Personnel Economics

Understanding People, Practices, and Productivity

Chris Stanton

Harvard Business School

April 2026

What Is Personnel Economics?

Broadly: the study of labor economics issues *inside* firms and organizations

- Applies economic tools to human resources (HR)
- Unit of emphasis: the *firm* or group of firms
- Workers' responses are inputs to firm decisions, not the primary outcome

Contrast with labor economics:

- Labor economist on parental leave → child outcomes, inequality
- Personnel economist on parental leave → policy design tradeoffs: substitution for other compensation, retention, sorting

Core topical areas:

- Incentives and compensation
- Hiring and selection
- Managers and peer effects
- Technology and time use
- Training

Intellectual lineage:
Pathbreaking work of Eddie Lazear; emerged from the Chicago tradition (Becker); now diffused worldwide

Key point: Personnel economics focuses on the firm side of the employment relationship—how practices affect firm performance and what that implies for policy design.

Why Personnel Practices Matter

Productivity differences between firms are enormous:

- Average TFP gap between the 75th and 25th percentile within a 4-digit industry: ≈45 log points (Syverson 2004)
- Dispersion is large even in commodity industries with limited product differentiation
- Management practices correlate strongly with productivity (Bloom & Van Reenen 2007, 2011)

Productivity differences within firms:

- Worker output varies dramatically at the *same job*
- These differences often exceed pay differences across job levels
- Sources: ability, effort, match quality, management

What Can Firms Do?

$$y_i = T_j \times h_{ij} \times e_i$$

T_j	firm technology / practices
h_{ij}	human capital in setting j
e_i	worker effort

Firms can influence all three levers:

- Better management (T)
- Better hiring / training (h)
- Better incentives / monitoring (e)

Key point: Understanding what drives productivity—and what practices matter —are the central agenda of personnel economics.

Main PPL Approach: "Insider Econometrics"

What it means:

- Researchers work in close collaboration with companies
- Access proprietary administrative records: individual productivity, wages, tenure, manager IDs, disciplinary actions
- Design field experiments (RCTs) *within* organizations
- Exploit natural experiments generated by firm policy changes
- Analyze digital trace data: time stamps, call records, clicks

Why this matters:

- Yields causal evidence on returns to HR practices
- Enables measurement of productivity at the individual level (not just wages)
- Allows researchers to observe the same worker across managerial assignments, incentive regimes, or team compositions

Contrast:

Traditional labor economics uses large administrative surveys or matched employer–employee datasets where output is rarely observed

Key point: The insider approach generates granular causal evidence that is impossible to obtain from standard datasets—but it raises serious questions about external validity that must be considered.

Roadmap for the Session

Part 1: Stylized facts

- How large is productivity dispersion?
- The production function framework
- Three channels through which policy acts

Part 2: Pay practices and worker heterogeneity

- Principal–agent framework
- Safelite Auto Glass (Lazear 2000, AER)
- Pakistani teacher incentives (Brown & Andrabi 2023)
- Oligopsony theory; efficiency wages (Emanuel & Harrington 2020)
- Heterogeneous responses to incentive pay (Sandvik et al. 2021, MSCI)
- Wages and different workers' effort over the business cycle (Lazear, Shaw & Stanton 2016, JoLE)
- Valuing amenities (Mas & Pallais 2017, AER)

Part 3: Non-pay Practices

- Why doesn't knowledge diffuse inside firms?
- Field experiment (Sandvik et al. 2020, QJE)
- Discretion in hiring (Hoffman, Kahn & Li 2018, QJE)
- Supervisors (Frederiksen, Kahn & Lange 2020, JPE)
- Talent hoarding and internal frictions (Haegele 2022)

The Central Puzzle: Workers Doing the Same Job Vary Enormously

Empirical regularity:

- Substantial, persistent productivity differences exist across workers doing *identical tasks* in *identical settings*
- Differences of comparable magnitude to those across firms or plants
- Productivity variation *dwarfs* pay variation within job titles — compensation systems compress wages relative to output
- Found at all skill levels: cashiers, bus drivers, sales agents, teachers, emergency physicians

Why this is surprising:

- Standard models predict competitive labor markets drive wages to marginal products
- If wages tracked output, dispersion in pay and output would match
- Instead: the market is not fully sorting on ability within a given firm

What it suggests:

Firms that close productivity gaps — through hiring, incentives, or management — can earn rents relative to competitors

Key point: Persistent, large productivity differences at the individual level are among the most robust findings in personnel economics.

Productivity Dispersion: Evidence Across Occupations

Hoffman & Stanton (2025), Table 1 — workers doing the same task, same setting

Paper	Setting	Statistic	Difference
Mas & Moretti (2009)	Grocery cashier checkout	P90−P10	21%
Soetevent & Romensen (2024)	Bus drivers (fuel, braking)	C.V.	0.05–0.86
Lazear et al. (2015)	Technology service agents	C.V.	0.13
Sandvik et al. (2020)	Revenue on random sales calls	P75−P25	48%
Staiger & Rockoff (2010)	Teacher value added	SD of VA	0.15
Chan & Chen (2022)	ED visit cost (NPs & MDs)	SD; P75−P25	21%; $650k
Chan et al. (2022)	Chest X-ray accuracy	P90−P10	22%

Notes: Estimates other than Mas and Moretti use shrinkage. C.V. = coefficient of variation.

A Framework for Thinking About Worker Output

Worker i's output in setting j:

$$\boxed{y_i = T_j \times h_{ij} \times e_i}$$

T_j	Firm technology: IT, management practices, managers, peers
h_{ij}	Human capital or ability; may be firm-specific (the j subscript)
e_i	Effort: worker's endogenous choice

Firm's objective:

$$\max_{w(\cdot)} f(\textstyle\sum_i y_i) - \sum_i w(y_i)$$

where $w(\cdot)$ is the compensation policy

Three channels for policy:

1. Incentive compatibility (effort)

How does $w(\cdot)$ affect e_i? Standard moral hazard / agency problem.

2. Individual rationality (sorting)

Who is attracted to or retained by a given $w(\cdot)$? Affects h_{ij} through selection.

3. On-the-job human capital

Does $w(\cdot)$ encourage learning? Least studied channel; likely important.

Most theory has focused on channels 1 and 2. Channel 2 is often as important as 1.

