Achievement Tradeoffs and No Child Left Behind (NCLB)

Dale Ballou and Matthew G. Springer

National Center on Performance Incentives
Peabody College of Vanderbilt University

August 12, 2009

Urban Institute
Washington, DC

*We thank the Northwest Evaluation Association for providing data for this study as well as the Smith Richardson Foundation and the federally-funded National Center on School Choice at Vanderbilt University for financial support.

Outline

1. What we are interested in learning.
2. Data, sample, and identification strategy.
3. Results.
4. Summary and implications.

1. What We are Interested in Learning

1. Previous Research on Distributional Effects of NCLB

- Conditions do not permit the conduct of a randomized field trial.
- Researchers have resorted to two strategies.
  - Pre- and post-NCLB accountability system comparisons.
    - Neal and Schanzenbach, 2009; Kreig, 2008.
  - Exploiting variation in the strength of incentives.
    - Schools facing sanctions (Rosaen, Schwartz, and Forbes, 2007).
    - Schools on either side of minimum n (Sims, 2007).
    - Students near the proficiency cutscore (Springer, 2007; Neal and Schanzenbach, 2008).
    - Study of instructional practices in failing and non-failing schools (Booher-Jennings, 2005; White and Rosenbaum, 2007).

*There are a number of studies that addressed distributional effects under pre-NCLB accountability programs. See, for example, Chakrabarti (2007) and Rouse et al (2007).

1.1. What Are We Interested in Learning?

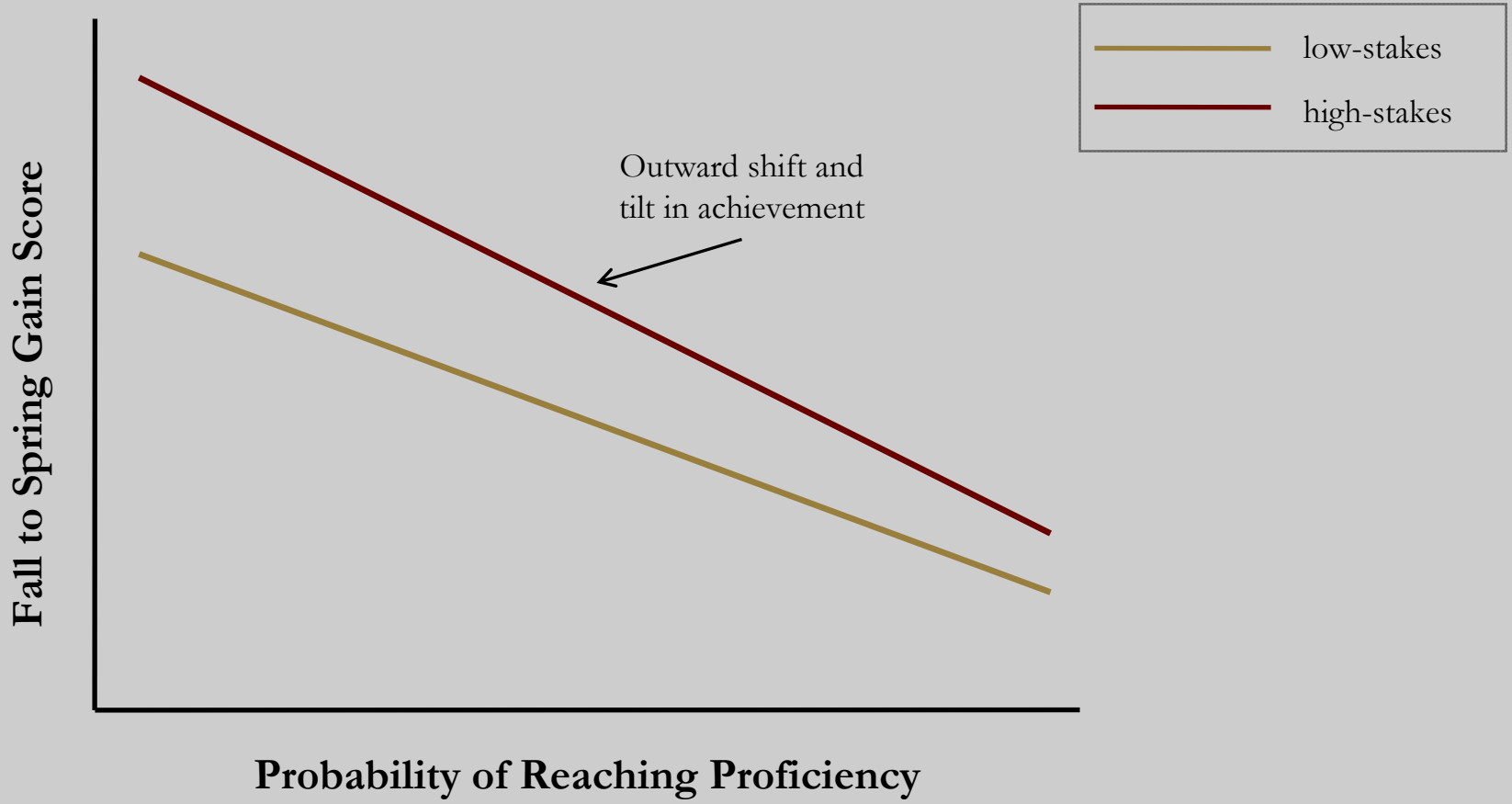
- **NCLB Main Effects.**
  - ***Does NCLB raise student achievement across the board? Is improvement targeted toward low achievers?***

- Urgency of Improvement Hypothesis.
  - Does the urgency with which schools need to focus on raising achievement of students comes at the expense of students that are already proficient or that are far below proficiency threshold?

