# The Dynamic Effects of Educational Accountability

Hugh Macartney
Duke University

Federal Reserve Bank of Cleveland
Policy Summit
June 2012

# Introduction

- Important incentive issue in economics:
  - if rewards condition on past behavior, forward-looking agents should take future rewards into account
  - may be less responsive today to avoid more onerous future targets - the ‘ratchet effect’

# Introduction

- Important incentive issue in economics:
  - if rewards condition on past behavior, forward-looking agents should take future rewards into account
  - may be less responsive today to avoid more onerous future targets - the 'ratchet effect'
- Key questions:
  1. Does such dynamic gaming occur in practice?
  2. If so, how important is it?
  3. What can be done to refine incentives?

# Introduction

- Important incentive issue in economics:
  - if rewards condition on past behavior, forward-looking agents should take future rewards into account
  - may be less responsive today to avoid more onerous future targets - the 'ratchet effect'
- Key questions:
  1. Does such dynamic gaming occur in practice?
  2. If so, how important is it?
  3. What can be done to refine incentives?
- Consider educational accountability reforms where ratchet effects may arise

# Educational Accountability

- Recent educational reforms have held teachers more accountable for student achievement
  - many states now attach pecuniary rewards to scores by awarding bonus pay to teachers who exceed a set target

# Educational Accountability

- Recent educational reforms have held teachers more accountable for student achievement
  - many states now attach pecuniary rewards to scores by awarding bonus pay to teachers who exceed a set target
- Key aspect of educational accountability:
  - choosing the target

# Educational Accountability

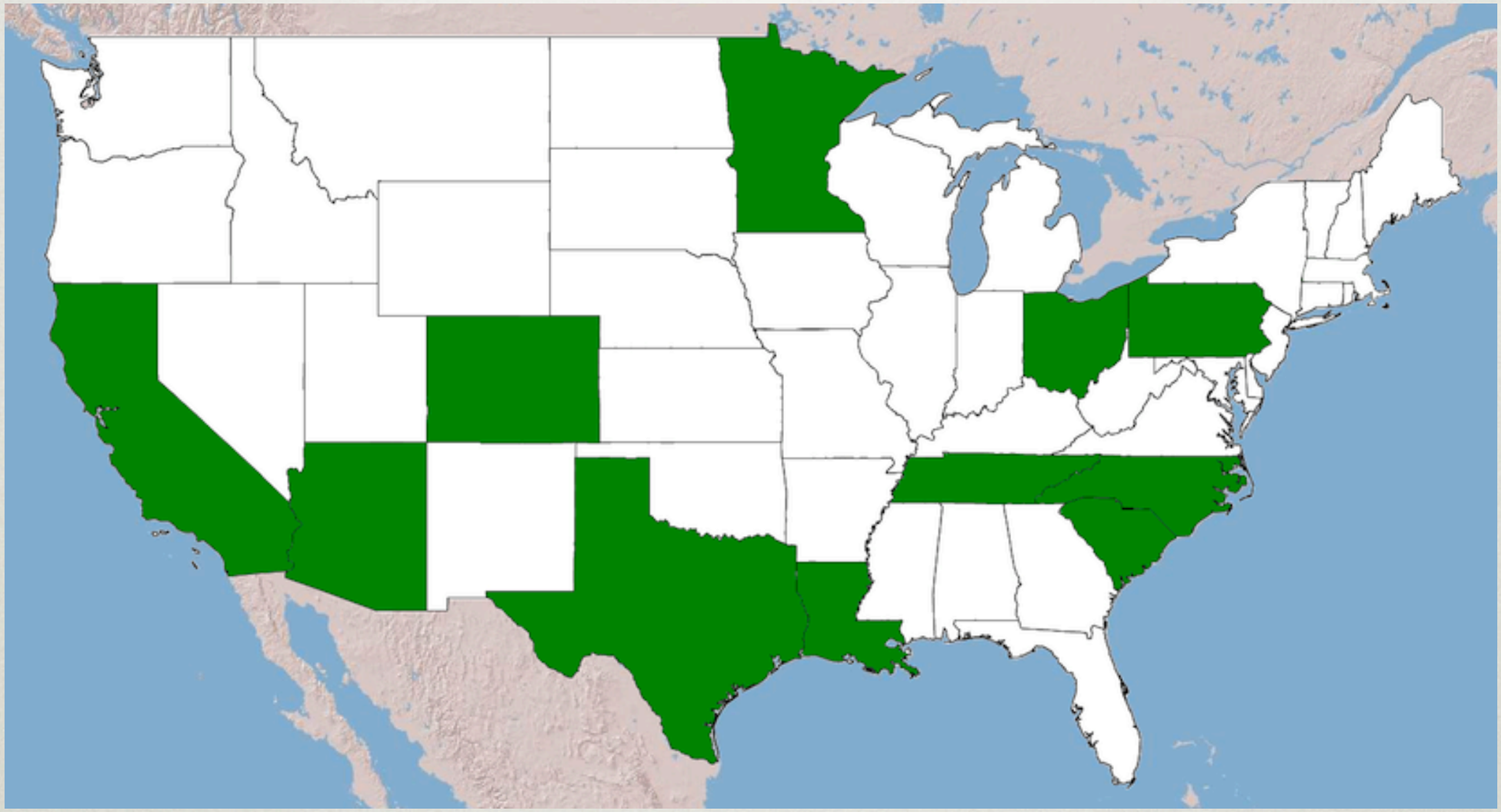
- Recent educational reforms have held teachers more accountable for student achievement
  - many states now attach pecuniary rewards to scores by awarding bonus pay to teachers who exceed a set target
- Key aspect of educational accountability:
  - choosing the target
- Two rival methods:
  1. Fixed scheme: $b(y_g - \alpha_g)$

# Educational Accountability

- Recent educational reforms have held teachers more accountable for student achievement
  - many states now attach pecuniary rewards to scores by awarding bonus pay to teachers who exceed a set target
- Key aspect of educational accountability:
  - choosing the target
- Two rival methods:
  1. Fixed scheme: $b(y_g - \alpha_g)$
  2. Value-added scheme: $b(y_g - \alpha y_{g-1})$

