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

Matteo Bobba† Tim Ederer‡ Gianmarco León-Ciliotta§
Christopher A. Neilson¶ Marco Nieddu‖

November 18, 2026

**Abstract**

We study the design of decentralized signal extraction in centralized assignment. Using data on Peru's nationwide teacher allocation mechanism, we estimate a rich model of applicants' behavior, teacher effectiveness and schools' discretionary screening accounting for heterogeneity in schools' preferences and signal-extraction ability. Decentralized screening improves the effectiveness of selected teachers on average relative to test-based hiring, but its returns vary substantially across schools. We find that efficiently allocating screening capacity according to schools' signal extraction ability is unfeasible. Instead, we show that using teacher value-added measures for centralized prescreening can generate large gains and outperforms disclosing that information directly to schools. However, prescreening too aggressively leaves more vacancies unfilled as schools increasingly compete for the same candidates.

*We thank the Ministry of Education of Perú for supporting rigorous policy evaluation efforts and continued collaboration. Arnaud Maurel, Elena Pastorino, Michela Tincani, and Andrea Salvati generously provided us with insightful feedback on the draft. Matteo Bobba acknowledges financial support from the French National Research Agency under grants ANR-17-EURE-0010 (Investissements d'Avenir program) and ANR-26-CE26-7389-01 (BRIDGEGAPS). Gianmarco León-Ciliotta recognizes financial support from the Spanish Ministry of Science, Innovation and Universities (through the Severo Ochoa Programme for Centres of Excellence in R&D CEX2019-000915-S), grant PID2023-150768NB-I00, and the ICREA Academia program.

†Toulouse School of Economics & CEPR & IZA, matteo.bobba@tse-fr.eu
‡Carnegie Mellon University, tederer@andrew.cmu.edu
§Universitat Pompeu Fabra & BSE & IPEG & BREAD & CEPR & IZA, gianmarco.leon@upf.edu
¶Yale University & NBER, christopher.neilson@yale.edu
‖University of Cagliari & CRENoS, mgnieddu@unica.it

# 1 Introduction

Information asymmetries are pervasive in matching markets and can undermine the effectiveness of centralized assignment mechanisms (Artemov, 2021; Fernandez et al., 2022; Friedrich et al., 2024). While centralization can help coordinate competition among agents for available matches, information about match quality can often only be acquired through costly decentralized evaluations, such as interviews (Lee and Schwarz, 2017). Allocating screening capacity efficiently thus requires directing it toward promising matches while preserving opportunities for decentralized signal extraction. This design problem can be particularly challenging to resolve when the returns to local discretion are heterogeneous and difficult to predict ex ante (Dessein, 2002; Frankel, 2021). This paper studies this trade-off in the context of centralized teacher recruiting.

Teachers are among the most important inputs in the education production function, and being exposed to effective teachers has important consequences for both contemporaneous achievement and long-run outcomes (Rockoff, 2004; Rivkin et al., 2005; Chetty et al., 2014b). Yet we still know comparatively little about how to allocate effective teachers to the schools where they matter most. Recent work studies how compensation and assignment policies can attract effective teachers to disadvantaged or hard-to-staff schools (Biasi et al., 2021; Bobba et al., 2021). These policies, however, can only improve teacher allocation if schools can identify and select effective teachers among the applicants they face (Bates et al., 2025). Screening policy design is thus essential to enhancing the efficiency and equity of teacher allocation.

Our analysis combines detailed information on applicants' choices, schools' discretionary evaluations at different stages of the recruitment process, and subsequent student outcomes within Peru's centralized teacher allocation mechanism. We first estimate teacher effectiveness and show that discretionary screening generates positive but modest average gains in expected student achievement relative to test-based hiring, with substantial heterogeneity across schools. We then develop and estimate a model of hiring under uncertainty and learning that separates schools' preferences from their ability to extract informative signals about teacher effectiveness. The estimates imply substantial heterogeneity along both dimensions and show that they jointly explain one quarter of the variance in gains from discretion across schools. We find that efficiently allocating screening capacity according to schools' signal

extraction ability is unfeasible, as the latter is difficult to predict from observables. Instead, we show that using teacher value-added measures for centralized prescreening can generate large gains and outperforms disclosing that information directly to schools. However, prescreening too aggressively leaves more vacancies unfilled as schools increasingly compete for the same candidates.

The Ministry of Education of Peru hires permanent teachers through a two-stage process that combines centralized screening with decentralized school evaluations. Applicants first take a national test and submit rank-ordered lists of schools. Applicants are then prescreened through serial dictatorship using their test score as priorities. The shortlisted applicants are then evaluated by each school in a decentralized stage. School committees, presided by the school principal, then have substantial discretion over candidate rankings through a CV review and a set of in-person interviews. Pre-interview school rankings are only moderately correlated with the centralized test-score ranking. Schools also substantially reshuffle candidates as they acquire additional information: about two-thirds change their rankings between the pre- and post-interview stages, with a rank correlation across the two stages of about 0.70. Schools thus actively use their discretionary power when selecting among shortlisted candidates.

To quantify the value of discretion, we estimate a flexible model of teacher effectiveness that allows teachers to differ in their average value-added and their comparative advantage with respect to students' past achievement and gender. To address endogenous teacher sorting across schools, we use a control function approach capturing selection on unobserved match quality, leveraging applicants' revealed preferences through their rank-ordered school choices. We then construct empirical-Bayes posterior measures of effectiveness for each applicant–school pair. Recovering applicant preferences from stable interview assignments rather than assuming truthful reporting yields substantially different preference estimates, but nearly identical selection-corrected estimates of teacher effectiveness. Our estimates reveal substantial variation in teacher effectiveness that is not explained by observable characteristics, suggesting large potential gains from decentralized information acquisition.

