# Signal Extraction in Centralized Assignment
## Evidence from Teacher Recruiting

Matteo Bobba (TSE)
Tim Ederer (CMU)
Gianmarco León-Ciliotta (UPF & BSE & IPEG)
Christopher Neilson (Yale)
Marco Nieddu (University of Cagliari)

LACEA-LAMES 2026, PUCP-Lima

November 2026

# Motivation

- Many matching markets are run centrally: school choice, medical residencies, public-sector hiring
  - ⇒ Centralization coordinates choices, but information about match quality is imperfect (Artemov, 2021; Fernandez et al., 2022; Friedrich et al., 2024)
  - ⇒ Often it can only be learned through costly local evaluations: interviews, demonstrations (Lee and Schwarz, 2017)
- Design problem: how to allocate scarce screening capacity
  - ⇒ Direct it toward promising matches...
  - ⇒ ...while letting local evaluations uncover matches that central screening would miss
- Hard when the returns to local discretion are difficult to predict ex ante (Dessein, 2002; Frankel, 2021)

# Motivation

- Many matching markets are run centrally: school choice, medical residencies, public-sector hiring
  - ⇒ Centralization coordinates choices, but information about match quality is imperfect (Artemov, 2021; Fernandez et al., 2022; Friedrich et al., 2024)
  - ⇒ Often it can only be learned through costly local evaluations: interviews, demonstrations (Lee and Schwarz, 2017)
- Design problem: how to allocate scarce screening capacity
  - ⇒ Direct it toward promising matches...
  - ⇒ ...while letting local evaluations uncover matches that central screening would miss
- Hard when the returns to local discretion are difficult to predict ex ante (Dessein, 2002; Frankel, 2021)

# Motivation

- Many matching markets are run centrally: school choice, medical residencies, public-sector hiring
  - ⇒ Centralization coordinates choices, but information about match quality is imperfect (Artemov, 2021; Fernandez et al., 2022; Friedrich et al., 2024)
  - ⇒ Often it can only be learned through costly local evaluations: interviews, demonstrations (Lee and Schwarz, 2017)
- Design problem: how to allocate scarce screening capacity
  - ⇒ Direct it toward promising matches...
  - ⇒ ...while letting local evaluations uncover matches that central screening would miss
- Hard when the returns to local discretion are difficult to predict ex ante (Dessein, 2002; Frankel, 2021)

# Motivation

- Two features of the local evaluator (Hoffman et al., 2018; Frankel, 2021)
  - ⇒ What they learn: how good are they at reading candidates? (signal extraction)
  - ⇒ What they want: do they care about productivity, or about other things? (objectives)
- Relevant for policy
  - ⇒ Poor readers ⇒ give them better information
  - ⇒ Misaligned objectives ⇒ change whom they choose from, or delegate less
- Separating them requires observing what evaluators know, learn, and choose, benchmarked against an objective measure of productivity — such data rarely exist

# Motivation

- Two features of the local evaluator (Hoffman et al., 2018; Frankel, 2021)
  - ⇒ What they learn: how good are they at reading candidates? (signal extraction)
  - ⇒ What they want: do they care about productivity, or about other things? (objectives)
- Relevant for policy
  - ⇒ Poor readers ⇒ give them better information
  - ⇒ Misaligned objectives ⇒ change whom they choose from, or delegate less
- Separating them requires observing what evaluators know, learn, and choose, benchmarked against an objective measure of productivity — such data rarely exist

# Motivation

- Two features of the local evaluator (Hoffman et al., 2018; Frankel, 2021)
  - ⇒ What they learn: how good are they at reading candidates? (signal extraction)
  - ⇒ What they want: do they care about productivity, or about other things? (objectives)
- Relevant for policy
  - ⇒ Poor readers ⇒ give them better information
  - ⇒ Misaligned objectives ⇒ change whom they choose from, or delegate less
- Separating them requires observing what evaluators know, learn, and choose, benchmarked against an objective measure of productivity — such data rarely exist

# Motivation

- Every employer hires with imperfect information about productivity
  - ⇒ A common design: a standardized screen (test, credentials) followed by a discretionary interview
  - ⇒ Firms routinely leave the final call to line managers (Autor and Scarborough, 2008; Hoffman et al., 2018)
- The stakes are higher in the public sector
  - ⇒ Rigid pay scales, rare dismissals, long careers: hiring is the main personnel margin (Oyer and Schaefer, 2011; Finan et al., 2017)
  - ⇒ No profit motive disciplining local evaluators: room for other objectives (Colonnelli et al., 2020)
  - ⇒ Large centralized systems: national exams + local committees for teachers, doctors, police, civil servants
  - ⇒ The state also cares about where good workers end up, and long tenures make hiring mistakes persistent

# Motivation

- Every employer hires with imperfect information about productivity
  - ⇒ A common design: a standardized screen (test, credentials) followed by a discretionary interview
  - ⇒ Firms routinely leave the final call to line managers (Autor and Scarborough, 2008; Hoffman et al., 2018)
- The stakes are higher in the public sector
  - ⇒ Rigid pay scales, rare dismissals, long careers: hiring is the main personnel margin (Oyer and Schaefer, 2011; Finan et al., 2017)
  - ⇒ No profit motive disciplining local evaluators: room for other objectives (Colonnelli et al., 2020)
  - ⇒ Large centralized systems: national exams + local committees for teachers, doctors, police, civil servants
  - ⇒ The state also cares about where good workers end up, and long tenures make hiring mistakes persistent

# This Paper: Teacher Recruitment

- Teachers are among the most important inputs into learning
  - ⇒ Effects on test scores and on long-run outcomes (Rockoff, 2004; Rivkin et al., 2005; Chetty et al., 2014b)
- Most policies focus on the supply side: pay and assignment rules to attract good teachers to hard-to-staff schools (Biasi et al., 2026; Bobba et al., 2026)
  - ⇒ They only work if schools identify and pick effective teachers among those who apply (Bates et al., 2025)
  - ⇒ Screening design is central to the efficiency and equity of teacher allocation
- Questions
  - ⇒ Does school discretion improve teacher selection, and where?
  - ⇒ Should the central system choose which schools screen, or which applicants get screened?
  - ⇒ Should better productivity measures be used for prescreening or disclosed to schools?

