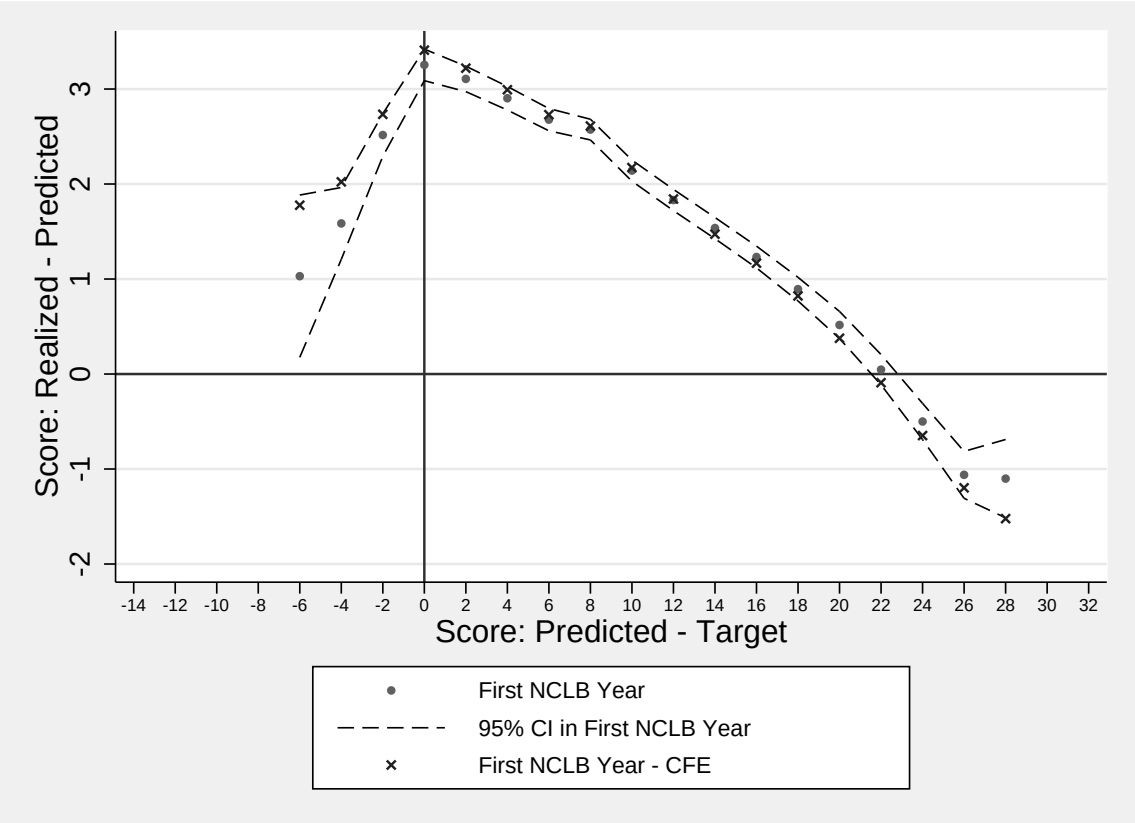
- More able teachers not being re-assigned to more marginal students, and more marginal students not being assigned to smaller classes (as we will see) suggests taking either option is too costly
  - *think*: taking best teachers away from high-SES students
- Further, NCLB-related incentives may not be over-riding
  - perhaps suggestive of *informal incentives*, as distinct from NCLB-related incentives
- **Next Question**: Are teachers changing how they direct their effort?

Student-Specific Targeting



- Profile from before

Student-Specific Targeting

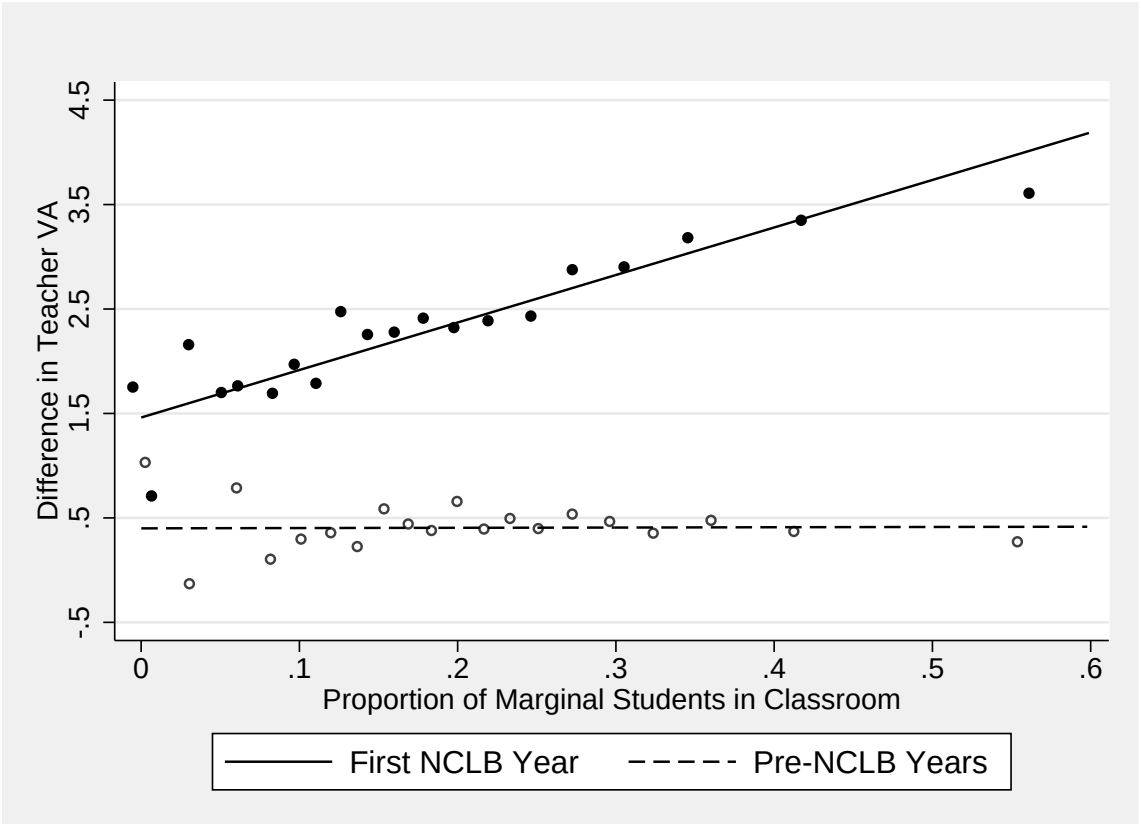


- Near-identical results using only *within-classroom* variation
  - suggests action occurs at the classroom level: teachers directing their effort toward more marginal students