- Bubble Student Hypothesis.
  - Do schools under particular pressure to make AYP focus on students near the margin and ignore those whose probability of passing is much higher or lower?

- Educational Triage Hypothesis (asymmetric).
  - Do schools treat students above the marginal student the same as students below the marginal student?

1.1.1. NCLB Main Effects



1.3. What Are We Interested in Learning?

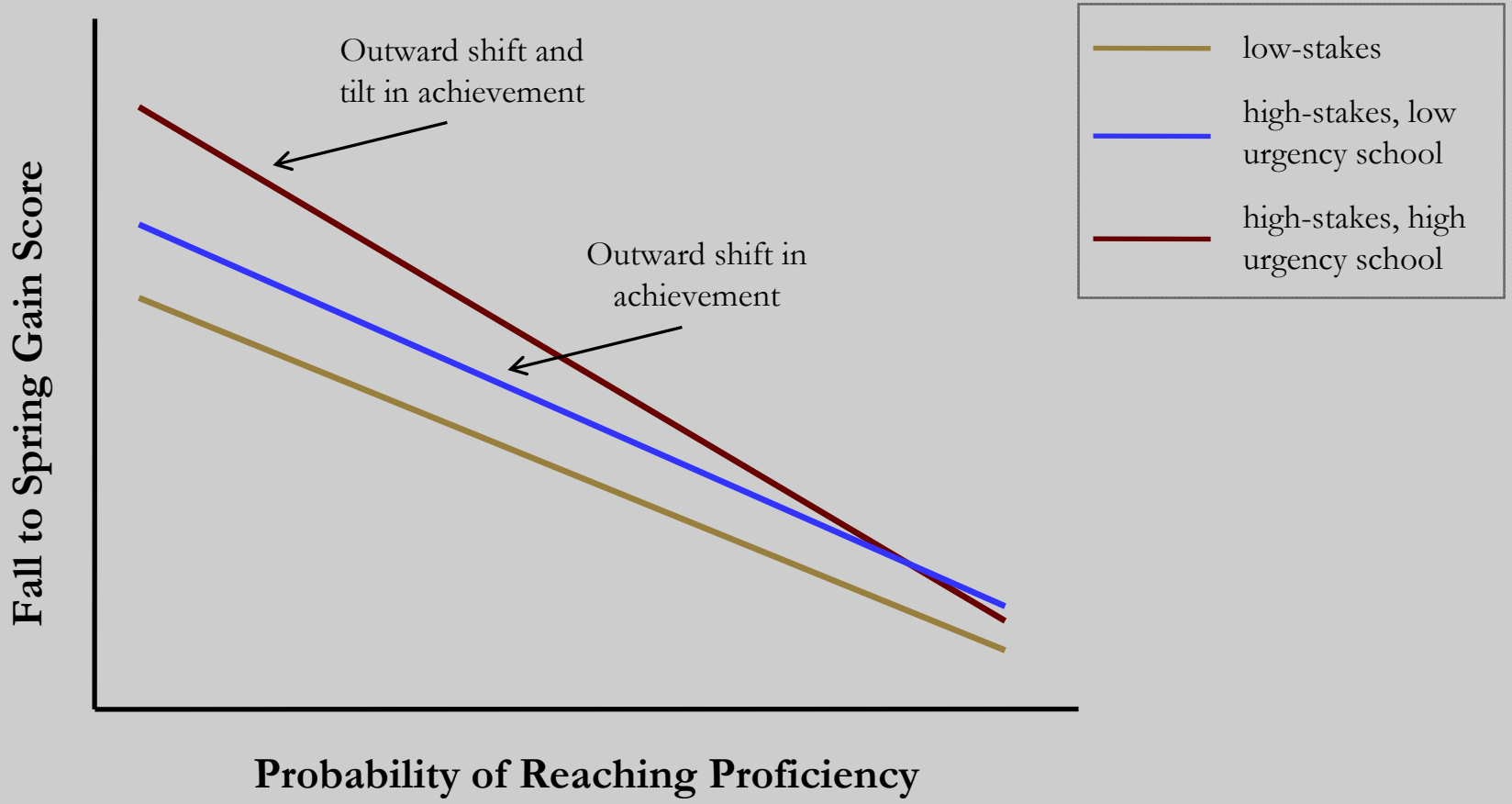
- NCLB Main Effects.
  - Does NCLB raise student achievement across the board? Is improvement targeted toward low achievers?

- **Urgency of Improvement Hypothesis.**
  - ***Does the urgency with which schools need to focus on raising achievement of students comes at the expense of students that are already proficient or that are far below proficiency threshold?***

- Bubble Student Hypothesis.
  - Do schools under particular pressure to make AYP focus on students near the margin and ignore those whose probability of passing is much higher or lower?

- Educational Triage Hypothesis (asymmetric).
  - Do schools treat students above the marginal student the same as students below the marginal student?