# Educational Accountability

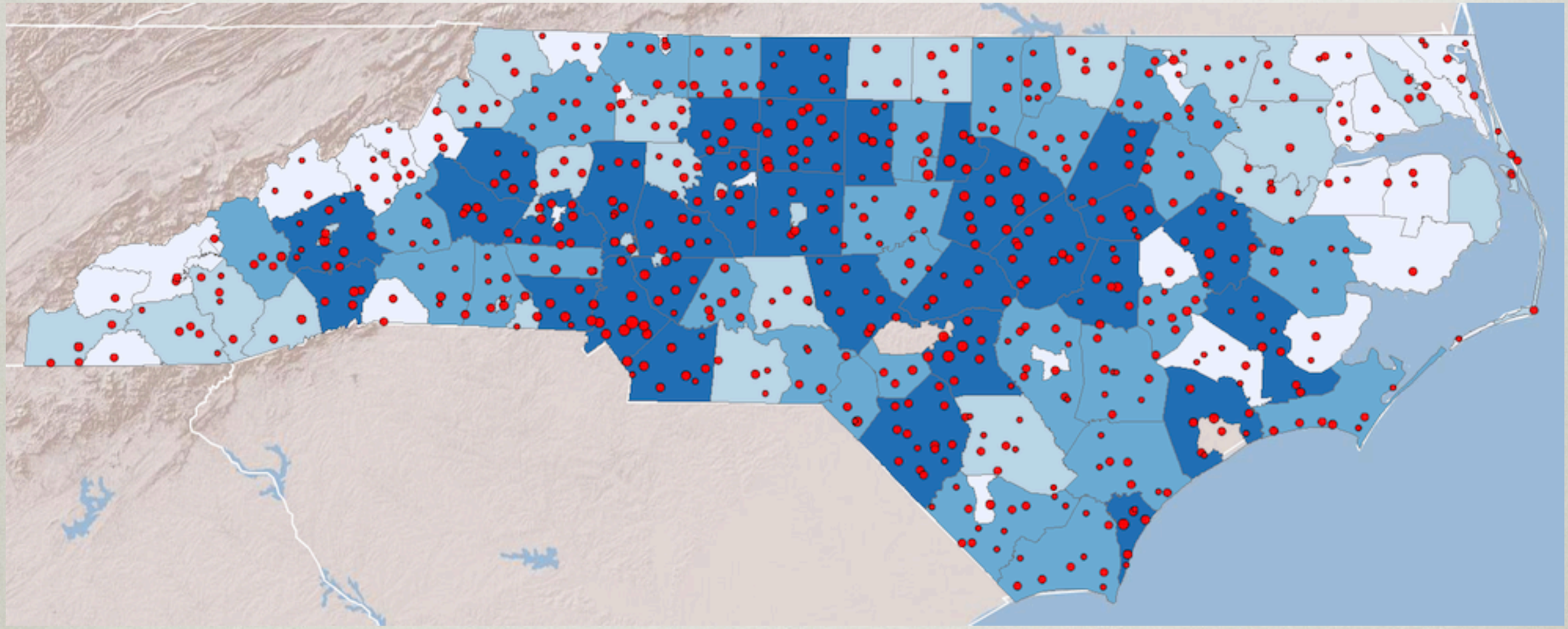
- Value-added increasingly widespread



- advantage: adjusts for unobserved heterogeneity
- disadvantage: targets are potentially manipulable

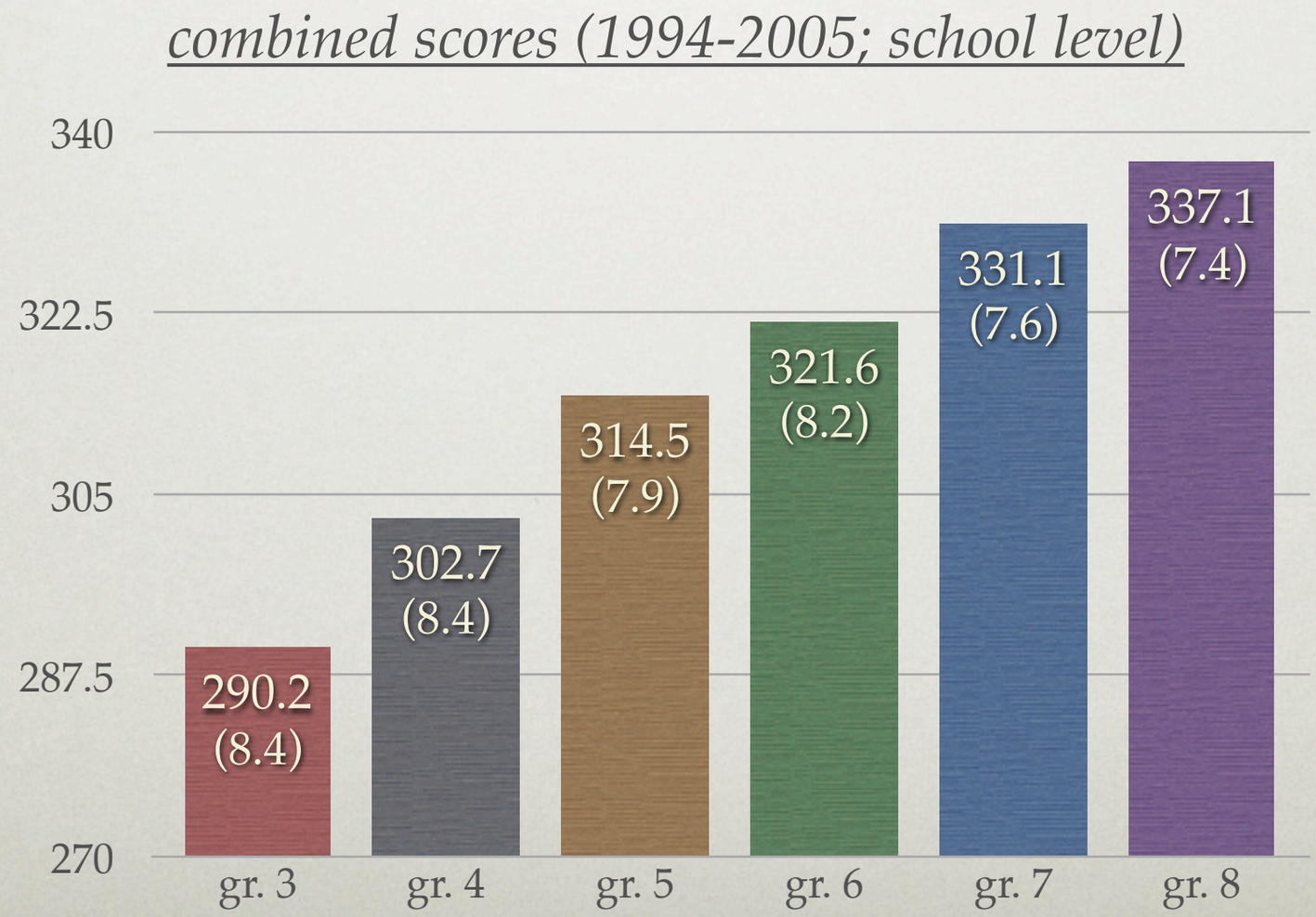
# Educational Accountability

- An instance of such a scheme:
  - the one covering all elementary schools in North Carolina beginning in 1996-97



# The NC Reform

- Based on standardized test scores
  - introduced in 1992-93
  - developmental scale in math and reading (gr. 3-8)
  - comparable across grades and years
  - each point designed to measure same amount of learning



