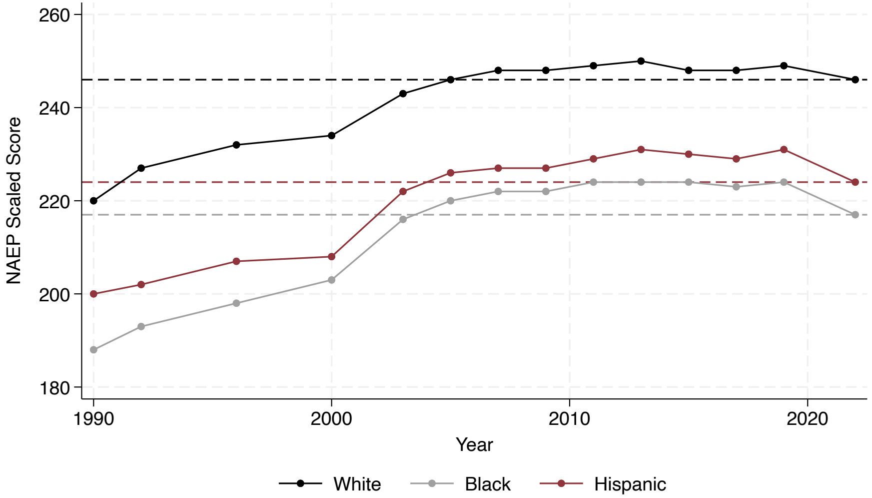
Education Production

Christopher Campos

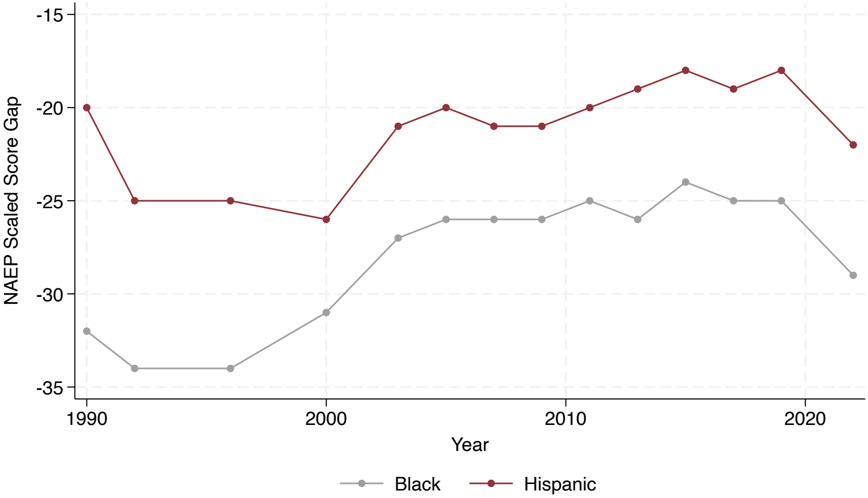
Economics of Education, Spring 2026

Motivation

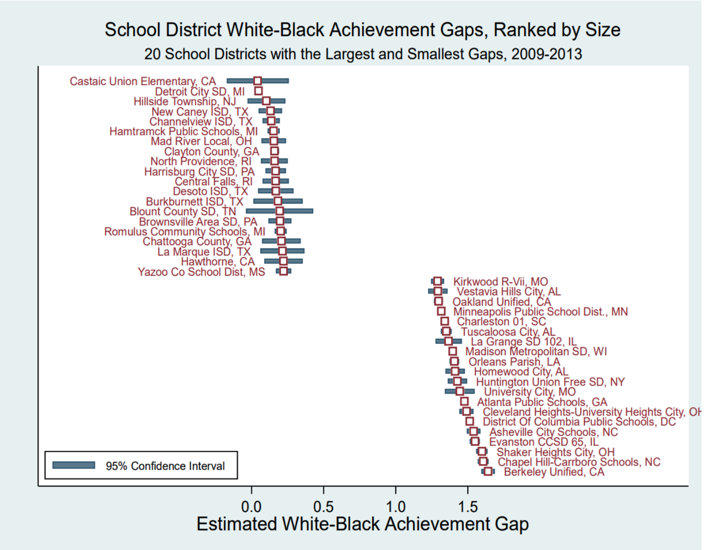
Trends in Educational Achievement from NAEP



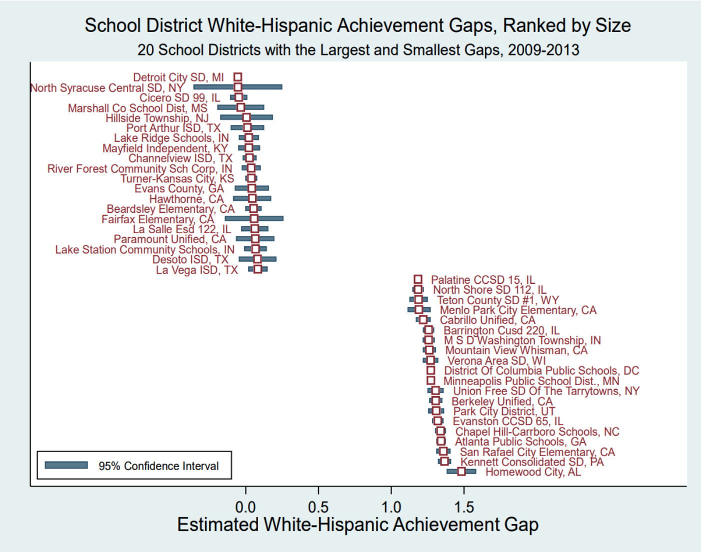
Trends in Educational Achievement Gaps from NAEP



Racial achievement gaps are pervasive and large (Reardon 2014)



Racial achievement gaps are pervasive and large (Reardon 2014)



The Coleman Report (1966)

- Formally commissioned by Congress as the Equality of Educational Opportunity Study as part of the 1964 Civil Rights Act
- James Coleman was the lead author, a sociologist with interests in mathematics and economics
- Data collected from 4000 schools, 66,000 teachers, and 600,000 students. Findings influenced decades of research
- **Key Findings (a series of descriptive facts):**
  1. Schools are extremely segregated.
     - 80 percent of White pupils attend schools that are 90-100 percent White
     - Roughly 70 percent of Black pupils attend schools that are 90-100 percent Black
  2. First quantification of the racial achievement gap in the United States
     - Black-White Nonverbal Achievement Gap: 1.1σ
     - Black-White Verbal Achievement Gap: 0.8σ
  3. Schools and inputs unrelated to outcomes once we condition for family background. From the report:

     It is known that socioeconomic factors bear a strong relation to academic achievement. When these factors are statistically controlled, however, it appears that differences between schools account for only a small fraction of differences in pupil achievement.
- Motivated numerous studies aiming to understand the relationship between schooling inputs, quality, and learning

Education Production Functions

Connection to Previous Lecture

- Last week: education is an *investment* — individuals choose schooling to maximize the present value of lifetime earnings
- We assessed *returns to schooling* using various research designs
- But schooling is a black box: a year of school in one setting may be very different from a year in another
- Today's question: What happens within and between schools that generates learning?
  - → Which inputs matter?
  - → How much does each input contribute?
  - → Does learning in childhood translate into better long-run outcomes?
- We approach this by analyzing the production process: write down a *production function* and develop methods to estimate it

Outline for Today

1. The Education Production Function
   - → Schools as producers: inputs, technology, and outputs
   - → Todd and Wolpin (2003): the value-added specification
2. Empirical Evidence
   - → Inputs on returns: Card and Krueger (1992)
   - → Class size: Krueger (1999), Angrist and Lavy (1999)
   - → School spending: Jackson, Johnson, and Persico (2016), Lafortune, Rothstein, and Schanzenbach (2018)
3. Teacher and School Value-Added
   - → Measuring teacher quality: Rockoff (2004), Chetty, Friedman, and Rockoff (2014a,b), Rothstein (2010, 2017)
   - → Lottery-based validation: Angrist, Hull, Pathak, and Walters (2017)
4. Putting it all together: Bruhn, Campos, and Chyn (2025)

Schools as “Producers”

- A firm takes inputs and produces output. A school does the same:
  - → Output: Student achievement (test scores, graduation, earnings (?))
  - → Inputs: Teachers, class size, instructional time, curriculum, facilities, peers, family background
- Write the education production function as:

$$A = f(\underbrace{T, S, M}_{\text{school inputs}}, \underbrace{P}_{\text{peers}}, \underbrace{H}_{\text{family}}, \underbrace{\mu}_{\text{ability}})$$

where A = achievement, T = teacher quality, S = class size (or its inverse), M = materials/spending, P = peer quality, H = family inputs, μ = innate ability

- Many empirical papers attempt to estimate something akin to the marginal product of different inputs
  - → Project STAR, Maimonides' Rule: class size
  - → Value-added models: teachers
- Schools (principals?) choose a combination of inputs to maximize achievement subject to a budget constraint
  - → Theory predicts that more spending tends to increase achievement, but does it?
  - → Hanushek (1987), Jackson et al. 2016, Lafortune et al. (2018)

A Two-Input Simplification

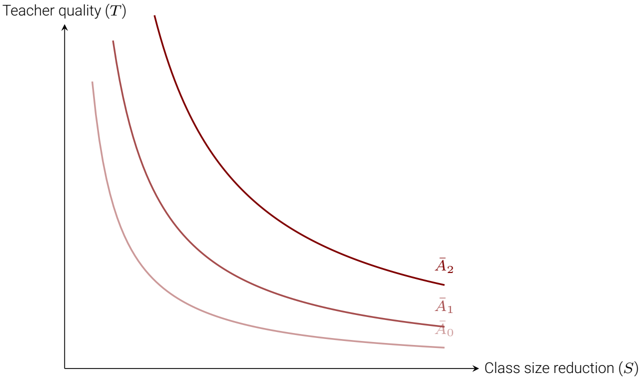
- To build intuition, consider a school choosing two inputs: teacher quality (T) and class size reduction (S)
- Achievement is produced according to:

$$A = f(T, S)$$

  with $f_T > 0$, $f_S > 0$ (more of each input raises achievement) and f concave (diminishing marginal returns)
- The school faces a budget constraint. Let:
  - → w = cost per unit of teacher quality (e.g., salary premium for better teachers)
  - → p = cost per unit of class size reduction (e.g., hiring additional teachers)
  - → B = total budget
- Budget constraint:

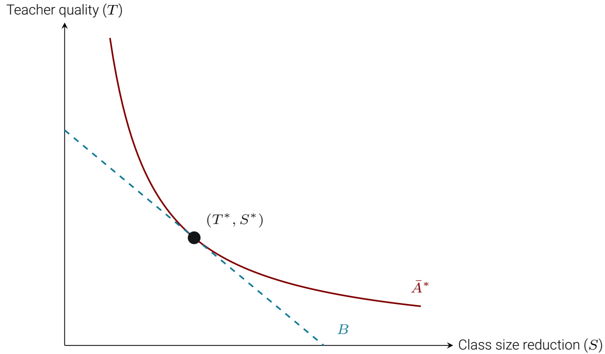
$$wT + pS = B$$

Combinations That Produce the Same Learning



- Each isoquant: all (T, S) combinations yielding the same achievement $\bar{A}$
- Slope = $-f_S\,/\,f_T$ = marginal rate of technical substitution (MRTS)
- Higher isoquants $\Rightarrow$ more learning
- Convexity reflects diminishing MRTS: as you substitute away from one input, you need increasing amounts of the other to maintain the same level of production

Cost Minimization

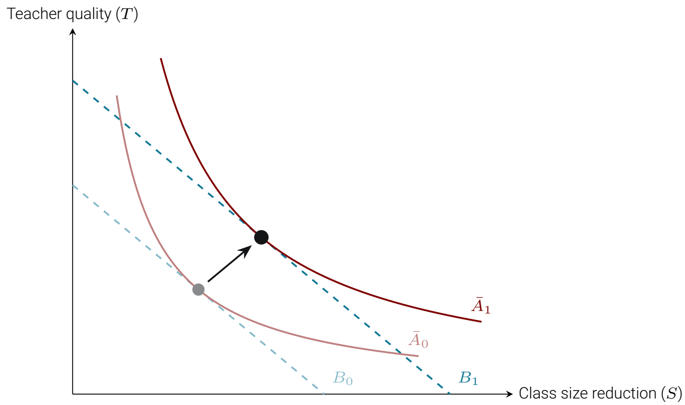


- Optimum: choose (T^*, S^*) where the isoquant is *tangent* to the isocost line
- Tangency (bang-for-buck) condition:

$$\underbrace{\frac{f_S}{f_T}}_{\text{MRTS}} = \underbrace{\frac{p}{w}}_{\text{price ratio}} \iff \frac{f_S}{w} = \frac{f_T}{p}$$

- Marginal product per dollar is equalized across inputs

Connection to a classic debate: what happens when the budget increases?