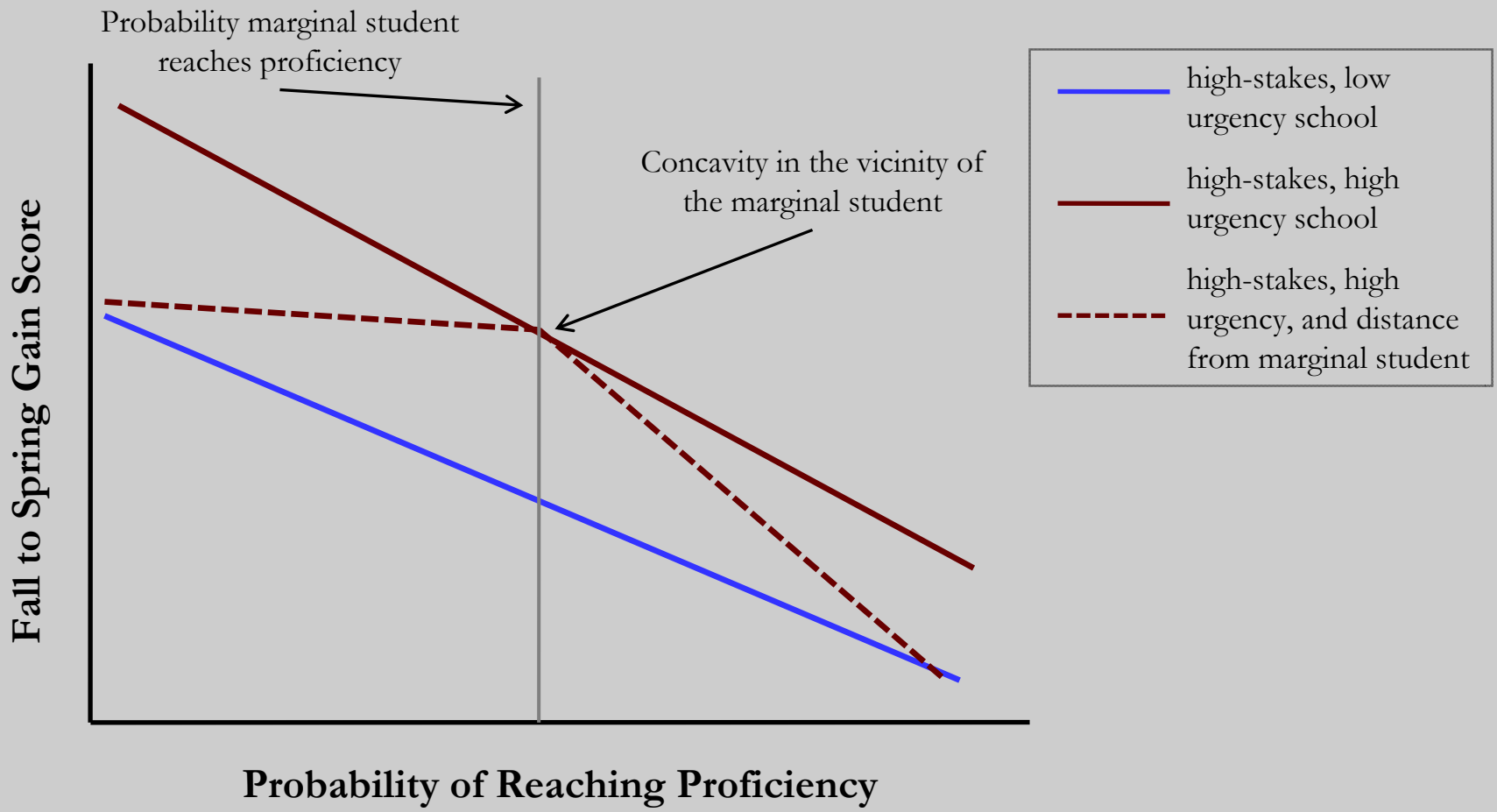
1.2.1. Urgency of Improvement Hypothesis



1.3. What Are We Interested in Learning?

- NCLB Main Effects.
  - Does NCLB raise student achievement across the board? Is improvement targeted toward low achievers?
- Urgency of Improvement Hypothesis.
  - Does the urgency with which schools need to focus on raising achievement of students comes at the expense of students that are already proficient or that are far below proficiency threshold?
- **Bubble Student Hypothesis.**
  - ***Do schools under particular pressure to make AYP focus on students near the margin and ignore those whose probability of passing is much higher or lower?***
- Educational Triage Hypothesis (asymmetric).
  - Do schools treat students above the marginal student the same as students below the marginal student?

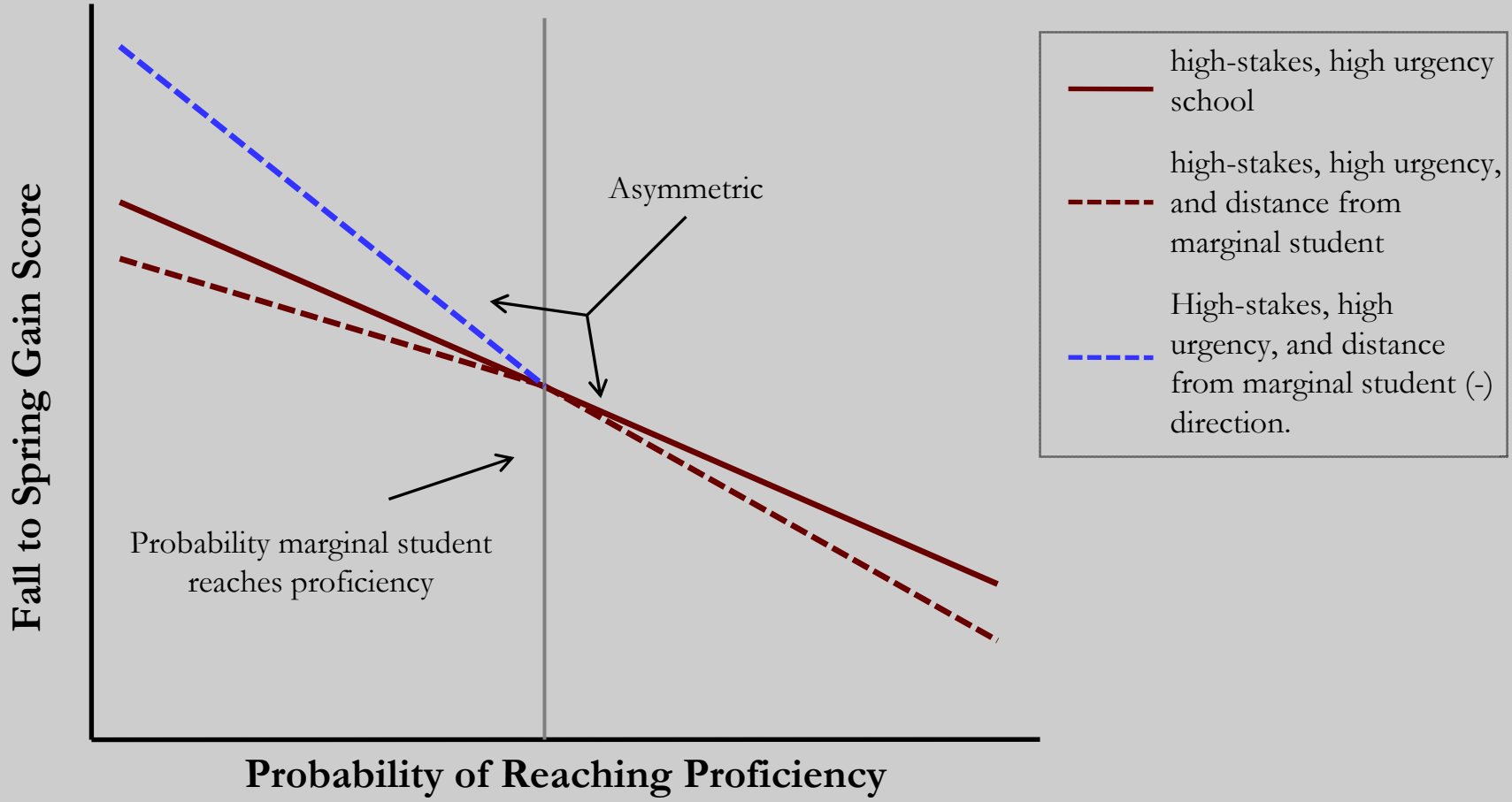
1.3.1. Model 3 - Bubble Student Hypothesis