Interpretation

- Evidence of teachers directing more effort to marginal students *within-classroom*
  - consistent with principals devolving effort decisions to agents most likely to have relevant local knowledge
- Motivates a *teacher*-level effort model as a fruitful simplification (discussed below)
- **Related**: Effect of NCLB on individual teacher VA

Change in VA vs. Proportion Marginal



- *Within-teacher*, teacher VA increases more under NCLB as proportion of marginal students in her classroom rises
  - using purely pre-reform variation, no such relationship exists

The Technology

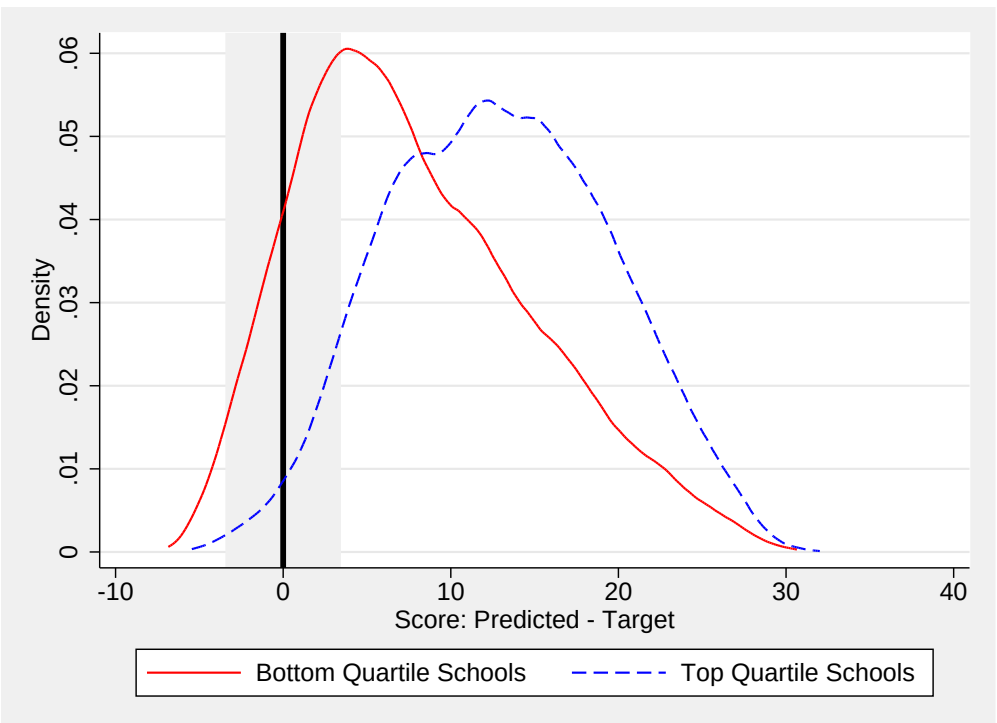
- Recalling: unknown form; many relevant inputs unobserved
- But here we can make some headway:
  1. Isolate *teacher effort* as a relevant input
     - 'effort' $\equiv$ incentive-related improvements in test scores
     - effort can be student-specific: e_i for student i
     - effort a function of incentive strength
     - **Question**: Can the effort-incentive mapping be identified?
     - Yes, we already have that, semi-parametrically
  2. Distinguish teacher effort and non-incentive-related ability in education production for teacher j:
     - $y_{ij} = f(effort_{ij}, ability_j, \mathbf{X}_{ij})$
     - **Question**: Can the relative effects of these distinct inputs be separately identified?
     - Yes: strategy in Macartney, McMillan and Petronijevic (2020b)

Heterogeneous School Responses

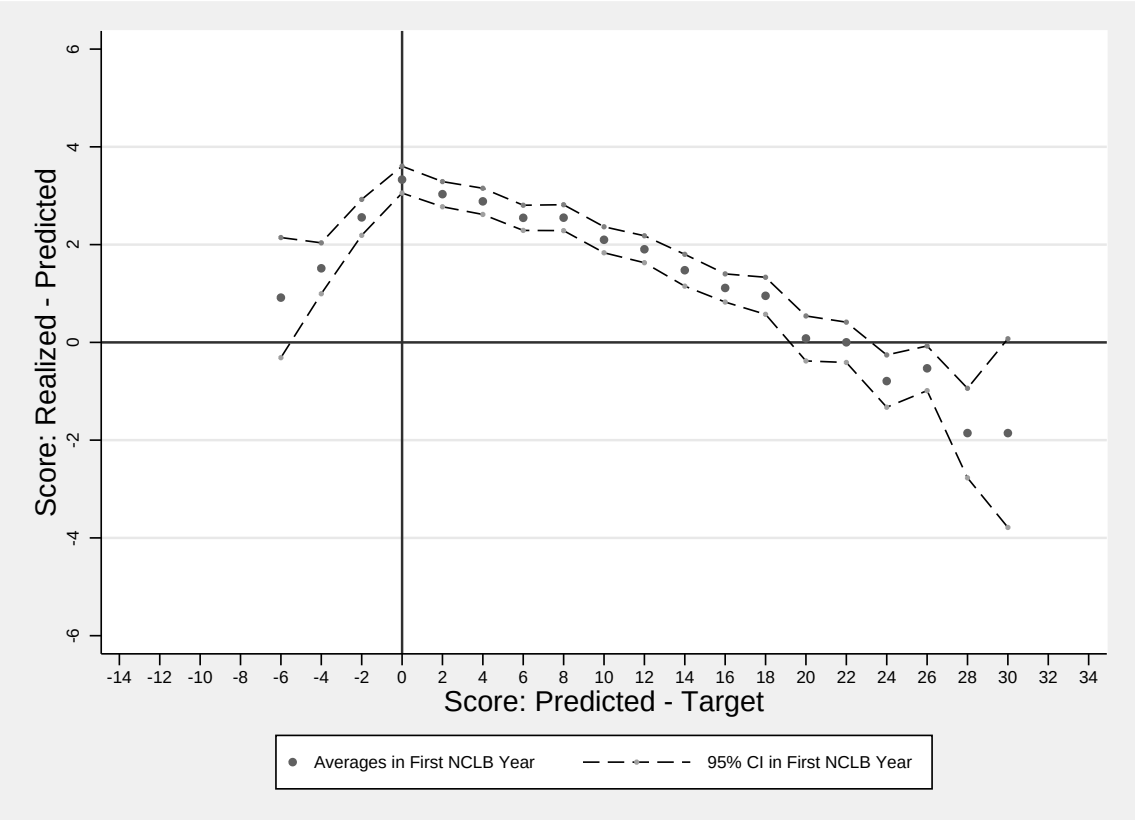
- Another dimension to learn more about *objectives*
- Consider a setting (e.g. North Carolina) in which
  - sophisticated VA scheme was operating (ABCs), and then
  - second scheme was introduced on top (NCLB)
- What should schools do? It depends:
  - for some schools, nearly all students would pass NCLB proficiency threshold without further effort
  - and responding to NCLB could result in losing future bonus
- One would predict that such schools would ignore NCLB (if they only cared about $$$)
- Let us see whether that was the case