- A spending increase shifts the isocost outward $\Rightarrow$ school reaches a higher isoquant
- More of *both* inputs $\Rightarrow$ higher achievement
- This is the theoretical basis for expecting money to matter
- Any ideas as to why the empirical literature was conflicted on this for so long?

Challenges in Estimating the Production Function

- Suppose we want to estimate:

$$A_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 S_{it} + \beta_3 H_{it} + \beta_4 P_{it} + \mu_i + \varepsilon_{it}$$

- Some problems:

  1. Omitted variables: Family inputs H and ability μ are correlated with school inputs — better-resourced families sort into better-funded schools
  2. Simultaneity: Inputs respond to outcomes — struggling students may receive more resources (remediation), biasing estimates downward
  3. Cumulative nature: Today's achievement depends on the *entire history* of inputs, not just current ones

- Todd and Wolpin (2003) formalize these issues and show that the standard "value-added" specification — conditioning on lagged test scores — is a special case requiring that lagged scores are a *sufficient statistic* for the input history

The Value-Added Specification (Todd and Wolpin 2003)

- Start from the general cumulative production function

$$A_{it} = f\left(\mathbf{X}_{it}, \mathbf{X}_{i,t-1}, \ldots, \mathbf{X}_{i,0}, \mu_i, \varepsilon_{it}\right)$$

  where $\mathbf{X}_{it}$ collects all inputs at time t

- Todd and Wolpin (2003) work through the specific assumptions about persistence of past inputs and endowed ability that allow us to simplify to the value-added model:

$$A_{it} = \alpha + \boldsymbol{\beta}'\mathbf{X}_{it} + \gamma A_{i,t-1} + \varepsilon_{it}$$

- The lagged score $A_{i,t-1}$ is a *sufficient statistic* for the history of inputs — it absorbs everything that happened before period t

- This is a workhorse in the empirical literature

From the cumulative model to the value-added model

- The cumulative model is:

$$A_{it} = X_{it}\alpha_1 + X_{i,t-1}\alpha_2 + \cdots + X_{i0}\alpha_t + \beta_t\mu_i + \delta_{it}$$

- To see the assumptions needed to arrive at the value-added model, subtract $\gamma A_{i,t-1}$:

$$A_{it} - \gamma A_{i,t-1} = X_{it}\alpha_1 + X_{i,t-1}(\alpha_2 - \gamma\alpha_1) + \cdots + X_{i0}(\alpha_t - \gamma\alpha_{t-1}) + (\beta_t - \gamma\beta_{t-1})\mu_i + \delta_{it} - \gamma\delta_{i,t-1}$$

- For this to reduce to

$$A_{it} = \alpha + \beta' X_{it} + \gamma A_{i,t-1} + \varepsilon_{it},$$

  all lagged-input terms and the endowment term must disappear. This assumes:

  1. $\alpha_{k+1} = \gamma\alpha_k \quad \forall k \iff \alpha_k = \gamma^{k-1}\alpha_1$: past inputs decay geometrically at an identical rate
  2. $\beta_t = \gamma\beta_{t-1} \iff \beta_t = \gamma^{t-1}\beta_1$: endowed ability must decay at the same rate

- Lagged achievement is sufficient only if both observed inputs and unobserved ability evolve with the same decay law. Reasonable?

Policy Effects versus Production Function Parameters

TW make an important and subtle point distinguishing *production function parameters* from *policy effects*. In a simple two period model, achievement evolves according to

$$\underbrace{A_1 = g_0(F_0, \mu_i)}_{Period\ 1\ Achievement} \quad \text{and} \quad \underbrace{A_2 = g_1(S_1, F_1, \mu_i)}_{Period\ 2\ Achievement}$$

Families choose their inputs. :

$$F_1 = \phi(A_1, W, \mu_i, S_1).$$

- A_t: achievement at the start of year t
- F_t: family inputs during year t
- S_t: school inputs during year t
- W: family resources
- μ_i: endowed ability

Key point: F_1 is not fixed. Families may respond to the school environment, so changing S_1 may also change F_1. Therefore, TW distinguish between policy effects that potentially have family responses and production function parameters

Policy Effects vs. Production Function Parameters

The ceteris paribus production function effect:

$$\frac{\partial g_1}{\partial S_1}$$

The total effect of a policy intervention that changes school inputs and allows family responses:

$$\frac{dA_2}{dS_1} = \frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial F_1}\frac{\partial F_1}{\partial S_1}$$

- Production function estimation targets $\partial g_1 / \partial S_1$
- Policy experiments typically identify the total derivative dA_2 / dS_1

Only if $\frac{\partial F_1}{\partial S_1} = 0$, meaning there is no behavioral response to the policy, do the two coincide

Empirical Evidence

The Hanushek Claim

- Hanushek (1986) surveyed 147 studies estimating the education production function
- Key conclusion: *“There is no strong or systematic relationship between school expenditures and student performance”*
- This became the “money doesn’t matter” conventional wisdom for over two decades
- The argument: most studies found statistically insignificant or inconsistent relationships between spending and outcomes
- More modern research uses more credible quasi-experimental variation in spending and the assignment of inputs

Class Size: Does It Matter?

- Class size is a salient input to parents
- But causal evidence on its effect is hard to come by:
  - → Smaller classes are often assigned to struggling students
  - → Wealthier communities have both smaller classes and higher scores
  - → OLS is badly confounded in both directions
- Three important papers attack this using different strategies:
  - → Krueger (1999): randomized experiment (Tennessee STAR)
  - → Angrist & Lavy (1999): regression discontinuity / IV (Israel)
  - → Angrist, Lavy, Leder-Luis, & Shany (2019): revisits the Israeli evidence with newer data

Krueger (1999): Project STAR — Design

- Tennessee's Student/Teacher Achievement Ratio experiment, 1985–1989
- Within each school, students and teachers randomly assigned to:
  - → Small classes (13–17 students)
  - → Regular classes (22–25 students)
  - → Regular + full-time aide (22–25 students)
- Design complications:
  - → Regular/aide students re-randomized after K (parental complaints)
  - → ≈10% of students switched class types across grades
  - → High attrition: only half of K entrants present in all grades K–3
  - → No baseline test scores available
- Within-school balance checks on race, free lunch, age: no significant differences conditional on school FEs

Smaller Class Size Leads to Improved Learning

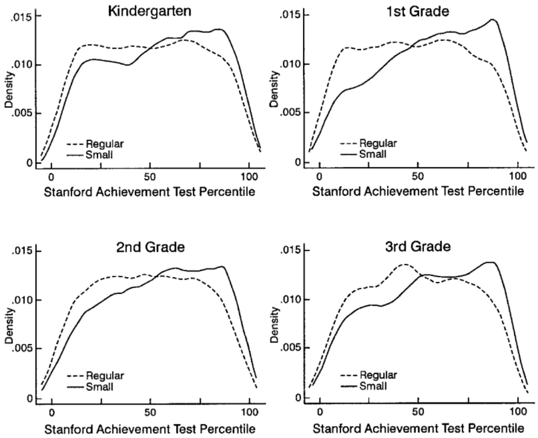


FIGURE I
Distribution of Test Percentile Scores by Class Size and Grade

Positive Effects in Most Classrooms (although class-specific effects are noisy)

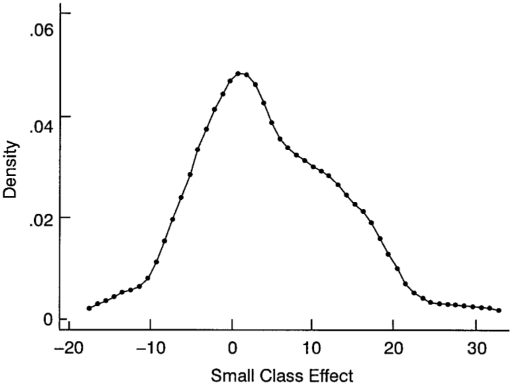


FIGURE II
Kernel Density of School-Level Small-Class Effects

Krueger (1999) Summary

- Outcome: average percentile rank on Stanford Achievement Test (math, reading, word recognition), normed against regular-class distribution
- Small class effect: performance improvement on standardized tests; robust to school FEs, student controls, and use of initial assignment as IV for actual class size
  - → Dynamics: ≈4 point discrete jump the *first year* in a small class, plus ≈1 point per additional year
  - → Implication: value-added specifications miss most of the gain
- Teacher aides: small, mostly insignificant effects
- Heterogeneity: larger effects for Black students, free-lunch students, and inner-city schools

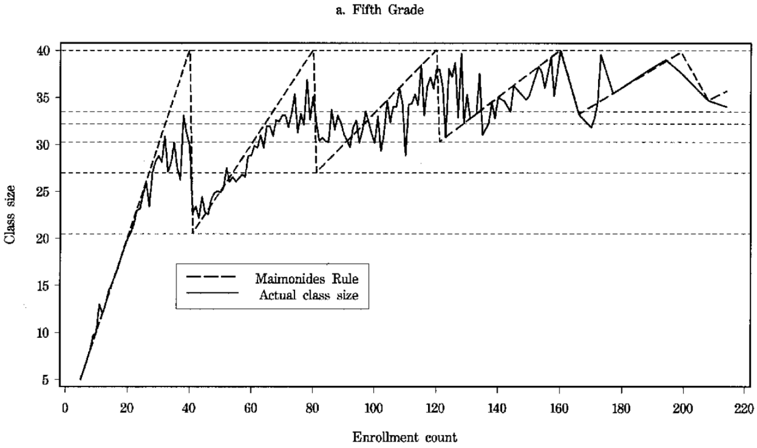
Angrist & Lavy (1999): Maimonides' Rule as an Instrument

Israeli public schools follow a maximum class size of 40 (from Maimonides' 12th-century ruling). This generates a sawtooth relationship between enrollment and class size:

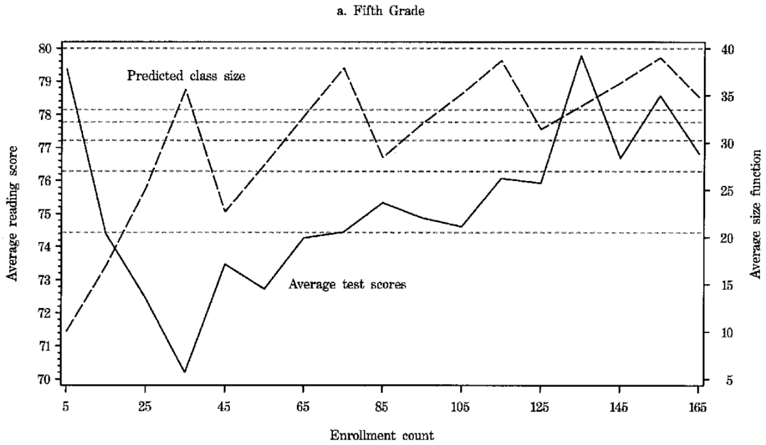
$$f_{sc} = \frac{e_s}{\lfloor (e_s - 1)/40 \rfloor + 1}$$

- Enrollment of 40 $\Rightarrow$ one class of 40; enrollment of 41 $\Rightarrow$ two classes of $\approx$20.5
- f_{sc} is a nonlinear function of enrollment $\Rightarrow$ controls for smooth enrollment effects leave the jumps as identifying variation
- This is a fuzzy RD / IV design: instrument is f_{sc} (or dummies for being above cutoffs), endogenous variable is actual class size
- Key assumption: conditional on smooth functions of enrollment and SES (percent disadvantaged), the only reason f_{sc} predicts test scores is through class size

A Visualization of Maimonides' Rule



Visualization of First Stage and Reduced Form



Angrist, Lavy, Leder-Luis, & Shany (2019): Redux

Revisit with 2002–2011 Israeli data (GEMS tests, ≈240,000 5th graders, 8,800 classes)

Finding 1: Precisely estimated zeros.