# This Paper: Teacher Recruitment

- Setting: Perú's nationwide, centralized allocation of permanent public-school teachers
  - ⇒ Test-based shortlists + school committees with substantial discretion
  - ⇒ Applicants' ranked lists of schools; full ranking of the same shortlist before and after the interview
  - ⇒ Link candidates to student outcomes ⇒ teacher value-added (TVA)
- Three steps
  1. Measure the value of discretion relative to random and test-based hiring
  2. Model hiring and learning: separate objectives from signal extraction
  3. Design screening: target discretion across schools; prescreen applicants on predicted TVA; disclose TVA; vary shortlist length

# This Paper: Teacher Recruitment

- Setting: Perú's nationwide, centralized allocation of permanent public-school teachers
  - ⇒ Test-based shortlists + school committees with substantial discretion
  - ⇒ Applicants' ranked lists of schools; full ranking of the same shortlist before and after the interview
  - ⇒ Link candidates to student outcomes ⇒ teacher value-added (TVA)
- Three steps
  1. Measure the value of discretion relative to random and test-based hiring
  2. Model hiring and learning: separate objectives from signal extraction
  3. Design screening: target discretion across schools; prescreen applicants on predicted TVA; disclose TVA; vary shortlist length

# Contributions

- Information acquisition and mechanism design in centralized matching
  - ⇒ Lee and Schwarz (2017), Echenique et al. (2022), Artemov (2021), Fernandez et al. (2022), Hakimov et al. (2023), Friedrich et al. (2024)
  - ⇒ How to allocate scarce decentralized screening: prescreening on predicted productivity vs. disclosure, and the cost in unfilled vacancies
- Learning and discretion in labor markets
  - ⇒ Jovanovic (1979), Farber and Gibbons (1996), Altonji and Pierret (2001), Kahn and Lange (2014), Pallais (2014), Carranza et al. (2022), Lessem and Miller (2026); Autor and Scarborough (2008), Hoffman et al. (2018), Estrada (2019), Colonnelli et al. (2020), Weaver (2021)
  - ⇒ Separately identify objectives vs. signal-extraction ability, benchmarked on a direct productivity measure
- Sorting in teacher labor markets
  - ⇒ Biasi et al. (2026), Bobba et al. (2026), Ederer (2023), Bates et al. (2025), Colas and Fu (2025); Jacob and Lefgren (2008), Staiger and Rockoff (2010)
  - ⇒ Demand side: schools' objectives and learning shape what screening policy can achieve

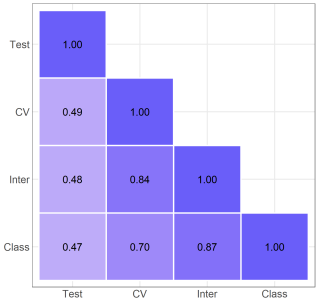
# Context and Descriptives

# Allocation Mechanism for Permanent Teachers in Peru

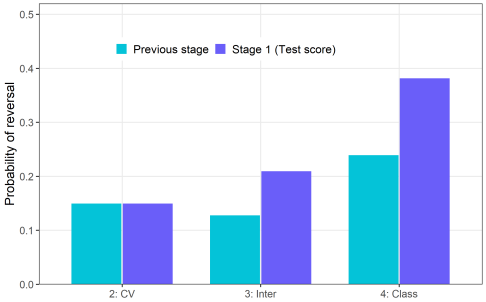
- Centralized stage
  - ⇒ Applicants take a national test (PUN): pedagogical skills and curricular knowledge
  - ⇒ Scoring above 60% in each component ⇒ eligible for permanent positions
  - ⇒ Submit rank-ordered lists (ROLs) of schools within a region
- Shortlists of at most 10 candidates per school (a "recruitment contest")
  - ⇒ Applicants ranked by test score and assigned sequentially to their top three schools
- School-based stage: committee led by the principal
  - ⇒ Three screening devices: CV review → interview → classroom demonstration
  - ⇒ Final ranking = test score + school-assigned points; offers made in that order

Data

# Schools Use Their Discretion




**a) Rank correlation across stages**




**b) Probability of a change in top-ranked**

⇒ Pre-interview (CV) ranking only moderately correlated with the test-score ranking (0.49)

⇒ By the final stage, the top candidate differs from the test-score ranking in ≈40% of contests

⇒ Interview + class demo reshuffle further: CV vs. final rank corr. 0.70; changes in 2/3 of contests

# Does Discretion Improve Teacher Selection?

# Measuring the Value of a Hiring Rule

- Potential outcome of student $\ell$ if taught by teacher $j$

$$Y_{\ell j} = \alpha_j + X'_\ell \beta_j + Z'_\ell \gamma + \epsilon_{\ell j}$$

  $\Rightarrow$ TVA of teacher $j$ in school $i$: $q_{ij} \equiv \alpha_j + X'_i \beta_j$ (ATE + match effects)

- Effect of screening rule $r$ in school $i$ relative to random selection

$$\tau_i(r) \equiv q_{ij_i(r)} - \frac{1}{|\mathcal{J}_i|} \sum_{j \in \mathcal{J}_i} q_{ij}$$

  $\Rightarrow$ Rankings at each stage are observed $\rightarrow$ the effect of school-specific hiring is observed

  $\Rightarrow$ Remaining challenge: recover $q_{ij}$ for shortlisted candidates

# Measuring Teacher Effectiveness

- We infer teacher quality by comparing the test scores of their students
  - ⇒ But teachers are not randomly assigned to schools
- We deal with this selection problem by letting teachers' own school choices reveal what they care about
  - ⇒ Preferences recovered from where teachers get interviews: robust to leaving out schools they cannot reach
- Distance and wages are used as excluded instruments
  - ⇒ Exclusion restriction: neither predicts school dropout, failure rates, or teacher retention Balance
- Result: selection on teacher preferences is small ($\hat{\varphi} = 0.022$, s.e. 0.021; 0.018 if ranked lists are taken at face value) Model Robustness

# What Principals Know Before the Interview (the "Prior") and Teacher Quality

- Each teacher's individual value-added is measured with noise, often based on one or two classrooms
- Estimate the distribution of true effectiveness across teachers
  - $\Rightarrow$ Separate real differences between teachers from estimation noise
  - $\Rightarrow$ Let what is on the application (test score, experience, gender, school choices) predict where a teacher likely sits
- The application-based prediction is what we assume a principal knows about a candidate *before* the interview
- Shrinkage: blend each teacher's own record with the prediction from her application $\Rightarrow$ our best prediction of true quality given the data

# Empirical Bayes (EB) Shrinkage of TVA Estimates

- Assume $\theta_j = (\alpha_j, \beta_j)$ drawn i.i.d. from a normal prior

$$\theta_j | z_j, \tau_j \sim \mathcal{N}(\mu(z_j, \tau_j), \Sigma_\theta)$$

  $\Rightarrow$ Estimate $\mu(\cdot)$ and $\Sigma_\theta$ by MLE from OLS estimates $\hat{\theta}_j$