We first use these estimates to quantify the effects of hiring discretion relative to alternative selection rules on student achievement, holding the set of shortlisted candidates fixed. Ranking candidates by the centralized test score raises expected math test scores by 0.010 standard deviations ($\sigma$), relative to random selection from the same shortlist. Schools' discre-

tionary rankings generate a similar gain of $0.011\sigma$ pre-interview but a larger gain of $0.025\sigma$ post-interview, indicating that schools extract informative signals about teacher effectiveness. This average effect masks substantial heterogeneity: the standard deviation of these gains across schools is about 0.19. More importantly, school rankings differ systematically from rankings based on predicted teacher effectiveness, even under alternative specifications of schools' prior beliefs and among more experienced principals. Schools also favor applicants previously employed at the school (internal candidates) without preferentially selecting those with higher value added. These patterns are difficult to rationalize by imperfect information alone and suggest that schools value candidate attributes beyond teacher effectiveness. This points to two sources of heterogeneity in screening performance: differences in schools' ability to acquire productivity-relevant information during the interview and differences in how much schools value teacher effectiveness.

To separate these channels and understand their influence on screening effectiveness, we develop an empirical model of discretionary hiring and learning. Schools begin with beliefs about applicant effectiveness and receive a private productivity signal during the interview. We also allow schools to value and acquire information about candidate attributes other than expected teacher effectiveness. We model preferences over these other traits as a function of observable candidate attributes and unobservable match-specific attributes containing both a persistent and transitory component. Importantly, we allow for unobserved heterogeneity in the weight schools place on expected teacher effectiveness and in the precision of the signal they extract during the interview, which we label as signal extraction ability.

Identification relies on observing complete rankings of the same candidate set before and after schools acquire additional information. We use teacher test scores as a shifter of prior expected teacher effectiveness excluded from preferences to identify how schools value productivity. Changes in rankings post-interview relative to estimated teacher value-added then identify the precision of schools' private signals. We show that common preference and signal precision parameters can be identified from a single cross-section without requiring parametric restrictions on preferences over other observed and unobserved traits. Repeated observations of schools across years provide additional variation to identify the distribution of heterogeneous preferences and signal-extraction ability.

Our estimates imply both limited alignment of schools' objectives with expected teacher effectiveness and signal extraction ability. However, we find substantial heterogeneity along

both dimensions. For schools in the top decile of the distribution of weight placed on effectiveness, the rank correlation between expected teacher effectiveness and the overall pre-interview ranking reaches about 0.21, while being close to zero among schools placing the least weight on effectiveness. Similarly, the rank correlation between post-interview beliefs and underlying teacher effectiveness reaches 0.76 among schools in the top decile of signal precision, compared with only 0.11 among schools in the bottom decile. We show that the estimated model closely reproduces rank persistence and reshuffling across stages, the cross-school dispersion in productivity gains, the value-added gain generated by learning, and changes in candidates selection based on observable characteristics.

Overall, heterogeneity in objective functions and signal extraction ability jointly account for 24.4 percent of the variance in selection gains relative to random selection across schools, while the remaining share comes from differences in applicant pools across schools. Heterogeneity in preferences alone accounts for 17 percent while heterogeneity in signal extraction ability accounts for 11.4 percent. Observable school and principal characteristics explain virtually none of the variation in either dimension. We evaluate policies that allocate discretion to schools with greater signal-extraction ability and find modest improvements in teacher selection, even in the extreme (oracle) scenario in which this ability is perfectly observed. A feasible policy that allocates discretion based on observed posterior means of signal extraction ability generates no gains relative to the status quo.

Finally, we study whether centralized prescreening can instead direct screening opportunities toward more productive applicants. We first show that random priorities lead to less effective hires than test-score priorities when screening capacity is restricted. However, making test score prescreening more restrictive by reducing the maximum shortlist length yields virtually no improvement in teacher effectiveness, reflecting their limitations as proxies for value-added. Using shrunken value-added estimates as priorities in the centralized prescreening stage instead generates substantial gains, outperforming a policy that provides this information directly to schools. We show that the relative effectiveness of these interventions depends on schools' preferences and signal-extraction ability. Centralized prescreening is particularly beneficial for schools that place little weight on teacher effectiveness, whereas disclosure can be more valuable for schools with strong productivity preferences but noisy signals. Reducing shortlist length further increases the effectiveness of hires, but decreases the share of filled vacancies as schools increasingly compete for the same candidates.

This paper first contributes to the literature on information acquisition and the design of centralized matching mechanisms. Previous work studies how interviews and other pre-matching interactions affect matching outcomes (Lee and Schwarz, 2017; Echenique et al., 2022), how assignment mechanisms shape incentives to acquire information (Artemov, 2021; Hakimov et al., 2023), and how asymmetric information about match quality affects strategic incentives (Friedrich et al., 2024). We study how centralized mechanisms should allocate scarce opportunities for decentralized screening when information about applicant productivity is imperfect. We find that efficiently allocating screening capacity is unfeasible when signal extraction ability is difficult to predict. Instead, using information about teacher effectiveness for prescreening can generate large gains. However, overly restrictive prescreening increases competition across schools for the same candidates, leaving more vacancies unfilled.

We also contribute to the literature on equilibrium sorting in teacher labor markets. Recent work studies how compensation and other non-pecuniary motives shape teacher sorting and student achievement (Biasi et al., 2021; Bobba et al., 2021; Ederer, 2026; Bates et al., 2025). Much less is known about the demand side of teacher allocation, despite evidence that supply-side interventions may have limited effects when schools do not hire teachers on the basis of effectiveness (Bates et al., 2025; Colas and Fu, 2025). We show that discretionary screening improves teacher selection on average relative to test-based hiring, but its returns vary substantially across schools. Heterogeneity in schools' objectives and signal-extraction ability explains nearly one quarter of the variation in these returns. We find that prescreening teachers based on shrunken teacher value-added estimates can generate large gains in student achievement and outperforms disclosing this information directly to schools.