- 2SLS estimates of class size effects range from -0.03 to $+0.03$, SEs $\approx 0.03–0.04$
- Robust across specifications, running variable controls, and subsamples

Finding 2: Enrollment manipulation.

- Clear spikes in November enrollment just *above* Maimonides cutoffs (41, 81)
- School administrators had financial incentives to push enrollment past cutoffs (class-based funding rule)
- MOE memos explicitly warn against this behavior

Finding 3: Manipulation is not driving the zeros.

- Birthday-based imputed enrollment (using Chanukah cutoff applied to 4th–6th graders) eliminates sorting $\Rightarrow$ still precisely estimated zeros
- Maimonides' rule is uncorrelated with school SES after minimal controls

Redux: What Changed?

- The 1991 data also show some evidence of sorting at the first cutoff, but:
  - → Maimonides' rule is uncorrelated with school SES in 1991
  - → Donut estimates (dropping enrollment 39–41) are similar to full-sample estimates
- So the old results appear internally valid — the *production function* changed, not the research design
- Possible explanations for disappearing effects:
  - → Median class size fell from 31 (1991) to 28 (2002–2011) — smaller marginal returns at lower class sizes?
  - → Schools increasingly hire supplementary staff (especially high-SES schools)
  - → Possible teaching-to-the-test on GEMS exams

Takeaways on Class Size

- **STAR** provides the cleanest experimental evidence: meaningful gains from reducing class size by ≈8 students, concentrated in the first year
- **Angrist & Lavy (1999)** showed that a clever quasi-experimental design (Maimonides' rule) could recover similar magnitudes from observational data
- **Angrist et al. (2019)** update to more recent years: the same research design in newer data yields statistical zeros, and the original results now look less robust than they once appeared
- Class size effects may depend on context: baseline class size, complementary inputs, testing regime

Does School Spending Matter?

- Since the Coleman Report (1966), a large literature has asked whether school spending improves student outcomes
- Hanushek (1986, 2003): cross-sectional and observational studies find little relationship between spending and achievement
- School spending is endogenous to a host of factors:
  - $\rightarrow$ Compensatory funding $\Rightarrow$ negative bias (more money flows to struggling districts)
  - $\rightarrow$ Tiebout sorting $\Rightarrow$ positive bias (rich families live where spending is high)
- Jackson, Johnson, and Persico (2016): exploit *court-ordered school finance reforms* (SFRs) as exogenous shifters of spending, and track children into adulthood using the PSID

Institutional Background: School Finance Reforms

- Prior to the 1970s, most K–12 funding came from local property taxes $\Rightarrow$ large within-state spending inequality across rich and poor districts
- Between 1971 and 2010, state supreme courts overturned school finance systems in 28 states
- Two waves of legal challenges:
  - → **Equity cases** (1971–mid-1980s): local financing violates equal protection
  - → **Adequacy cases** (late 1980s onward): state fails to provide constitutionally adequate education
- Reforms changed funding formulas in varied ways: foundation plans, spending limits, reward-for-effort plans, equalization plans
- Common thread: reforms reduced spending inequality by *increasing* spending in previously low-spending districts

Data

- Outcomes: Panel Study of Income Dynamics (PSID)
  - → 15,353 individuals born 1955–1985, followed through 2011
  - → Adult outcomes observed ages 20–45: years of education, HS graduation, wages, family income, poverty status
- Spending: Historical Database on Individual Government Finances (INDFIN) + CCD F-33
  - → Annual district-level per-pupil spending, 1967–2010
- Key "treatment" variable: $\ln(\text{PPE}_{d,\text{age } 5\text{–}17})$ — average per-pupil spending during an individual's school-age years in their childhood district

Empirical Strategy: 2SLS Difference-in-Differences

First stage:

$$\ln(\text{PPE}_{517})_{idb} = \alpha_1(\text{Exp}_{idb} \times \text{Dosage}_d) + \alpha_2\text{Exp}_{idb} + C_{idb} + \delta_d + \gamma_b + \nu_{idb}$$

Second stage:

$$Y_{idb} = \beta \ln\widehat{(\text{PPE}_{517})}_{idb} + C_{idb} + \delta_d + \gamma_b + \varepsilon_{idb}$$

- Exp_{idb}: number of school-age years after the court order (0–12)
- Dosage_d: predicted reform-induced spending change for district d

- District FEs δ_d: compare cohorts *within* the same district
- Birth cohort FEs γ_b: account for national trends across cohorts

- Key assumption: timing of court orders is exogenous to district-level changes that directly affect outcomes

- Rich controls C_{idb}: parental SES, race $\times$ division $\times$ cohort FEs, school desegregation, Head Start, Title I, Medicaid, AFDC, food stamps, tax limits, ...

Two Approaches to Measuring "Dosage"

- **Approach 1**: Prereform spending level
  - → Quartile of district in state distribution of per-pupil spending in 1972
  - → Low-spending districts should gain more from equalizing reforms
- **Approach 2**: Leave-out predicted spending change ($\widehat{\Delta \text{Spend}}_d$)
  - → Predict dosage using experiences of similar districts *in other states*
  - → "Similar" = same prereform spending quartile, same prereform income quartile, same reform type
  - → Jackknife IV: excludes all data from own state $\Rightarrow$ avoids mechanical endogeneity
- Both approaches yield very similar 2SLS estimates; Approach 2 is more precise (stronger first stage)

First Stage: Reforms Do Not Increase Spending in High-Spending Districts

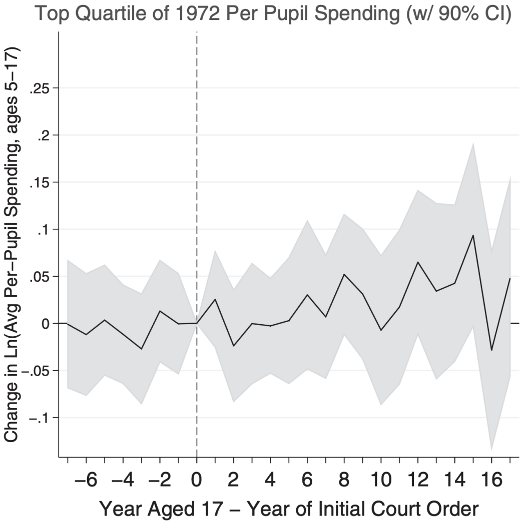


FIGURE I

First Stage: Reforms *Do* Increase Spending in *Low*-Spending Districts

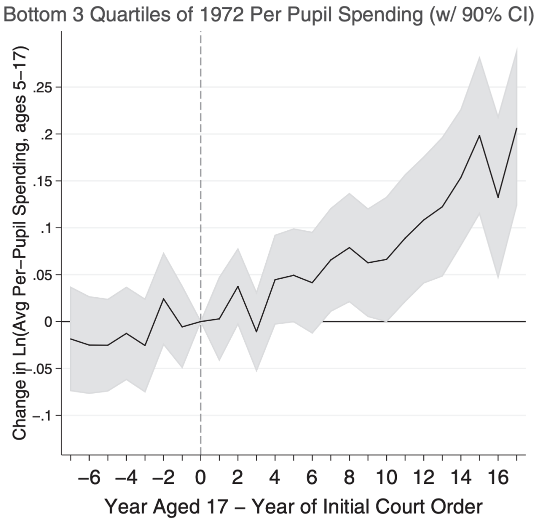
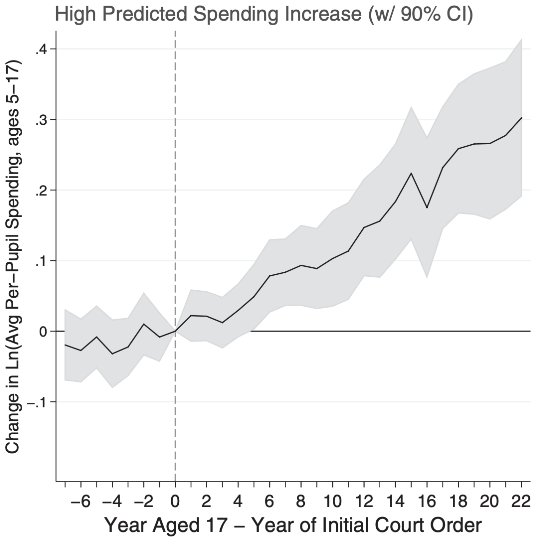


FIGURE I

First Stage: Dosage Design Yields a Stronger First Stage



Reduced-Form Event Study: Years of Education

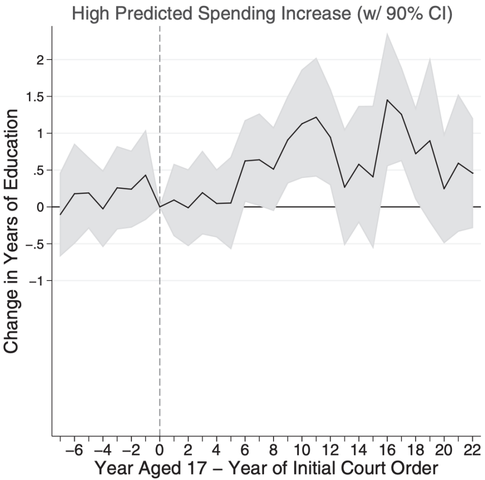


FIGURE III

Reduced-Form Event Study: Adult Wages

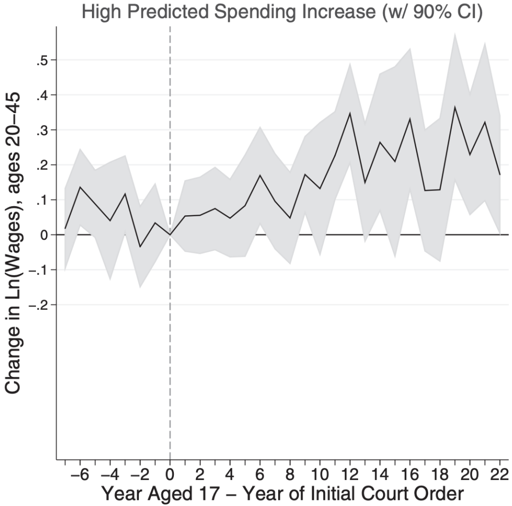
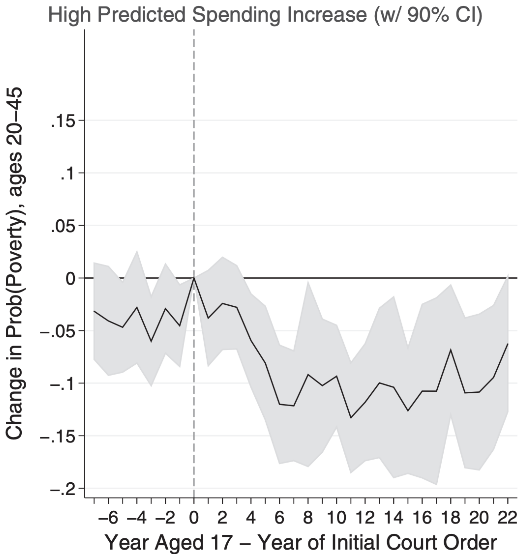


FIGURE IV

2SLS Results: Educational Attainment



Mechanisms: How Was the Money Spent?