External Validity: The Limits of Single-Firm Studies

Why single-firm studies dominate personnel economics:

- Output is measurable at the individual level only in specific settings
- Firms share proprietary data only when they perceive value
- Field experiments require firm cooperation

Key threats to external validity:

- Study firms may be unusual (innovative, large, cooperative)
- Measurable output $\Rightarrow$ less complex tasks; complex tasks may face different incentive problems
- Worker composition in participating firms may be atypical
- Scaling a treatment to a full labor market changes equilibrium wages, sorting, and competition

What the field is doing about it:

- Triangulating across many firms and industries
- Studies that follow workers *across* firms (portability of human capital)
- Large-scale platform data (gig economy, online labor markets)
- Theory to extrapolate from the studied setting to others

A recurring theme

Throughout this module, ask: does this finding generalize? What are the boundary conditions?

The Principal–Agent Problem

The setup:

- Worker (agent) takes effort e; effort unobservable (moral hazard)
- Output: $y = e + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ is exogenous noise
- Firm (principal) observes y, sets $w(y) = a + by$

Worker maximizes:

$$E[w(y)] - \frac{c}{2}e^2 - \tfrac{1}{2}rb^2\sigma^2$$

where r is the coefficient of risk aversion. Subject to $e = \arg\max$ and participation constraint.

Optimal incentive slope:

$$b^* = \frac{1}{1 + rc\sigma^2}$$

The risk–incentive tradeoff:

- High $b \rightarrow$ strong incentives; worker bears risk σ^2
- Low $b \rightarrow$ weak incentives (fixed wage); firm bears all risk
- b^* falls with: risk aversion r, output noise σ^2

Two broad contract types:

	Incentives	Worker risk
Fixed wage	Weak	Low
Piece rate	Strong	High

Three Effects of Performance Pay

1. Effort effect

Given employment, workers exert *more effort* under output-based pay.

Agent's payoff is linked to output $\Rightarrow$ higher e is worth more.

Acts on e_i in $y_i = T_j h_{ij} e_i$

2. Sorting effect

More productive workers *prefer* output-based pay; less productive workers prefer fixed wages.

Performance pay attracts high-h workers and repels low-h workers.

Acts on h_{ij} through selection

3. Learning effect

High-powered incentives may cause workers to invest more in human capital on the job.

Career and tournament incentives create "rat race" dynamics.

Least studied; likely important in early-career jobs

These effects compound over time:

- Effort is immediate; sorting takes time to adjust
- A firm switching to performance pay improves *both* the workforce composition and the effort of stayers
- Empirically separating the two requires clever design

Safelite Auto Glass: The Canonical Study

Lazear (2000, AER)

Setting:

- Nationwide chain of windshield installers (≈3,000 workers)
- Policy change: switch from *hourly wage* to *piece rate* (per windshield successfully installed)
- Administrative data; pre/post variation staggered across locations.
- Workers paid $\max(\text{piece rate earnings}, \text{base}) \Rightarrow$ no one worse off

Estimating equation:

$$\log(\text{output})_{it} = \alpha + \beta\,\text{PR}_t + \gamma_i + X_{it}\delta + \varepsilon_{it}$$

where γ_i is a worker fixed effect and PR_t is the piece-rate indicator.

Key results:

- **44% total productivity increase**
- Effort effect: ∼22% (same workers install faster)
- Sorting effect: ∼22% (new hires more productive; leavers less productive)
- Workers who exit under piece rate were systematically below average — confirming adverse selection into fixed wages

Wage distribution:

- Average pay rose: workers capture about half the gains
- *Variance* of pay rose
- Profitable for firm: productivity gain > wage cost

Performance Pay in Non-Routine Jobs: Motivation

Brown & Andrabi (2023)

Why look beyond manual tasks?

- Safelite and early studies: piece-rate workers doing routine, countable tasks
- Most workers perform *non-routine* tasks: teaching, medicine, law, management
- Multi-task environments: incentives may crowd out effort on hard-to-measure dimensions
- Intrinsic motivation may be crowded out by financial rewards

Two open questions:

1. Do non-routine workers respond to incentives?
2. Do worker *preferences* for pay structures matter?

Why teachers?

- Output (value added) measurable and comparable across settings
- Non-routine task drawing on intrinsic motivation
- No variation in complementary inputs across schools in a district $\Rightarrow$ relatively clean
- Teacher performance pay increasingly common in policy

Design innovation

Brown & Andrabi design a two-phase experiment that separately identifies the *sorting* and *effort* channels using individual-level contract elicitation.

Brown & Andrabi (2023): Setting and Model

Setting:

- Chain of ≈150 private schools in Pakistan
- Teachers earn ≈$4,000/year; within-school mobility possible
- Pre-experiment SD of teacher value added: **0.15** — remarkably close to New York City (Staiger & Rockoff 2010)

Two contract types:

$$u_i = \begin{cases} w_0 + \varepsilon_{iF} & j = j_F \\ p(\hat{\theta}_i + \hat{\beta}_i) - 0.5p\hat{\beta}_i + \varepsilon_{iP} & j = j_P \end{cases}$$

Parameters:

p	performance pay rate
$\hat{\theta}_i$	teacher's belief about baseline ability
$\hat{\beta}_i$	effort responsiveness to incentives
ε_i	non-wage amenity (job-specific)

Output differences between contracts decompose into:

1. Sorting on ability (θ)
2. Sorting on effort responsiveness (β)
3. Average effort increase under performance pay

$\hat{\beta}_i$ roughly maps to the inverse of the cost-of-effort function; uncertainty ($\hat{}$) reflects teachers' prior beliefs about their own type.

Brown & Andrabi (2023): Experimental Design

Phase 1 — Contract Elicitation:

- Each teacher reports preferred fraction of raise to be performance-linked vs. fixed
- Elicitation is *incentive-compatible*: 1-in-3 chance their choice is implemented
- Teachers preferring performance pay have value-added **0.05 SD higher** $\Rightarrow$ high-θ teachers select performance pay
- This 0.05 SD difference represents $\approx \frac{1}{3}$ of the SD of baseline productivity

Phase 2 — School-Level Randomization:

- Schools randomized to three contract regimes:
  1. Fixed raise: $+5\%$ base salary
  2. Performance pay: pay-for-percentile within school
  3. Choice contract: teacher's preferred contract
- Teacher mobility across schools allows identification of sorting

Identification strategy:

Two phases generate *four comparisons*:

1. Did teachers who want performance pay sort to performance schools? (*sorting*)
2. Did teachers who want performance pay respond differently when they got it? (*effort response heterogeneity*)
3. Did teachers who preferred fixed pay but received performance pay also respond? (*average effort effect*)
4. Did classroom env deteriorate, especially under contract mismatch? (*multitasking/crowding out*)