1.4. What Are We Interested in Learning?

- NCLB Main Effects.
  - Does NCLB raise student achievement across the board? Is improvement targeted toward low achievers?
- Urgency of Improvement Hypothesis.
  - Does the urgency with which schools need to focus on raising achievement of students comes at the expense of students that are already proficient or that are far below proficiency threshold?
- Bubble Student Hypothesis.
  - Do schools under particular pressure to make AYP focus on students near the margin and ignore those whose probability of passing is much higher or lower?
- **Educational Triage Hypothesis (asymmetric).**
  - **Do schools treat students above the marginal student the same as students below the marginal student?**

1.4.1. Educational Triage Hypothesis



3. Data, Sample, and Identification Strategy

3.1. Data

- Data from Northwest Evaluation Association's Growth Research Database (GRD). Longitudinal, student level test score data from 3,400 districts in 45 states.
- Tests used for diagnostic and formative assessment purposes, results should be unaffected by teachers and administrators trying to game the system (except Idaho).
- Vertically linked to measure student growth (RIT scale, using Rasch scaling methodologies).
- Availability of fall-to-spring gain scores in mathematics and reading, can calculate intra-year gains.
- Schools receive feedback on students shortly after students take test.

3.2. Sample

- Use data from districts in seven states between 2002-03 and 2005-06 school years.
  - Arizona
  - Colorado
  - Idaho
  - Michigan
  - Minneapolis
  - Minnesota
  - Wisconsin
- Approximately 2,000,000 student observations.
- NWEA districts and schools compared to districts and schools in state with similar grades.

3.3.1. Identification Strategy

- Identify an NCLB effect by comparing outcomes across low- and high-stakes years within a grade.
- Not all grades counted toward AYP when NCLB took effect in 2002-03 school year.
- States were required to test in reading and mathematics once at the elementary level and once at the middle school level.
- Starting in 2005-06 school year, schools were required to test in grades 3 through 8 in reading and mathematics.
- Calculated probability student i scores proficient on next test administration

3.3.2. Low- and High-Stakes Grades

Table 1. Number of Observations in Indiana by Grade and Year*

Grade	2002-03	2003-04	2004-05	2005-06	All Years
3rd	***21,570***	***23,259***	***23,106***	***24,391***	92,326
4th	25,931	23,830	***22,914***	***24,257***	96,932
5th	26,249	24,455	***23,212***	***24,548***	98,464
6th	***22,819***	***22,927***	***23,646***	***24,218***	93,610
7th	25,795	23,066	***21,684***	***23,497***	94,042
8th	21,108	20,887	***21,861***	***22,468***	86,324

*Entries in **boldface** and *italics* are high-stakes grades and years.

3.3.3. Probability Student *i* Scores Proficient on Next Test Administration

$$\hat{\pi}_{igst} = Prob\ (c_{gst} - (f_{igst} + m_{igst}))$$

- Identify an NWEA equivalent to each state's proficiency cutscore ($\boldsymbol{c_{gst}}$).
- Model the probability that student *i* reaches proficiency on the next state test as the probability that student *i's* fall test score on the NWEA assessment ($\boldsymbol{f_{igst}}$) plus the expected gain for that student ($\boldsymbol{m_{igst}}$) exceeds the NWEA cutscore-equivalent.
- $\boldsymbol{m_{igst}}$ is the mean gain for a given state and grade within the sample period, obtained by regressing the observed fall score and next NWEA score closest to high-stakes administration on student *i's* fall score and a set of state and grade effects.
- NWEA pass probabilities are highly correlated with our pass probabilities.

3.4. Identification Strategy

- Testing dates vary between fall and spring assessments by state and year.
  - Mean days between tests is 135 with standard deviation of approx. 7 days.