We also contribute to the literature on learning and information in labor markets and bridge it to work in personnel economics on discretion and worker selection. A large literature studies how employers learn about worker productivity and how information frictions affect labor-market outcomes (Farber and Gibbons, 1996; Altonji and Pierret, 2001; Kahn and Lange, 2014; Pallais, 2014; Carranza et al., 2022) while other work studies the role of managers and discretion in hiring (Autor and Scarborough, 2008; Hoffman et al., 2018; Colonnelli et al., 2020; Weaver, 2021). We study hiring under uncertainty as a signal-extraction problem and separately identify schools' objectives from their signal extraction ability. We find that heterogeneity in both dimensions explains a substantial share of the variation in returns to discretion across schools. Separating these two channels is therefore important

for understanding why discretion performs differently across organizations.

# 2 Context and Data

## 2.1 The Allocation Mechanism for Permanent Teachers

There are two types of teaching positions in Peru's public schools. Short-term contracts last for up to two years and pay a flat wage. Permanent positions are of indefinite length, pay wages that increase with experience, and constitute the career track of the public teaching profession. Since 2015, all positions are assigned through nationwide, centralized recruitment drives administered by the Ministry of Education (MoE). We focus on the recruitment process of permanent teachers, which combines a centralized prescreening mechanism with a decentralized school-level evaluation.[1]

All applicants take a standardized test (Prueba Única Nacional, PUN) evaluating skills and curricular knowledge. Only candidates scoring above 60% in each component of the PUN are eligible for permanent positions, while lower-scoring applicants may pursue short-term contract positions. Once the Ministry of Education makes the list of vacant permanent positions publicly available, applicants eligible for a permanent position choose an administrative region and submit a rank-ordered list (ROL) of schools within that region. The maximum length of the rank-ordered lists submitted by applicants was five in 2015 but became unrestricted from 2017 onward.

Applicants are then pre-screened and assigned to interview slots generating school-specific shortlists subject to year-specific capacities.[2] Prescreening is done via serial dictatorship: applicants are processed in order of their national test score rank and are assigned to their preferred schools among those with remaining available interview slots, subject to a maximum of two interviews per applicant in 2015 and three from 2017 onward. We refer to each school-year shortlist as a recruitment contest.

Each school then evaluates its shortlisted candidates. School committees are led by the school principal, and include the deputy principal (or a designated teacher) alongside another

[1] Bobba et al. (2021) provide a description of the assignment mechanism and the structure of compensation for teachers who apply for short-term positions.

[2] The maximum shortlist size was 20 candidates in 2015; in 2017 it was 10 candidates for single-vacancy contests and 15 for contests with multiple vacancies; from 2018 onward it was 10.

appointed teacher.[3] The committee sequentially deploys three screening devices: (CV) an evaluation of the applicant's curriculum vitae, (Inter) a personal interview assessing the candidate's overall fit with the school, and (Class) a live classroom demonstration session assessing pedagogical skills. The final ranking, which determines the sequence in which offers are made through the platform, integrates the discretionary screening scores and the standardized national test score.

## 2.2 Data Sources and Sample Composition

We combine several administrative datasets containing detailed information on the national recruitment drives for permanent, public-teacher positions for the years 2015, 2017, 2018, and 2019. We observe the universe of applicants and job postings, the rank-ordered lists of schools submitted by applicants, the set of short-listed candidates evaluated by each school and the candidate rankings at each stage of the decentralized evaluation. Our analysis focuses on public primary education, which offers the broadest geographic coverage of any schooling level in Peru.

We link this rich dataset on the recruitment process of permanent teachers to additional data sources. Employment records track all teachers and school principals in the Peruvian public school system from 2013 to 2019. This information allows us to identify the assigned teachers in any given school-year as well as who was the principal in that school-year. We additionally obtain applicants' residential locations from administrative household records, which we use to construct the distance between applicants' residence and schools. The employment histories also allow us to identify applicants already working at the recruiting school, whom we refer to as internal candidates. The school census provides us with detailed information on teacher composition, infrastructure, and access to school-level amenities such as electricity, water, internet, banking, and public libraries. Student academic performance is obtained from a national standardized assessment (Evaluación Censal de Estudiantes, ECE) that measures proficiency in mathematics and Spanish language. This evaluation provides individual test scores for second- and fourth-grade students attending public primary schools from 2014 to 2018. Importantly, we are able to link students across years and match them

[3] Alternative committee compositions exist for single-teacher or multi-grade schools, where the committee is led by the Head of the Pedagogical Management Area and includes the Director of the Educational Network (or a UGEL specialist) and a teacher from the candidate's level or modality. Single-teacher schools are excluded from our main analysis sample.

with their respective classroom and teacher identifiers through an administrative teacher-classroom dataset (Sistema de Información de Apoyo a la Gestión de la Institución Educativa, SIAGIE).

We construct three samples used in different parts of our empirical analysis. The *applicant choice sample* is used to estimate applicants' preferences over schools and construct the applicant types and control-function terms used to estimate teacher effectiveness. We retain applicants and vacancies for which the applicant, school, and geographic characteristics entering the choice model described in Section 3.2 are observed. The resulting sample contains 10,908 applicants observed over 15,400 applicant-year observations and 20,114 vacancies in 6,417 schools.

The *outcome sample* is used to estimate teacher effectiveness. We link teachers in the applicant choice sample to fourth-grade classrooms observed in the 2016 and 2018 student assessments. We retain students for which the achievement, demographic, classroom, and school characteristics included in the value-added model are observed as well as teachers with sufficient within-teacher variation in student characteristics to estimate the match-effect coefficients described in Section 3.3. The sample contains 28,283 students taught by 1,425 teachers, corresponding to 1,527 teacher-year observations in 1,273 schools.

Finally, the *interview sample* is used to study the effects of discretionary screening and estimate the hiring model described in Section 4. We retain shortlisted applicant-school pairs for which prior expected teacher effectiveness can be constructed. We exclude bilingual and single-teacher schools and retain school-year contests with at least two candidates after imposing these restrictions. The final sample contains 14,479 evaluated applicant-school pairs involving 7,062 applicants in 3,642 school-year contests and 2,874 schools.