Performance pay structure:
Teachers in top 10% of school $\rightarrow +10\%$ raise; 61st–90th $\rightarrow +7\%$; 16th–60th $\rightarrow +5\%$; below 15th $\rightarrow$ 0–2%

Brown & Andrabi (2023): Results

Effort effects:

- Teachers who *wanted* and *received* performance pay: $+$**0.09 SD** value added
- Teachers who preferred fixed pay but got performance pay: $+$**0.01 SD** — near zero
- Difference = effort response heterogeneity; $\hat{\beta}_i$ is real and large

Sorting effects:

- Teachers moving from flat to performance-pay schools: **0.064 SD above mean** value added
- Re-sorting accounts for **0.022 SD** difference in baseline value added between school types (averaged over movers and stayers)
- Long-run sorting likely much larger (1-year horizon)

Multitasking and side effects:

- Some deterioration in classroom environment (more yelling, more student stress) *among teachers who preferred fixed pay*
- Allowing mismatched teachers to sort out would have improved aggregate outcomes
- No evidence of widespread test manipulation or gaming

Takeaways:

1. Non-routine workers respond strongly to incentives
2. But responses are highly heterogeneous
3. Both sorting *and* effort matter quantitatively
4. Matching workers to their preferred contracts matters for aggregate output *and* worker welfare

Key point: $\hat{\beta}_i$ is private information. Giving workers contract choice allows us to partially infer it.

Connecting Pay Practices to the External Labor Market

A key question behind the previous papers:

- But firms operate in labor markets. Labor economists would look at idiosyncratic contracts and may conclude firms have wage-setting power
- How does that monopsony interact with the incentive and sorting effects we have been discussing?

The monopsony literature (Naidu & Dube, NBER Reporter 2024):

- Even in thick urban labor markets, the share of workers likely to leave in response to a 10% wage cut is often only 20–30%

Policy context:

- Minimum wages: monopsony theory predicts a wage floor may *raise* employment if firms markdown pay
- 2024 FTC rule banning non-competes: limits firms' ability to reduce workers' outside options, raising ε
- 2023 DOJ Merger Guidelines (Guideline 10): mergers examined for effects on *labor market* competition

The link to personnel economics:
Does the standard monopsony inference survive once we account for worker heterogeneity and compensation design?

The Oligopsony Model: How Firms Set Wages

Setup. Firm chooses employment N facing upward-sloping labor supply $w = w(N)$:

$$\pi(N) = Y(N) - w(N)\,N$$

First-order condition:

$$\underbrace{Y'(N)}_{\text{MRP}} = \underbrace{w(N) + Nw'(N)}_{\text{Marginal wage cost}}$$

Elasticity form. Let $\varepsilon \equiv \frac{\partial N}{\partial w}\frac{w}{N}$:

$$w \;=\; Y'(N) \cdot \underbrace{\frac{\varepsilon}{1+\varepsilon}}_{\text{Markdown}<1}$$

Steady-state link to quit rates: $\varepsilon \approx 2\varepsilon_{\text{Departures}}$ (Manning 2003).

Interpretation:

- $\varepsilon \to \infty$: competitive; markdown $\to 0$
- Small ε: workers face few alternatives; firm pays below MRP
- Empirically estimated from quit elasticities

Concentration link. With homogeneous jobs, the market-level markdown $= HHI/\varepsilon$ — where HHI is the Herfindahl index of employer concentration. This bridges firm-level and merger-analysis approaches (Azar, Marinescu & Steinbaum 2024).

The problem: departure elasticities conflate market competition with compensation design, matching, and sorting rationales, ignoring worker heterogeneity.

Worker Heterogeneity Limits Simple Monopsony Inference

Two extensions from personnel economics:

1. Wages as an incentive device (efficiency wages).
Let $e(w)$ be effort as a function of the wage. Effective labor $E(w) = e(w) \cdot N(w)$. Optimal wage:

$$w = Y'(E)\, e(w)\, \frac{2\varepsilon_{\text{Depart}} + \varepsilon_{\text{Effort}}}{1 + 2\varepsilon_{\text{Depart}}}$$

Markdown shrinks as $\varepsilon_{\text{Effort}} \to 1$.

2. Wages to sort heterogeneous workers.
Let $\bar{\theta}(w)$ be the average type recruited at wage w. Effective labor $E(w) = \bar{\theta}(w) \cdot N(w)$. Then:

$$w = Y'(E)\, \bar{\theta}(w)\, \frac{2\varepsilon_{\text{Depart}} + \varepsilon_{\theta}}{1 + 2\varepsilon_{\text{Depart}}}$$

Markdown shrinks as $\varepsilon_{\theta} \to 1$.

The punchline:

- Departure elasticities are *not* a sufficient statistic for market power or firm payoffs
- A firm that appears monopsonistic based on departure data may be paying efficiently to retain and attract high-type workers
- Marking down pay raises the per-worker surplus extracted on paper — but destroys value through sorting and effort channels

Preview: Sandvik et al. (2021, MSCI)

A commission cut at a sales firm raised retention of *low*-productivity workers (due to relative performance evaluation) but caused the *best* performers to exit. Standard departure elasticity: near zero — looks like a monopsonist. But the type-composition channel reversed the firm's payoff entirely.

Efficiency Wages in Practice

Emanuel & Harrington (2020)

Setting: Fortune 500 retailer (warehouse & call center)

- Workers paid flat hourly wages — no piece rates, no explicit performance bonuses
- **Identification:** the firm uses sticky national pay rates that do not continuously adjust to local labor markets, plus discrete pay recalibrations at specific sites
- This creates variation in how far the firm's wage sits above or below local outside options

Results:

- Higher wages $\rightarrow$ higher productivity (elasticity ≈ 1.1–1.2)
- Lower attrition; recruitment improves
- Mechanism: W enters the worker's FOC directly — a higher wage raises the cost of job loss, increasing e^*

Connection to the oligopsony model:

- The standard markdown formula treats wages as a pure cost
- E&H show the effort elasticity $\varepsilon_{\text{Effort}}$ is real and large: the optimal wage exceeds the "raw" monopsonist price even when departure elasticities are low
- Productivity and turnover gains exceed the wage premium

Changes on Intensive Margin: Setting and Theory Sandvik, Saouma, Seegert & Stanton (2021, MSCI)

Setting:

- Inbound sales call center; one division surprises workers with a commission cut — average total pay fell 7%, commission pay fell 18% from a \$318/week baseline
- Calls randomly allocated ⇒ each agent faces the same revenue opportunity distribution
- Pre-change: P75 agent earns 48% more per call than P25 — enormous within-firm heterogeneity

The commission change:

- Eligible revenue for the most commonly sold bundles was cut sharply; raised on rarely sold big-ticket items
- Framed as a "rebalancing" but was in fact a net cut
- Post-announcement surveys show workers expected ≈13% reductions in pay

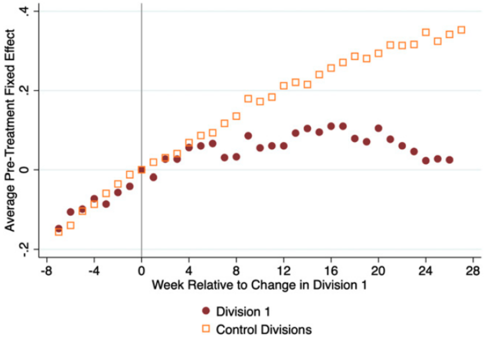
Theory predictions:

- **Effort:** Most theory ⇒ effort falls. Building in income effects ⇒ moves in opposite direction of substitution — net prediction ambiguous
- **Sorting:** top performers have the most to lose; their outside options become relatively more attractive
- **Product substitution:** agents should shift toward rarely sold big-ticket items where commission rates rose

Why this matters for monopsony

Non-neutral departure rates change workforce composition, so retention elasticities are not a sufficient statistic for the firm's market power.

Result 1: The Best Workers Leave



Sandvik et al. (2021, MSCI)

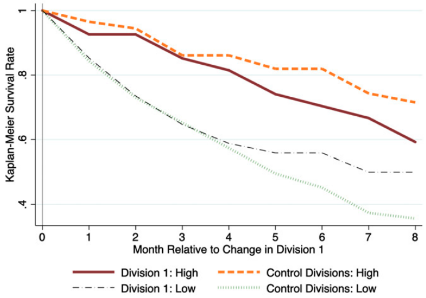
What the figure shows:

- Survival rates in treated vs. control divisions, by pre-treatment sales fixed effects
- The highest-ability workers in Division 1 exit, but with a lag
- No differential attrition among median workers

Financial consequences: the cut was not profitable

- Revenue per call fell \$0.58 in week 8
- By week 18, the revenue loss from composition *swamped* the compensation savings
- Effort changes were near zero for stayers. All the damage came through sorting.

Result 2: Standard Monopsony Test Gives the Wrong Answer Sandvik et al. (2021, MSCI)



Survival Rates Relative to Month 0

What the figure shows:

- Departure elasticity estimated from the commission change: *near zero*
- Retention of lower-type workers actually *rose* (relative performance evaluation makes low types look better once top performers exit)

The inference problem:

- Without productivity data, the standard test says: departure elasticity $\approx 0 \Rightarrow$ this firm is a monopsonist
- But the high-type *composition* elasticity reverses the payoff calculation entirely

Motivation: An Aggregate Puzzle

What happened during the Great Recession?

- US real GDP fell ≈7.2% from Q4 2007 to Q3 2009
- Total hours worked fell ≈10.0%
- Aggregate labor productivity *rose* ≈3.2%
- This gain exceeded productivity growth in both the period immediately before *and* after the recession

Two candidate explanations:

1. **Effort channel**: workers faced higher job-loss risk and worked harder to keep their jobs
2. **Composition channel**: lower-productivity workers were laid off, raising the average quality of the retained workforce

Lazear, Shaw & Stanton (2016, JoLE)

Why this question matters:

- It speaks to whether fixed-wage workers respond to incentives — key to efficiency wage-like mechanisms
- It has implications for aggregate productivity measurement over the business cycle
- It connects the macro literature on "labor hoarding" to micro evidence on personnel choices

The paper's strategy

Use individual-level daily productivity data from a single firm, with workers in many US states, to separately identify effort changes from workforce composition changes.

The Model: Worker's Choice of Effort

LSS (2016) — see Appendix for full derivation

Setup:

- Workers indexed by ability k (higher = more able)
- Cost of effort e for type k: $c(e)/k$
- Firm announces wage W and minimum performance standard $x \sim G(x)$
- If output falls below x, worker is dismissed
- Dismissed worker finds a new job with probability $(1-u)$, yielding rent R

Worker's surplus:

$$\max_e \; G(e)\left[W - \frac{c(e)}{k}\right] + [1 - G(e)](1-u)R$$

where $G(e)$ is the probability that the standard is met at effort e.

Key predictions:

1. $\partial e/\partial k > 0$: more able workers exert more effort at any given wage and unemployment rate.
2. $\partial e/\partial u > 0$: effort *rises* as unemployment rises because job loss becomes more costly — the outside option $(1-u)R$ worsens.
3. Lower-ability workers (for whom the standard is binding) respond most to higher unemployment — they have the most to gain from avoiding dismissal.

Full first-order condition and comparative statics in Appendix A. Composition predictions (who survives the recession) in Appendix B.

Key point: The model predicts that effort rises when unemployment rises, even with *fixed wages* — because the cost of job loss increases. This is a sharply testable prediction.

Data

Source: One large technology-based services (TBS) firm

What is a TBS job?

- Insurance claims processing, call centers, retail cashiers, airline gate agents, IT specialists
- Jobs that are labor-intensive but supported by a structured IT system that records each transaction

Sample:

- 23,580 unique workers across multiple US states
- ≈5.6 million worker-day observations
- Period: June 2006 – May 2010 (pre-recession / recession / post-recession)
- Workers all perform *the same task* across establishments in different states

Key features enabling identification:

- Computer tracks time to process each transaction ⇒ individual daily productivity measured precisely
- Workers paid *hourly*, not by piece rate ⇒ effort changes reflect pure incentive response to job-loss risk, not changed pay structure
- Same task across geographically dispersed establishments ⇒ local labor market variation (state unemployment rate) is exogenous to individual effort choices
- Firm did not change task assignment or performance standards during the recession period

Workers are sent home when there are too many to accommodate demand ⇒ downtime is not in the measured productivity denominator

Productivity Measure and Summary Statistics

Output per hour (OPH):

$$\text{OPH}_{it} = \frac{\text{transactions completed by worker } i \text{ on day } t}{\text{hours of active processing time}}$$

- Active processing time: clock time spent actually processing transactions; *excludes* slack time
- This means OPH measures the *speed* of processing, not how much demand there is — abstracting from demand shocks
- Workers increase OPH by processing transactions faster

Summary statistics:

	Pre-rec.	Rec.+Post
Log OPH	2.21	2.33
	(0.29)	(0.28)
Uptime	0.96	0.96
Tenure (days)	568	695
Unemp. rate	4.55%	8.05%
N (obs.)	1,883,328	3,744,345

Means and (SD). Peak unemployment: 9.71% in Q2 2009.