Response by School Type

- **Prediction**: Response should increase in proportion of marginal students in school (giving stronger incentives)
- To test, we group schools by quartile of NCLB *predicted* passing probability (similar to Deming *et al.* 2016)
  - as NC target set low in performance distribution, higher proportion of marginal students in bottom quartile schools



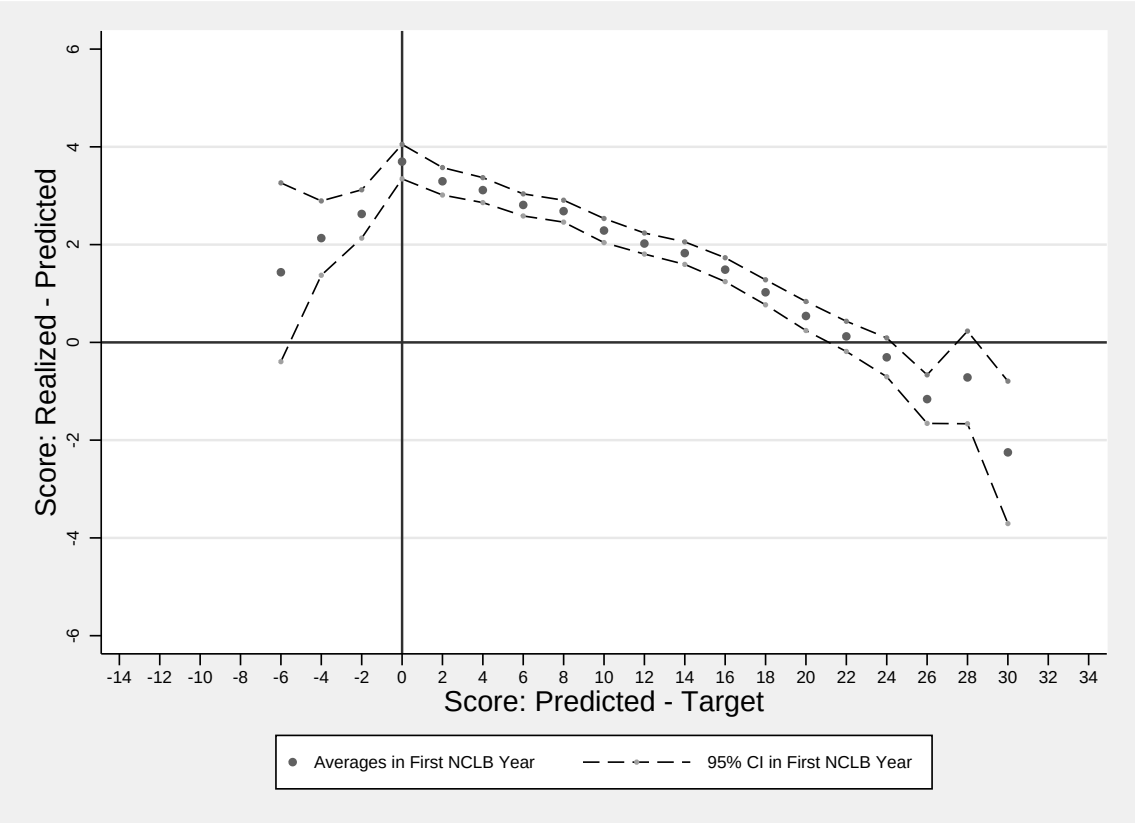
Response by School Type (2)



Bottom Quartile

- Against prediction, *similar* response by quartile
  - marginal students always receive disproportionate effort

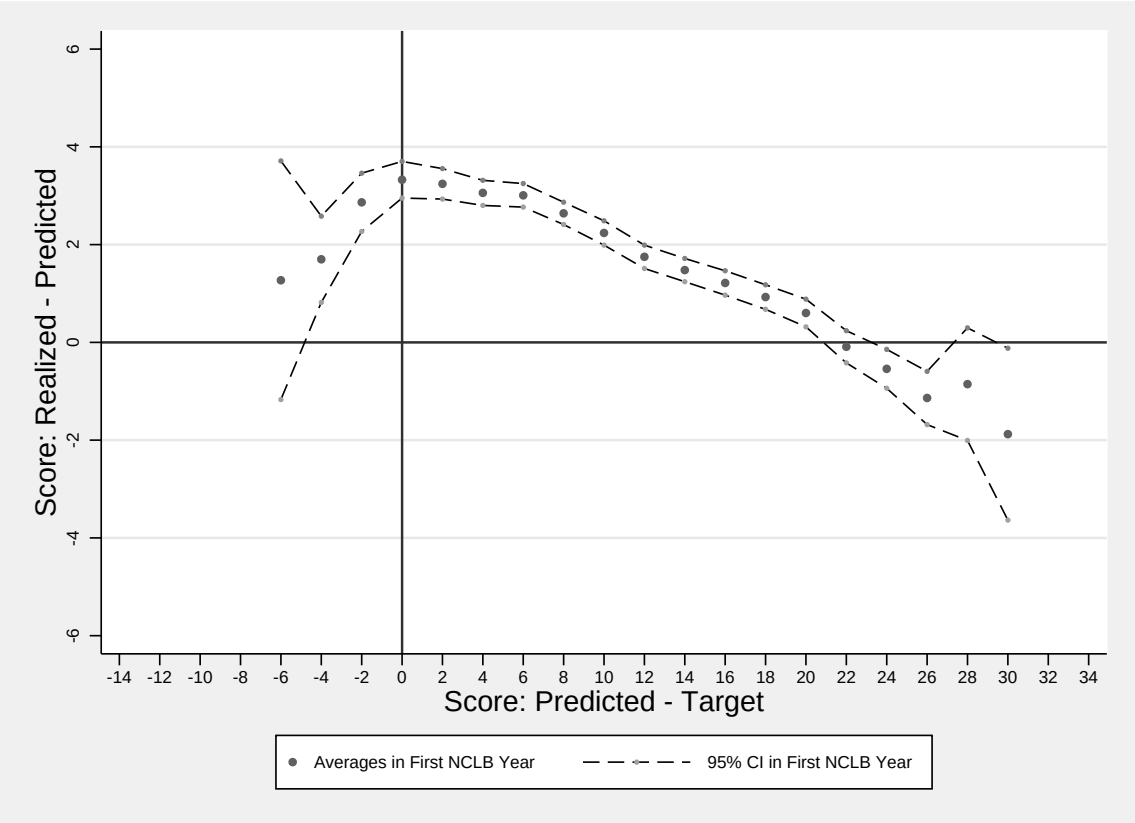
Response by School Type (2)