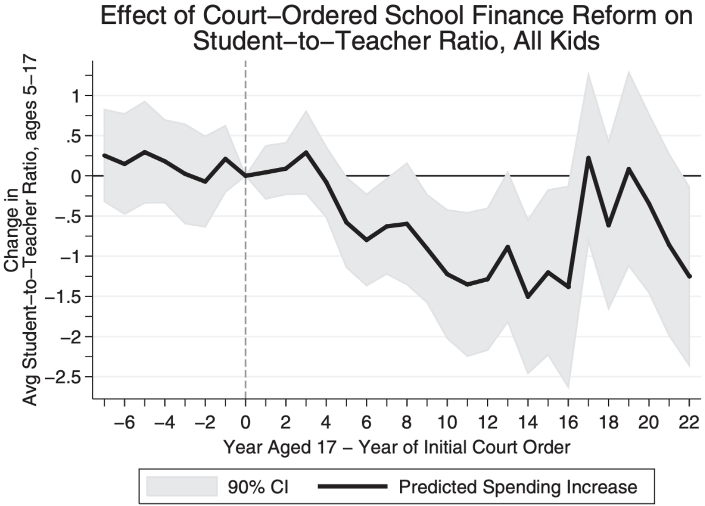


FIGURE VI

Takeaways from JJP (2016)

- School spending *does* matter, especially for low-income children
  - → Raises completed education by 0.31 years
  - → Increases adult earnings by $\sim$7%
  - → Reduces adult poverty by 3.2 percentage points
- The Coleman Report finding reflects endogeneity bias, not a true null effect — OLS is biased *against* finding positive spending effects
- How money is spent matters: reform-induced spending went to productive inputs (smaller classes, higher teacher pay, longer school years)
- Why hasn't doubled spending since 1970 closed gaps?
  - → Countervailing forces: rising income segregation, single-parent families, crack epidemic, mass incarceration
  - → Endogenous spending increases may not go to productive inputs

More Spending and Test Scores: Lafortune, Rothstein, and Schanzenbach (2018)

- JJP show that court-ordered SFRs raised spending in low-income districts and improved long-run outcomes (education, earnings, poverty)
- But JJP rely on a small PSID sample and focus on pre-1990 **equity** reforms
- Open questions:
  - → Do the post-1990 **adequacy** reforms—enacted at higher baseline spending levels—also raise achievement?
  - → Can we see effects on contemporaneous test scores using nationally representative data?
- LRS use state NAEP microdata (1990–2011) covering a much larger sample with math and reading test scores in 4th and 8th

First Stage in the Adequacy Era

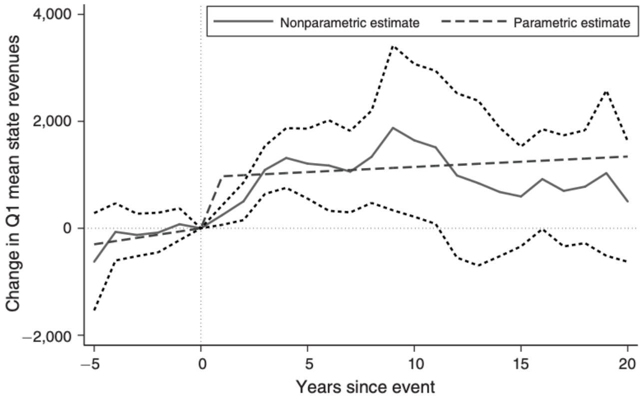


Figure 3. Event Study Estimates of Effects of School Finance Reforms on Mean State Revenues in Lowest Income Districts

Impacts on test scores in Low-Income Districts

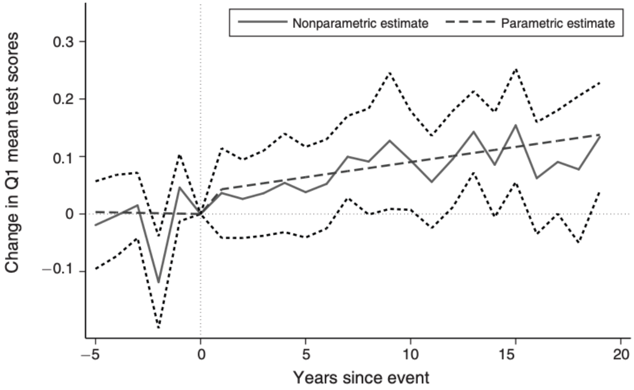


FIGURE 7. EVENT STUDY ESTIMATES OF EFFECTS OF SCHOOL FINANCE REFORMS ON MEAN TEST SCORES IN LOWEST INCOME SCHOOL DISTRICTS

Impacts on test scores in High-Income Districts

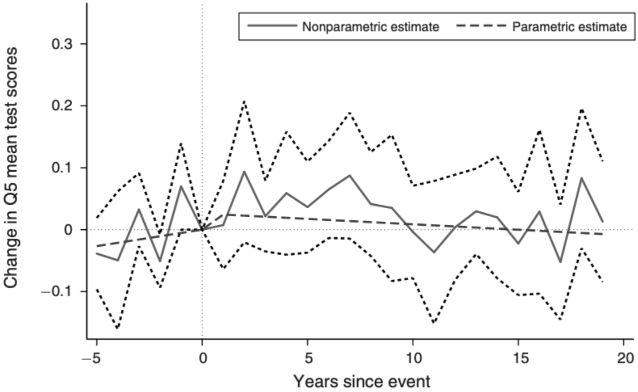


FIGURE 8. EVENT STUDY ESTIMATES OF EFFECTS OF SCHOOL FINANCE REFORMS ON MEAN TEST SCORES IN HIGHEST INCOME SCHOOL DISTRICTS

Does Money Matter? Yes — But Open Questions Still Remain

- The modern evidence is clear: money matters
- School spending increases raise long-run earnings, especially for disadvantaged students (JJP) and achievement (LRS)
- In production function terms: the isoquant map is not flat. Shifting the isocost outward *does* move you to higher achievement
- Class size reduction raises achievement (Krueger, Angrist and Lavy) and associated with better long-term outcomes (Chetty et al. 2011)
- But this raises the follow-up questions:
  - → What does the money *buy*?
  - → Which specific inputs have the largest marginal products?
  - → The literature overwhelmingly points to **teachers**

Teacher Value-Added

Teachers Are the Arguably the Most Important School Input

- The production function has many arguments: class size, spending, curriculum, facilities, peers, ...
- The literature has found that teacher quality variation is large and dwarfs other within-school inputs
- But teacher quality is not predicted by observable credentials
  - $\rightarrow$ Master's degrees, certifications, experience (beyond the initial years) explain very little of the variance
- This means we cannot identify effective teachers from their resumes and mostly rely on their *value-added*

Bias in Teacher VA

- Teacher VA models estimate each teacher's causal impact on test scores by controlling for student observables (especially lagged scores) and estimating teacher-specific effects
- Key policy question: do VA estimates reflect *causal* teacher effects, or are they contaminated by student sorting?
- If VA is biased, policies that reward/penalize teachers based on VA will misallocate incentives
- Two notions of bias:
  - → **Forecast bias**: is teacher VA unbiased *on average*?
  - → **Teacher-level bias**: does each individual teacher's VA estimate converge to her true effect?

Chetty, Friedman, and Rockoff (2014): Evaluating Teacher VA

Student i's test score in year t:

$$A^*_{it} = \beta X_{it} + \mu_{jt} + \varepsilon_{it}$$

- X_{it}: observables (lagged scores, demographics, class/school characteristics)
- μ_{jt}: teacher j's value-added in year t
- $\varepsilon_{it} = \theta_c + \tilde{\varepsilon}_{it}$: class shock + idiosyncratic student error
- Both X_{it} and ε_{it} may be correlated with μ_{jt}

Teacher quality drifts over time:

$$\text{cov}(\mu_{jt}, \mu_{j,t+s}) = \sigma_{\mu s}$$

- Depends only on lag s (stationarity), and allowed to decay
- A teacher who is effective this year may be somewhat less effective in five years (new curriculum, burnout, growth, etc.)

How Do We Estimate Teacher VA?

Strip out the effect of observables to get residual scores:

$$A_{it} = A^*_{it} - \beta X_{it} = \mu_{jt} + \varepsilon_{it}$$

Let $\bar{A}_{jt}$ denote the mean residual score in teacher j's class in year t. This is a noisy signal of μ_{jt}—it contains the teacher effect plus class-level noise

Idea: use a teacher's track record from *other* years to predict her quality in year t

$$\hat{\mu}_{jt} = \sum_{s \neq t} \psi_s \bar{A}_{js}$$

- This is a leave-year-out estimator: year t is excluded to avoid putting the same noise on both sides when we validate
- The weights ψ_s are chosen to minimize mean-squared prediction error
- Two forces shape the weights:
  - $\rightarrow$ **Noise**: past class means are noisy $\Rightarrow$ shrink toward the grand mean
  - $\rightarrow$ **Drift**: recent years are more informative $\Rightarrow$ weight them more heavily
- To compute optimal weights, we need to know how $\bar{A}_{jt}$ covaries across years for the same teacher

Estimation in Practice

Step 1: Residualize test scores.
Regress A^*_{it} on X_{it} with teacher fixed effects:

$$A^*_{it} = \alpha_j + \beta X_{it} \qquad \Rightarrow \qquad A_{it} = A^*_{it} - \hat{\beta} X_{it}$$

- Within-teacher variation identifies β, so coefficients on X aren't contaminated by sorting to teachers
- Baseline X_{it}: cubics in lagged math/English scores $\times$ grade, demographics, class/school means, class size, grade/year FEs

Step 2: Predict VA.
Regress mean class residuals $\bar{A}_{jt}$ on the vector of mean residuals from other years $\mathbf{A}_j^{-t} = (\bar{A}_{j1}, \ldots, \bar{A}_{j,t-1}, \bar{A}_{j,t+1}, \ldots)$:

$$\hat{\mu}_{jt} = \psi' \mathbf{A}_j^{-t}, \qquad \psi = \arg\min \sum_j \left(\bar{A}_{jt} - \sum_{s \neq t} \psi_s \bar{A}_{js} \right)^2$$

- This is just OLS—the optimal weights ψ are $\Sigma_A^{-1}\gamma$, where γ is the autocovariance vector and Σ_A is the VCV of $\mathbf{A}_j^{-t}$
- Stationarity (σ_{As} depends only on lag s) reduces the number of parameters: all teacher-year pairs s years apart are pooled to estimate each σ_{As}
- Leave-year-out: year t data excluded to avoid mechanical correlation between $\hat{\mu}_{jt}$ and outcomes in year t

Intuition for the Prediction Weights

The weights ψ solve a standard best linear prediction problem:

$$\psi = \arg\min \sum_j \left(\bar{A}_{jt} - \sum_{s \neq t} \psi_s \bar{A}_{js} \right)^2$$

Two special cases:

Case 1: One prior year only.

$$\hat{\mu}_{jt} = \frac{\sigma_{A1}}{\sigma_A^2} \bar{A}_{j,t-1}$$

- Single shrinkage factor incorporating both noise and drift

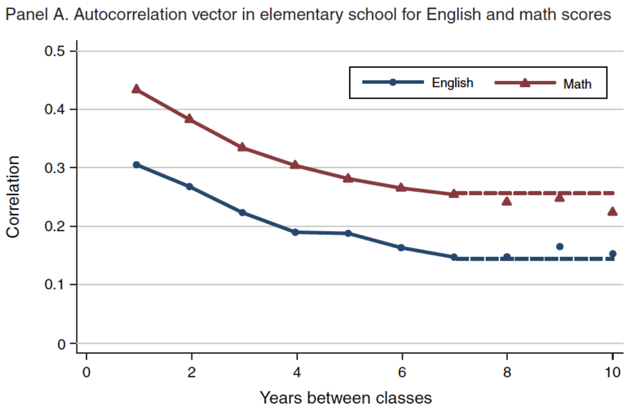
Case 2: Fixed quality ($\mu_{jt} = \mu_j$ for all t), i.i.d. errors.