The three samples are constructed for different empirical purposes and can thus differ in their composition. For example, the applicant choice sample includes applicants and vacancies that remain unmatched at the end of assignment mechanism, whereas the outcome sample necessarily requires teachers to be matched to a classroom for which student achievement is observed. Similarly, the interview sample only contains schools for which a shortlist of evaluated candidates is observed and thus excludes those that receive no applications. Differences in observable characteristics across the three samples thus reflect these sampling requirements rather than selected attrition. Appendix Table A.1 reports summary statistics for selected applicant and school characteristics across the three samples, together with the

corresponding sample sizes.

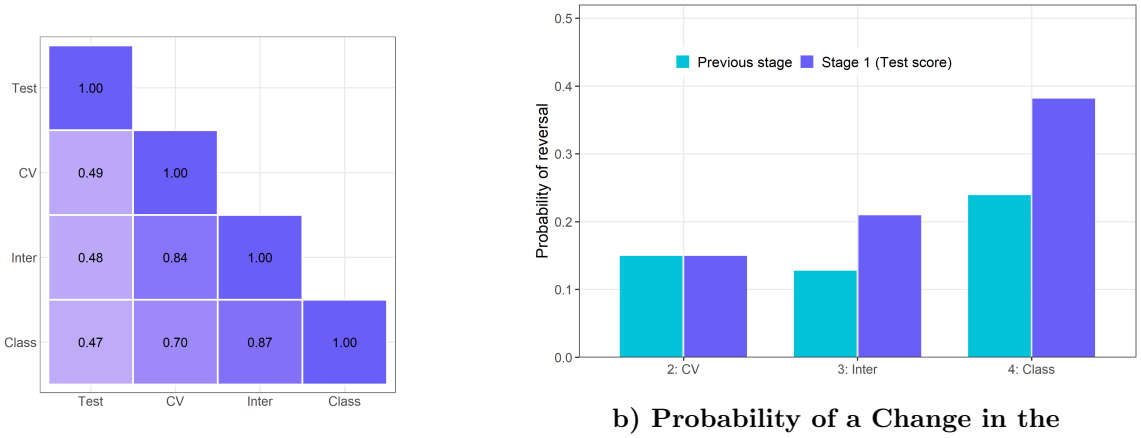
## 2.3 Discretion and the Interpretation of School Rankings

Our analysis uses the rankings generated by school committees as revealing how schools value candidates given the information available at each stage of the recruitment process. In particular, we distinguish in the rest of the analysis between a pre-interview ranking, based on the CV score assigned by the school, and a post-interview ranking, based on the cumulative score assigned across all three stages of discretionary evaluation—the CV review, interview, and classroom demonstration. We interpret the change between these rankings as reflecting schools' objective functions and the additional information acquired during the decentralized evaluation. This interpretation requires that the institutional scoring system leave schools enough latitude to reorder candidates to reflect their preferences. We provide evidence in this section that the resulting rankings do not reflect strategic behavior or mechanical features of the assignment mechanism.

Figure 1 shows that schools make substantial use of this discretion. Panel (a) display the rank correlation across candidates' rankings at successive stages of the decentralized evaluation. The pre-interview ranking is only moderately correlated with the centralized test-score ranking, and subsequent evaluations further change candidates' relative positions. Panel (b) tracks how often the identity of the top-ranked candidate under a cumulative score that adds the school-assigned scores to the centralized test score changes as the CV, interview, and classroom-observation scores are sequentially incorporated. The frequent changes in the leading candidate show that the discretion exercised by schools is consequential for the outcome of the selection process.

The variation in school rankings is also not simply a mechanical consequence of the scoring rubric. Table A.2 relates the CV score to the applicant characteristics observed by the committee, such as test scores, experience and gender. These characteristics explain part of the variation in CV scores, but leave substantial residual variation, indicating that committees use their discretion even at the first stage of the evaluation. Appendix Figure A.1 provides additional evidence ruling out strategic motives underlying schools' ranking decisions. Panel (a) shows that the scores assigned during the different stages of school evaluation are not systematically related to candidates' initial positions in the centralized test-score ranking. Panel (b) similarly shows that, comparing the same applicant across

**Figure 1:** The Role of School Committees in the Selection Process




**a) Rank correlation across stages**

**b) Probability of a Change in the Top-Ranked Candidate**

NOTES: Panel (a) reports pooled Pearson correlations between candidates' rank positions across stages of the school-level selection process. The test-score ranking is based on the centralized test score; the subsequent school rankings are based on the CV score, the cumulative CV and interview scores, and the cumulative CV, interview, and classroom-observation scores, respectively. Panel (b) reports the probability that the identity of the top-ranked candidate changes as successive school-evaluation components are incorporated into the official cumulative merit score. The merit scores at the four stages are the centralized test score; test score plus CV score; test score plus CV and interview scores; and test score plus CV, interview, and classroom-observation scores. The light bars compare each stage with the preceding stage, while the dark bars compare it with the initial test-score ranking. Exact ties in cumulative merit scores are broken first by the centralized test score and then using a common fixed tie-breaker.

schools, evaluation scores are largely unrelated to the applicant's position in competing shortlists. Thus, schools do not appear to adjust their evaluations strategically to preserve or overturn the initial ordering of candidates, or in response to candidates' prospects elsewhere.

Taken together, these patterns support treating the pre- and post-interview rankings as revealed preferences over candidates by schools under different information sets. We first use these rankings to study whether decentralized screening is effective at selecting higher value-added teachers (Section 3). We then leverage these rankings to estimate a model of hiring under uncertainty and learning about unobserved teacher effectiveness (Sections 4–5).

# 3 Discretionary Hiring and Teacher Effectiveness

## 3.1 Conceptual Framework

Our object of interest is the effect of alternative hiring rules on the achievement of the students who will be taught by the selected teacher. Consider a set of schools $i = 1, \ldots, N$ and a set of applicants $j = 1, \ldots, J$. Let $Y_{\ell j}$ denote the potential outcome of student $\ell$ if

taught by teacher $j$, and suppose

$$Y_{\ell j} = \alpha_j + X'_\ell \beta_j + Z'_\ell \gamma + \epsilon_{\ell j}, \tag{1}$$

where $E[\epsilon_{\ell j} \mid X_\ell, Z_\ell] = 0$ by construction. The vector $X_\ell$ contains student characteristics along which we allow teacher effectiveness to vary, while $Z_\ell$ contains additional student, classroom, and school controls. We normalize $X_\ell$ to have mean zero, so that $\alpha_j$ can be interpreted as the average treatment effect (ATE) of teacher $j$ and $\beta_j$ captures comparative advantage across students with different characteristics.