# The NC Reform

- Consists of monetary rewards for outperformance
  - teachers and principal each receive $1500 or $750 bonus if <u>school-level</u> score exceeds high- or expected-growth target
  - score and target are composites of grade (gr. 3-8) and subject (reading & math) quantities averaged across students
  - targets condition on <u>student</u> prior test scores:
    $\hat{y}_{gt} = \alpha y_{g\text{-}1t\text{-}1}$ *(for grade g and year t)*

# The NC Reform

- Consists of monetary rewards for outperformance
  - teachers and principal each receive $1500 or $750 bonus if school-level score exceeds high- or expected-growth target
  - score and target are composites of grade (gr. 3-8) and subject (reading & math) quantities averaged across students
  - targets condition on student prior test scores:
    $\hat{y}_{gt} = \alpha y_{g\text{-}1t\text{-}1}$ *(for grade g and year t)*
- Since school-level targets condition on student prior test scores:
  - school's investment in grade g children today affects the target it faces tomorrow when they are in grade g+1
  - makes NC reform environment a suitable place to look for instances of dynamic gaming

# Data

- Rich longitudinal dataset to investigate dynamic incentive issues:
  - student-level scores for the years 1993-94 through 2004-05 and grades 3 through 8
    - all tested grades in pre- and post-reform periods
  - track students, teachers and schools over time using unique identifiers
  - student, teacher and school characteristics
  - important: school grade span (e.g. K-5, K-8)

# Descriptive Statistics

- School level (1994-2005)

Student Variables

| Variable | Mean | St. Dev. |
|---|---|---|
| Pared - No HS | 0.11 | 0.10 |
| Pared - HS Graduate | 0.45 | 0.17 |
| Pared - Trade School | 0.09 | 0.09 |
| Pared - Com. College | 0.12 | 0.09 |
| Pared - 4yr College | 0.19 | 0.15 |
| Pared - Grad Degree | 0.05 | 0.07 |

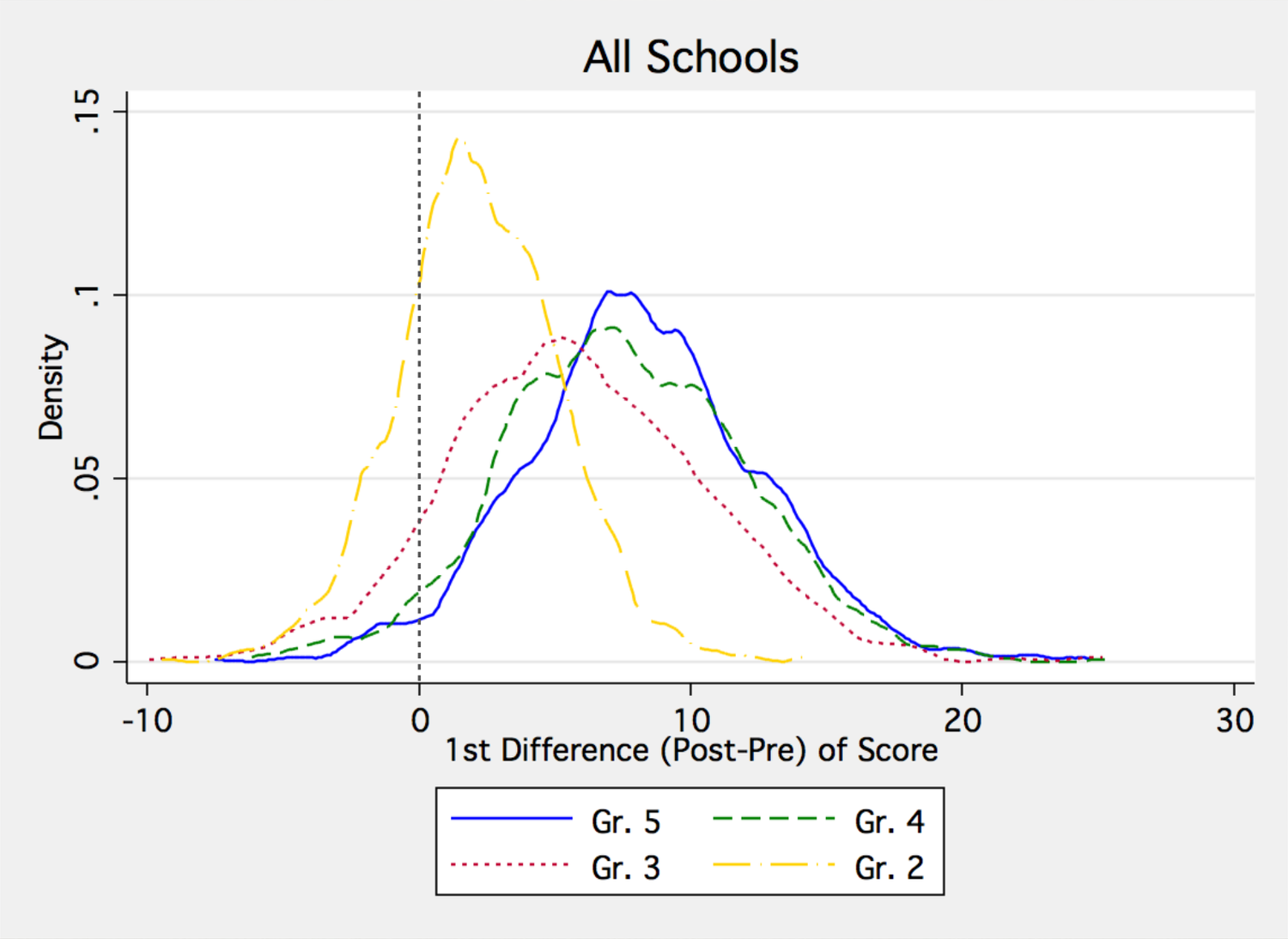
| Variable | Mean | St. Dev. |
|---|---|---|
| Ethnic - White | 0.63 | 0.29 |
| Ethnic - Black | 0.28 | 0.26 |
| Ethnic - Other | 0.09 | 0.14 |
| Learn Impairment | 0.12 | 0.07 |
| No Special Label | 0.76 | 0.12 |
| Gifted | 0.13 | 0.11 |

School Variables

| Variable | Mean | St. Dev. |
|---|---|---|
| Teacher Experience | 13.2 | 6.5 |
| License Test Score | 0.01 | 0.58 |
| % Free Lunch Eligible | 0.44 | 0.22 |
| # of Classes (gr. 3-5) | 3.5 | 1.4 |