- Construct shrunken estimates of $\theta_j$ as posterior means after observing $\hat{\theta}_j$

$$\hat{\theta}_j^* = \sum_{k=1}^{K} P(\tau_j = k | z_j, d_j, R_j) \, (\hat{\Omega}_j^{-1} + \hat{\Sigma}_\theta^{-1})^{-1} \Big( \underbrace{\hat{\Omega}_j^{-1} \hat{\theta}_j}_{\text{own record}} + \underbrace{\hat{\Sigma}_\theta^{-1} \hat{\mu}(z_j, k)}_{\text{prediction}} \Big)$$

  $\Rightarrow$ Evaluate at each school's students: $q_{ij}^* = \hat{\alpha}_j^* + X_i' \hat{\beta}_j^*$

# Estimates of the Prior Mean

| | ATE | Lagged score | Female |
|---|---|---|---|
| Baseline ($\tau = 1$) | -0.036 (0.051) | 0.525 (0.026) | -0.057 (0.032) |
| Baseline ($\tau = 2$) | 0.005 (0.071) | 0.593 (0.033) | -0.046 (0.044) |
| Baseline ($\tau = 3$) | -0.122 (0.056) | 0.539 (0.028) | -0.059 (0.036) |
| Female | 0.057 (0.031) | -0.032 (0.016) | -0.045 (0.021) |
| Test score (std.) | 0.096 (0.024) | 0.008 (0.012) | -0.001 (0.014) |
| Quechua | 0.007 (0.050) | -0.009 (0.022) | 0.013 (0.035) |
| Exp public $> 0$ | 0.028 (0.037) | -0.012 (0.018) | 0.010 (0.022) |
| Exp private $> 0$ | 0.065 (0.030) | 0.019 (0.015) | -0.003 (0.020) |

- Female, high-scoring, and experienced teachers have higher average effectiveness (ATE)
  - $\Rightarrow$ Latent preference types (revealed by school choices) also predict the ATE: what teachers want says something about how good they are
  - $\Rightarrow$ Observables explain little of comparative advantage

# Estimates of the Prior Variance

| | ATE | Lagged score | Female |
|---|---|---|---|
| SD | 0.383 | 0.151 | 0.108 |
| | (0.010) | (0.007) | (0.015) |
| Corr. with ATE | | -0.075 | -0.066 |
| | | (0.053) | (0.107) |
| Corr. with lagged score | | | 0.076 |
| | | | (0.119) |

- Large residual variance even after conditioning on the application: $0.38\sigma$ in average effectiveness, $\approx 4\times$ the effect of $+1$ SD in the test score; $0.15\sigma$ / $0.11\sigma$ in comparative advantage

  $\Rightarrow$ Most of what makes a teacher good is invisible on paper: lots of room for schools to learn during the evaluation

# Which Hiring Rule Picks Better Teachers?

| | Test score | Pre-interview | Post-interview | Prior TVA |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Mean gain in expected achievement | 0.010 | 0.011 | 0.025 | 0.016 |
| | (0.008) | (0.008) | (0.008) | (0.008) |
| SD of gains across contests | 0.192 | 0.189 | 0.193 | 0.190 |
| | (0.007) | (0.006) | (0.007) | (0.007) |

Gains $\tau_i(r)$ relative to random selection from the same shortlist, in student test-score SD. 526 contests with $\geq 2$ candidates with TVA.

- Before the interview, schools do no better than the test score (0.011 vs. 0.010)
- After the interview: $0.025\sigma \Rightarrow$ the evaluation reveals productivity-relevant information ($+0.014\sigma$)
- But the average hides a lot: SD across contests $\approx 0.19\sigma$, about $8\times$ the post-interview mean
  $\Rightarrow$ Discretion improves selection in some schools and worsens it in others

# Who Gets Picked Under Each Rule?

| | Random | Δ top-ranked candidate vs. random | | | |
|---|---|---|---|---|---|
| | selection | Test score | Pre-interview | Post-interview | Prior TVA |
| | (1) | (2) | (3) | (4) | (5) |
| Female | 0.802 | -0.009 | 0.006 | -0.024 | 0.065 |
| Test score (std.) | 1.653 | 0.334 | 0.017 | 0.012 | 0.158 |
| Quechua | 0.089 | 0.003 | 0.007 | 0.004 | 0.009 |
| Exp public $> 0$ | 0.900 | 0.013 | 0.047 | 0.039 | 0.032 |
| Exp private $> 0$ | 0.712 | 0.000 | -0.027 | -0.014 | 0.104 |
| Internal candidate | 0.112 | 0.012 | 0.012 | 0.040 | 0.006 |

- After the interview, the top candidate is 2.4 pp less likely female and 4.0 pp more likely internal
- Vs. prior TVA: too little weight on test scores, women, private experience; too much on insiders
- Not because schools know more: pre-interview picks no better than prior TVA (0.011 vs. 0.016); insiders favored whatever their TVA Priors Insiders

  $\Rightarrow$ To steer screening where it pays off, need a model separating what schools want from what they learn

# A Model of Discretionary Hiring and Learning

# Beliefs and Learning

- Schools are heterogeneous along two margins: how much they value TVA ($\delta_i$) and how noisy their interview read is ($\gamma_i$)
- Before the interview, schools have a prior we estimated
  - $\Rightarrow$ Rational expectations given the application: same information on everyone
  - $\Rightarrow$ Exception: internal candidates, whom the school may already know; beliefs about them move a fraction $\rho$ toward their true track record
- During the interview and class demo, the committee gets a noisy signal
  - $\Rightarrow$ Move from the prior toward the signal
  - $\Rightarrow$ Move more when the prior is uncertain; less when the signal is noisy
- Noise of the signal is school-specific

# Beliefs and Learning

- Prior on $q_{ij}$ before the interview, with extra information on insiders ($I_{ij} = 1$)

$$q_{ij} \mid z_j, R_j, d_j, X_i \sim \mathcal{N}\big(\mu_q(z_j, R_j, d_j, X_i), \sigma_q^2(X_i)\big), \qquad \mathbb{E}[q_{ij}|\mathcal{S}_{ij}^1] = \mu_q + \rho I_{ij}\,(q_{ij}^* - \mu_q)$$

- Schools receive a private signal during the interview

$$\nu_{ij} \mid \theta_j, X_i \sim \mathcal{N}\big(q_{ij}, \exp(\gamma_i)\big), \qquad \gamma_i \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2)$$