- Create annualized gain score.
  - (Student's test score gain / # of days between fall and spring assessment) * 180 days.

- Restrict sample to schools that test at least 80 percent of their students.

4. Results

4.1. Summary of Results

- What we are interested in learning
  - Main effect
  - Urgency to improve
  - Bubble kids
  - Educational triage

- Relationship between annualized student test score gains and fall test scores
  - Urgency to improve

- Series of regression models
  - Urgency to improve

4.1. Relationship between Annualized Gains and Fall Scores

$$Y_{istg} = \Sigma_{s=1,7}\,\mu_{sg} + \Sigma_{t=1,4}\,\mu_{tg} + \Sigma_{j=1,2}\,\psi_{jg}(F_{istg} - C_{stg}) + \delta\, hs_{gst} + u_{istg}$$

g = grade (3…7), j = hs, ls

s = state (1,...7)

$\mathbf{Y}_{istg}$ is fall-to-spring gain for student i in state s and year t.

$\boldsymbol{\psi}_{jg}(\mathbf{F}_{istg} - \mathbf{C}_{stg})$ is an unknown function of the fall score/cutscore difference.

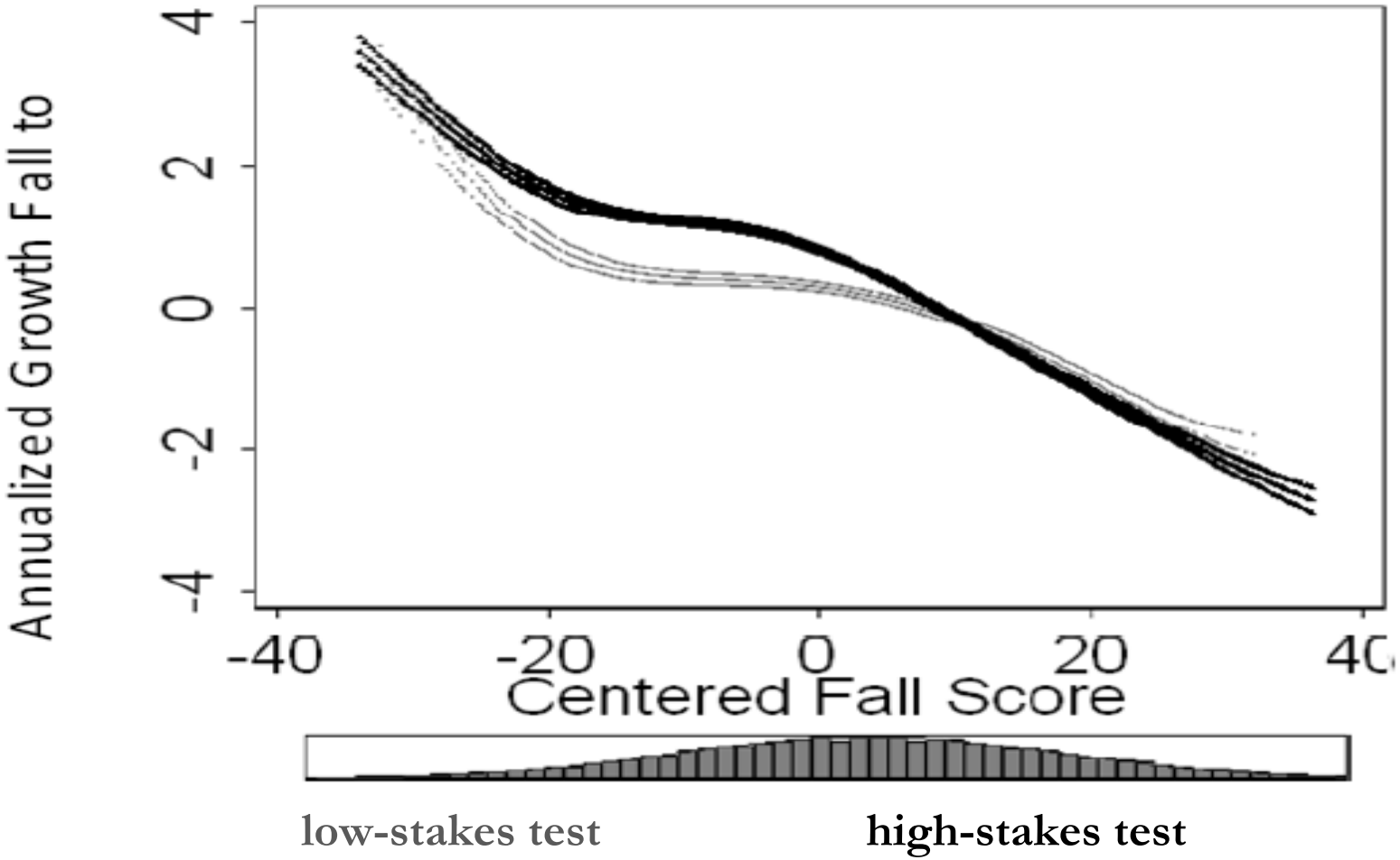
$\boldsymbol{\psi}_{jg}$ are approximated by cubic splines, with penalties for departures from

smoothness.

$\boldsymbol{hs}_{gst}$ dummy variable indicating whether year t was a high-stakes year for grade g in state s.