Let $L_i$ denote the population of students in school $i$. We define the expected school-level potential outcome associated with assigning teacher $j$ to school $i$ as

$$Y_{ij} \equiv E\left[\frac{1}{|L_i|}\sum_{\ell \in L_i} Y_{\ell j} \,\middle|\, \{X_\ell, Z_\ell\}_{\ell \in L_i}\right] = \alpha_j + X'_i \beta_j + Z'_i \gamma,$$

where $X_i$ and $Z_i$ denote school averages of the corresponding student characteristics. We define teacher $j$'s expected effectiveness in school $i$ as

$$q_{ij} \equiv \alpha_j + X'_i \beta_j. \tag{2}$$

Consider now a school contest $i$ with realized candidate set $J_i$. For any selection rule $r$, let $j_i(r)$ denote the candidate ranked first among the set $J_i$ under that rule. The corresponding expected school-level potential outcome is

$$Y_i(r) = Y_{ij_i(r)} = q_{ij_i(r)} + Z'_i \gamma. \tag{3}$$

We use random selection from the same realized candidate set as our benchmark selection rule. Under random selection, each candidate in $J_i$ is selected with equal probability, so the expected school-level outcome is

$$Y_i(0) = \frac{1}{|J_i|}\sum_{j \in J_i} Y_{ij} = \frac{1}{|J_i|}\sum_{j \in J_i} q_{ij} + Z'_i \gamma.$$

The treatment effect of selection rule $r$, relative to random selection, is therefore

$$\tau_i(r) \equiv Y_i(r) - Y_i(0) = q_{ij_i(r)} - \frac{1}{|J_i|} \sum_{j \in J_i} q_{ij}. \tag{4}$$

The gain in teacher effectiveness relative to random selection therefore measures the effect of the hiring rule on expected student achievement, holding fixed the realized candidate pool and the student population of the school.

Provided that $q_{ij}$ is known for every candidate in the shortlist, equation (4) identifies the treatment effect associated with each selection rule for every school contest. We can therefore identify both the average of these effects across contests and their cross-school distribution. Our data contain the full shortlist $J_i$ and the rankings generated at each stage of the recruitment process, so the candidate selected under the observed rules is known and the selected candidate under alternative rules can be constructed. The remaining empirical challenge is to recover $q_{ij}$ for the candidates in each shortlist from realized student outcomes.

We denote $S_\ell \in \{1, \ldots, J\}$ as the realized assignment of student $\ell$ to teachers and define $Y_\ell = \sum_j Y_{\ell j} \mathbf{1}\{S_\ell = j\}$ as the observed realized outcome of student $\ell$. We take 4th grade standardized math test scores at the national exam as our measure of achievement. The vector $X_\ell$ in equation (1) includes the student's lagged math test score (2nd grade) and gender. The vector $Z_\ell$ contains squared lagged math achievement, ethnolinguistic background, and age at the student level; school and classroom averages of lagged math achievement, gender, ethnolinguistic background, and age; classroom size and its square; and school-level controls for local amenities, travel time to the provincial capital, log population, and the interaction between travel time and log population.

Although $\epsilon_{\ell j}$ has mean zero before conditioning on the realized assignment, this may not be true once we condition on $S_\ell$. In particular, the realized teacher–school matches may be selected either because applicants may sort across schools on unobserved match-specific preferences that are correlated with teacher–school productivity. We address selection generated by applicants' choices over schools by modeling their application behavior and using their submitted rank-ordered lists to construct control functions.[4] We then estimate teacher effectiveness using the resulting selection correction.

[4] Abdulkadiroğlu et al. (2020); Bobba et al. (2021); Laverde et al. (2025) use similar empirical strategies to control for selection in teacher/school value-added estimates.

## 3.2 Applicant Behavior

We specify the indirect utility that applicant $j$ gets from being matched with school $i$ as

$$U_{ij} = u(z_j, x_i, d_{ij}, \tau_j) + \eta_{ij}, \tag{5}$$

where $u(\cdot)$ is a flexible function of observed applicant characteristics $z_j$, school characteristics $x_i$, distance between applicants and schools $d_{ij}$, an unobserved type $\tau_j = 1, \ldots, K$, and an i.i.d. type-I extreme value unobserved shock $\eta_{ij}$. We denote the outside option as $i = 0$, corresponding to the option value of participating in later rounds of the centralized admission platform allocating temporary teaching positions. We parameterize its systematic utility through an indicator whose coefficient varies with latent type, observed applicant characteristics, and recruitment years.

We consider two different identification strategies to recover $u(\cdot)$ from applicants' application behavior. The first uses the submitted rank-ordered lists (ROLs) assuming that applicant reveal their preferences truthfully. The second exploits the fact that the prescreening mechanism assigns interview slots through serial dictatorship. In particular, we assume that schools where applicants get shortlisted are their preferred options among those that still have available interviewing slots once higher scoring teachers have been assigned. This approach is akin to the stability-based identification strategy of Fack et al. (2019) extended to many-to-many matching and allows us to recover preferences without requiring applicants to rank schools they cannot reach.