$$\hat{\mu}_{jt} = \bar{A}_j^{-t} \cdot \frac{\sigma_\mu^2}{\sigma_\mu^2 + (\sigma_\theta^2 + \sigma_{\tilde{\varepsilon}}^2/n)/(T-1)}$$

- All years weighted equally—no drift, so year 1 is as informative as year $t-1$
- Standard Empirical Bayes shrinkage (Kane & Staiger 2008)

The general case nests both: recent years get more weight and there is still some shrinking to the mean

Some findings



1. More recent cohorts contain more signal for contemporaneous teacher VA
2. Sizable dispersion of teacher VA: A 1 SD improvement in teacher VA raises test scores by $\approx 0.10 - 0.14$ SD (Math and ELA)

Two Notions of Bias in VA Estimates

Forecast bias (CFR's emphasis):

$$B = 1 - \lambda, \qquad \lambda = \frac{\text{cov}(\mu_{jt}, \hat{\mu}_{jt})}{\text{var}(\hat{\mu}_{jt})}$$

- Does a 1-unit increase in estimated VA predict a 1-unit increase in scores *on average*?
- Not informative about teacher-level bias

Teacher-level bias (Rothstein (2017) emphasis):

- Does each *individual* teacher's VA estimate converge to her true causal effect?
- Write $\hat{\mu}_{jt} = \mu_{jt} + b_j + e_{jt}$, where b_j is systematic bias for teacher j
- Teacher-level unbiasedness: $b_j = 0$ for all j

Key: forecast unbiasedness ($B = 0$) does *not* imply teacher-level unbiasedness—some teachers could be systematically overrated and others underrated, yet the forecast bias could be zero

Relationship between Forecast Bias and Teacher-Level Bias

Decompose estimated VA: $\hat{\mu}_{jt} = \mu_{jt} + b_j + e_{jt}$

- μ_{jt}: true causal effect
- b_j: systematic (teacher-level) bias from sorting
- e_{jt}: estimation noise

Forecast coefficient:

$$\lambda = \frac{\text{var}(\mu_{jt}) + \text{cov}(\mu_{jt}, b_j)}{\text{var}(\hat{\mu}_{jt})} = \frac{\text{var}(\mu_{jt}) + \text{cov}(\mu_{jt}, b_j)}{\text{var}(\mu_{jt}) + \text{var}(b_j) + 2\text{cov}(\mu_{jt}, b_j)}$$

- Forecast unbiasedness ($\lambda = 1$) can occur if:
  - $\rightarrow$ $b_j = 0$ for all j (no teacher-level bias): Cov$(\mu_{jt}, b_j) = 0$ and var$(b_j) = 0$ and cov$(\mu_{jt}, b_j) = 0$
  - $\rightarrow$ var$(b_j) > 0$ (with teacher-level bias): $(b_j) > 0$ and cov$(\mu_{jt}, b_j) < 0$

More on forecast and teacher-level bias

- **Forecast bias** answers: "If we raise estimated VA by 1 unit, do we see a commensurate increase in test scores on average?"
  - → Relevant for aggregate gains from VA-based selection
- **Teacher-level bias** answers: "Is teacher j being correctly rated?"
  - → Relevant for high-stakes individual decisions (firing, bonuses, tenure)
  - → Even small forecast bias could coexist with many individually misclassified teachers
- Sometimes debate and research focuses too narrowly on forecast bias, while individual teachers face career consequences based on their *individual* VA estimates
- CFR provide evidence that teacher VA is forecast unbiased leveraging quasi-experimental teacher mobility

Validating VA: The Teacher Switching Quasi-Experiment

There is substantial teacher turnover: 30% switch grades within school, 6% switch schools, 6% leave the district each year

Idea: when a high-VA teacher leaves school s in grade g and is replaced by a lower-VA teacher, scores for the next cohort should fall—by roughly the predicted amount if VA is unbiased

Let Q_{sgt} denote the student-weighted mean of *leave-two-year-out* VA estimates $\hat{\mu}_{jt}^{-\{t,t-1\}}$ across teachers in school s, grade g, year t and let $\bar{A}_{sgt}$ be the cohort's average test scores. Estimate:

$$\Delta \bar{A}_{sgt} = a + b\,\Delta Q_{sgt} + \Delta \chi_{sgt}$$

Assumption: cov$(\Delta Q_{sgt},\, \Delta \chi_{sgt}) = 0$

- High-frequency teacher turnover is orthogonal to changes in student quality
- Plausible: families are unlikely to move neighborhoods because a single teacher leaves
- Analysis is at the school-grade-cohort level $\Rightarrow$ nonrandom assignment to classrooms *within* a school-grade is not a threat

Under exogenous teacher switching, $b = \lambda = 1 - B$.

Event Studies of Teacher Entry and Exit

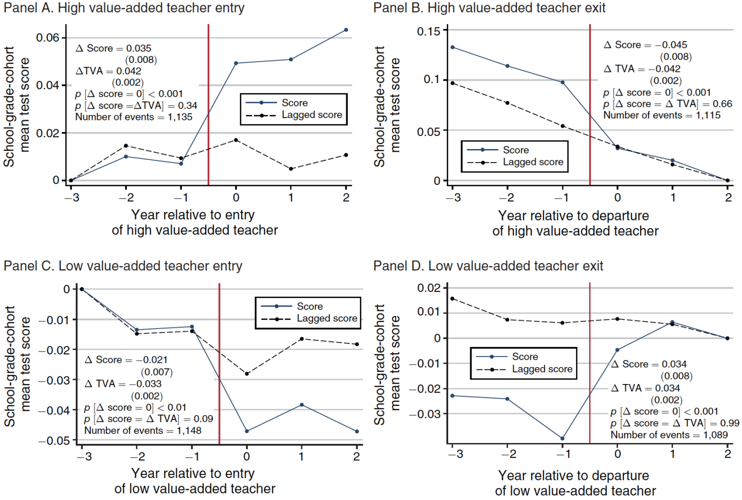
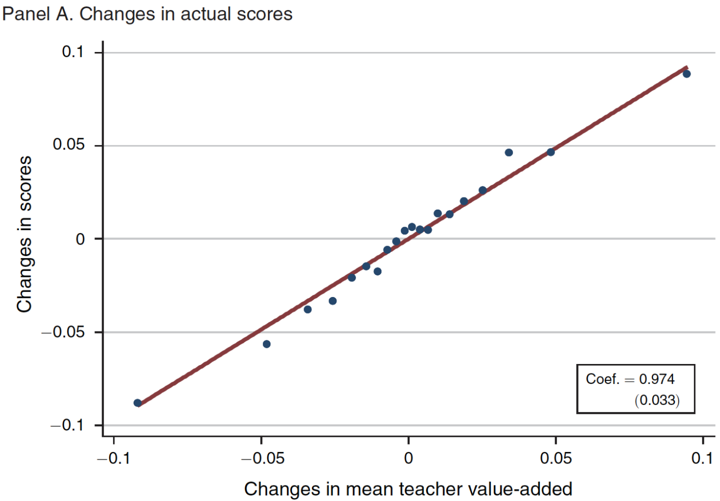


FIGURE 3. IMPACTS OF TEACHER ENTRY AND EXIT ON TEST SCORES

Estimate of Forecast Bias in the Entire Sample



Summarizing CFR

- Baseline quasi-experimental estimate: $\lambda = 0.974$ (SE $= 0.033$)
  - $\rightarrow$ Implied forecast bias: 2.6%, not statistically different from 0
  - $\rightarrow$ Upper bound of 95% CI: 9.1%
- Placebo tests support identification:
  - $\rightarrow$ Changes in predicted scores (parent chars) uncorrelated with Δ teacher VA
  - $\rightarrow$ In middle school, Δ teacher VA in math does not affect English scores (different teachers)
- Which controls matter? (Table 6)
  - $\rightarrow$ With lagged scores: bias $\approx 5\%$ (not significant)
  - $\rightarrow$ Without any score controls: bias $\approx 66\%$
  - $\rightarrow$ Lagged test scores are important

Chetty, Friedman, and Rockoff (2014b): VA and Adult Outcomes

- Question: what are the *long-run consequences* of having a good teacher?
- Link teacher VA estimates to students' adult outcomes using tax records (earnings, college attendance, neighborhood quality, teen birth)
- Results:
  - → A 1 s.d. increase in teacher VA associated with increase earnings at age 28 $\sim$1.3%
  - → Positive effects/association on college attendance, teen birth rates, and neighborhood quality