Estimating Equations

Equation 1: Recession effect (effort vs. composition)

$$\log(q_{ijt}) = \gamma_1 R_t + X_{it}\beta + \zeta_j + \varepsilon_{ijt}$$

- R_t: recession dummy (Dec 2007 – May 2009)
- X_{it}: 5th-order polynomial in tenure, 11 month FEs
- ζ_j: establishment fixed effect

Progressively add:

- α_i: worker fixed effect — if $\hat{\gamma}_1$ is unchanged, composition is unlikely to drive results
- ϕ_m: manager (boss) fixed effect
- Balanced panel: only workers present throughout

Equation 2: Local unemployment rate

$$\log(q_{ijt}) = \gamma_2 U_{jt} + X_{it}\beta + \zeta_j + \alpha_i + \tau_t + \varepsilon_{ijt}$$

- U_{jt}: state monthly unemployment rate matched to establishment j
- τ_t: year $\times$ month fixed effect — removes common time trends; identified only from cross-establishment variation in state U within a given month
- Standard errors clustered at the state level or with wild cluster bootstrap

Results: The Recession Effect (Table 2)

	(1) Est FE	(2) +Wkr FE	(3) +Boss FE	(4) Bal.
Recession	0.054	0.054	0.053	0.048
	(0.007)	(0.006)	(0.006)	(0.004)
N		5,627,673		1,104,889
R^2	0.071	0.251	0.256	0.05

Dep. var.: $\log(\text{OPH})$. All models include 5th-order polynomial in tenure, establishment FEs, and 11 month FEs. SE clustered at state level.

Decomposing the composition effect:

- Leavers during recession: no quality differential
- Hires during recession: $+$**1.5%** more productive (better available workers)
- But new hires are only 30% of the workforce $\Rightarrow$ aggregate composition effect $\approx 0.5\%$
- Composition explains at most $0.5/5.4 \approx$ **9%** of the total gain

Bottom line:
The large productivity increase during the recession comes from workers *processing transactions faster*, not from the workforce becoming more skilled on average.

Results: Local Unemployment Rate (Table 3)

	(1) Est FE	(2) +Wkr FE	(3) +Boss FE	(4) Bal.
Monthly U (state)	0.008 (0.003)	0.008 (0.003)	0.007 (0.003)	0.009 (0.003)
5 pp ΔU	3.9%	4.1%	3.7%	4.4%
N		5,627,673		1,104,889

Dep. var.: $\log(\text{OPH})$. Year×month FEs remove aggregate time trends. SE clustered at state level.

Magnitude:
The national unemployment rate rose ≈5 pp from just before the recession to its peak ⇒ predicted productivity gain of ≈4%, in the ballpark of the 5.4% total estimate

What this delivers:

- At the *same establishment*, at the *same point in time*, workers in high-unemployment states work measurably harder than workers in low-unemployment states
- Year×month FEs remove any common time trend or aggregate firm-level shock
- Establishment FEs absorb persistent local demand differences

Heterogeneity (consistent with the model):

- Effort response is *largest for low-productivity workers* — those for whom the minimum standard is most binding and job loss is most costly

Valuing Amenities: Experimental Design

Mas & Pallais (2017, AER)

Core question:

- How much do workers value non-wage amenities?
- Standard market wages don't reveal preferences if workers lack bargaining power or amenity supply is thin
- Solution: run a discrete choice experiment *inside* a real hiring process

Setting:

- National call center advertises data entry positions
- ≈7,000 job applicants; real job offers
- Applicants randomly offered one of several work arrangements with a randomly assigned wage
- Wage variation identifies willingness to accept (WTA) to switch arrangements

Work arrangements tested:

1. Standard 9–5, in-office (*baseline*)
2. Work from home (fixed hours)
3. Flexible hours (choose own schedule)
4. **Employer-controlled scheduling** (hours set by firm at short notice)

Key design features:

- Real job offers ⇒ no hypothetical bias
- Wage randomization: cross-sectional and within-applicant
- Actual take-up decisions reveal true preferences
- Large sample across demographics

Valuing Amenities: Main WTP Estimates

Mas & Pallais (2017, AER)

Average willingness to pay:

Arrangement	WTP (as % of wages)
Work from Home	+8%
Flexible hours	+3%–5%
Employer-controlled scheduling	−20%

- Workers *value* WFH and flexibility, but not by enormous amounts on average
- Workers *strongly dislike* employer scheduling discretion — would require a **20% wage premium** to accept it
- Firms effectively get a large subsidy if they can impose scheduling discretion without compensating workers

Why employer scheduling discretion matters:

- Very common in retail, food service, call centers: irregular or on-call scheduling imposed on hourly workers
- Workers bear substantial cost; firms pay little or nothing extra in wages
- Labor market fails to price this amenity correctly because workers have limited outside options
- Policy implication: scheduling rules (advance notice laws) may improve efficiency

Key takeaway

Non-wage amenities are valued, but their market prices are distorted — especially when they favor the employer.

Valuing Amenities: Heterogeneous WTP

Mas & Pallais (2017, AER)

WTP for WFH varies dramatically by worker type:

Worker group	WTP for WFH
All applicants (average)	8% of wages
Women with young children	∼25% of wages
Women without young children	∼8% of wages
Men (all)	∼3% of wages

- The value of WFH is *not* uniform: it is concentrated in workers with young children
- Mothers value WFH as much as a 25% wage increase — nearly an order of magnitude above the male average

Why heterogeneity matters for labor markets:

- If the average WTP is modest but variance is high, *total surplus from offering WFH* can still be large
- Firms that offer WFH attract a *different* pool of workers (sorting on amenity preferences)
- Standard wages don't reveal who values which amenities: market wages reflect the *marginal* worker, not the high-WTP worker
- Wage compression around amenities can create large efficiency losses

Key point: Most workers don't value flexible *scheduling* much on average, but those who do value it *enormously*. Employer scheduling discretion is a large, uncompensated cost imposed on workers

Why Doesn't Knowledge Flow Freely Inside Firms? Sandvik, Saouma, Seegert & Stanton (2020, QJE)

The puzzle:

- Best workers substantially outperform peers: P75 sells 48% more per call than P25
- Manager and agent surveys attribute this to *knowledge*: how to bundle products, handle objections, use the computer system
- If knowledge diffused, it would raise the floor of the productivity distribution

Two candidate frictions:

1. **Initiation costs (knowledge seekers)**: social discomfort, reluctance to appear incompetent, coordination costs
2. **Contracting costs (knowledge providers)**: commission depends on relative performance ⇒ sharing information helps rivals

Why this matters beyond one firm:

- Marshall (1890) argued agglomeration benefits operate through knowledge spillovers
- This is the same mechanism at the *micro* level, inside a single employer
- Peer effects literature documents that *who* you work with matters — but whether this is effort externalities or knowledge transfer is rarely identified

Research question

Can management practices overcome the frictions that limit knowledge flows? And which friction binds more?