Second Quartile

- Against prediction, *similar* response by quartile
  - marginal students always receive disproportionate effort

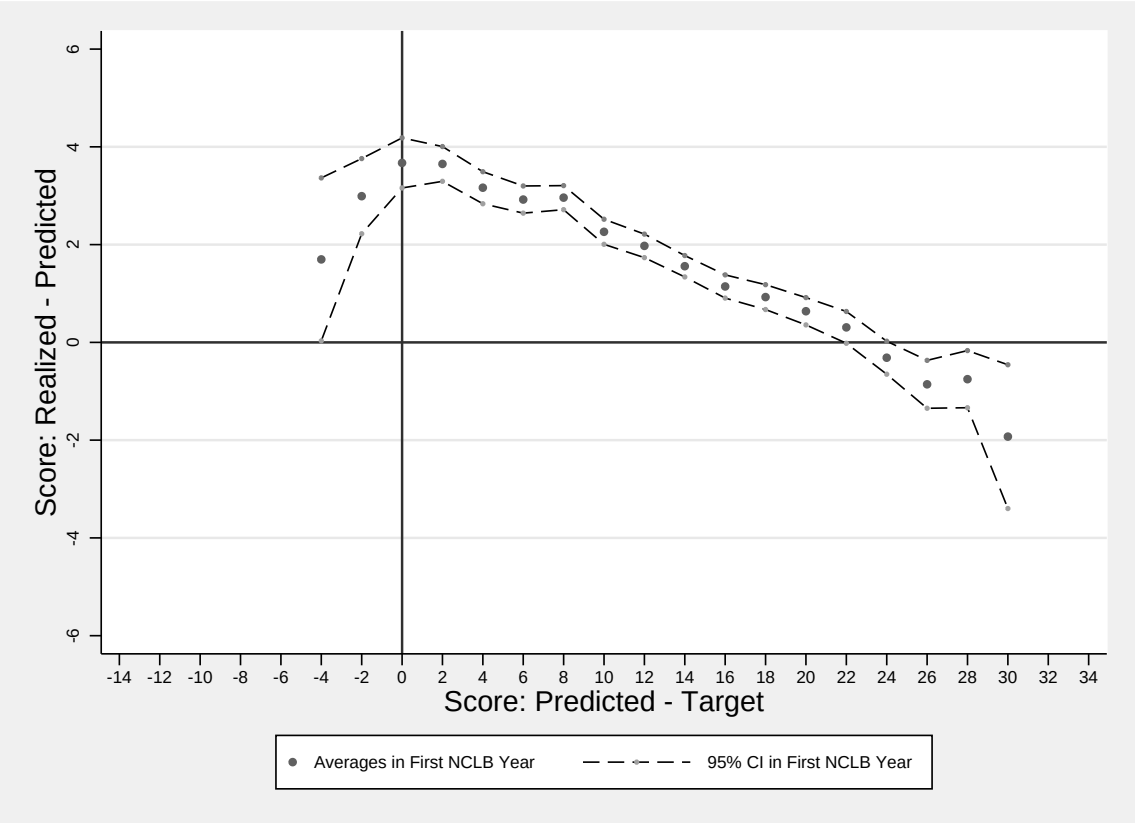
Response by School Type (2)



Third Quartile

- Against prediction, *similar* response by quartile
  - marginal students always receive disproportionate effort

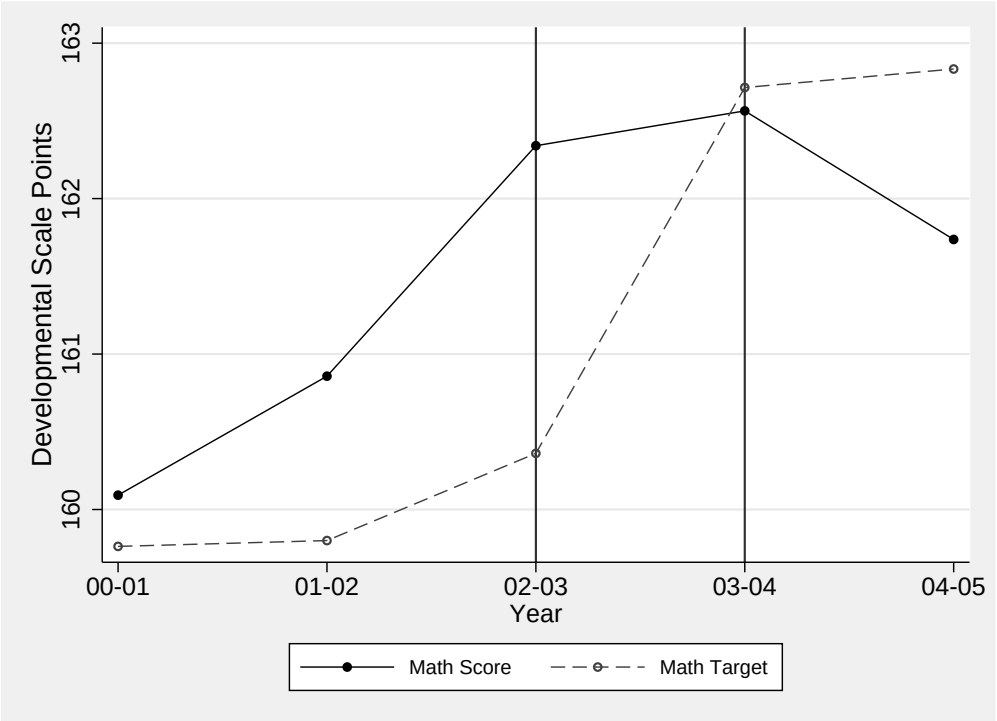
Response by School Type (2)