- and update their beliefs by Bayes' rule

$$\mathbb{E}[q_{ij}|\mathcal{S}_{ij}^2] = \mathbb{E}[q_{ij}|\mathcal{S}_{ij}^1] + \underbrace{\frac{(1-\rho I_{ij})\sigma_q^2(X_i)}{(1-\rho I_{ij})\sigma_q^2(X_i) + \exp(\gamma_i)}}_{w_{ij}\text{: weight on the signal}} \Big(\nu_{ij} - \mathbb{E}[q_{ij}|\mathcal{S}_{ij}^1]\Big)$$

$\Rightarrow$ $\exp(\gamma_i)$: noise in school $i$'s signal; low $\gamma_i$ = good signal extraction; $\sigma_\gamma^2$: heterogeneity across schools

$\Rightarrow$ Information sets: $\mathcal{S}_{ij}^1 = (z_j, d_j, R_j, X_i, I_{ij})$, $\mathcal{S}_{ij}^2 = (\mathcal{S}_{ij}^1, \nu_{ij})$

# School Preferences: What Schools Want

- Schools rank candidates by how much they value them, which mixes:
  - ⇒ Expected effectiveness — with a weight that differs across schools
  - ⇒ Observable traits: experience, gender, ethnicity, being an insider
  - ⇒ Things we don't see: "fit", personality, skills not captured by test scores
- We are agnostic about why other traits matter
  - ⇒ Could be productive in ways TVA misses, taste, or bias
  - ⇒ Weights on observable traits can change after the interview (new info on other traits, interview performance, biased reading, . . . )
- Key point: re-ranking counts as learning only if it tracks candidates' value-added; shifts on observable traits and random reshuffling are accounted for separately

# School Preferences

- Value to school $i$ of hiring teacher $j$ at stage $g \in \{1 \text{ (pre)}, 2 \text{ (post)}\}$

$$V_{ij}^{g} = \underbrace{\exp(\delta_i)\,\mathbb{E}[q_{ij}|\mathcal{S}_{ij}^{g}]}_{\text{expected effectiveness}} + \underbrace{v^{g}(z_j, l_{ij})}_{\text{observable traits}} + \underbrace{\xi_{ij}}_{\text{persistent unobs.}} + \underbrace{\varepsilon_{ij}^{g}}_{\text{transitory}}$$

  - $\Rightarrow$ $\delta_i \sim \mathcal{N}(\mu_\delta, \sigma_\delta^2)$: weight on effectiveness, heterogeneous across schools; independent of $\gamma_i$
  - $\Rightarrow$ $v^{g}(z_j, l_{ij}) = z_j'\psi_v^{g} + \psi_l^{g} l_{ij}$: stage-specific; test score excluded
  - $\Rightarrow$ $\xi_{ij} \sim \mathcal{N}(0, \sigma_{\xi,l_{ij}}^2)$: persistent across stages, different dispersion for insiders
  - $\Rightarrow$ $\varepsilon_{ij}^{g}$: EVT1, i.i.d. across $i, j, g$

- Schools rank candidates by $V_{ij}^{g}$ $\Rightarrow$ rank-ordered logit at each stage

# Identification: Separating Preferences from Learning

- Key for identification: the same candidates ranked twice — before and after the interview
- Ranking before the interview ⇒ what schools want
  - ⇒ Do candidates with higher *expected* TVA get ranked higher, holding other traits fixed?
  - ⇒ Test scores move expected TVA but not other valued traits ⇒ clean variation
- Re-ranking after the interview ⇒ what schools learn
  - ⇒ Candidates who turn out better than expected should move up
  - ⇒ The more re-ranking lines up with TVA *surprises*, the more precise the signal
- Schools appear in several contests with full rankings ⇒ recover each school's "dials", not just averages
- Assumptions: no strategic ranking by schools; shortlists formed from ROLs and test scores before screening Formal Estimation

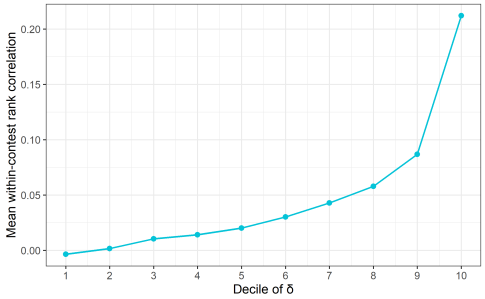
# Estimates

# Heterogeneity in Objectives and in Signal Extraction

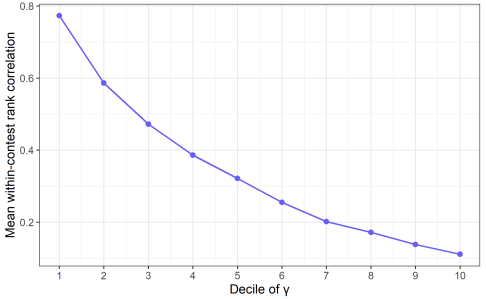
| | Mean | SD |
|---|---|---|
| *Structural parameters* | | |
| Log weight on effectiveness ($\delta_i$) | 0.237<br>(0.243) | 1.122<br>(0.139) |
| Log signal variance ($\gamma_i$) | -0.059<br>(0.524) | 2.081<br>(0.878) |
| *Implied functionals* | | |
| Weight on effectiveness, exp($\delta_i$) | 2.378<br>(0.350) | 3.774<br>(0.847) |
| Weight on interview signal, $w_{ij}$ | 0.234<br>(0.067) | 0.259<br>(0.078) |

- Objectives: weight on effectiveness is $\approx 18\times$ larger at the 90th vs. 10th pctile (5.3 vs. 0.3)
- Learning: on average, schools put 23% the signal from the interview (SD 0.26)
- Insiders: no extra information about TVA ($\hat{\rho} \approx 0$), yet the preference for them quadruples after the interview Other parameters

# What the Heterogeneity Means




**a) Expected TVA vs. pre-interview utility**
by decile of $\delta_i$




**b) Post-interview beliefs vs. true TVA**
by decile of $\gamma_i$

⇒ Even schools that care most: TVA explains little of their ranking (rank corr. $\approx 0.21$; $\approx 0$ at the bottom)

⇒ Best readers' beliefs correlate 0.76 with true TVA after the interview; worst readers only 0.11

Model fit

# Why Does Discretion Work in Some Schools but Not Others?

# Variance Decomposition of Gains from Discretion

- Drop each school "type" $(\delta, \gamma)$ into every shortlist: how much of the variance in gains comes from the decision maker rather than the applicant pool?

| | Variance | Share of total |
|---|---|---|
| Total variance of selection gain | 0.083 | 1.000 |
| Applicant pools | 0.063 | 0.756 |
| Joint heterogeneity in $\delta_i$ and $\gamma_i$ | 0.020 | 0.244 |
| Heterogeneity in $\delta_i$ only (objectives) | 0.014 | 0.170 |
| Heterogeneity in $\gamma_i$ only (signal extraction) | 0.010 | 0.114 |

One-at-a-time rows fix the other primitive at its median; they need not sum to the joint share.