**Approach 1: Truth-telling.** We first assume that applicants rank schools by order of preferences. Let $R_j$ denote applicant $j$'s ranking, where $iR_jk$ means that school $i$ is ranked above school $k$ by applicant $j$. We treat unranked schools as being ranked below the outside option 0. Conditional on $(z_j, d_j, \tau_j)$, the likelihood of observing $R_j$ is the standard exploded logit formula:

$$\mathcal{L}^{\text{TT}}(R_j \mid z_j, d_j, \tau_j) = \prod_{i:\, iR_j0 \text{ or } i=0} \frac{\exp\{u(z_j, x_i, d_{ij}, \tau_j)\}}{\exp\{u(z_j, x_i, d_{ij}, \tau_j)\} + \sum_{k:\, iR_jk} \exp\{u(z_j, x_k, d_{kj}, \tau_j)\}}, \tag{6}$$

where the denominator sums over the alternative at the current position and all alternatives ranked below it.[5]

**Approach 2: Stability of interview assignments.** We next account for the fact that applicants may strategically omit schools they regard as infeasible from their ROL. We consider an alternative identification strategy based on the stability of interview assignments under the centralized prescreening mechanism. Under this approach, applicants are assumed to be shortlisted at their preferred schools among those feasible for them, so that observed assignments are discrete choices from personalized feasible sets.

The prescreening mechanism allocates interview slots through serial dictatorship, with applicants considered in order of their national test scores and subject to limits on the number of interviews per applicant and school. Let $s_j$ be applicant $j$'s test score and $c_i$ the score of the last ranked shortlisted candidate at school $i$. We set the cutoff $c_i$ to zero when the school is undersubscribed. The set of schools feasible for applicant $j$ is

$$\mathcal{F}_j = \{i : s_j \geq c_i\} \cup \{0\}, \tag{7}$$

where 0 denotes the outside option. Let $A_j$ be the set of schools at which applicant $j$ obtains interviews. The assigned schools retain the relative ordering implied by the observed ROL, and every assigned school is preferred to the feasible schools not assigned to the applicant. If an applicant receives fewer interviews than the maximum permitted, the outside option is preferred to all remaining feasible schools; if the applicant reaches that maximum, we impose no restriction on the relative ordering of the outside option and schools not assigned to the applicant. Let $A_j^+$ denote the assigned schools augmented with the outside option when the applicant receives fewer interviews than permitted, and the assigned schools otherwise. Conditional on $\mathcal{F}_j$, the likelihood is

$$\mathcal{L}^{\mathrm{S}}(A_j \mid z_j, d_j, \tau_j, \mathcal{F}_j) = \prod_{i \in A_j^+} \frac{\exp\{u(z_j, x_i, d_{ij}, \tau_j)\}}{\exp\{u(z_j, x_i, d_{ij}, \tau_j)\} + \sum_{k \in \mathcal{F}_j : iR_jk} \exp\{u(z_j, x_k, d_{kj}, \tau_j)\}}, \tag{8}$$

where the denominator sums over the remaining feasible alternatives.

[5] Although applicants formally choose an administrative region before submitting a rank-ordered list of schools within that region, we do not condition the choice set on the realized chosen region. Schools outside that region enter the truth-telling likelihood as unranked alternatives.

We estimate $u(\cdot)$ under both approaches by maximizing the corresponding sample log-likelihood after integrating out the unobserved types $\tau_j$. We do so using the expectation–maximization (EM) algorithm with $K = 3$ and under the following specification for $u(\cdot)$:

$$u(z_j, x_i, d_{ij}, \tau_j) = z_j'\Theta_1 x_i + z_j'\Theta_2 d_{ij} + \Theta_{3,\tau_j}' x_i + \Theta_{4,\tau_j}' d_{ij},$$

where the vector $x_i$ contains the wage offered in school $i$, an amenity/infrastructure index,[6] indicators for whether school $i$ belongs to specific categories that are eligible for wage bonuses such as multi-grade, single-teacher, bilingual, and/or specific geographic areas, the locality's population and travel time (in hours) to the provincial capital, and school averages of student characteristics including their 2nd grade test scores in math and Spanish, gender, age, and ethnolinguistic background. The vector $d_{ij}$ contains linear splines of the geodesic distance between school $i$ and teacher $j$'s home location, and an indicator for whether teacher $j$ was working in school $i$ in the previous year. The vector $z_j$ contains an intercept, gender, competency score, indicators for experience in the public and private sector, and ethnolinguistic background. We report preference estimates in Table A.3 and Table A.4.

Note that this approach does not account for strategic behavior where applicants would omit schools for which they anticipate that, conditional on entering the interview, their probability of getting hired at the end of the decentralized evaluation is low. Appendix Figure A.2 shows that application probabilities respond only modestly to whether an applicant's test score would place her inside rather than outside a school's shortlist. There is also no evidence that application behavior depends on the predicted test-score rank within shortlists conditional on feasibility. Strategic behavior thus appears quantitatively limited and, if anything, takes the form of omitting schools that applicants are unlikely to get interviewed by, which is explicitly allowed by the stability-based estimator described here.

## 3.3 Measuring Teacher Effectiveness

We allow for selection on unobserved teachers' preferences for their matched school in the student outcome equation in the form of $\mathbb{E}[\epsilon_{\ell j}|X_\ell, Z_\ell, S_\ell = j, \eta_{i(\ell)j}] = \varphi(\eta_{i(\ell)j} - \mu_\eta)$ where $i(\ell)$ denotes the school where student $\ell$ is enrolled in and $\mu_\eta = \mathbb{E}[\eta_{i(\ell)j}]$. The parameter

[6] The amenity index is constructed by aggregating different indicators measuring the quality of infrastructure in the locality as well as an asset-based measure of poverty at the individual level that we aggregate at the locality level (which enters with a negative sign in the overall index) by PCA.

$\varphi$ captures linearly the relationship between unobserved applicant preferences and residual productivity at that school. Applying the law of iterated expectations, we can thus rewrite expected potential outcomes conditional on the realized teacher-school assignment as follows

$$\mathbb{E}[Y_{\ell j}|X_\ell, Z_\ell, S_\ell, R_j, z_j, d_j] = \alpha_j + X'_\ell\beta_j + Z'_\ell\gamma + \varphi\lambda_{i(\ell)}(R_j, z_j, d_j), \quad j = 1, ..., J, \tag{9}$$

where $\lambda_{i(\ell)}(R_j, z_j, d_j) := \mathbb{E}[\eta_{i(\ell)j} - \mu_\eta|R_j, z_j, d_j]$ are control function terms to be constructed.