Top Quartile

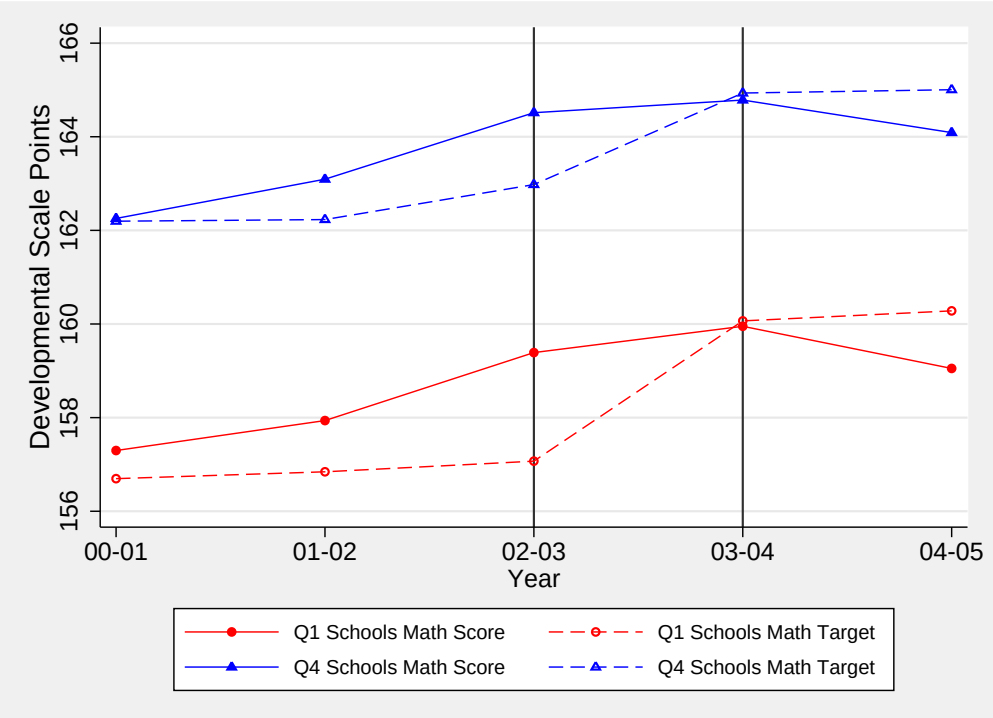
- Against prediction, *similar* response by quartile
  - marginal students always receive disproportionate effort

NCLB-ABCs Interaction



- **Dynamic link**: NCLB response *raised* subsequent ABCs targets, leading to higher probability of failure in 03-04
  - e.g., fourth grade response in 02-03 raised average fifth grade target in 03-04, exceeding average score in 03-04 for first time

NCLB-ABCs Interaction by School Type



- Occurred for *both* bottom- and top-quartile schools
  - suggests top-quartile schools over-responded
  - what were the financial consequences of doing so?

Costs of 2002-03 NCLB Response

		Quartile School Passing Prob.			
	All	Q 1	Q 2	Q 3	Q 4
ΔPr(Pass ABCs)	-0.17	-0.24	-0.18	-0.16	-0.12
Expected Loss	-$121	-$177	-$135	-$116	-$90
N	988	233	259	263	233

- In expectation, and before responding in 2003-04, schools most at risk of failing NCLB in 2002-03 stood to lose $177 per teacher in 2003-04 because of their response the prior year
- Even top-quartile schools least at risk jeopardized ABCs payments by responding to NCLB, standing to lose $90 per teacher