Teacher VA Predicts Better College Outcomes

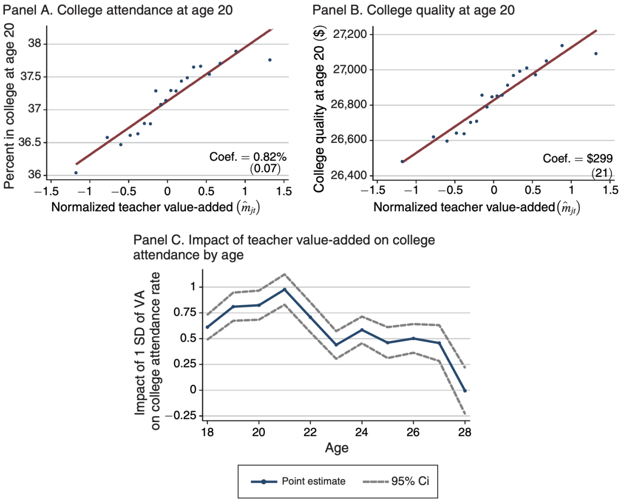


FIGURE 1. EFFECTS OF TEACHER VALUE-ADDED ON COLLEGE OUTCOMES

Teacher VA Predicts Higher Earnings at Age 28

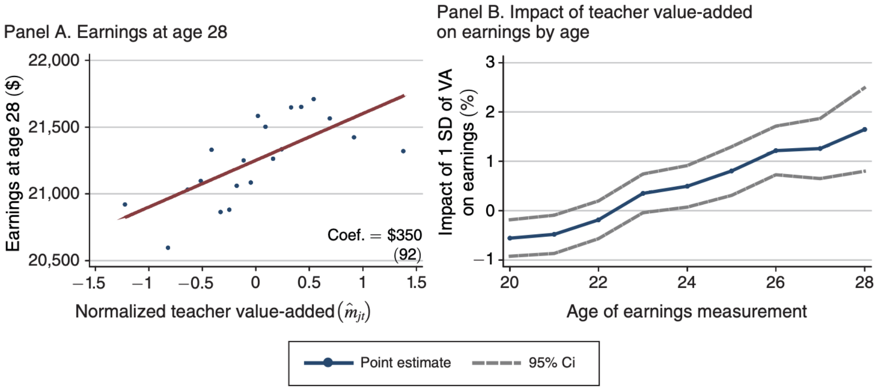


Figure 2. Effect of Teacher Value-Added on Earnings

Teacher VA Predicts Improvements in Other Outcomes

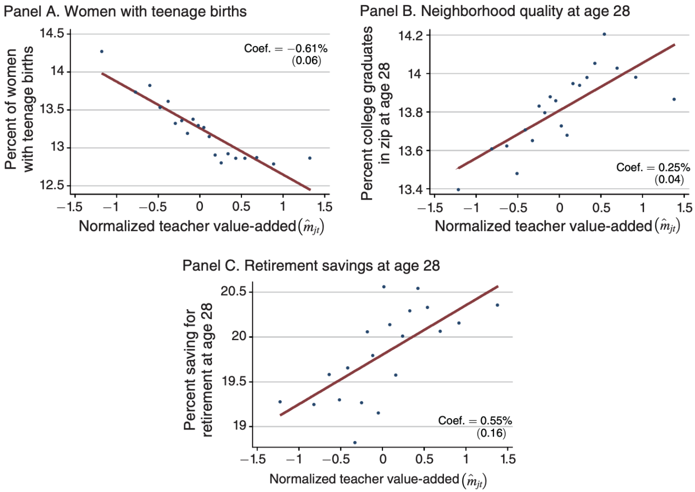
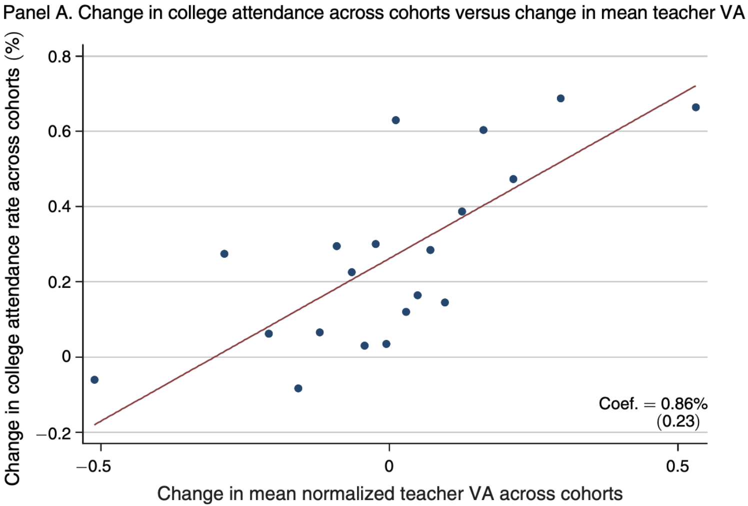


FIGURE 3. EFFECTS OF TEACHER VALUE-ADDED ON OTHER OUTCOMES IN ADULTHOOD

College Effects Robust to Quasi-Experimental Teacher Switching Approach



School Value-Added

From Teacher VA to School VA

- Teacher quality seems to matter with some reasonable caveats
- Can we do the same at the *school* level?
- This is both harder and perhaps more policy-relevant:
  - → Harder because students *sort across schools* far more actively than across teachers within a school (neighborhood choice, school choice, selective admissions)
  - → More policy-relevant because school choice, accountability, and competition policies all require measuring school effectiveness
- Rothstein concerns are even more severe for schools: if students select into schools based on unobservable characteristics correlated with achievement, school VA is biased
- But we do have credible policy variation to validate school VA? **Lotteries**

Constant Effects Value-Added Framework

$$Y_{ij} = \mu_j + a_i$$

- μ_j: mean potential outcome at school j
- a_i: student i's latent achievement potential
- Additive separability $\Rightarrow$ causal school effects are constant across students

Let D_{ij} indicate attendance at school j. Then observed outcomes satisfy

$$Y_i = Y_{i0} + \sum_{j=1}^{J} (Y_{ij} - Y_{i0}) D_{ij} = \mu_0 + \sum_{j=1}^{J} \beta_j D_{ij} + a_i, \qquad \beta_j \equiv \mu_j - \mu_0.$$

Interpretation: β_j is school j's value-added relative to an omitted reference school

Proxy for ability using observables

Project latent ability on observables (including lagged test scores):

$$a_i = X_i'\gamma + \varepsilon_i, \qquad E[X_i\varepsilon_i] = 0.$$

Substituting into observed outcome equation:

$$Y_i = \mu_0 + \sum_{j=1}^{J} \beta_j D_{ij} + X_i'\gamma + \varepsilon_i.$$

This equation is still causal but an empirical estimate of β_j need not equal the causal effect without additional assumptions

What OLS estimates

- Define the population projection

$$Y_i = \alpha_0 + \sum_{j=1}^{J} \alpha_j D_{ij} + X_i'\zeta + v_i.$$

- Because this is a projection,

$$E[v_i D_{ij}] = 0, \qquad E[v_i X_i] = 0.$$

- OLS has a causal interpretation only if school choices are unrelated to the unobserved component of ability:

$$E[\varepsilon_i \mid D_{ij}] = 0, \qquad j = 1, \ldots, J. \tag{*}$$

- If (*) is true, $\alpha_j = \beta_j$.
- Otherwise define school-specific bias by

$$\alpha_j = \beta_j + b_j.$$

  Can we formally assess if OLS estimates recover β_j reasonably well?

Angrist, Hull, Pathak, and Walters (QJE 2017): Overview

- Just as CFR used teacher switches to validate teacher VA, AHPW use *admission lotteries* to validate school VA
- Oversubscribed schools in Boston (charters, exam schools) use lotteries for admission
- The lottery provides an *experimental benchmark*: the causal effect of attending oversubscribed school j
- The question: does observational school VA line up well with these lottery-based causal effects?
- Paper makes two broad contributions:
    1. A formal framework for testing VA validity using lotteries
    2. An empirical Bayes combined estimator that optimally weights observational and lottery-based information; partly addresses the individual/school level bias concerns related to the Rothstein critique of teacher-level VA bias

Lottery-based validation setup

- Let C_i represent lottery strata indicators plus baseline controls and let lottery offers for L oversubscribed schools be represented by

$$Z_i = (Z_{i1}, \ldots, Z_{iL})'$$

- Lottery offers are conditionally random, so

$$E[\varepsilon_i \mid C_i, Z_i] = \lambda_0 + C_i'\lambda_c. \tag{7}$$

Interpretation:

- Lotteries are valid instruments for school attendance
- Even when $L < J$, the extra lottery moments become overidentifying restrictions once OLS has pinned down the J school coefficients
- This turns VAM validation into an IV overidentification problem

LM omnibus test of VAM validity

- Let $\hat{\varepsilon}$ be residuals from the OLS VAM. The LM statistic is

$$\hat{T} = \frac{\hat{\varepsilon}' P_{\tilde{Z}} \hat{\varepsilon}}{\hat{\sigma}^2_{\tilde{\varepsilon}}}, \qquad \hat{\sigma}^2_{\tilde{\varepsilon}} = \frac{\hat{\varepsilon}' M_C \hat{\varepsilon}}{N},$$

  where

$$P_C = C(C'C)^{-1}C', \qquad M_C = I - P_C,$$

  and the projection onto residualized lottery offers:

$$P_{\tilde{Z}} = M_C Z(Z' M_C Z)^{-1} Z' M_C$$

- Under the joint null,
  1. selection-on-observables holds, and
  2. lottery offers are valid instruments,

  we have

$$\hat{T} \overset{a}{\sim} \chi^2_L.$$

- Intuition: after conditioning on lottery strata C_i, lottery offers should explain none of the OLS residual (because OLS and TSLS coincide when D_{ij} is in the instrument set)

Decomposing the omnibus test

Let $\hat{Y}_i$ denote fitted values from the OLS VAM, and define the IV forecast coefficient

$$\hat{\phi} = (\hat{Y}' P_{\tilde{Z}} \hat{Y})^{-1} \hat{Y}' P_{\tilde{Z}} Y.$$

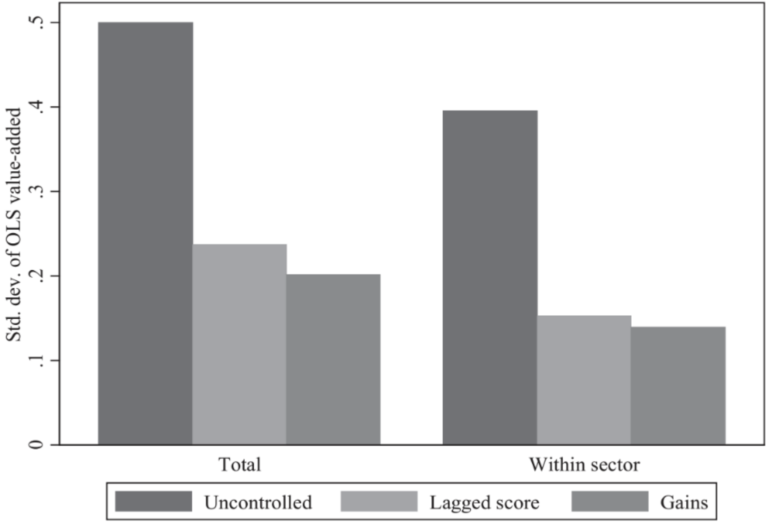
Since $\hat{\varepsilon} = Y - \hat{Y} = (Y - \hat{\phi}\hat{Y}) + (\hat{\phi} - 1)\hat{Y}$, the omnibus test becomes

$$\hat{T} = \frac{(\hat{\phi} - 1)^2}{\hat{\sigma}^2_{\tilde{\varepsilon}} (\hat{Y}' P_{\tilde{Z}} \hat{Y})^{-1}} + \frac{(Y - \hat{\phi}\hat{Y})' P_{\tilde{Z}} (Y - \hat{\phi}\hat{Y})}{\hat{\sigma}^2_{\tilde{\varepsilon}}}. \tag{9}$$

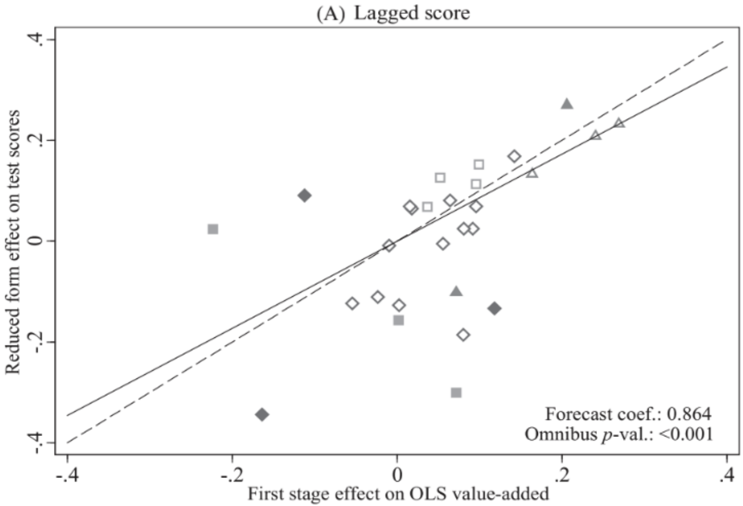
So the omnibus test combines:

1. **Forecast-bias** test: $H_0 : \phi = 1$;
2. **Overidentification** test: does predictive validity of OLS VAM hold for each lottery?