Identification of $\varphi$ requires shifters of applicants' school choices that are excluded from the student achievement equation conditional on $X_\ell$ and $Z_\ell$. We use two excluded preference shifters: the distance between the applicant's residence and the school and the wage attached to the position. Distance to home has been previously used as a preference shifter in other school-choice and teacher-assignment settings (Walters, 2018; Abdulkadiroğlu et al., 2017; Bobba et al., 2021; Laverde et al., 2025). Wages are determined by a known compensation schedule based on predetermined school and locality characteristics that we flexibly control for such that residual wage variation comes from quasi-exogenous shifts at arbitrary thresholds of a location's population.[7]

Table A.5 assesses these exclusion restrictions using predetermined school outcomes that are not included in the teacher value-added (TVA) equation. We regress student dropout rates, end-of-year examination failure rates, and year-to-year teacher retention on distance and wages, conditional on the school-level covariates entering the outcome equation. The estimated coefficients are uniformly small, supporting the assumption that distance and wages affect student achievement mainly through applicants' school choices.

We construct the control-function terms following Abdulkadiroğlu et al. (2020), extending their rank-ordered logit selection correction to allow for unobserved applicant preference types. For each type, we compute the expected idiosyncratic preference shock of each ranked alternative conditional on the applicant's entire reported ROL. We then integrate these type-specific control functions over the posterior distribution of applicant types:

$$\mathbb{E}[\eta_{ij} - \mu_\eta|R_j, z_j, \boldsymbol{d_j}] = \sum_{k=1}^{K} P(\tau_j = k|z_j, \boldsymbol{d_j}, R_j)\mathbb{E}[\eta_{ij} - \mu_\eta|R_j, z_j, \boldsymbol{d_j}, \tau_j = k],$$

[7]See Bobba et al. (2021) for a detailed discussion on the determinants of compensation bonuses in this setting.

where $\pi_k = P(\tau_j = k)$ denotes the population probability of type $k$ and the posterior type probabilities are given by

$$P(\tau_j = k|z_j, \boldsymbol{d_j}, R_j) = \frac{\pi_k \mathcal{L}(R_j|z_j, \boldsymbol{d_j}, \tau_j = k)}{\sum_{l=1}^{K} \pi_l \mathcal{L}(R_j|z_j, \boldsymbol{d_j}, \tau_j = l)}. \tag{10}$$

Appendix B.1 provides details on the construction of the control-function terms for ranked and unranked alternatives.

Provided that there is sufficient within-teacher variation in $X_\ell$, a regression of $Y_\ell$ on teacher fixed effects, $X_\ell$ interacted with teacher fixed effects, $Z_\ell$ and the control function terms yields consistent estimates of $(\alpha_j, \beta_j, \gamma, \varphi)$ denoted as $(\hat{\alpha}_j, \hat{\beta}_j, \hat{\gamma}, \hat{\varphi})$.[8] The estimated control-function coefficients are consistently small and statistically insignificant under our truth-telling and stability-based estimation strategies: $\widehat{\varphi}^{TT} = 0.018$ and $\widehat{\varphi}^{S} = 0.022$ with standard errors of 0.014 and 0.021, respectively (Table A.6). These estimates imply that teacher selection on unobserved teacher-school match effects is negligible in our context.

We also consider a second source of potential selection: school committees may possess private information about match-specific productivity that is not captured by $q_{ij}$. To assess the importance of this channel, we average the residuals from the baseline TVA regression at the teacher–school–year level and regress them separately on the candidate's CV, interview, and classroom-observation scores, each standardized within the original school contest. Table A.7 shows that the estimated coefficients are uniformly small, providing limited evidence that school evaluations select candidates based on unobserved teacher-student determinants of student achievement.[9]

We then use these estimates to estimate the prior distribution of the average teacher effectiveness and comparative advantage parameters $\theta_j = (\alpha_j, \beta_j)$. Assuming that the value-added estimates are approximately normal $\hat{\theta}_j|\theta_j, \hat{\Omega}_j \sim \mathcal{N}(\theta_j, \hat{\Omega}_j)$, we specify a normal joint prior on $\theta_j$ conditional on $(z_j, \tau_j)$

$$\theta_j|z_j, \tau_j \sim \mathcal{N}(\mu_\theta(z_j, \tau_j), \Sigma_\theta). \tag{11}$$

[8] This approach does not address selection on unobservables generated by endogenous sorting of students across schools or classrooms. Addressing these additional margins would require separate sources of exogenous variation affecting students' school choices or within-school classroom assignments.

[9] Consistently with this evidence, Ederer (2026), in the same setting, estimates schools and teachers preferences and shows that controlling for two-sided endogenous selection has little effect on the resulting teacher value-added estimates.

We assume that latent types are unobserved both by the econometrician and the school but that the submitted ROLs $R_j$ are observed such that the prior distribution of $\theta_j$ conditional on $(z_j, \boldsymbol{d_j}, R_j)$ is

$$\theta_j | z_j, \boldsymbol{d_j}, R_j \sim \sum_{k=1}^{K} P(\tau_j = k | z_j, \boldsymbol{d_j}, R_j) \mathcal{N}(\mu_\theta(z_j, \tau_j = k), \Sigma_\theta).$$

We estimate the prior distribution by maximum likelihood (ML) after specifying $\mu_\theta(\cdot)$ as:

$$\mu_\theta(z_j, \tau_j) = \Lambda_1 z_j + \Lambda_{2,\tau_j}.$$

Finally, we construct shrunken estimates of $\theta_j$ denoted as $\theta_j^*$ corresponding to posterior beliefs about $\theta_j$ after observing the estimates $\hat{\theta}_j$

$$\hat{\theta}_j^* = \sum_{k=1}^{K} P(\tau_j = k | z_j, d_j, R_j)(\hat{\Omega}_j^{-1} + \hat{\Sigma}_\theta^{-1})^{-1}(\hat{\Omega}_j^{-1}\hat{\theta}_j + \hat{\Sigma}_\theta^{-1}\hat{\mu}_\theta(z_j, \tau_j = k)). \tag{12}$$