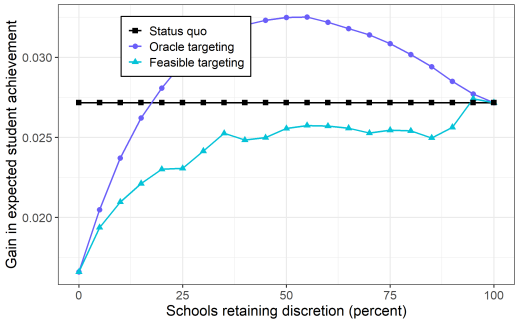
- Decision makers explain one quarter of the variance; both margins matter (objectives 17%, signal extraction 11%) Heatmap
- Neither is predictable from school or principal observables ($R^2 \leq 0.002$) Table $\Rightarrow$ Hard to target delegation

# Designing Screening in Centralized Assignment

# Two Design Margins for the Central Platform

- Discretion adds information on average, but its value varies across schools and is hard to predict
- Margin 1: which schools screen?
  - $\Rightarrow$ Keep discretion only where schools read candidates well; the rest hire the top test scorer
- Margin 2: which applicants get screened?
  - $\Rightarrow$ Who makes the shortlist (random, test score, or predicted TVA) and number of slots ($K$)
  - $\Rightarrow$ Alternative: disclose TVA estimates to schools and let them keep interviewing
- Simulation: shortlists $\rightarrow$ schools rank with their estimated preferences and interview signals $\rightarrow$ school-proposing deferred acceptance; applicants accept according to their estimated preferences

# Targeting Discretion to Schools that Read Candidates Well




⇒ Schools ranked by signal-extraction ability ($\gamma_i$); schools without discretion hire the top test scorer

⇒ Oracle (true $\gamma_i$): gains peak at $0.033\sigma$ with half the schools screening, vs. $0.027\sigma$ in the status quo

⇒ Feasible (estimated $\gamma_i$): below the status quo almost everywhere

⇒ Allocating screening across schools offers little: ability is hard to predict, and even perfectly observed it buys little

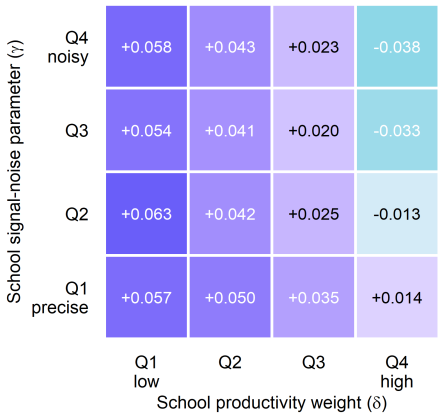
# Who Gets Screened: Prescreening vs. Disclosure

| | Interviews | | | TVA disclosed + interviews | | |
|---|---|---|---|---|---|---|
| Shortlist capacity | Test score | Random | ATE | Test score | Random | ATE |
| Status quo | 0.180 | 0.181 | 0.179 | 0.199 | 0.199 | 0.198 |
| $K = 5$ | 0.182 | 0.176 | 0.220 | 0.194 | 0.189 | 0.228 |
| $K = 3$ | 0.181 | 0.168 | 0.283 | 0.187 | 0.173 | 0.286 |
| $K = 1$ | 0.182 | 0.163 | 0.340 | 0.182 | 0.163 | 0.340 |

Mean relative TVA of assigned teachers (net of student composition). Columns: centralized priority rule. $K$: maximum shortlist length (at least the number of vacancies).

- Priorities matter only when slots are scarce: random $\approx$ test score in the status quo; at $K = 1$ random falls to 0.163
- Tighter test-score shortlists buy nothing (0.180 $\rightarrow$ 0.182): the test is a weak proxy for productivity
- Predicted-TVA shortlists: 0.220 ($K = 5$) $\rightarrow$ 0.283 ($K = 3$) $\rightarrow$ 0.340 ($K = 1$)
- Disclosing TVA to schools: 0.180 $\rightarrow$ 0.199, below TVA prescreening at $K = 5$; combining both adds little (0.228)

# Who Benefits from Prescreening vs. Disclosure?




- Each cell: assigned TVA under TVA prescreening ($K = 5$) minus TVA disclosure (status quo shortlists); positive favours prescreening
- Prescreening wins where schools care little about TVA ($+0.04$ to $+0.06\sigma$), whatever their signal quality
- Disclosure wins only where schools care a lot but read candidates poorly (high $\delta_i$, noisy $\gamma_i$)

$\Rightarrow$ Better information does not help schools that want other things; prescreening changes the menu they choose from

# The Cost of Aggressive Prescreening

| | Mean TVA | | Unfilled vacancies (%) | | Var. of school TVA | |
|---|---|---|---|---|---|---|
| Shortlist capacity | Test score | ATE | Test score | ATE | Test score | ATE |
| Status quo | 0.180 | 0.179 | 37.5 | 37.5 | 0.117 | 0.117 |
| $K = 5$ | 0.182 | 0.220 | 35.8 | 35.8 | 0.116 | 0.117 |
| $K = 3$ | 0.181 | 0.283 | 42.9 | 43.0 | 0.122 | 0.116 |
| $K = 1$ | 0.182 | 0.340 | 56.8 | 56.9 | 0.122 | 0.118 |

Interviews, no disclosure. Variance of average assigned TVA across schools with at least one hire. Full table

- Shorter shortlists ⇒ schools compete for the same candidates with fewer fallbacks: unfilled vacancies rise from 37% to 43% ($K = 3$) and 57% ($K = 1$)
- The vacancy cost is the same under every priority rule: it comes from restricting capacity, not from using TVA
- Cross-school variance of assigned TVA stays $\approx 0.117$: higher average quality, not less inequality

  ⇒ Trade-off: direct scarce screening to promising matches vs. keep enough options for matches to form

# Discussion

- Rich data on how Peruvian schools rank the same candidates before and after the interview, linked to student outcomes
  - ⇒ Discretion improves teacher selection on average through learning, but gains are highly unequal across schools
- Model separates objectives from signal-extraction ability
  - ⇒ Large heterogeneity in both; together 1/4 of the variance in gains
  - ⇒ Neither is predictable from observables
- Counterfactuals: how should the central platform organize screening?
  - ⇒ Targeting discretion to schools that read candidates well: modest gains even with an oracle, none in practice
  - ⇒ Centralized priorities matter once screening slots are scarce; tighter test-score prescreening adds little
  - ⇒ Using predicted TVA to decide who gets screened beats disclosing it to schools, especially where schools care little about TVA
  - ⇒ Prescreening too aggressively leaves more vacancies unfilled; inequality across schools barely moves