# Outcome of Reform

- Did reform improve scores overall?
  - evidence in line with expectations



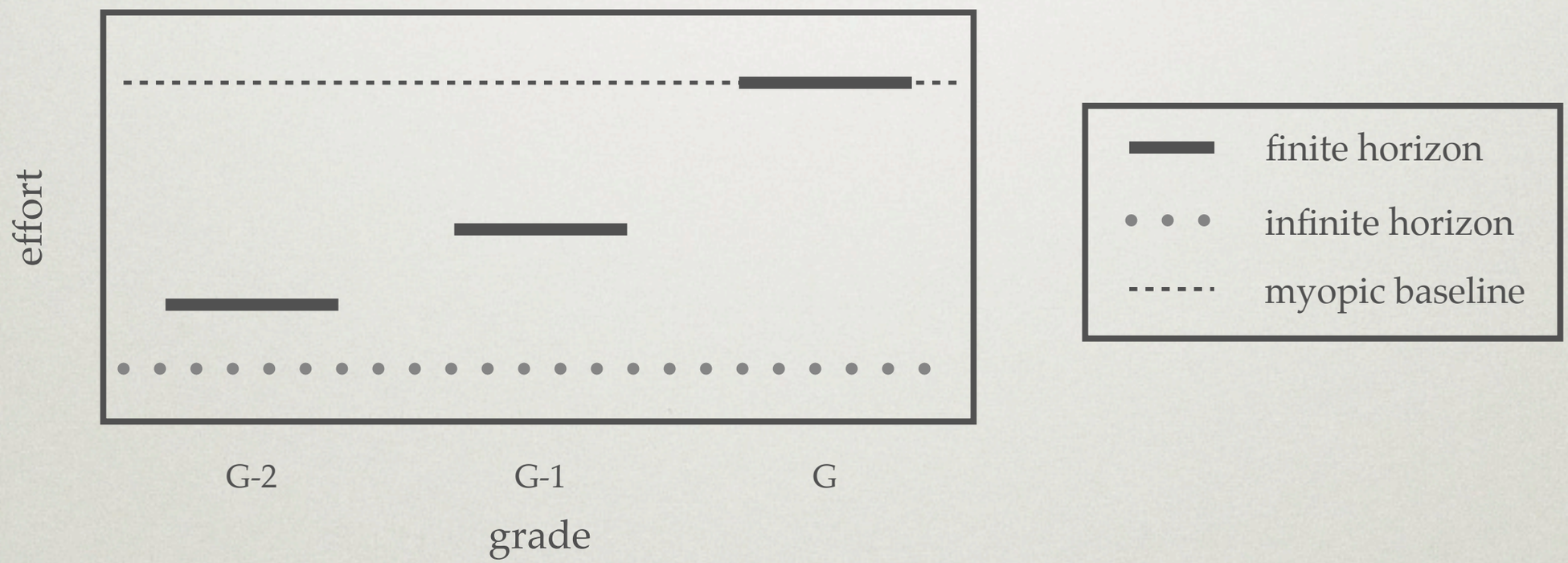

- How can we detect dynamic gaming?

# A Stylized Model

- Extension of dynamic moral hazard literature
  - finite-horizon ratchet effects
- School-level incentives
  - school principal as agent
  - monitor and coordinate teacher effort (input in production)
- Target depends on student prior score
  - students in school for fixed period of time
  - horizon dictated by number of grades until they graduate

# A Stylized Model

- Intuitive idea:
  - outperformance today makes future targets more difficult
  - agents anticipate by responding less to reform
  - more targets to consider under a longer horizon
  - dynamic distortion increases with horizon

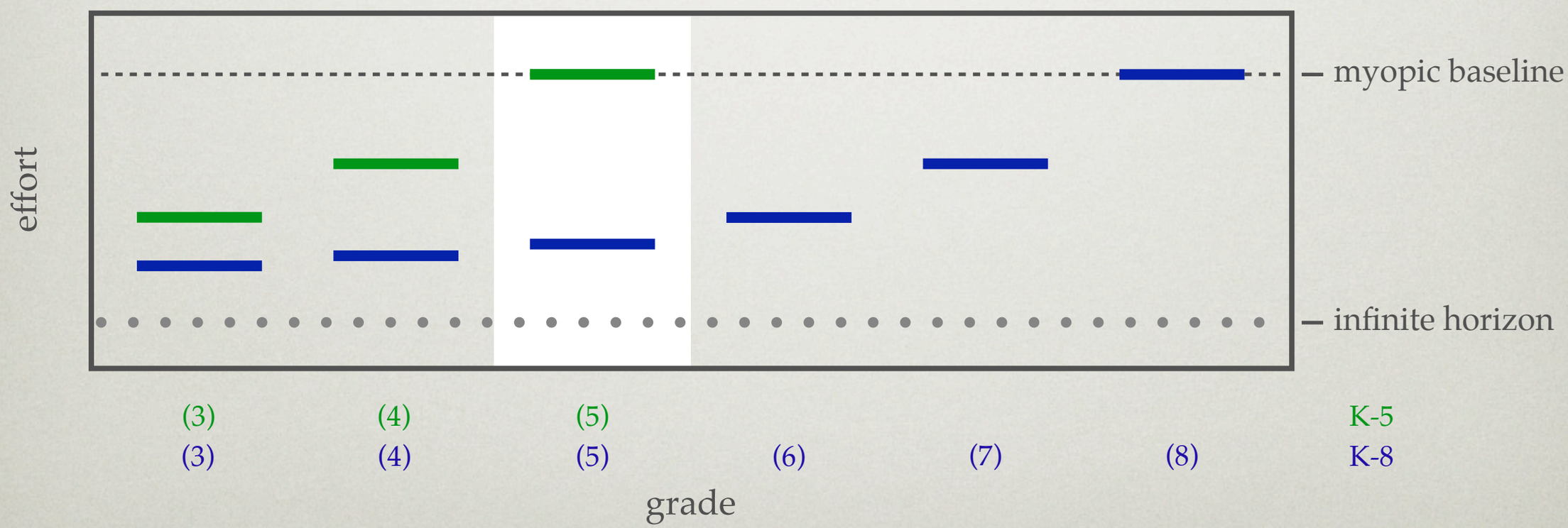


# Research Design

- Model maps into empirics
  - observe grade span of school in data (e.g. K-5, K-8)
  - grade and span dictate horizon
  - exploit variation in scores across grades and spans

# Research Design

- Model maps into empirics
  - observe grade span of school in data (e.g. K-5, K-8)
  - grade and span dictate horizon
  - exploit variation in scores across grades and spans
- Simple example: gr. 5 effort in K-5 versus K-8 school



# Research Design

- All else equal, scores should follow a similar pattern
  - suggests comparing scores across configurations
  - but all else is not necessarily equal...