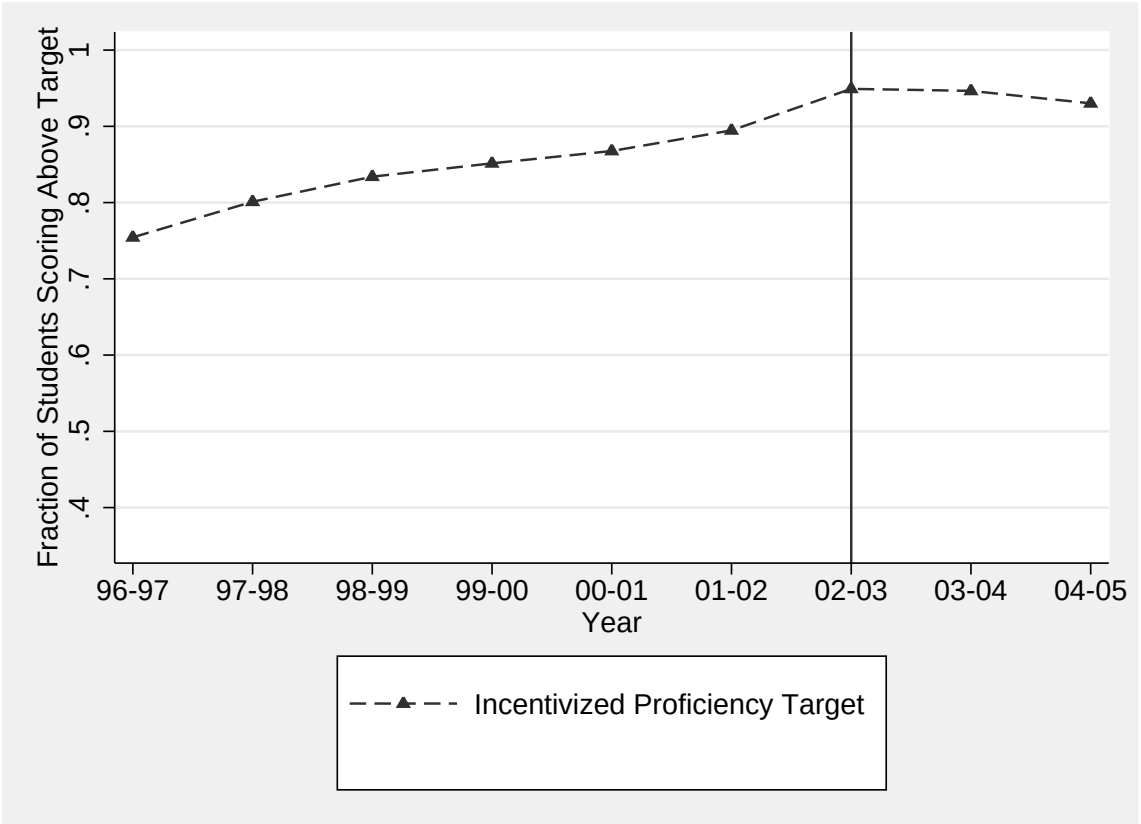
Interpretation

- Expect schools with most marginal students – those in bottom quartile – make most effort
  - strong incentive to avoid failing NCLB
  - in process, forgo monetary rewards under ABCs
- Response can be used to monetize the NCLB sanction
  - see Macartney, McMillan and Petronijevic (2020b)
- Top-quartile schools also raise effort and forgo monetary rewards
- Take as indication that they care about other matters (instead of their making a mistake)
  - important when understanding the effort-setting decision
  - (more on this next)

Non-Incentivized Targets

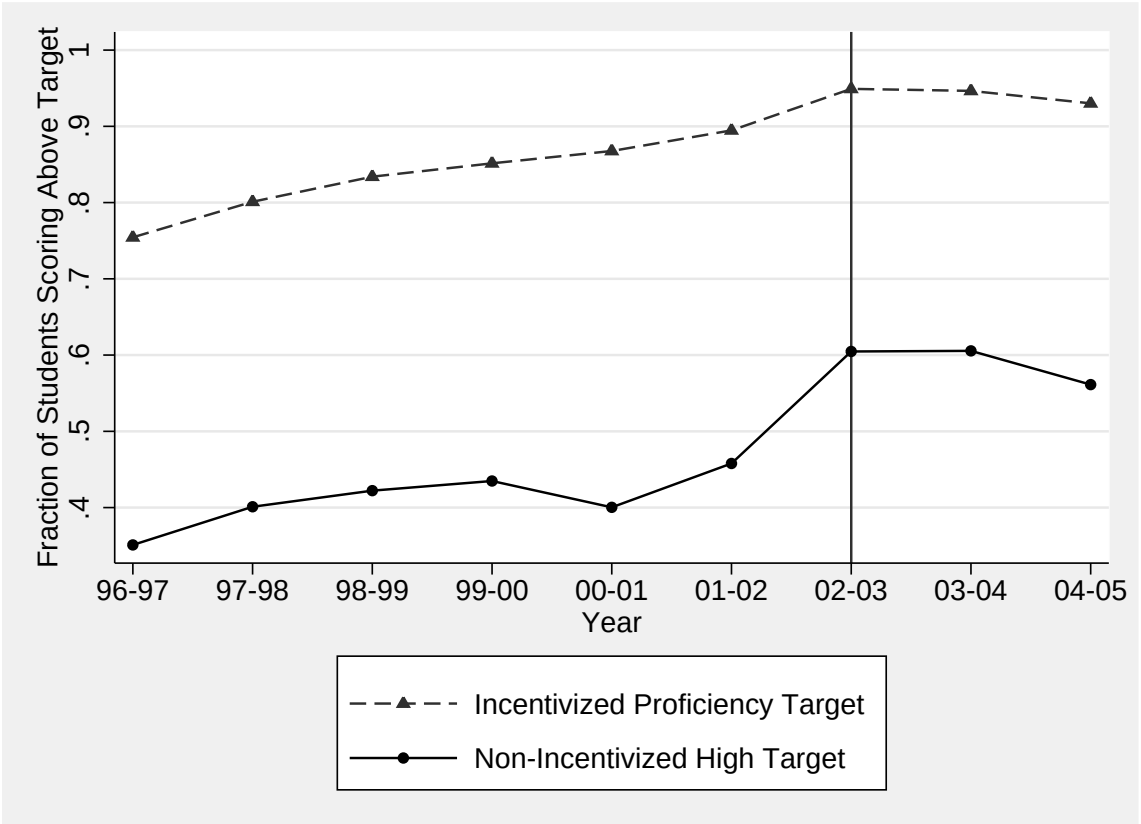
- Under the ABCs, a three-tier system of recognizing school performance in a low-stakes way operated:
  - low, middle, high
- When NCLB was introduced, the middle threshold under the pre-existing low-stakes scheme became the NCLB passing threshold
  - high target was not incentivized
- So what happened when NCLB came in?

Response to Incentivized and Non-Incentivized Targets



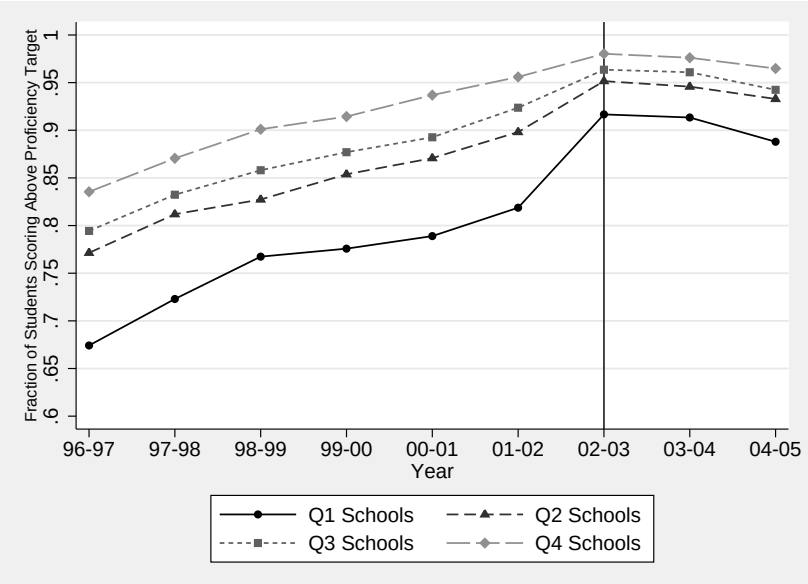
- In 2002-03, proportion of students achieving proficiency target increased (compared to 2001-02)