# Discussion

- The value of information depends on how it is used, not only on its quality
  - ⇒ The same TVA estimates are worth more for deciding who gets screened than when handed to schools
  - ⇒ Especially where school objectives are poorly aligned with productivity
- Supply and demand sides of teacher allocation are complements
  - ⇒ Better screening raises average quality but leaves inequality across schools unchanged: that is decided by who applies where
  - ⇒ Pay and incentives shape the pool; screening determines who is picked from it (Bobba et al., 2026; Biasi et al., 2026; Bates et al., 2025)
- Open questions
  - ⇒ Can signal-extraction ability be measured or improved (screening protocols, evaluator training)?
  - ⇒ Same design problem wherever central exams meet local evaluations: doctors, police, civil servants, residency matches with interviews

# Backup Slides

# Data

- Centralized recruitment drives for permanent positions: 2015, 2017, 2018, 2019
  - $\Rightarrow$ Universe of applicants, vacancies and applicants' ROLs
  - $\Rightarrow$ Shortlists and candidate ranking at each stage of school evaluation

- Linked administrative sources
  - $\Rightarrow$ Payroll 2014–2023: who was hired, who was principal, internal candidates
  - $\Rightarrow$ Applicants' home location $\Rightarrow$ distance to each school
  - $\Rightarrow$ Student test scores (2nd & 4th grade, 2016 & 2018) linked to classrooms and teachers
  - $\Rightarrow$ School census: infrastructure, amenities, locality characteristics

- Main hiring sample: 14,479 evaluated applicant–school pairs, 7,062 applicants, 3,642 contests in 2,874 schools Samples

Back

# Analysis Samples

| | Applicant choice | Outcome (TVA) | Interview |
|---|---|---|---|
| Applicants / teachers | 10,908 | 1,425 | 7,062 |
| Applicant / teacher-years | 15,400 | 1,527 | 8,755 |
| Schools | 6,417 | 1,273 | 2,874 |
| School-year observations | 11,466 | 1,413 | 3,642 |
| Vacancies | 20,114 | | |
| Students | | 28,283 | |
| Shortlisted applicant–school pairs | | | 14,479 |

- Samples serve different purposes $\Rightarrow$ composition differs by construction, not attrition

Back

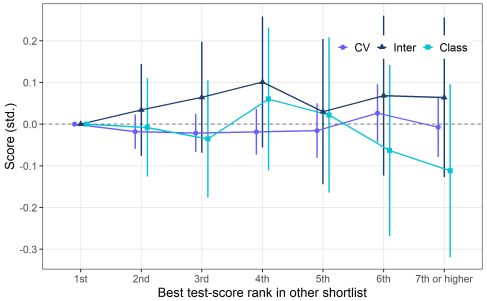
# Reading School Rankings as Revealed Preferences

- We treat each ranking as what the school wants, given what it knows at that point
  - ⇒ Pre-interview ranking: based on CV score
  - ⇒ Post-interview ranking: cumulative CV + interview + classroom score
- There is real discretion in how committees score CV:
  - ⇒ Observables (test score, experience, gender) explain only 22% of the CV score
- . . . and no strategic gaming
  - ⇒ Scores unrelated to a candidate's initial position in the shortlist
  - ⇒ Scores unrelated to the candidate's prospects in other schools Details

# School Scores Are Not Strategic




**a) Initial position in the focal shortlist**




**b) Position in competing shortlists**

$\Rightarrow$ CV, interview and class scores unrelated to the candidate's test-score rank in the shortlist, or to her best rank elsewhere

Back

# Selection Correction of TVA Estimates

- Utility of teacher $j$ from school $i$, with latent preference type $\tau_j \in \{1, 2, 3\}$

$$U_{ij} = u(z_j, x_i, d_{ij}, \tau_j) + \eta_{ij}, \qquad \eta_{ij} \sim \text{EVT1}$$

  - $\Rightarrow$ $x_i$: wage, amenities, remoteness, students; $z_j$: gender, ethnicity, experience, test score; $d_{ij}$: distance
  - $\Rightarrow$ Baseline: stability of the interview assignment (serial dictatorship by test score) $\Rightarrow$ choices from each applicant's feasible set; alternative: truth-telling ROLs (exploded logit)
  - $\Rightarrow$ Types integrated out with the EM algorithm ($K = 3$)
- Allow teachers' unobserved taste for their school to correlate with productivity there

$$\mathbb{E}[Y_{\ell j} | X_\ell, Z_\ell, S_\ell, R_j, z_j, d_j] = \alpha_j + X_\ell' \beta_j + Z_\ell' \gamma + \varphi \, \lambda_{i(\ell)}(R_j, z_j, d_j)$$

  - $\Rightarrow$ $\lambda(\cdot)$: Dubin–McFadden control function, integrating over posterior types; built from ROLs $R_j$ (truth-telling) or interview assignments $R_j^S$ (stability)
  - $\Rightarrow$ Excluded preference shifters: distance $d_{ij}$ and wages
  - $\Rightarrow$ $\hat{\varphi}^S = 0.022$ (0.021), $\hat{\varphi}^{TT} = 0.018$ (0.014); shrunken ATEs nearly identical across the two (corr. $= 0.992$) Figure

Back

# Balance Tests for Excluded Preference Shifters

| | Dropout | Failing rate | Teacher retention |
|---|---|---|---|
| Distance (1,000 km) | 0.0005 | 0.0030 | -0.0219 |
| | (0.0030) | (0.0040) | (0.0196) |
| Wage (1,000 soles) | 0.0030 | 0.0012 | -0.0105 |
| | (0.0012) | (0.0017) | (0.0087) |

Separate regressions for distance and wage; all control for school characteristics in the outcome equation.

- Coefficients uniformly small $\Rightarrow$ shifters move preferences, not achievement determinants

Back

# Do Schools Select on Unobserved Match Quality?

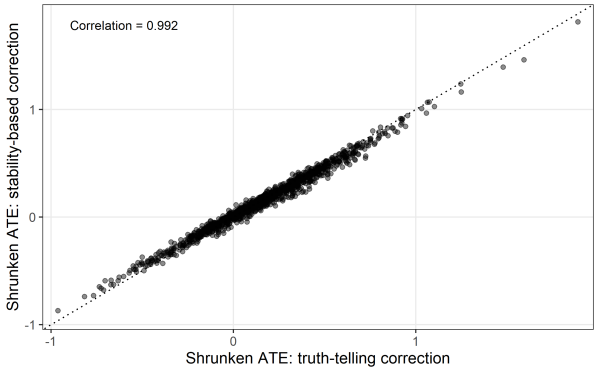
| | CV | Interview | Classroom |
|---|---|---|---|
| Screening score | -0.0028 | -0.0054 | 0.0029 |
| | (0.0049) | (0.0048) | (0.0052) |
| $R^2$ | 0.001 | 0.003 | 0.001 |

Dep. var.: teacher–school–year average residual from the TVA equation. Scores standardized within contest.