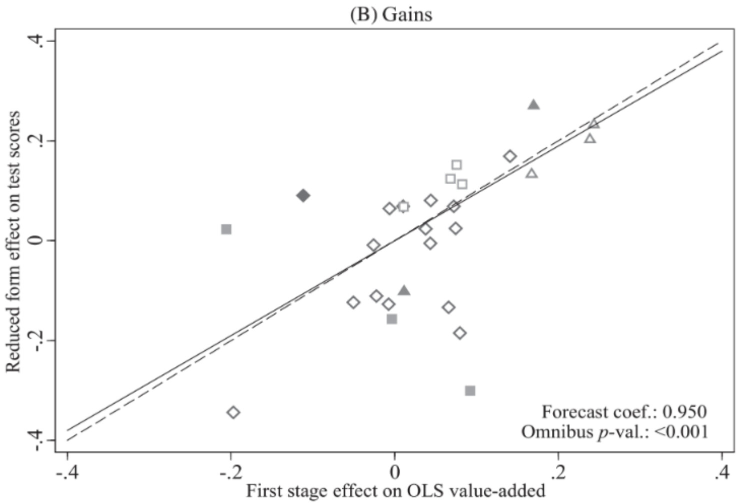
Dispersion in school effectiveness



Visual Instrumental Variables Plot Summarizing Validity Test



Visual Instrumental Variables Plot Summarizing Validity Test



OLS VAM does reasonably well, on average

TABLE III

TESTS FOR BIAS IN CONVENTIONAL VALUE-ADDED MODELS

	All lotteries		Excluding charter lotteries	
	Lagged score (1)	Gains (2)	Lagged score (3)	Gains (4)
Panel A: Sixth grade				
Forecast coefficient (φ)	0.864	0.950	0.549	0.677
	(0.075)	(0.084)	(0.164)	(0.193)
First stage F-statistic	29.6	26.6	11.2	9.3
p-values:				
Forecast bias	0.071	0.554	0.006	0.095
Overidentification	0.003	0.006	0.043	0.052
Omnibus test χ^2 statistic (d.f.)	77.7 (28)	72.1 (28)	48.0 (23)	41.7 (23)
p-value	<0.001	<0.001	<0.001	0.010
N	8,718		6,162	
Panel B: All middle school grades				
Forecast coefficient (φ)	0.880	0.924	0.683	0.726
	(0.055)	(0.060)	(0.124)	(0.133)
First stage F-statistic	14.7	15.0	7.6	7.8
p-values:				
Forecast bias	0.028	0.204	0.011	0.039
Overidentification	0.011	0.011	0.062	0.065
Omnibus test χ^2statistic (d.f.)	172.8 (75)	167.0 (75)	111.6 (60)	107.9 (60)
p-value	<0.001	<0.001	<0.001	<0.001
N	20,935		15,027	

Well on average does not rule out school-specific bias

"How do we combine OLS with lotteries to create a hybrid estimator to account for school-specific bias?"

Estimator	Strength	Weakness
OLS VAM	precise, available for all schools	biased
Lottery-based estimates	credible	noisy, incomplete coverage

AHPW propose a model for the joint distribution of value-added, bias, and lottery compliance, then form EB-style posterior predictions of school effectiveness.

Intuition for where we are heading: Create a combined estimator that is a weighted average of lottery estimates and OLS estimates

$$\hat{\theta}_j^{\text{combined}} = \lambda_j\,\hat{\theta}_j^{\text{LOT}} + (1-\lambda_j)\,\hat{\theta}_j^{\text{OBS}}$$

Random coefficients lottery model

The model considers reduced-form, and first-stage estimates:

$$Y_i = \tau_0 + C_i'\tau_c + Z_i'\rho + u_i, \qquad D_{ij} = \phi_{0j} + C_i'\phi_{cj} + Z_i'\pi_j + \eta_{ij}.$$

For lottery ℓ,

$$\rho_\ell = \sum_{j=1}^{J} \pi_{\ell j}\beta_j.$$

So each lottery identifies a *weighted average* of school effects (with weights depending on fallback option shares), not just the target school's effect. Observed estimates are noisy measurements of latent school parameters:

$$\hat{\alpha}_j = \beta_j + b_j + e_j^{\alpha},$$

$$\hat{\rho}_\ell = \sum_{j=1}^{J} \pi_{\ell j}\beta_j + e_\ell^{\rho}, \qquad \hat{\pi}_{\ell j} = \pi_{\ell j} + e_{\ell j}^{\pi}.$$

- β_j: causal value-added
- b_j: OLS bias
- $\pi_{\ell j}$: attendance shifts induced by lottery ℓ

Parameterizing first stages: compliance and fallback schools

The own-school first stage is parameterized as

$$\pi_{\ell\ell} = \frac{\exp(\delta_\ell)}{1 + \exp(\delta_\ell)}.$$

Cross-school first stages are

$$\pi_{\ell j} = -\pi_{\ell\ell} \cdot \frac{\exp(\xi_j + \nu_{\ell j})}{1 + \sum_{k \neq \ell} \exp(\xi_k + \nu_{\ell k})}, \qquad j \neq \ell.$$

Parameter intuition

- δ_ℓ: how strongly students comply with an offer at school ℓ
- ξ_j: how attractive school j is as a fallback option
- $\nu_{\ell j}$: lottery-specific fallback shock
- Logit structure disciplines first stage effects to the relevant intervals

Prior distribution

Each school has four latent parameters:

$$(\beta_j, b_j, \delta_j, \xi_j).$$

Conditional on whether lottery data are available,

$$(\beta_j, b_j, \delta_j, \xi_j)' \mid Q_j \sim N\Big((\beta_0 + \beta_Q Q_j,\ b_0,\ \delta_0,\ \xi_0)', \Sigma\Big),$$

and

$$\nu_{\ell j} \mid Q_j \sim N(0, \sigma_\nu^2).$$

Key Assumptions:

1. Mean value-added can differ between lottery and non-lottery schools: $\beta_Q \neq 0$
2. The distribution of bias and the covariance structure in Σ is the same across lottery and non-lottery schools

Connection to conventional empirical Bayes framework

If OLS estimates were unbiased, the standard EB posterior mean (recall from Chetty et al. 2014a) would be

$$E[\alpha_j \mid \hat{\alpha}_j] = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \mathrm{Var}(e_j^\alpha)}\hat{\alpha}_j + \left(1 - \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \mathrm{Var}(e_j^\alpha)}\right)\alpha_0.$$

Interpretation:

- noisy school estimates are shrunk toward the grand mean;
- precise estimates are shrunk less.

AHPW consider noise e_j^α *and* bias b_j

$$\hat{\alpha}_j = \beta_j + b_j + e_j^\alpha,$$

so their model implies a different posterior mean

Hybrid empirical Bayes posterior

Suppose the first-stage matrix Π is known. Then the posterior mean of β is

$$E[\beta \mid \hat{\alpha}, \hat{\rho}] = W_\alpha(\hat{\alpha} - b_0\iota) + W_\rho\hat{\rho} + (I - W_\alpha - W_\rho\Pi)\beta_0\iota.$$

The hybrid estimator is an MMSE combination of

- Reduced form estimates via $W_\rho\hat{\rho}$,
- OLS net of average bias via $W_\alpha(\hat{\alpha} - b_0\iota)$,
- the prior mean via $(I - W_\alpha - W_\rho\Pi)\beta_0\iota$

Overall, the hybrid EB estimator uses lotteries to calibrate posteriors, again depending on key assumption that the distribution of bias is the same across lottery and non-lottery schools

Model Estimates: Sizable dispersion in estimated VAM bias!

TABLE VI

ESTIMATES OF THE JOINT DISTRIBUTION OF CAUSAL VALUE-ADDED AND VAM BIAS FOR SIXTH-GRADE MATH SCORES

		Models without sector effects			Models with sector effects	
		Uncontrolled	Lagged score	Gains	Lagged score	Gains
Parameters		(1)	(2)	(3)	(4)	(5)
σ_β	Std. dev. of causal VA	0.195	0.220	0.222	0.171	0.170
		(0.024)	(0.021)	(0.023)	(0.028)	(0.023)
σ_b	Std. dev. of VAM bias	0.501	0.182	0.166	0.148	0.133
		(0.061)	(0.048)	(0.048)	(0.029)	(0.030)
$\sigma_{\beta b}$	Covariance of	−0.018	−0.014	−0.017	−0.016	−0.013
	VA and bias	(0.010)	(0.003)	(0.004)	(0.006)	(0.003)
r_α	Regression of VA	0.078	0.644	0.753	0.694	0.783
	on OLS (reliability ratio)	(0.204)	(0.066)	(0.072)	(0.152)	(0.122)
VA shifters	Charter				0.426	0.396
					(0.104)	(0.106)
	Pilot				0.130	0.111
					(0.129)	(0.129)
	Lottery school (β_Q)	0.040	−0.024	−0.033	0.104	0.066
		(0.127)	(0.061)	(0.054)	(0.042)	(0.041)
Bias shifters	Charter				−0.005	−0.063
					(0.103)	(0.099)
	Pilot				−0.121	−0.089
					(0.124)	(0.121)
	χ^2statistic (d.f.):	10.9 (7)	10.8 (7)	9.1 (7)	9.0 (13)	6.0 (13)
	Overid. p −value:	0.145	0.147	0.247	0.773	0.946

Notes. This table reports simulated minimum distance estimates of parameters of the joint distribution of causal school value-added and OLS bias for sixth-grade math scores. The moments used in estimation are functions of OLS value-added, lottery reduced-form, and first-stage estimates, as described in Online Appendix B. Uncontrolled estimates come from an OLS regression that includes year effects. The notes to Table III describe the other value-added models. Simulated moments are computed from 500 samples constructed by drawing school-specific parameters from the random coefficient distribution along with estimation errors based on the asymptotic covariance matrix of the estimates. The estimates in columns (4) and (5) are from models allowing the means of the random coefficients distribution to depend on school sector. Moments are weighted by an estimate of the inverse covariance matrix of the moment conditions, calculated from a first-step estimate using an identity weighting matrix. The weighting matrix is produced using 1,000 simulations, drawn independently from the samples used to simulate the moments.

Application to Remote Schooling

Who Benefits from Remote Schooling? Self-selection and Match Effects

Jesse Bruhn, Christopher Campos, Eric Chyn, and Anh Tran

Motivation

- School choice expansions hypothesized to improve student-school match quality (Hoxby 2003)
- Match effects are theoretically important but evidence is thin
  - → Do parents know their match quality? Limited information (Abdulkadiroglu et al. 2020)
  - → Weak evidence of sorting on match quality (Campos and Kearns 2024)
- The pandemic provides a unique context: families *compelled* to try remote learning and learned about their match quality
- Post-pandemic demand for remote learning is large and sustained
  - → Enrollment in exclusively virtual schools increased ∼75% nationally relative to 2019
  - → All 40 largest U.S. districts currently offer a remote option

This Paper

- Leverage the pandemic as a shock to families' awareness of their remote learning match quality
- Focus on Los Angeles, where families self-selected into remote options in 2021–2022
- Collect novel survey data with conjoint choice experiments (Mas and Pallais 2017; Wiswall and Zafar 2018)
- **Methodological innovation:** Link experimentally derived preference estimates to a potential outcomes model to identify treatment effects and characterize selection
- **Questions:**
  1. What are families' preferences for remote learning?
  2. What are the causal effects of remote learning on cognitive and non-cognitive outcomes?
  3. Is there evidence of sorting on match quality (selection on gains)?
  4. Can choice experiments be a useful tool for program evaluation?