Response to Incentivized and Non-Incentivized Targets

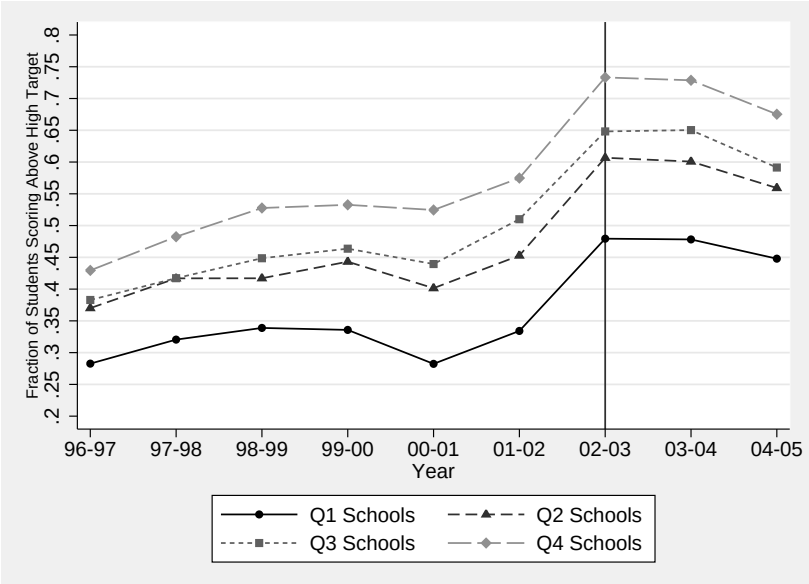


- In 2002-03, proportion of students achieving proficiency target increased (compared to 2001-02), but proportion achieving non-incentivized high target increased even more

Response to Incentivized and Non-Incentivized Targets (2)



Incentivized Proficiency Target



Non-Incentivized High Target

- NCLB proficiency-target response was particularly pronounced for schools most likely to be sanctioned
- High-target response relatively uniform across schools

Interpreting the Response

- Clear evidence that, just as schools sought to get students 'above' the NCLB threshold, so they also made strenuous efforts to get them above the high target
- **Question**: How might this fact be explained?
- One rationalization:
  - NCLB brought renewed attention to academic achievement, for all schools
  - because of extra salience, schools had **informal** incentive to boost achievement at higher levels than the proficiency mandate under NCLB, even though that would have dynamic consequences (along lines already rehearsed)
- In sum, plausible that objectives incorporate *informal* incentives to perform

The Objective

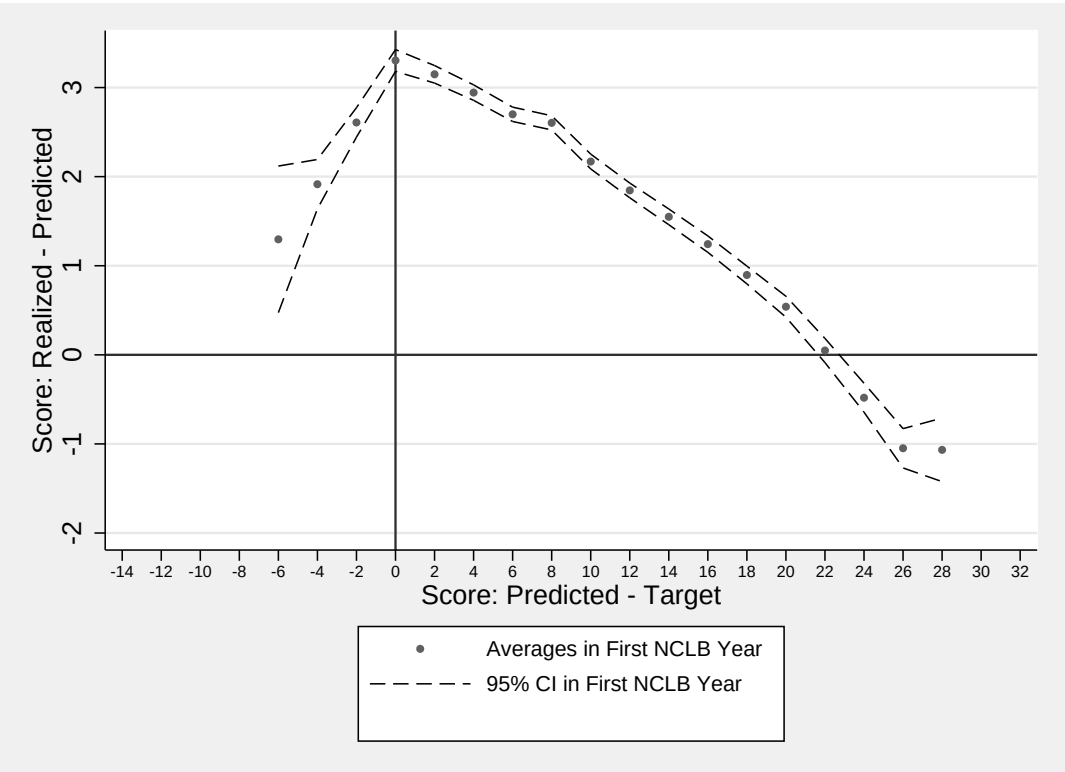
- Based on the evidence we have presented, objective should:
  - reflect agency at **teacher** (not school) level
  - focus on **effort setting**, allowing for student targeting
  - include **formal** incentives associated with NCLB's introduction, but ignore prior ABCs reform
  - include **informal** incentives – specifically, the high target
- Form embodying these criteria: $U = B(e_1, \ldots, e_{N_c}) - C(e_1, \ldots, e_{N_c})$
  - where formal and informal incentives affect **benefit** of effort:
    $B = b_M \sum_{i=1}^{N_c} \left[1 - F(y_M^T + d_M - \hat{y}_i - e_i)\right] + b_H \sum_{i=1}^{N_c} \left[1 - F(y_H^T + d_H - \hat{y}_i - e_i)\right]$
  - **cost** of effort with spillovers: $C = \frac{\psi}{2}\left[\sum_{i=1}^{N_c} e_i^2 + \theta\left(\sum_{j=1}^{N_c} e_j\right)^2\right]$

The Objective (2)

- Able to bring objective to the data and identify parameters of benefit and cost functions, with excellent model fit
  - see Macartney, McMillan and Petronijevic (2020a)