- Schools' CV/interview/class scores do not predict TVA residuals
- Preferences under truth-telling instead of stability $\Rightarrow \hat{\varphi} = 0.018$ (0.014); corr. of shrunken ATEs $= 0.992$ Figure

Back

# Robustness: Strategic Omission of Infeasible Schools




⇒ Shrunken ATEs under truth-telling vs. stability-based preference estimates (Fack et al., 2019): correlation = 0.992, although the preference estimates themselves differ substantially

Back

# Misalignment Under Coarser Priors

| | Pre-interview | Prior TVA | Coarse: no $R_j$ | Coarse: no $R_j$, gender, ethn. |
|---|---|---|---|---|
| Female | 0.006 | 0.065 | 0.136 | 0.008 |
| Test score (std.) | 0.017 | 0.158 | 0.239 | 0.268 |
| Exp public $> 0$ | 0.047 | 0.032 | -0.004 | -0.012 |
| Exp private $> 0$ | -0.027 | 0.104 | 0.189 | 0.203 |
| Internal candidate | 0.012 | 0.006 | 0.002 | 0.005 |
| Shrunken TVA | 0.011 | 0.016 | -0.002 | 0.004 |

- Pre-interview rankings remain weakly aligned with predicted TVA under any of these priors
- Same misalignment among experienced and high-skill principals

Insiders

# Selection of Internal Candidates

| | Random | Pre-interview | Post-interview |
|---|---|---|---|
| Internal candidate | 0.121 | 0.005 (0.010) | 0.033 (0.010) |
| Shrunken TVA | 0.272 | 0.011 (0.008) | 0.025 (0.008) |
| High-VA internal | 0.059 | 0.003 (0.008) | 0.017 (0.008) |
| Low-VA internal | 0.061 | 0.001 (0.007) | 0.017 (0.007) |

- Insider premium after the interview is the same for high- and low-VA insiders

# Identification: Formal

- Assumptions
  - $\Rightarrow$ Schools rank by $V_{ij}^g$: no strategic ranking (consistent with scores unrelated to positions elsewhere)
  - $\Rightarrow$ $(\varepsilon_{ij}^g, \xi_{ij}, \delta_i, \gamma_i, \nu_{ij}) \perp \mathcal{J}_i$: shortlists formed from ROLs and test scores before screening

- Weight on effectiveness $\delta_i$: for two candidates $j, k$ in the same contest, $P(j \succ k)$ increasing in

$$\exp(\delta_i)\,\Delta\mathbb{E}[q|\mathcal{S}^1]_{jk} + \Delta v^1_{jk}$$

  - $\Rightarrow$ Test score shifts $\Delta\mathbb{E}[q]$ but not $\Delta v^1$ (checked: residual test score $\approx$ no effect on school scores) Check

- Signal precision $\gamma_i$: belief revisions respond to productivity surprises with slope

$$w_{ij}(\gamma_i) = \frac{s_{ij}^2}{s_{ij}^2 + \exp(\gamma_i)} \quad \text{strictly decreasing in } \gamma_i$$

Back

# Test-Score Exclusion Restriction

| | CV | Interview | Classroom |
|---|---|---|---|
| Residualized test score (1 SD) | 0.020 (0.010) | 0.022 (0.010) | 0.009 (0.010) |
| Partial $R^2$ | 0.0006 | 0.0006 | 0.0001 |
| Strict pairwise reversals | 0.00% | 0.00% | 0.00% |

Scores and test score residualized on prior expected TVA, other $v^g$ observables, and school×year FE.

- Test score affects school scores only through expected effectiveness

Back

# Estimation: Intuition

- We never see a school's $\delta_i$, $\gamma_i$, or the signals it received
- Simulation approach
  - $\Rightarrow$ For each school, imagine many hypothetical versions of it: different weights on TVA, different learning ability, different interview signals
  - $\Rightarrow$ Signals drawn consistent with what we know about each candidate's TVA
  - $\Rightarrow$ For each version, compute how likely it is to produce both observed rankings (pre and post)
  - $\Rightarrow$ Average over versions; pick the parameters that make the observed rankings most likely
- By-product: a posterior over each school's $(\delta_i, \gamma_i)$, used in all simulations

Back

# Estimation

- Likelihood of observing ranking $R_i^g$ (rank-ordered logit)

$$\mathcal{L}(R_i^g|\delta_i,\gamma_i,\boldsymbol{\xi_i},\boldsymbol{\mathcal{S}_i^g}) = \prod_{j\in\mathcal{J}_i} \frac{\exp\{\exp(\delta_i)\mathbb{E}[q_{ij}|\mathcal{S}_{ij}^g] + v^g(z_j, l_{ij}) + \xi_{ij}\}}{\sum_{k:jR_i^g k}\exp\{\exp(\delta_i)\mathbb{E}[q_{ik}|\mathcal{S}_{ik}^g] + v^g(z_k, l_{ik}) + \xi_{ik}\}}$$

- Maximize simulated joint likelihood (25,000 draws)

$$\sum_{i=1}^{N}\log\int\prod_{g=1}^{2}\mathcal{L}(R_i^g|\delta_i,\gamma_i,\boldsymbol{\xi_i},\boldsymbol{\mathcal{S}_i^g})\,dF(\delta_i,\gamma_i,\boldsymbol{\xi_i},\boldsymbol{\nu_i}|\hat{\boldsymbol{\theta}})$$

  - $\Rightarrow$ Draw $\nu_{ij}$ conditional on value-added estimates $\hat{\theta}_j$: extra variation for learning parameters
  - $\Rightarrow$ $(\delta_i,\gamma_i)$ fixed across contests of the same school
  - $\Rightarrow$ Parameters: $\mu_\delta,\sigma_\delta,\mu_\gamma,\sigma_\gamma,\psi_v^1,\psi_v^2,\psi_l^1,\psi_l^2,\rho,\sigma_{\xi,0},\sigma_{\xi,1}$