Experimental Setting and Design

Sandvik et al. (2020, QJE)

Setting:

- Inbound sales call center; July–August 2017
- $>$730 salespeople in three geographically separate offices
- Calls allocated by *random assignment* of sales opportunities to agents
- Performance metric: revenue per call (RPC)
- Pay: hourly wage + commission based on quintile of relative RPC within division

Treatment cells (N = 653 in two main offices):

1. *Internal Control*: pairs displayed, no intervention
2. *Structured-Meetings*: worksheets, guided self-reflection, catered lunch; targets *initiation costs*
3. *Pair-Incentives*: joint RPC bonus for pairs; targets *contracting costs*
4. *Combined*: both elements simultaneously

Common elements for all treatment agents:

- Paired with a randomly chosen partner from same treatment, division, and office
- Joint performance scores published daily on TV monitors and internal messaging
- Individual sales data shared with researchers

External Control:

- Third office, 600 miles away ($N = 83$)
- Unaware of the experiment
- Provides estimate of counterfactual RPC trends

Data: 4 weeks pre-intervention, 4-week intervention, 20 weeks post-intervention; pre-registration AEARCTR-0002332

Results: Persistence Through 34 Weeks

Sandvik et al. (2020, QJE, Online Appendix)

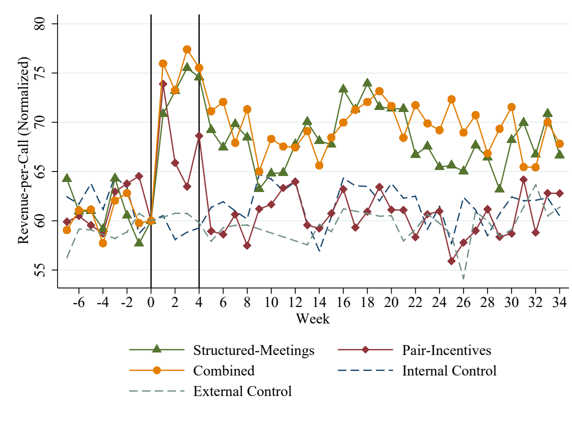


Figure A.2. Extended 34-week window. Structured-Meetings and Combined treatment effects remain stable through week 34, consistent with permanent knowledge acquisition rather than temporary effort.

Results

Sandvik et al. (2020, QJE)

Interpretation:

- Effects are robust to using Internal Control or External Control as the comparison group ⇒ not Hawthorne effects
- Structured-Meetings induced durable *knowledge acquisition* — not just short-run effort
- Pair-Incentives induced short-run effort change that *dissipated* once incentives ended
- Productivity dispersion *fell* under Structured-Meetings: lower-tail agents improved most
- Structured-Meetings ($\approx$ Combined) $\gg$ Pair-Incentives ⇒ **social initiation costs** (knowledge seekers) are the binding constraint, not financial incentives
- Individual weekly earnings rose \$35–\$43; firm revenue rose \$580–\$720 per agent-week

Mechanism check — worksheets:

- >80% of Structured-Meetings worksheets contain examples of *contextual knowledge* on improving sales: product bundles, handling credit-check difficulties, selective discounting strategies
- Worksheets with pure motivational statements ("stay positive") predict *smaller* gains
- Agents with knowledge-content worksheets have the largest persistent sales gains

Broader lessons:

1. Management practices can raise productivity without changing the pay structure
2. Spillovers operate even for tasks with no *direct* production interdependencies
3. Firms need to pay attention to skills ⇒ hiring and training will matter.

Discretion in Hiring: Setting and Design

Hoffman, Kahn & Li (2018, QJE)

Research question:

- When firms use algorithmic hiring tests, do managers who override the algorithm exercise valuable private information — or costly bias?
- Answer has direct implications for how much *discretion* to give hiring managers

Setting:

- 15 client firms sharing the same job-testing provider
- Low-skill service sector (call centers, retail, etc.)
- Applicants receive algorithmic score: **Green** / **Yellow** / **Red**
- ≈266,000 hires; ≈400,000 post-testing applicants; 445 managers

Measuring managerial discretion:

- **Exception** = hiring a Yellow when a Green was available, or hiring a Red when better-scored candidates existed
- Exception rate varies dramatically across managers: mean ≈ 22%, large variance
- High-exception-rate managers: do their "overrides" reflect private information (good) or systematic biases (bad)?

Identification

Within-firm, within-time variation in exception rates across managers; outcome = worker tenure (long tenure = good hire). Testing improves average tenure by ≈25% (0.23 log points).

Discretion in Hiring: Bias Dominates Private Information Hoffman, Kahn & Li (2018, QJE)

Main finding:

- Managers with *higher* exception rates hire workers with *worse* subsequent tenure
- This pattern is inconsistent with private information: if exceptions reflected genuine insight, high-exception managers should do *better*, not worse
- The pattern is consistent with **taste-based bias** or **systematic hiring mistakes**

Model: Manager utility over hire i:

$$U_i = (1-k)\underbrace{a_i}_{\text{applicant quality}} + k\underbrace{b_i}_{\text{idiosyncratic preference}}$$

where $k =$ bias parameter. Exceptions rise with k.

Additional evidence:

- Exception rates are stable over time within manager $\Rightarrow$ a persistent managerial trait, not random noise
- Firms that *restrict* exceptions (push toward algorithm-recommended hires) outperform those that allow more discretion
- The penalty for discretion is largest when good Green applicants are abundant

Key point: "Fit" with a job is predictable from testing. Formalized processes can improve outcomes.