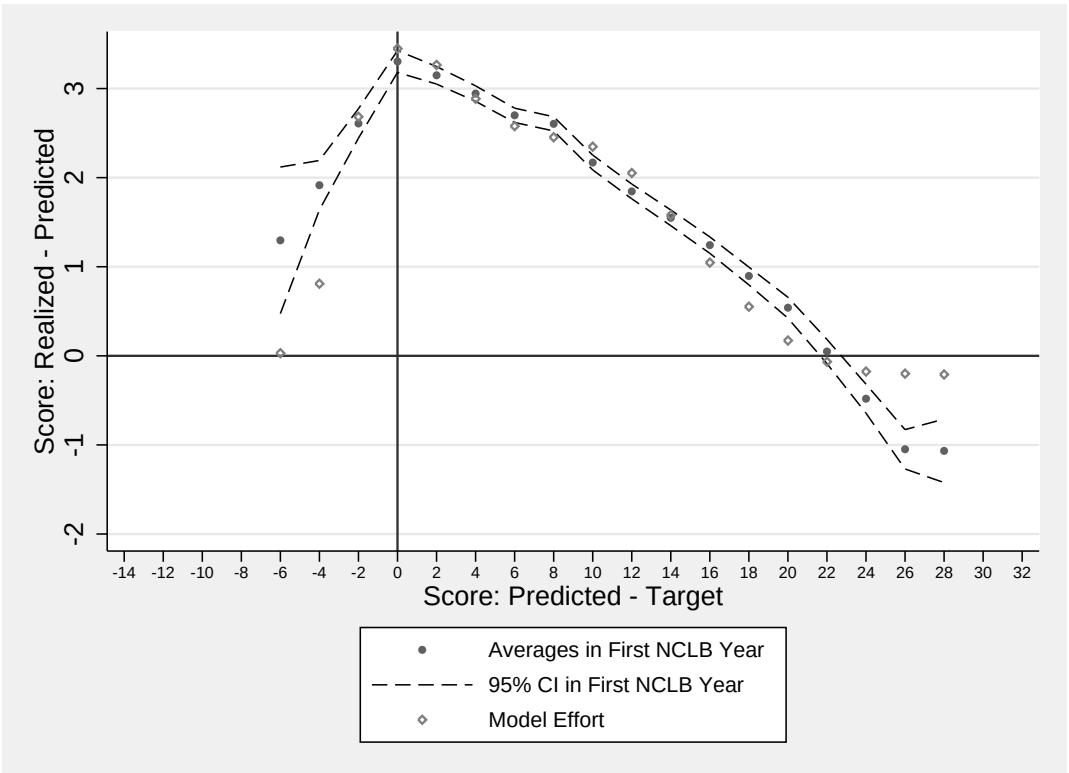
The Objective (2)

- Able to bring objective to the data and identify parameters of benefit and cost functions, with excellent model fit
  - see Macartney, McMillan and Petronijevic (2020a)



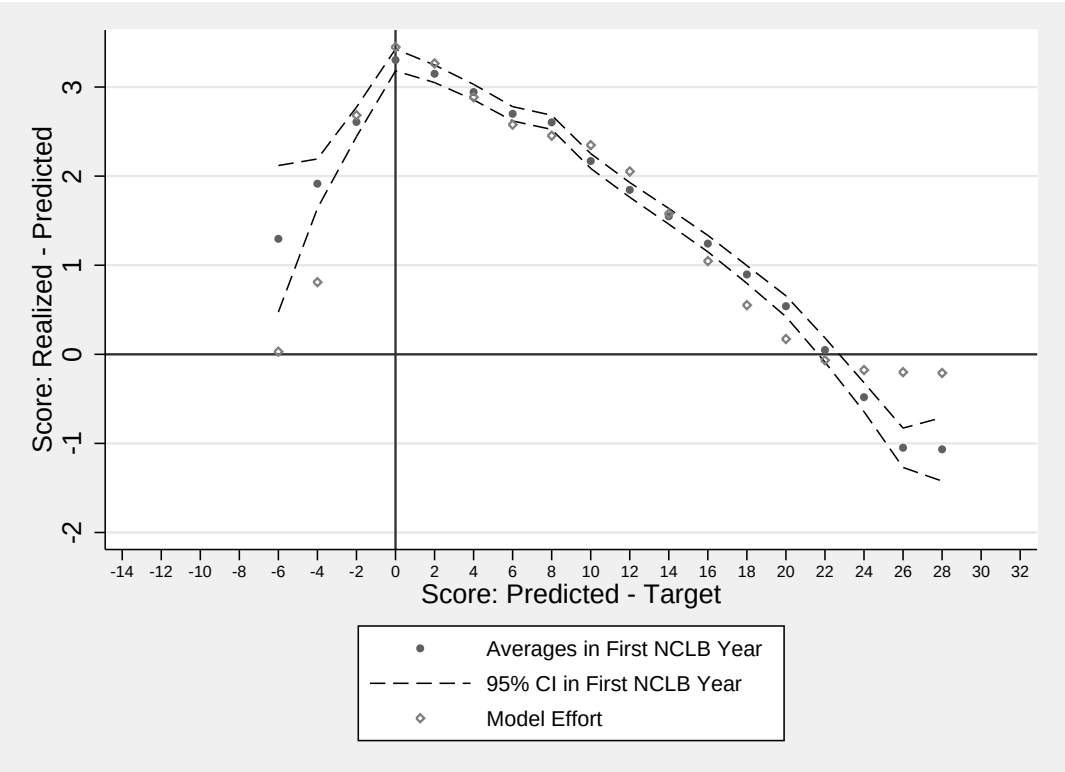
The Objective (2)

- Able to bring objective to the data and identify parameters of benefit and cost functions, with excellent model fit
  - see Macartney, McMillan and Petronijevic (2020a)



The Objective (2)

- Able to bring objective to the data and identify parameters of benefit and cost functions, with excellent model fit
  - see Macartney, McMillan and Petronijevic (2020a)



- Also able to separately identify teacher **ability** and **effort** inputs for first time
  - see Macartney, McMillan and Petronijevic (2020b)

Policy Implications

- Many informative exercises made possible having identified parameters and inputs of parsimonious model
- Not least, the following policy exercises:
  - Compare different incentive schemes in terms of their average effort and variance, including schemes yet to be enacted
    - sheds light on novel efficiency-inequality tradeoff that arises
    - (see 'incentive design' exercise in Macartney *et al.* 2020a)
  - Compare different types of policies – incentive-related policies relative to teacher hiring policies – in cost-benefit terms
    - (see Macartney *et al.* 2020b)
- And more ...

For further detail about how we leverage the facts discussed today, see Macartney, McMillan and Petronijevic (2020a, 2020b), available at hughmacartney.com

Thank you!

Questions and comments are most welcome.