Back

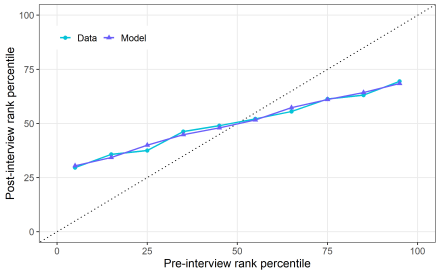
# Other Determinants of Rankings

| | Pre ($v^1$) | Post ($v^2$) |
|---|---|---|
| Female | 0.128 (0.053) | -0.361 (0.053) |
| Quechua | 0.212 (0.087) | 0.230 (0.092) |
| Public-sector experience | 1.593 (0.061) | 0.794 (0.062) |
| Private-sector experience | -0.315 (0.048) | -0.164 (0.050) |
| Internal candidate | 0.192 (0.067) | 0.811 (0.069) |
| Private info on insiders, $\rho$ | 0.002 (0.157) | |
| SD persistent unobs.: external / internal | 1.257 (0.025) / 1.279 (0.069) | |

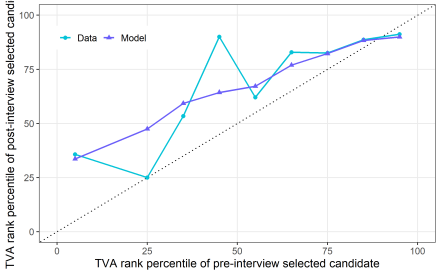
Also includes applicant latent-type probabilities (not shown).

- Insiders: no extra information about their TVA ($\rho \approx 0$), yet preference for them quadruples after the interview
- Women: penalized after the interview — bias, interview performance, or other traits? (we don't take a stand)
- Large weight on traits we don't observe ($\sigma_\xi \approx 1.26$)

# Model Fit




**a) Pre- vs. post-interview rank percentiles**




**b) TVA rank of selected candidates**

⇒ Reshuffling: any rank change 0.66 (data) vs. 0.67 (model); rank corr. 0.71 vs. 0.70

⇒ Learning gain $+0.014\sigma$ (data) vs. $+0.009\sigma$ (model); dispersion of gains 0.19 in both — not targeted

⇒ Also reproduces the drop in women and the rise in insiders after the interview Table

Back

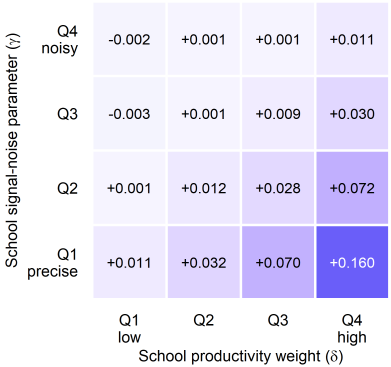
# Model Fit: Who Gets Picked

| | Random | Pre-interview | | Post-interview | |
|---|---|---|---|---|---|
| | | Data | Model | Data | Model |
| Female | 0.802 | 0.006 | 0.010 | -0.024 | -0.023 |
| Test score (std.) | 1.653 | 0.017 | 0.012 | 0.012 | 0.011 |
| Quechua | 0.089 | 0.007 | 0.006 | 0.004 | 0.007 |
| Exp public $> 0$ | 0.900 | 0.047 | 0.057 | 0.039 | 0.035 |
| Exp private $> 0$ | 0.712 | -0.027 | -0.026 | -0.014 | -0.016 |
| Internal candidate | 0.112 | 0.012 | 0.013 | 0.040 | 0.038 |

- Model reproduces the drop in women and the rise in insiders after the interview

Back

# Gains Concentrate Where Schools Both Care and Read Well




$\Rightarrow$ Gains from the interview concentrated at high $\delta_i$ and low $\gamma_i$

$\Rightarrow$ Schools below the median on both: essentially no gain vs. test-based hiring

# Observable Predictors of $\delta_i$ and $\gamma_i$

| | Posterior mean $\delta_i$ | | Posterior mean $\gamma_i$ | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Avg. student test score | 0.000 | 0.005 | 0.021 | 0.019 |
| | (0.009) | (0.012) | (0.012) | (0.014) |
| Female student share | -0.006 | 0.009 | 0.036 | 0.040 |
| | (0.034) | (0.065) | (0.038) | (0.073) |
| Urban | -0.007 | -0.008 | 0.009 | 0.010 |
| | (0.010) | (0.014) | (0.013) | (0.016) |
| Principal experience (yrs) | | 0.004 | | -0.006 |
| | | (0.004) | | (0.005) |
| Principal test score (std.) | | 0.001 | | 0.006 |
| | | (0.006) | | (0.009) |
| Observations | 2874 | 1479 | 2874 | 1479 |
| $R^2$ | 0.000 | 0.001 | 0.002 | 0.002 |

# Centralized Prescreening and Information Disclosure: Full Table

| | Interviews | | | TVA disclosed + interviews | | |
|---|---|---|---|---|---|---|
| Shortlist capacity | Test score | Random | ATE | Test score | Random | ATE |
| *Panel A. Mean VA of assigned teachers* | | | | | | |
| Status quo | 0.180 | 0.181 | 0.179 | 0.199 | 0.199 | 0.198 |
| $L = 5$ | 0.182 | 0.176 | 0.220 | 0.194 | 0.189 | 0.228 |
| $L = 3$ | 0.181 | 0.168 | 0.283 | 0.187 | 0.173 | 0.286 |
| $L = 1$ | 0.182 | 0.163 | 0.340 | 0.182 | 0.163 | 0.340 |
| *Panel B. Unfilled vacancies (%)* | | | | | | |
| Status quo | 37.47 | 37.49 | 37.48 | 37.50 | 37.52 | 37.50 |
| $L = 5$ | 35.82 | 35.89 | 35.75 | 35.87 | 35.93 | 35.77 |
| $L = 3$ | 42.90 | 43.07 | 42.97 | 42.92 | 43.10 | 43.01 |
| $L = 1$ | 56.84 | 57.03 | 56.86 | 56.84 | 57.03 | 56.86 |
| *Panel C. Variance of school-level assigned VA* | | | | | | |
| Status quo | 0.117 | 0.117 | 0.117 | 0.118 | 0.118 | 0.119 |
| $L = 5$ | 0.116 | 0.116 | 0.117 | 0.117 | 0.116 | 0.117 |
| $L = 3$ | 0.122 | 0.121 | 0.116 | 0.122 | 0.121 | 0.116 |
| $L = 1$ | 0.122 | 0.121 | 0.118 | 0.122 | 0.121 | 0.118 |

**Restricted simulation market (interview sample); final assignments use schools' discretionary rankings; school-proposing deferred acceptance. Averages across posterior-predictive simulations.**