Remote Schooling in Los Angeles

Pandemic era

- LAUSD closed March 2020; stayed remote until April 2021
- Families experienced $\sim$1 year of remote instruction, providing ample time to assess their relative suitability

Post-pandemic era (our focus)

- California mandates schools offer a remote option for 2021–2022
- LAUSD creates “City of Angels” virtual school
- 4.7% of students ($\sim$14,000) enroll in the remote option
- LAUSD subsequently opens six permanent virtual academies for 2022–2023

Data and Survey Design

LAUSD Administrative Data (2019–2022)

- Demographics, addresses, standardized test scores
- School Experience Survey: bullying and grit indices (Jackson et al. 2020)

Novel Survey Data

- $\sim$3,400 survey responses; 1,171 complete the conjoint experiment
- Conjoint design: respondents rank 3 hypothetical schools across $K = 10$ choice trials, varying three attributes:
  - $\rightarrow$ Academic quality (peer achievement rate)
  - $\rightarrow$ Travel time
  - $\rightarrow$ Learning modality (in-person vs. remote)
- Respondents are positively selected $\rightarrow$ we model preference heterogeneity parametrically and extrapolate to the full sample

Preference Estimation

- Indirect utility of school j for respondent i:

$$U_{ij} = \underbrace{\omega_Q Q_j + \omega_R \mathit{Remote}_j + \omega_d d_{ij}}_{V_{ij}} + \varepsilon_{ij}$$

  → Q_j: Academic quality of hypothetical school j
  → Remote_j: Remote learning indicator
  → d_{ij}: Travel time ($= 0$ for remote options)

- Logit assumption on ε_{ij} → estimate via exploded logit on rank-ordered lists

- Experimental variation in attributes credibly identifies preferences

- Two summary measures:

  → Willingness to travel: $-\omega_Q/\omega_d$

  → Achievement compensation for remote: $-\omega_R/\omega_Q$

Conceptual Framework: The Selection Problem

- Outcome equation:

$$Y_i = \alpha + X_i'\gamma + \beta D_i + e_i$$

- D_i: remote enrollment indicator; X_i: baseline covariates
- Selection into remote is not random: $E[e_i|D_i] \neq 0$
- Two schooling options: in-person ($j = 0$) and remote ($j = 1$). Normalize $V_{i0} = 0$, so relative indirect utility of remote is:

$$u_i = v_i + \varepsilon_i$$

  where v_i is the systematic (preference-driven) component

- **Assumption 1:**

$$Y_i(1), Y_i(0) \perp D_i \mid v_i$$

  Conditional on the systematic component of preferences, treatment assignment is as-good-as-random

Conceptual Framework: Using Choice Experiments

Key Assumption (Assumption 2): Let $s_i = (Q_i, d_i)$ be observed school attributes and $\boldsymbol{\omega}_i = (\omega_{Qi}, \omega_{Ri}, \omega_{di})$ the preference vector identified by the conjoint:

$$Y_i(1), Y_i(0) \perp D_i \mid v_i = v(s_i, \boldsymbol{\omega}_i)$$

- The choice experiments generate random variation that identifies $\boldsymbol{\omega}_i$ — the marginal rates of substitution governing choices
- Why are experiments necessary?
  - → Observational estimates of $\boldsymbol{\omega}_i$ would be confounded
  - → Conditioning on s_i alone is insufficient — two families at the same school can have different preferences and thus different v_i
  - → ω_{Ri} is a remote-schooling taste shifter with no interaction with observable data
- Bottom line: Experimentally identified preferences allow us to learn about a key determinant of selection that would otherwise be unobservable

From Preferences to Treatment Effects

- The logit distributional assumption implies a propensity score:

$$P(v_i) = \frac{\exp(v_i)}{1 + \exp(v_i)}$$

- **Testable implication** (Rosenbaum and Rubin 1983): Conditional on $P(v_i)$, baseline characteristics are balanced: $E[X_i | D_i = 1, P(v_i)] = E[X_i | D_i = 0, P(v_i)]$
- **Empirical specification:**

$$Y_i = \alpha_c + \gamma' X_i + \beta D_i + \theta \hat{P}(v_i) + \psi \hat{P}(v_i) \times D_i + e_i$$

→ β: average treatment effect of remote learning

→ θ: selection on levels (do high-taste students do well regardless?)

→ ψ: selection on gains / match effects (do high-taste students benefit *more* from remote?)

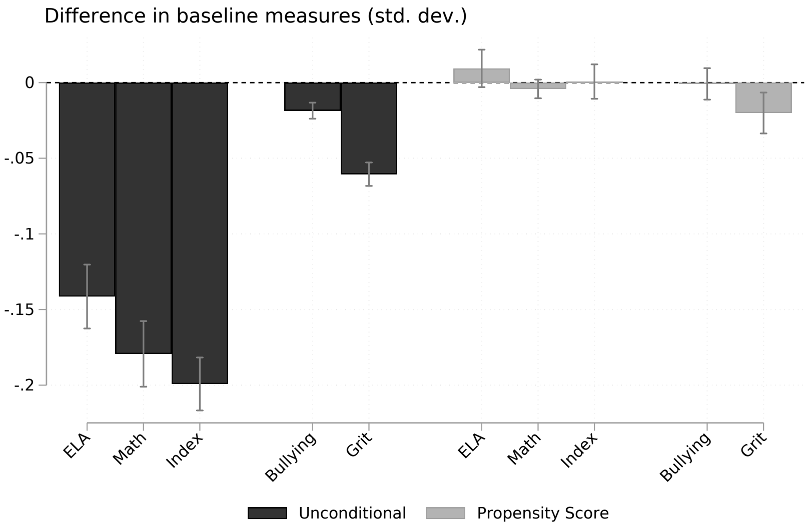
Estimation Details

- Preferences vary by covariate cell $c(X_i)$ (baseline achievement $\times$ poverty $\times$ URM $\times$ sub-district):

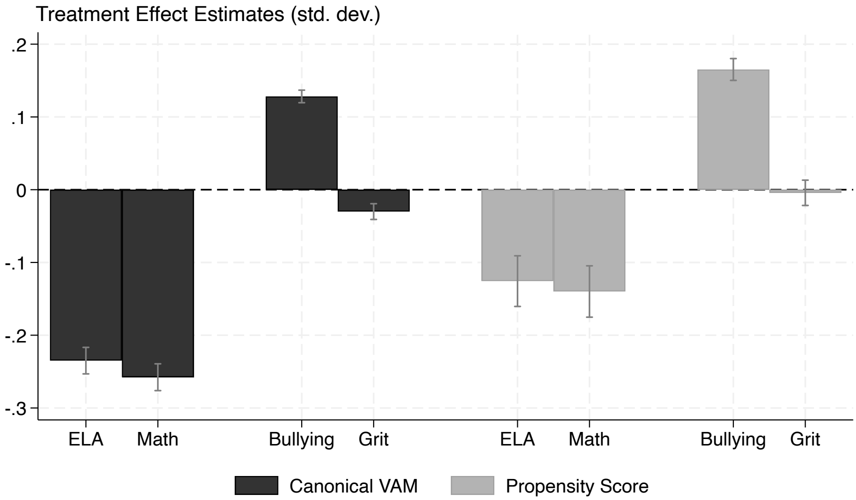
$$U_{ij} = \omega_{Qc(X_i)} Q_j + \omega_{Rc(X_i)} \mathit{Remote}_j + \omega_{dc(X_i)} d_{ij} + \varepsilon_{ij}$$

- Estimate separately by cell via MLE on rank-ordered lists from conjoint respondents
- **Extrapolation:** Apply cell-specific $\hat{\boldsymbol{\omega}}_{c(X_i)}$ to actual school attributes for *all* LAUSD students
- Covariate cell fixed effects α_c in the outcome equation ensure identification comes from within-cell variation in how students trade off quality and distance — not from across-cell demographic differences
- **Validation:**
  - → Extrapolated propensity scores are forecast unbiased (slope ≈ 0.96)
  - → Propensity scores predict real-world remote enrollment ($\hat{\beta} = 1.11$, $R^2 = 0.09$ within cells)

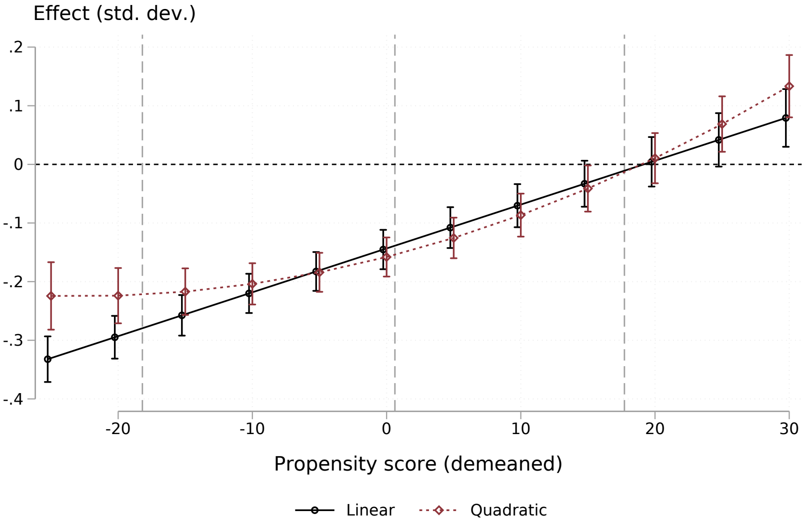
Baseline Balance



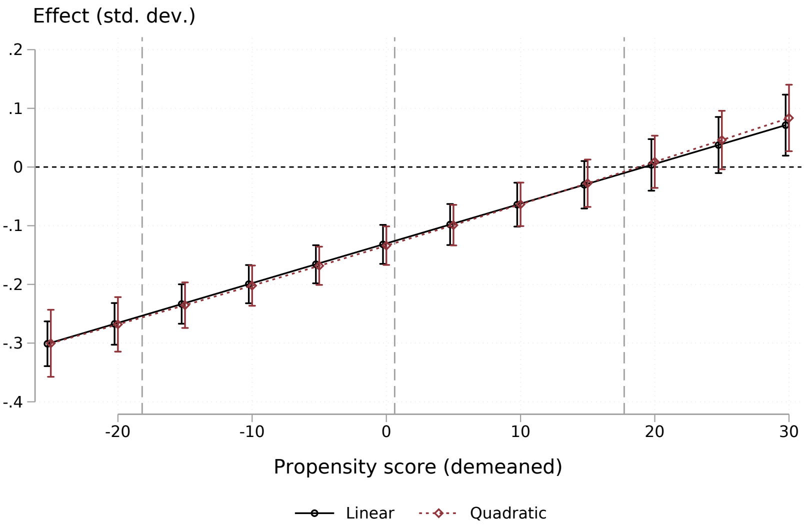
Average Remote Schooling Effects



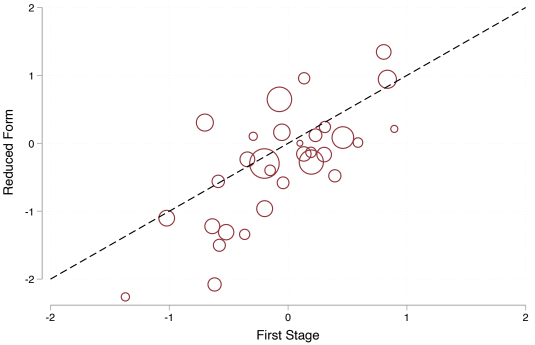
Evidence of Selection on Gains



Evidence of Selection on Gains



Lottery-Based Validation



- 32 oversubscribed LAUSD programs where remote is the fallback option
- Use lottery offers as instruments: does predicted $\tilde{Y}_i$ from our model forecast observed Y_i?
- Forecast coefficient: 1.03 (math), 0.67 (ELA); fail to reject $\phi = 1$
- Reassuring evidence that conjoint-based estimates capture the relevant selection

Concluding Thoughts

Findings

- Remote learning reduces achievement on average (-0.13 to -0.14σ), but positive match effects imply high-demand students fare no worse or benefit
- Bullying outcomes improve substantially ($+0.17\sigma$), suggesting a compensating differential for academic losses
- Families appear to sort on match quality: Roy-style selection on gains

Methodological contribution

- Choice experiments can characterize selection into treatment when observational approaches fail
- Validated against lottery variation — forecast unbiased
- Bridges conjoint/stated preference literature and program evaluation

Summing up the lecture

- **Motivating observation:** There is enormous variation in student outcomes across schools. What explains it?
- **One lens: schools as producers.** The education production function organizes the key empirical questions: What are the marginal products of different inputs? Can they substitute for each other? Does money matter?
- **Money matters — but how it's spent matters more:** Modern quasi-experimental evidence (JJP 2016, LRS 2018) overturns the Hanushek "money doesn't matter" claim.
- **Teachers are arguably the most important school input.** Teacher VA—with lagged test scores—is forecast unbiased (CFR, $\hat{\lambda} \approx 0.97$) and predicts long-run gains: 1 SD $\Rightarrow$ 1.3% higher earnings at 28.
- **Forecast unbiasedness $\neq$ individual-level unbiasedness.** VA can be forecast unbiased on average while systematically over- or under-rating specific teachers or schools. This matters for high-stakes individual decisions (firing, bonuses, accountability).
- **Next week:** What can policy do about the enormous variation in student outcomes across schools?