Supervisors and Performance Management Systems: Overview Frederiksen, Kahn & Lange (2020, JPE)

Core question:

- Supervisors vary widely in how they rate workers of similar underlying quality. What drives this?
- Is variation in ratings due to differences in supervisors' *managerial ability* (eliciting more output), or to differences in *leniency* (giving generous ratings regardless of output)?

Why it matters:

- If leniency: ratings are noisy signals; promotions and pay are misallocated; firms cannot use ratings to rank workers
- If ability: high-rating supervisors genuinely improve output; matching workers to better supervisors creates value

Setting:

- Large Scandinavian service-sector firm with formal annual performance reviews
- Panel data linking workers, supervisors, ratings, and (some) objective performance measures over many years
- Supervisors rated by the same firm-wide scale; workers can move across supervisors over time

Key punchline

Supervisor heterogeneity reflects, at least in part, *real differences in managerial ability*. Firms appear to know this: high-rating supervisors are paid more.

Theoretical Framework: Ability vs. Leniency

Frederiksen, Kahn & Lange (2020, JPE)

Three actors: workers, supervisors, firm. Neither firms nor supervisors directly observe effort.

Worker output (eq. 2): $q_i = e_i + \alpha_i + \varepsilon_i^q$.
Output is effort e_i plus worker type α_i plus noise. No multiplicative ability term.

Effort cost (eq. 3): $c(e) = -\frac{1}{2\mu_s} e^2$.
μ_s is *managerial ability*: higher $\mu_s \Rightarrow$ lower marginal cost of effort $\Rightarrow$ workers choose to exert *more* effort.

Supervisor's optimal rating (eq. 5): $r = q + \beta_s$.
$\beta_s \equiv \tilde{\beta}_s / \tilde{\gamma}_s$ is *leniency bias*: supervisor trades off accurate reporting against taste for favorable ratings.

What is estimable (eq. 6):

$$r_{it} = \alpha_i + \underbrace{(e_s + \beta_s)}_{\phi_s} + \varepsilon_{it}^q$$

Identification (Table 4 in paper):

	Leniency ($\sigma_\beta^2 > 0$)	Ability ($\sigma_\mu^2 > 0$)
Informed firm:		
Worker wages	0	≈ 0
Piece rate strength	0	> 0
Productivity	0	> 0
Supervisor wages	0	> 0
Uninformed firm:		
Worker wages	> 0	> 0
Piece rate strength	0	0
Productivity	0	> 0
Supervisor wages	0	> 0

Identification Strategy

Frederiksen, Kahn & Lange (2020, JPE)

Step 1 — Estimate ϕ_s (eq. 1):

$$p_{it} = \alpha_i + \phi_{s(i,t)} + \beta' X_{it} + \gamma' Y_{s(i,t)t} + \varepsilon^p_{it}$$

Double fixed effects (worker + supervisor), identified by workers moving across supervisors. Bias-corrected via Kline, Saggio & Sølvsten (2018).

Step 2 — Regress outcomes on $\hat{\phi}_s$ (eq. 7):

$$\log(w_{it}) = \beta_0 + \beta_1 \phi_s + \beta_2 \alpha_i + \beta_3 \varepsilon^p_{it} + \cdots$$

Split-sample IV (even vs. odd years) instruments $\hat{\phi}_s$ to correct for measurement error.

Key logic:

- *Pure leniency* (β_s only) never raises productivity or supervisor wages — regardless of whether the firm is informed
- *Managerial ability* (μ_s) raises productivity and supervisor wages in all cases

Identification insight

Productivity *and* supervisor pay both rise with ϕ_s in the data. This is consistent only with *managerial ability*. Worker wages also rise, suggesting the firm is at most *partially informed* — consistent with workers earning rents from high-rater matches.

Main Findings

Frederiksen, Kahn & Lange (2020, JPE)

Supervisor heterogeneity is large and real:

- Supervisors vary substantially in how they rate observationally similar workers
- High-rating supervisors manage teams with *better objective performance* — ability, not just leniency, is driving the dispersion
- High-rating supervisors are themselves *paid more* by the firm, consistent with the firm recognizing the ability differences

Worker career effects are large:

- Workers matched to a 90th- vs. 10th-percentile supervisor for *just one year* gain a present discounted value of earnings equivalent to **6–14% of annual pay**
- Subordinates of high-raters receive higher pay and stronger pay–performance alignment

Firm information:

- Evidence is consistent with the firm being at least *partially informed* about supervisor quality — it does not treat all supervisors symmetrically
- This has important design implications: if the firm knows who the good supervisors are, it can deploy them strategically

Implications for the production function:

$$y_i = T_j \times h_{ij} \times e_i$$

- Supervisor ability enters through T_j (who manages the team) *and* h_{ij} (coaching, task assignment)
- Leniency distorts incentives

Talent Hoarding and Internal Labor Market Frictions

Talent hoarding: Haegele (2022)

- Middle managers *strategically block* the promotion of high-performing subordinates to protect their own team productivity
- Internal labor markets are not fully efficient even inside a single firm
- Creates frictions in h_{ij}: the right worker is not always in the right role

Why this matters:

- The efficiency wage result (Emanuel & Harrington) shows that pay above market is a productivity investment
- But even if pay design is right, *internal* allocation frictions can offset those gains
- Talent goes unrewarded and under-deployed not because the market fails to price it, but because managers have incentives to hoard it

Implications for practice:

- Promote-from-within vs. external hiring: the right choice depends on whether internal hoarding distortions exceed the information advantages of internal promotion
- Monitoring and transparency about promotion criteria reduce hoarding incentives
- Firms that combine strong external pay with poor internal mobility will still lose top performers to competitors

Connection to pay design

Monopsony-based analyses of wage gaps often focus on *external* market power. But internal allocation frictions (hoarding, compression, rating bias) can generate similar patterns and require different remedies.

Synthesis: The Production Function Revisited

What shapes worker output?

$$y_i = T_j \times h_{ij} \times e_i$$

T_j	Management practices; supervisor quality; knowledge systems; appraisal design (Frederiksen et al. 2020; Sandvik et al. 2020)
h_{ij}	Hiring and selection; training; internal mobility frictions (Haegele 2022); portability of human capital
e_i	Incentive design (Lazear 2000; Brown & Andrabi 2023); job-loss risk (LSS 2016); efficiency wages (Emanuel & Harrington 2020)

Across the papers today:

- Effort and sorting are both quantitatively large — and often comparable in magnitude (Safelite: 50/50)
- Workers respond to implicit incentives (job-loss risk) not just explicit piece rates (LSS)
- Knowledge barriers are real; management can lower them at low cost (Sandvik QJE)
- Supervisor behavior and rating systems distort incentives from the top down (Frederiksen et al.)
- Heterogeneity in worker responses is large and should guide contract design (Brown & Andrabi)

Key point: Personnel economics has moved well beyond "does piece rate raise output?" The frontier is understanding *who* responds, to *what*, and through *which* channels.