# Research Design

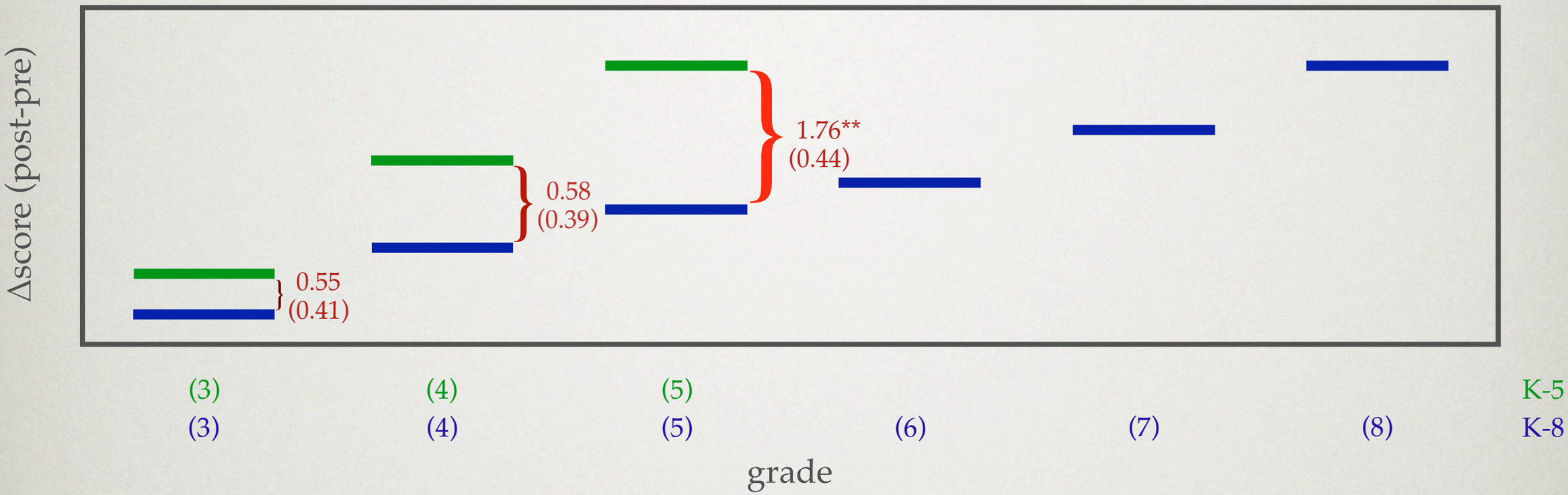
- All else equal, scores should follow a similar pattern
  - suggests comparing scores across configurations
  - but all else is not necessarily equal...
- Differences-in-differences strategy
  - exploit pre- and post-reform variation
  - accounts for time-invariant differences in inputs and shared trends across configurations

# Research Design

- All else equal, scores should follow a similar pattern
  - suggests comparing scores across configurations
  - but all else is not necessarily equal...
- Differences-in-differences strategy
  - exploit pre- and post-reform variation
  - accounts for time-invariant differences in inputs and shared trends across configurations
- Prediction:

$$\Delta\Delta y_{K5\text{-}K8,post\text{-}pre,g} = (y_{K5,post,g} - y_{K5,pre,g}) - (y_{K8,post,g} - y_{K8,pre,g}) > 0$$

# Reduced-Form Results




| K-5 vs K-8 | Grade 3 | Grade 4 | Grade 5 |
|---|---|---|---|
| $\Phi_{K5,post\text{-}pre}$ | 7.60** (0.17) | 9.01** (0.16) | 10.29** (0.18) |
| $\Phi_{K8,post\text{-}pre}$ | 7.05** (0.39) | 8.42** (0.38) | 8.52** (0.43) |
| $\Phi_{K5\text{-}K8,post\text{-}pre}$ | 0.55 (0.41) | 0.58 (0.39) | 1.76** (0.44) |

Estimates from joint F-tests of interaction dummies included in the estimating equation. Standard errors adjusted for clustering at school level are reported in parentheses. Significance levels: † 10 percent, * 5 percent, ** 1 percent

# Structural Estimation

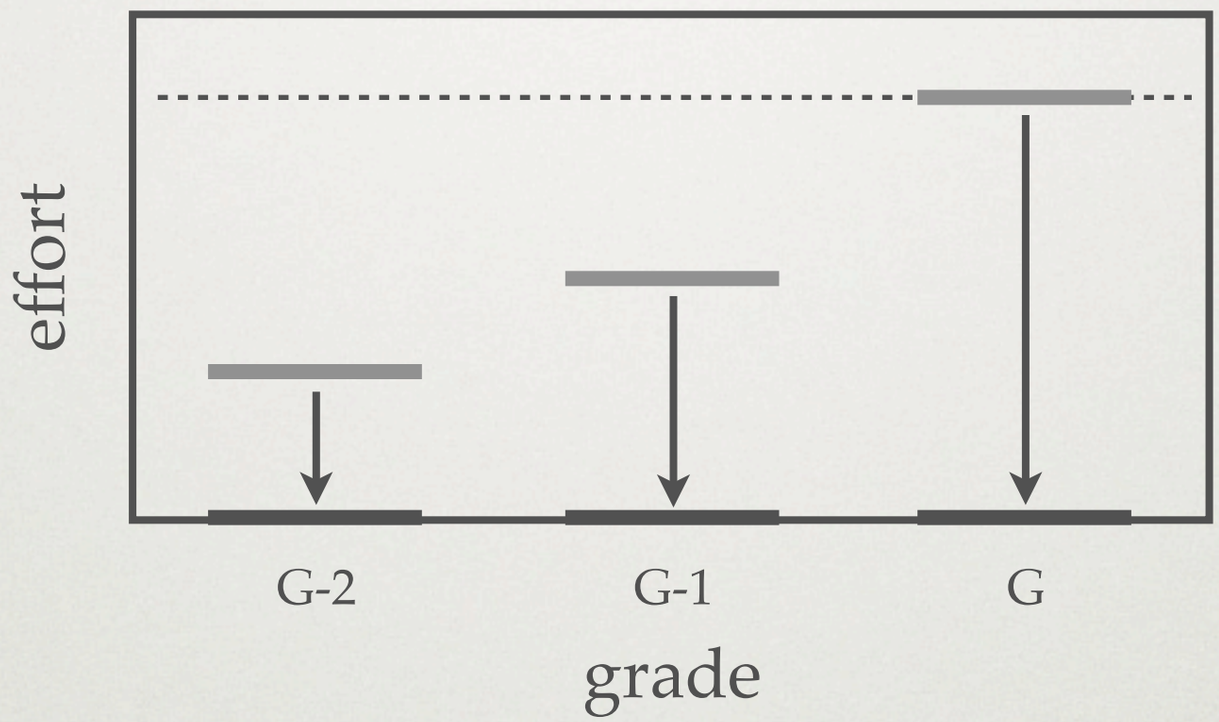
- Estimate the model parameters using MLE
  - insight into educational production process
    - allow for nonlinear interaction between inputs
    - evidence that teacher effort and student ability are complements
  - counterfactual policy experiments