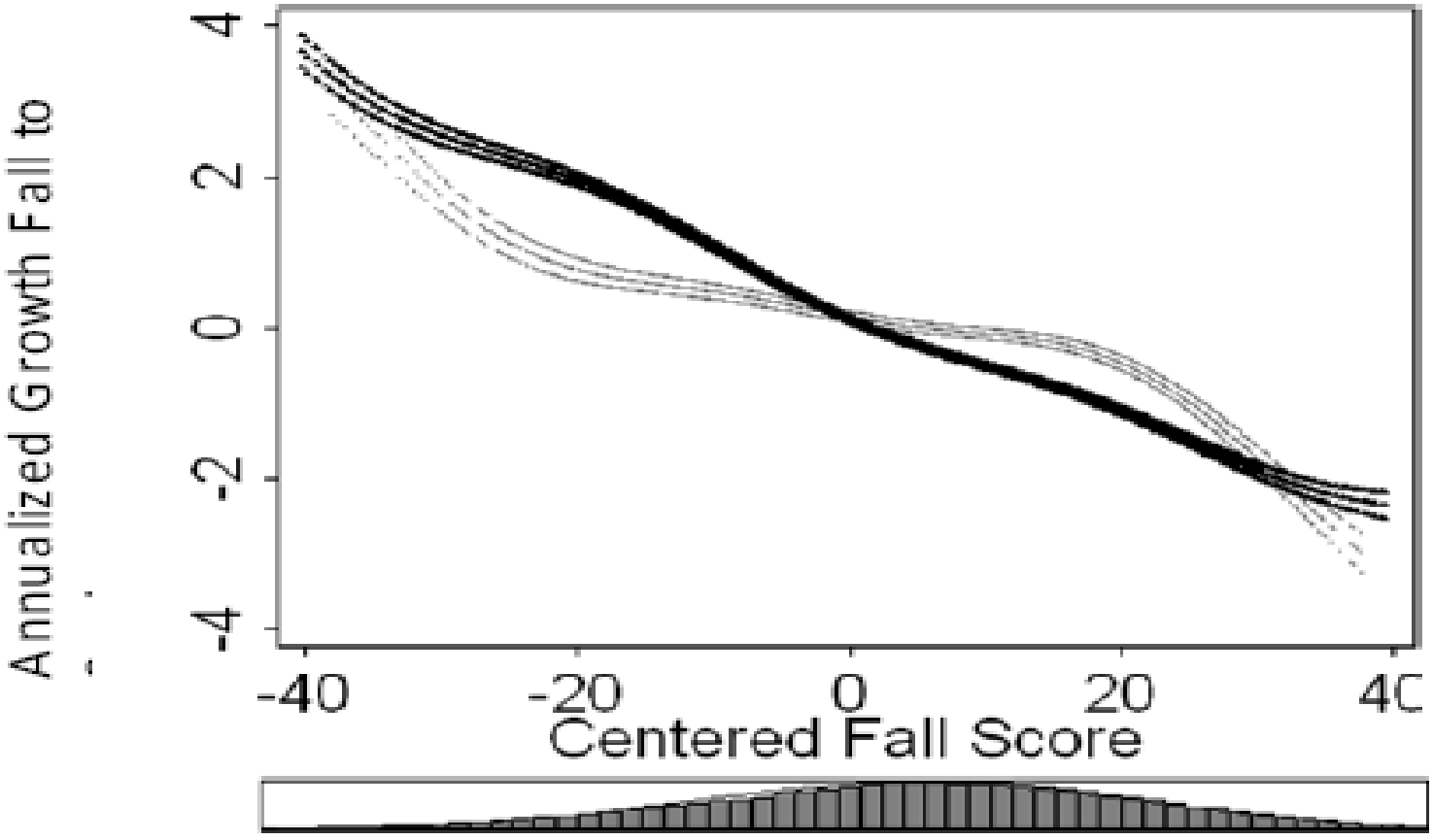
4.1.1. Relationship between Annualized Gains and Fall Scores

Figure 1c. Effect of NCLB, All States, Grade 5



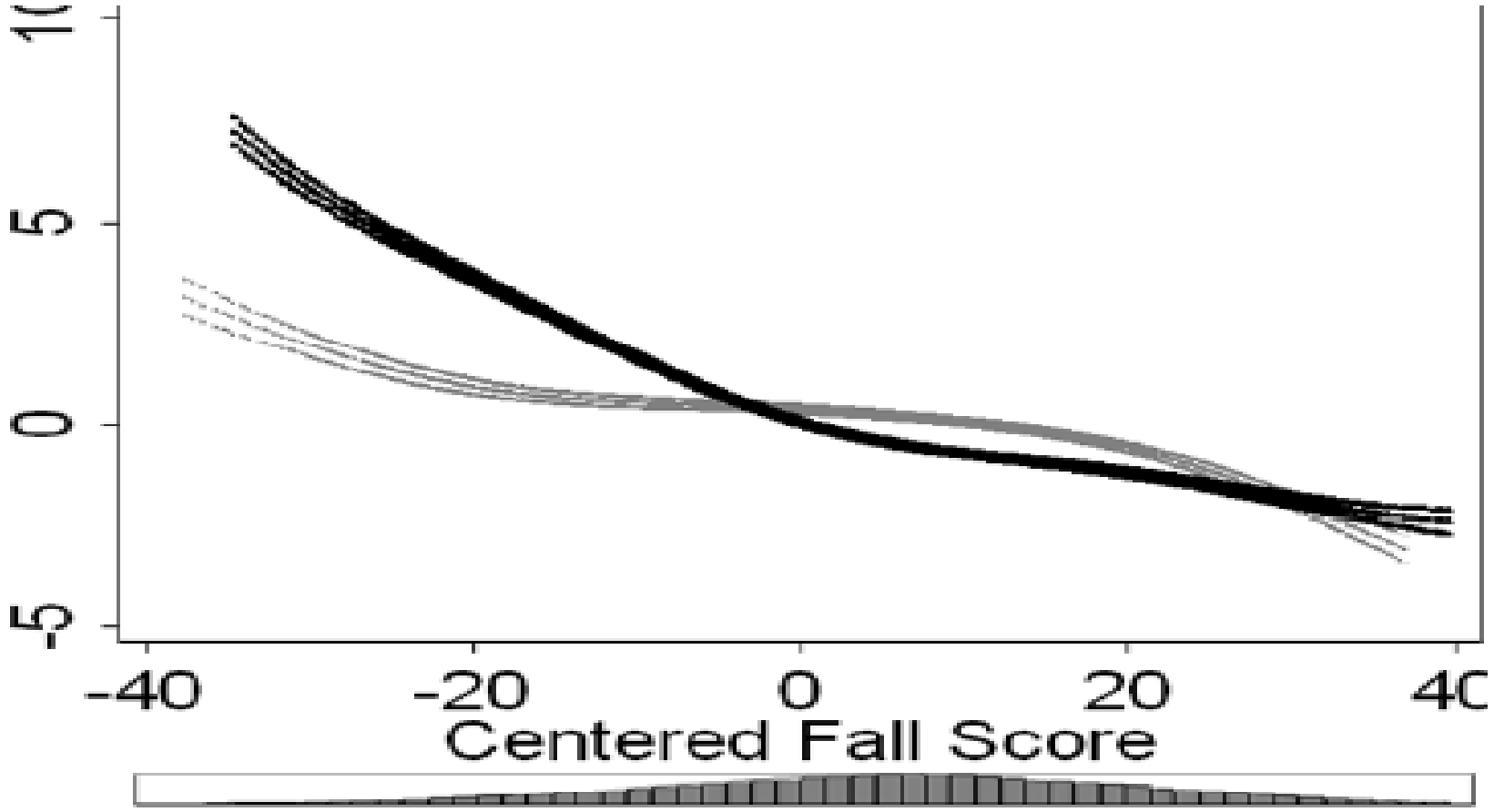
4.1.3. Relationship between Annualized Gains and Fall Scores