External Validity in Personnel Economics

Hoffman & Stanton (2025)

The core challenge from insider studies of individual firms: Are results specific to that trucking firm, fruitpicker, or courier, or do they generalize?

Two complementary responses:

- **Accumulate evidence across settings.** Several results—e.g., referred workers are more likely to be hired and less likely to quit—replicate across industries, building confidence in generality.
- **Embrace firm-level heterogeneity.** Prominent firms (e.g., GE) warrant study precisely because they are influential. Growing evidence on firm effects on wages and productivity (Abowd et al., 1999; Card et al., 2013) means understanding leading-firm practices is valuable even if effects don't generalize to laggards immediately.

Scaling and general equilibrium:

- RCT estimates at one firm may miss complementary practice changes when a policy scales (e.g., a firm raises effort standards after adopting training).
- Single-firm estimates of performance pay answer a different question than "what happens if all firms adopt performance pay"—researchers should be explicit about which question they address.

Improving External Validity: A Checklist

Hoffman & Stanton (2025)

Researchers can reduce external validity concerns through better research design and reporting:

1. **Choose workers & firms well-suited to your question.** Long-distance truckers are a natural fit for studying monitoring; they would be a poor fit for studying discretionary bonuses.
2. **Characterize the firm.** Is it a low-cost or product-differentiation firm? Typical or exceptional in its industry? Report demographics comparing sample workers to the broader industry.
3. **Disaggregate by worker type.** One firm often employs engineers, store workers, and back-office staff. Results may vary across these groups in informative ways.
4. **Explain why the firm is working with you.** This addresses both selection and multiple-testing concerns; report what the firm hoped to learn and how it responded to results.
5. **Vary conditions.** Where possible, examine results by geography, business cycle, or time period; check whether effects are complementary or contingent on other organizational changes.
6. **Return to theory.** Theory predicts how effects vary with firm type or worker characteristics, allowing external-validity claims to go beyond the data at hand.

Appendix A: LSS (2016) — Full Model Setup

Worker's problem:

Worker of type k chooses effort e to maximize expected surplus

$$\Omega(e, k, u) = G(e)\left[W - \frac{c(e)}{k}\right] + [1 - G(e)](1 - u)R$$

where:

- $G(e) = \Pr(\text{standard met} \mid e)$; $G' = g > 0$
- W: firm wage (constant; no piece rate)
- $c(e)/k$: cost of effort; $c' > 0$, $c'' > 0$
- $(1 - u)R$: expected value of outside option when unemployed with probability u

First-order condition:

$$g(e^*)\left[W - \frac{c(e^*)}{k} - (1 - u)R\right] = G(e^*)\,\frac{c'(e^*)}{k}$$

Comparative statics:

$$\frac{\partial e^*}{\partial u} = \frac{g(e^*)R}{-\partial^2\Omega/\partial e^2} > 0$$

$$\frac{\partial e^*}{\partial k} = \frac{c(e^*)/k^2 \cdot g(e^*) + G(e^*)c'(e^*)/k^2}{-\partial^2\Omega/\partial e^2} > 0$$

Both comparative statics hold as long as the worker's problem has an interior maximum (second-order condition satisfied) and $W - c(e^*)/k > (1 - u)R$ at the optimum.

Appendix B: LSS (2016) — Composition Predictions

Separations during recessions:
A worker is retained if the firm's expected revenue from keeping the worker exceeds the wage cost. During a recession:

- Quit rate falls (outside options worse): $\downarrow$
- Firm may also change dismissal threshold if productivity target adjusts

New hires:
A larger pool of workers is available at the same wage during a recession $\Rightarrow$ firm can hire from a higher-quality pool.

Empirical implementation:

$$\begin{aligned}\log(q_{ijt}) &= 0.053\, R_t \\ &\quad - 0.0002\,\text{Left during rec.} \\ &\quad + 0.015\,\text{Hired during rec.}\end{aligned}$$

What the data show:

- No quality differential for those who *leave* during the recession
- Hires during the recession are $+1.5\%$ more productive
- But new hires are only 30% of the workforce $\Rightarrow$ aggregate effect ≈ 0.5 pp
- Total composition effect $\leq 0.5/5.4 \approx 9\%$
- Cumulative effect on aggregate productivity due to attrition: ≈ 0.01 (less than one-fifth of total gain)

Balanced panel check:
Using only the 1,623 workers present throughout all three phases, the recession coefficient is 4.8% vs. 5.4% for the full sample. This small difference is further evidence that composition is not the main story.

Appendix C: Brown & Andrabi (2023) — Model Detail

Teacher utility under each contract:

Fixed contract $(j = j_F)$: $u_i = w_0 + \varepsilon_{iF}$
Perf. contract $(j = j_P)$:

$$u_i = p(\hat{\theta}_i + \hat{\beta}_i) - \frac{1}{2}p\hat{\beta}_i + \varepsilon_{iP}$$

Parameters:

- p: performance pay rate
- $\hat{\theta}_i$: teacher's belief about own baseline ability
- $\hat{\beta}_i$: teacher's belief about own effort responsiveness
- $\varepsilon_{iF}, \varepsilon_{iP}$: non-wage amenity shocks

Note: $\hat{\beta}_i$ roughly maps to the inverse of $c_i''(e)$ evaluated near the normal effort level — a teacher with low cost convexity will respond a lot to performance pay.

Output decomposition:
The difference in mean output between performance-pay and fixed-pay schools decomposes into:

1. **Ability sorting:** Teachers with high $\hat{\theta}_i$ prefer performance pay $\Rightarrow$ mean θ rises in perf-pay schools
2. **Effort sorting:** Teachers with high $\hat{\beta}_i$ prefer performance pay $\Rightarrow$ mean responsiveness rises in perf-pay schools
3. **Aggregate effort effect:** Even teachers with low $\hat{\beta}_i$ exert somewhat more effort under performance pay

Phase 1 evidence: Teachers who choose performance pay have value-added 0.05 SD higher ex ante — direct evidence of ability sorting.