    a. shut down the reform

    b. eliminate dynamic distortions using prediction of model

# Policy Experiments

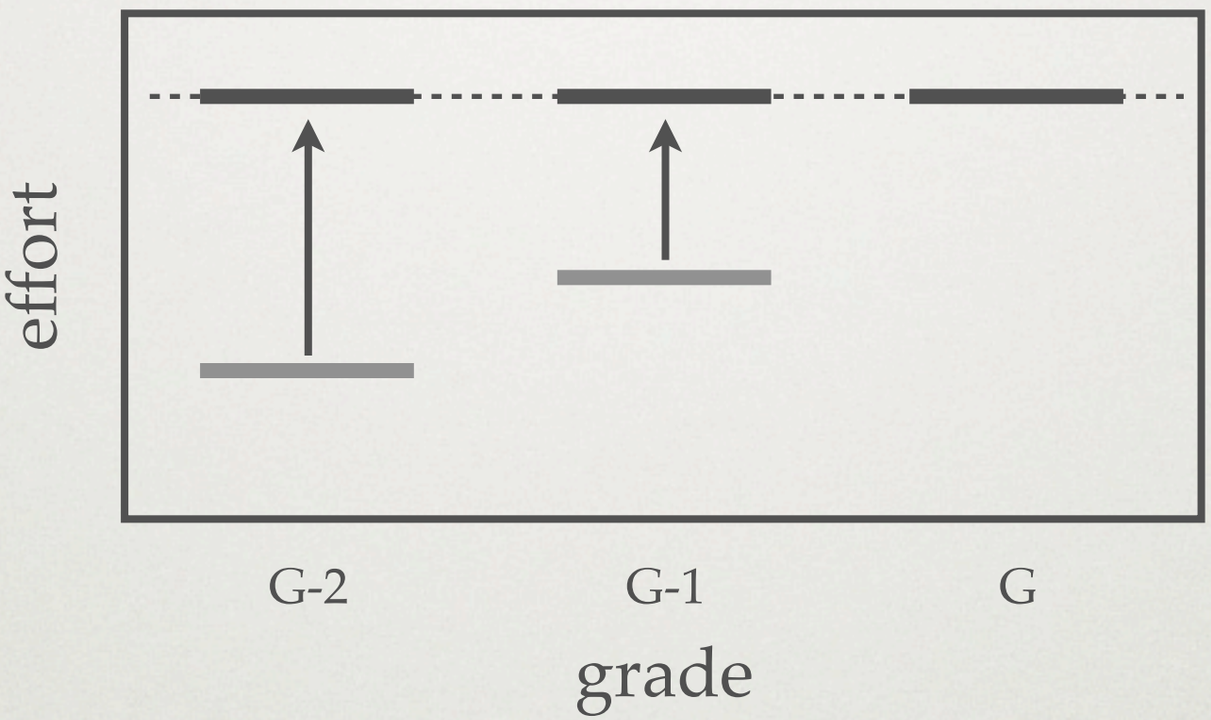
1. Shut down the reform
   - reveal effect of scheme with distortions




- cumulative grade 5 score for K-5 schools: about one s.d. lower (about 80% of black-white score gap in NC)
- linear technology restriction overstates cumulative effect of policy by about 2%

# Policy Experiments

2. Eliminate distortions
   - reveal cumulative effect




- cumulative grade 5 score for K-5 schools: 4% of a s.d. higher, but 38% more costly to implement
- linear technology restriction overstates cumulative distortion by about 5%

# Conclusion

- Dynamic incentive issue with broad relevance
- Theoretical contribution
  - finite-horizon extension with empirically-relevant prediction
- Evidence of dynamic gaming
  - greater grade 5 response in K-5 than in K-8 schools
- Structural estimation
  - glimpse into underlying technology
  - counterfactual exercises
    a. reform had large positive effect on scores
    b. cumulative extent of distortions quantified
- important step in determining the optimal scheme

# Theoretical Framework

- School's problem: choose $\{\{e_{scgt}\}_{g\in\mathcal{G}_c}\}_{t=1}^{\infty}$ to max

$$U_{sct} = \sum_{t=1}^{\infty} \delta^{t-1}\Big\{b\Big[1 - F\Big(\sum_{g=1}^{G_c}((\alpha-\gamma)y_{scg-1t-1} - e_{scgt} - a_{scgt})\Big)\Big] - \sum_{g=0}^{G_c} C(e_{scgt})$$

  - <u>subscripts</u>: school $s$, config $c$, grade $g$, time $t$, highest grade $G_c$
  - <u>quantities</u>: output $y$, effort $e$, teacher ability $a$, noise $u$, reward $b$, discount rate $\delta$, natural growth rate $\gamma$, target $\alpha$

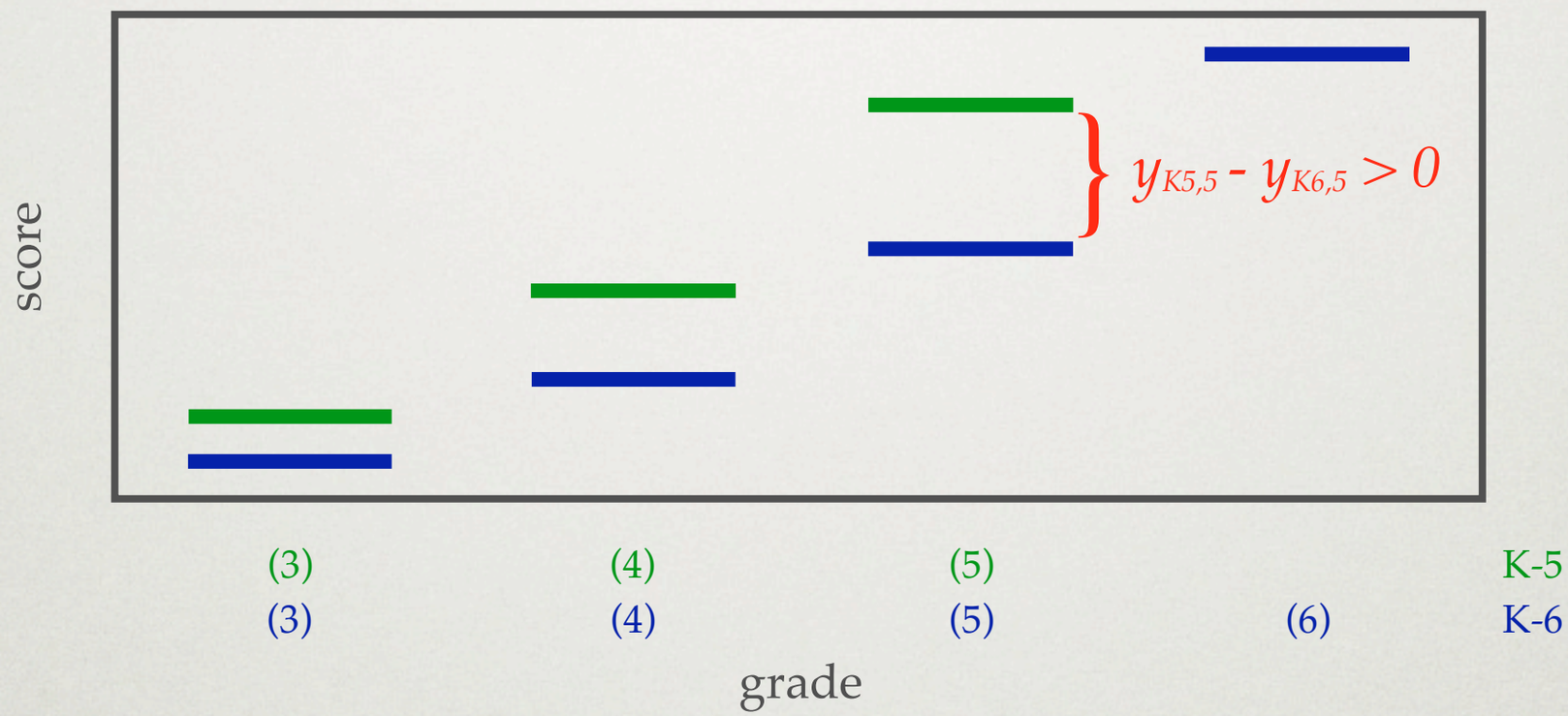
- For now, assume linear scheme: $b[\sum_{g=1}^{G_c}(y_{scgt} - \alpha y_{scg-1t-1})]$ and linear technology $y_{scgt} = \gamma y_{scg-1t-1} + e_{scgt} + a_{scgt} + u_{scgt}$