Figure 1e. Effect of NCLB, All States, Grade 7



4.1.2. Relationship between Annualized Gains and Fall Scores

Figure 1e. Effect of NCLB, Grade 7



low-stakes test **high-stakes test**

4.2.1. NCLB Main Effect

$$Y_{istg} = \sum \mu_{sg} + \sum \mu_{tg} + \hat{\pi}_{igst} + hs_{gst} + \hat{\pi}_{igst} \times hs_{gst} + u_{istg}$$

$\mathbf{Y}_{istg}$ is fall-to-spring gain for student *i* in state *s* and year *t*.

$\hat{\pi}_{igst}$ is the predicted probability that student *i* in grade *g*, state *s*, and year *t* achieves proficiency when next tested.

$\boldsymbol{hs}_{gst}$ dummy variable indicating whether year *t* was a high-stakes year for grade *g* in state *s*.

$\hat{\pi}_{igst} \times hs_{gst}$ interaction of high-stakes indicator with student *i's* probability of passing. A negative slope would indicate instruction focused on lower-achieving students when NCLB stakes are high.

$\boldsymbol{\mu}_{sg}$ is a state fixed effect. $\boldsymbol{\mu}_{tg}$ is a year fixed effect. $\boldsymbol{u}_{istg}$ is random disturbance term.

4.2.2. Model 1

Variable	Hypoth. Sign	Grade					
		3rd	4th	5th	6th	7th	8th
Probability of achieving proficiency. ($\hat{\pi}_{igst}$)	(-)	-7.01 (0.11)	-3.57 (0.07)	-2.08 (0.08)	-1.83 (0.07)	-1.86 (0.08)	-3.04 (0.11)
High-stakes grade. (hs_{gst})	(+)	-1.87 (0.10)	0.93 (0.09)	0.76 (0.09)	0.66 (0.08)	1.00 (0.08)	0.71 (0.10)
Interaction of probability of achieving proficiency and high-stakes grade. ($\hat{\pi}_{igst} \times hs_{gst}$)	(+)	1.19 (0.12)	-1.47 (0.09)	-0.98 (0.10)	-1.30 (0.08)	-1.50 (0.09)	-1.35 (0.12)

4.3.1. Introduce Summary Measure to Proxy for Pressure a School Faces to Improve Student Performance

$$\hat{\pi}_{Mst}$$

$\hat{\pi}_{Mst}$ is the predicted probability that marginal student *M* in grade *g*, state *s*, and year *t* achieves proficiency when next tested.

- We first calculate the number of students who must reach proficiency for the school to make AYP.
- We then use the distribution of fall test scores to identify the marginal student – the student ranked M^{th} in a school where M students must score proficient or above.

4.3.2. Model 2

$$Model\ 1 + \hat{\pi}_{Mst} + (1 - \hat{\pi}_{Mst}) \times hs_{gst} + (1 - \hat{\pi}_{Mst}) \times hs_{gst} \times \hat{\pi}_{igst}$$

$\hat{\pi}_{Mst}$ is the predicted probability that marginal student *M* in grade *g*, state *s*, and year *t* achieves proficiency when next tested.

$(1 - \hat{\pi}_{Mst})$ is a measure of the urgency with which schools need to focus on raising achievement.

$(1 - \hat{\pi}_{Mst}) \times hs_{gst}$ to the extent this is an impetus to improve overall efficiency, the expected sign is positive.

$(1 - \hat{\pi}_{Mst}) \times hs_{gst} \times \hat{\pi}_{igst}$ if greater urgency causes schools to focus more on raising achievement of lower-performing students, the expected sign is negative.