- Given a convex cost of effort $C(e) = de^2$, the FOCs are

$$e_{cg} = \begin{cases} \frac{b}{2d}\left[1 + \delta(\gamma-\alpha)\sum_{i=0}^{G_c-g-1}\delta^i\gamma^i\right] & \text{for } 1 \le g < G_c \\ \frac{b}{2d} & \text{for } g = G_c \end{cases}$$

# Theoretical Framework

- **Proposition 1**: If all else is equal for two configurations $c$ and $c'$, such that $G_{c'} > G_c$, then $y_{cg} > y_{c'g}$, $\forall\ g \in \mathcal{G}_c = \{0,\ldots,G_c\}$.

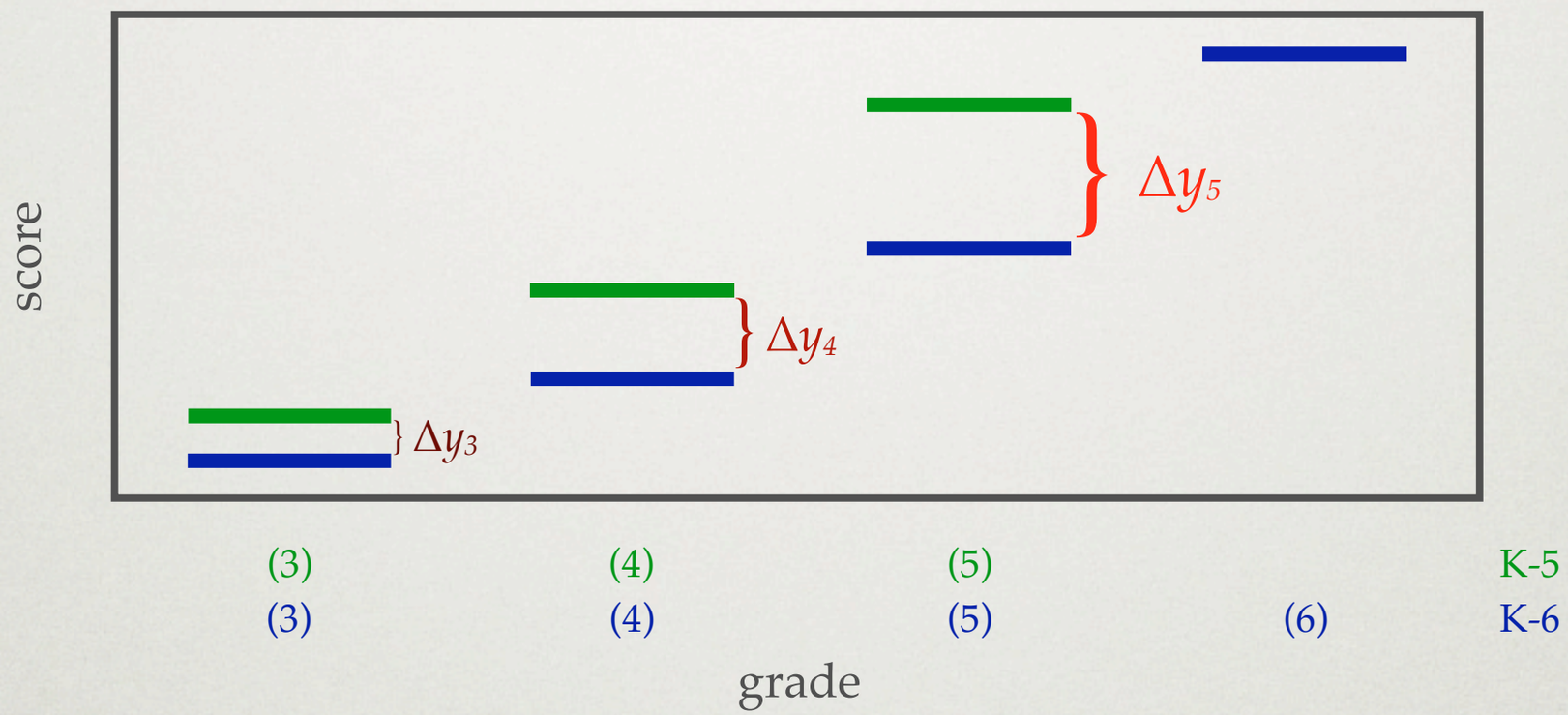



- effort levels from FOCs: $e_{K5,5} = \frac{b}{2d}$

$$e_{K6,5} = \frac{b}{2d}(1 + \delta(\gamma - \alpha))$$

# Theoretical Framework

- **Proposition 2**: If all else is equal for two configurations $c$ and $c'$, such that $G_{c'} > G_c$, then $y_{cg} - y_{c'g}$ is increasing in $g$.




- differences in effort from FOCs:

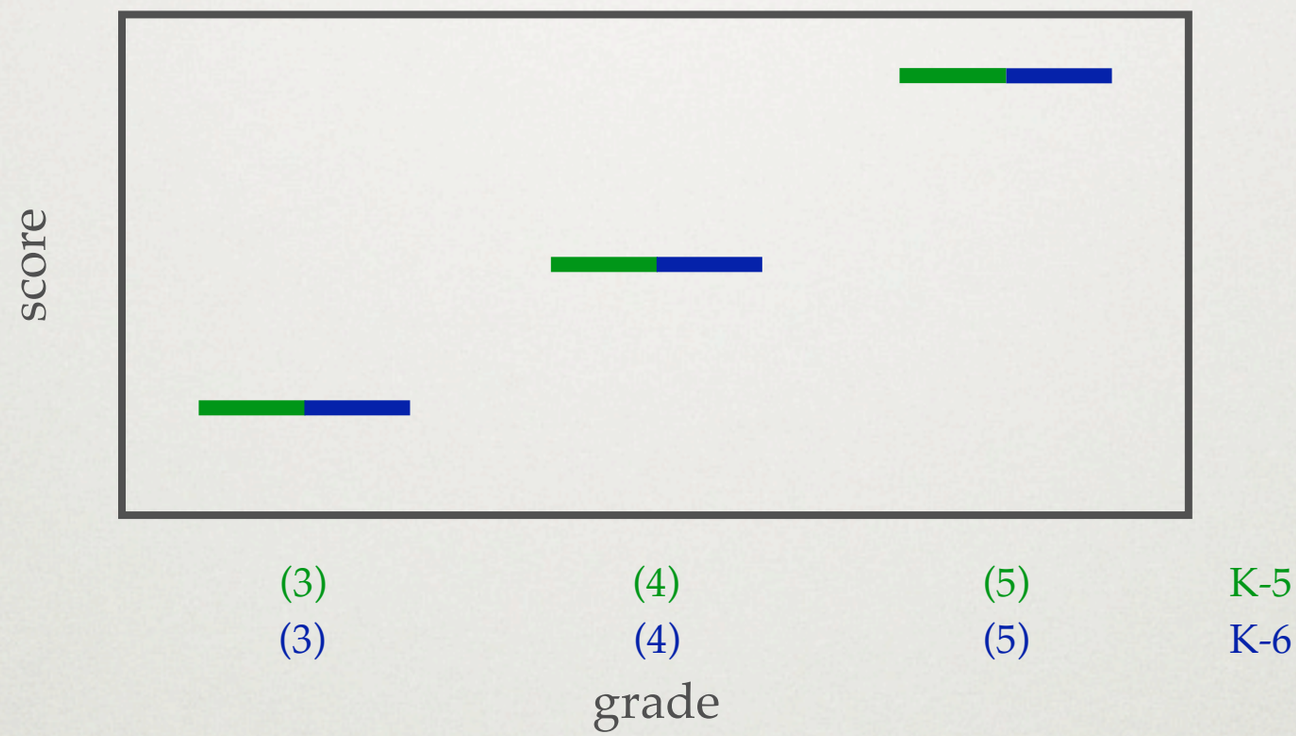
$$\Delta e_5 = \frac{b}{2d}\delta(\alpha - \gamma)$$

$$\Delta e_4 = \frac{b}{2d}\delta^2\gamma(\alpha - \gamma) = \delta\gamma\Delta e_5$$

$$\Delta e_3 = \frac{b}{2d}\delta^3\gamma^2(\alpha - \gamma) = \delta\gamma\Delta e_4$$

# Theoretical Framework

- **Proposition 4**: Dynamic distortions are eliminated if the target is set to the natural growth rate.




- from FOCs, if $\alpha = \gamma$, then $e_{cg} = \frac{b}{2d} \ \forall \ c, g$

# Econometric Strategy

- Estimating equation:

$$y_{scgt} = X'_{sgt}\beta_g + \sum_{c=1}^{C} \sum_{g \in \mathcal{G}_c} (\phi_{pre,c,g} + \phi_{post,c,g}) + \varepsilon_{scgt}$$

- DinD estimates from F-tests of $\phi$ coefficients:

$$\Phi_{\mathrm{K5-K8},post-pre,g} = (\phi_{post,\mathrm{K5},g} - \phi_{pre,\mathrm{K5},g}) - (\phi_{post,\mathrm{K8},g} - \phi_{pre,\mathrm{K8},g})$$

  - a finding of $\Phi_{\mathrm{K5-K8},post-pre,g} > 0$ is in line with theory

# Robustness Check

| K5 vs K8 | Grade 3 | Grade 4 | Grade 5 |
|---|---|---|---|
| All Schools | 0.55<br>(0.41) | 0.58<br>(0.39) | 1.76**<br>(0.44) |
| No Switchers | -0.07<br>(0.49) | -0.13<br>(0.44) | 1.61**<br>(0.55) |
| No Switch - All yrs | 0.33<br>(0.50) | -0.43<br>(0.45) | 1.24*<br>(0.55) |

| K5 vs K6 | Grade 3 | Grade 4 | Grade 5 |
|---|---|---|---|
| All Schools | 1.92**<br>(0.38) | 1.82**<br>(0.35) | 2.48**<br>(0.37) |
| No Switchers | 0.80<br>(0.90) | 0.96<br>(0.67) | 1.51*<br>(0.60) |
| No Switch - All yrs | 0.34<br>(1.09) | 0.21<br>(0.69) | 0.53<br>(0.66) |

Estimates from joint F-tests of interaction dummies included in the estimating equation.
Standard errors adjusted for clustering at school level are reported in parentheses.
Significance levels: † 10 percent, * 5 percent, ** 1 percent

# Triple Difference

**Additional Difference: Grade 4 to 5**

| | K5 - K8 | K5 - K6 |
|---|---|---|
| All Schools | 1.18* (0.48) | 0.66† (0.36) |
| No Switchers | 1.74** (0.60) | 0.56 (0.56) |
| No Switch - All years | 1.67** (0.63) | 0.33 (0.64) |

**Additional Difference: Grade 3 to 4**

| | K5 - K8 | K5 - K6 |
|---|---|---|
| All Schools | 0.04 (0.40) | -0.10 (0.38) |
| No Switchers | -0.06 (0.48) | 0.16 (0.83) |
| No Switch - All years | -0.10 (0.48) | -0.14 (0.92) |

# Supporting Evidence

| K5 vs K8 | Grade 3 | Grade 4 | Grade 5 |
|---|---|---|---|
| Combined | -0.07<br>(0.49) | -0.13<br>(0.44) | 1.61**<br>(0.55) |
| Mathematics | -0.19<br>(0.30) | -0.17<br>(0.28) | 1.12**<br>(0.38) |

| K5 vs K6 | Grade 3 | Grade 4 | Grade 5 |
|---|---|---|---|
| Combined | 0.80<br>(0.90) | 0.96<br>(0.67) | 1.51*<br>(0.60) |
| Mathematics | 0.74<br>(0.57) | 0.70<br>(0.45) | 0.78†<br>(0.43) |

Estimates from joint F-tests of interaction dummies included in the regression for schools that do not switch configuration during period.
Coefficient for reading is the difference between the values for Φ Combined and Φ Mathematics.
Standard errors adjusted for clustering at school level are reported in parentheses.
Significance levels: † 10 percent, * 5 percent, ** 1 percent

# Descriptive Statistics

- School count by type (1994-2005)

All Locales

| | K-5 | K-6 | K-8 |
|---|---|---|---|
| All Schools | 849 | 104 | 97 |
| No Switchers | 661 | 36 | 78 |
| No Switch - All years | 489 | 24 | 71 |

Rural Schools

| | K-5 | K-6 | K-8 |
|---|---|---|---|
| All Schools | 396 | 71 | 87 |
| No Switchers | 297 | 30 | 69 |
| No Switch - All years | 226 | 20 | 62 |