4.3.3. Model 2

Variable	Hypoth. Sign	Grade					
		3rd	4th	5th	6th	7th	8th
Probability of marginal student achieving proficiency. ($\hat{\pi}_{Mst}$)	(+)	2.23 (0.15)	1.81 (0.11)	1.40 (0.12)	2.67 (0.13)	3.49 (0.17)	0.61 (0.25)
Interaction of urgency measure with high-stakes grade. ($hs_{gst} \times (1 - \hat{\pi}_{Mst})$)	(+)	-2.22 (0.20)	-2.67 (0.22)	-3.20 (0.21)	-2.84 (0.20)	-1.82 (0.22)	-3.22 (0.28)
Interaction of probability of achieving proficiency with high-stakes grade and urgency measure. ($hs_{gst} \times \hat{\pi}_{igst} \times (1 - \hat{\pi}_{Mst})$)	(-)	2.06 (0.20)	3.86 (0.25)	4.81 (0.19)	2.04 (0.17)	3.12 (0.16)	0.54 (0.16)

4.4.1. Model 3

$$Model\ 2 + \left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|\left(1 - \hat{\pi}_{Mst}\right) + \left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|\left(1 - \hat{\pi}_{Mst}\right) \times hs_{gst}$$

$\left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|\left(1 - \hat{\pi}_{Mst}\right)$ absolute difference of student *i's* probability of scoring proficient and the marginal student's probability, times our measure of urgency.

$\left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|\left(1 - \hat{\pi}_{Mst}\right) \times hs_{gst}$ if schools under particular pressure to make AYP focus instruction on students near the margin and ignore those who probability of passing is much higher or much lower, the expected sign is

negative

in high-stakes grade-year combinations.

4.4.2. Bubble Student Hypothesis

Variable	Hypoth. Sign	Grade					
		3rd	4th	5th	6th	7th	8th
Interaction of urgency with distance from marginal student. $\lvert\hat{\pi}_{igst} - \hat{\pi}_{Mst}\rvert(1 - \hat{\pi}_{Mst})$	(?)	0.17 (0.31)	-0.86 (0.22)	-0.65 (0.34)	-0.97 (0.34)	0.08 (0.60)	-1.64 (0.40)
Interaction of urgency, distance from marginal student, and high-stakes grade. $\lvert\hat{\pi}_{igst} - \hat{\pi}_{Mst}\rvert(1 - \hat{\pi}_{Mst}) \times hs_{gst}$	(-)	-0.26 (0.38)	1.21 (0.37)	-2.69 (0.42)	4.04 (0.42)	3.36 (0.64)	8.74 (0.46)

4.4.1. Model 4 - Educational Triage Hypothesis

$$Model\ \ 3 + \left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|(1 - \hat{\pi}_{Mst}) \times \delta_{pos} + \left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|(1 - \hat{\pi}_{Mst}) \times \delta_{neg}$$
$$+ \left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|(1 - \hat{\pi}_{Mst}) \times \delta_{pos} \times hs_{gst} + \left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|(1 - \hat{\pi}_{Mst}) \times \delta_{neg} \times hs_{gst}$$

$\left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|(1 - \hat{\pi}_{Mst}) \times \delta_{pos}$

$\left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|(1 - \hat{\pi}_{Mst}) \times \delta_{neg}$

Relax the assumption that students well above the marginal student are treated the same as students well below the marginal student.

$\hat{\pi}_{igst} - \hat{\pi}_{Mst}\left|(1 - \hat{\pi}_{Mst}) \times \delta_{pos} \times hs_{gst}\right.$

$\left|\hat{\pi}_{igst} - \hat{\pi}_{Mst}\right|(1 - \hat{\pi}_{Mst}) \times \delta_{neg} \times hs_{gst}$

Under the triage hypothesis, interactions of distance from the marginal student with urgency and the high-stakes indicator will have a negative effect on achievement.

4.4.2. Educational Triage Hypothesis

Variable	Hypoth. Sign	Grade					
		3rd	4th	5th	6th	7th	8th
Interaction of urgency with positive distance from marginal student. $\left\|\hat{\pi}_{igst}-\hat{\pi}_{Mst}\right\|(1-\hat{\pi}_{Mst})\times\delta_{pos}$	(**0 or ?**)	0.09 (0.37)	-3.46 (0.28)	-0.84 (0.37)	-0.84 (0.37)	-0.23 (0.66)	0.65 (0.66)
Interaction of urgency with negative distance from marginal student. $\left\|\hat{\pi}_{igst}-\hat{\pi}_{Mst}\right\|(1-\hat{\pi}_{Mst})\times\delta_{neg}$	(**0 or ?**)	0.68 (1.43)	11.08 (0.82)	0.75 (0.99)	0.75 (0.99)	0.98 (0.92)	-7.01 (1.31)
Variable 1, interacted with high-stakes grade. $\left\|\hat{\pi}_{igst}-\hat{\pi}_{Mst}\right\|(1-\hat{\pi}_{Mst})\times\delta_{pos}\times hs_{gst}$	(**-**)	0.98 (0.48)	3.38 (0.44)	-2.88 (0.48)	-2.88 (0.48)	4.60 (0.73)	5.24 (0.75)
Variable 2, interacted with high-stakes grade. $\left\|\hat{\pi}_{igst}-\hat{\pi}_{Mst}\right\|(1-\hat{\pi}_{Mst})\times\delta_{neg}\times hs_{gst}$	(**-**)	-3.45 (1.51)	-9.49 (1.09)	-3.06 (1.14)	-3.06 (1.14)	0.80 (1.03)	15.75 (1.37)

5. Summary and Implications

5. Summary and Implications

- Small increase in test scores in high-stake grade-year combinations (NCLB main effect).
- Small "redistribution" of achievement gains away from above-average students toward below-average students. Redistribution is more likely to take place in schools that are relatively assured of making AYP (Urgency of improvement).
- No evidence that students near the margin are learning more than students far away from threshold (Bubble students).
- No evidence of schools writing off the lowest-achieving students (Educational triage).