We keep the posterior type probabilities inferred from applicants' ROLs fixed when constructing these empirical-Bayes posterior means. Finally, for each applicant–school pair, we evaluate the shrunken coefficients at the student composition of school $i$ to obtain

$$q_{ij}^* = \widehat{\alpha}_j^* + X_i' \widehat{\beta}_j^*.$$

For comparisons of teacher effectiveness across schools, we also construct the relative value-added measure

$$\widetilde{q}_{ij}^* = q_{ij}^* - X_i' \bar{\beta} = \widehat{\alpha}_j^* + X_i'(\widehat{\beta}_j^* - \bar{\beta}),$$

where $\bar{\beta}$ is the mean of the teacher-specific slope coefficients in our estimation sample. This normalization removes the common contribution of student composition.

We assess the sensitivity of the TVA estimates to the identification strategy used to recover applicants' preferences. Although the truth-telling and stability-based specifications generate substantially different preference estimates, they yield nearly identical estimates of teacher effectiveness. Figure A.3 shows that the shrunken ATE estimates obtained under the two approaches lie almost exactly on the 45-degree line, with a correlation of 0.992. We therefore use the stability-based estimates throughout the remainder of the paper. This

**Table 1:** Distribution of Teacher Effectiveness

| | ATE | Lagged score | Female |
|---|---|---|---|
| *Panel A. Prior mean* | | | |
| Baseline ($\tau = 1$) | -0.036 | 0.525 | -0.057 |
| | (0.051) | (0.026) | (0.032) |
| Baseline ($\tau = 2$) | 0.005 | 0.593 | -0.046 |
| | (0.071) | (0.033) | (0.044) |
| Baseline ($\tau = 3$) | -0.122 | 0.539 | -0.059 |
| | (0.056) | (0.028) | (0.036) |
| Female | 0.057 | -0.032 | -0.045 |
| | (0.031) | (0.016) | (0.021) |
| Test score | 0.096 | 0.008 | -0.001 |
| | (0.024) | (0.012) | (0.014) |
| Quechua | 0.007 | -0.009 | 0.013 |
| | (0.050) | (0.022) | (0.035) |
| Exp public $> 0$ | 0.028 | -0.012 | 0.010 |
| | (0.037) | (0.018) | (0.022) |
| Exp private $> 0$ | 0.065 | 0.019 | -0.003 |
| | (0.030) | (0.015) | (0.020) |
| *Panel B. Prior variance* | | | |
| Standard deviation | 0.383 | 0.151 | 0.108 |
| | (0.010) | (0.007) | (0.015) |
| Correlation with ATE | | -0.075 | -0.066 |
| | | (0.053) | (0.107) |
| Correlation with lagged score | | | 0.076 |
| | | | (0.119) |

NOTES. This table reports estimates of the prior distribution of the teacher-effectiveness parameters $\theta_j = (\alpha_j, \beta_j)$ defined in Section 3.3. The columns correspond to average teacher effectiveness (ATE) and comparative advantage with respect to students' lagged achievement and gender, respectively. Panel A reports the estimated prior mean function $\mu_\theta(z_j, \tau_j)$. The first three rows report the type-specific baseline means for $\tau_j = 1, 2, 3$; the remaining rows report shifts in the prior mean associated with observed applicant characteristics. Panel B reports the standard deviations and pairwise correlations implied by the estimated prior variance-covariance matrix $\Sigma_\theta$. Standard errors are reported in parentheses.

choice is consequential for simulating applicant behavior under counterfactual prescreening mechanisms (Section 6), despite having little effect on the estimates of teacher effectiveness.

Panel A of Table 1 displays the estimates of the prior mean of $\theta_j$. Female, higher-scoring, and more experienced applicants have higher average TVA, while observable characteristics explain relatively little of the variation in comparative advantage. The latent preference types generate meaningful variation in average effectiveness, indicating that heterogeneity

in applicants' preferences over job attributes capture unobserved traits that are correlated with TVA (Bobba et al., 2021). Panel B reports the estimated prior variance-covariance matrix. Average effectiveness varies substantially across teachers even after conditioning on covariates, with a standard deviation of 0.383. The standard deviations of the teacher-specific coefficients on lagged achievement and female are 0.151 and 0.108, respectively. The correlations between average effectiveness and these two dimensions of comparative advantage are small.

## 3.4 Measuring the Gains from Discretion

Table 2 reports the average treatment effects defined in (4) for several ranking rules, together with the standard deviation of these effects across school contests. Test-score ranking generates a 0.010 standard-deviation gain in expected student achievement, while pre-interview school rankings generate a similarly small gain of 0.011 standard deviations. We next exploit the change between the pre- and post-interview rankings to measure the effect of learning during the decentralized screening stage. The gain relative to random selection increases to 0.025 standard deviations after the interview. The additional information acquired during discretionary evaluation therefore raises the expected student achievement generated by the selected candidate by 0.014 standard deviations on average. This provides supporting evidence that schools learn about candidates' latent teacher effectiveness. While the average productivity gain may seem small, it masks substantial heterogeneity across schools. The standard deviation of the contest-specific gains is 0.189 under the pre-interview ranking and 0.193 under the post-interview ranking, both orders of magnitude above the average effects.

Table 3 reports the corresponding selection effects on observed candidate characteristics. Test-score ranking generates a large gain in applicants' standardized test scores while having little effect on selection on other observed attributes such as gender, ethnicity, and experience. Learning also changes selection along other dimensions. For most observable characteristics, the selection effect moves closer to zero between the pre- and post-interview rankings. Two notable exceptions are gender and internal-candidate status. Relative to the pre-interview ranking, learning during the discretionary evaluation reduces the probability that the top-ranked candidate is female by 3.0 percentage points and increases the probability that the candidate is internal to the school by 2.8 percentage points. These changes may reflect biased interpretation of interview signals, differences in candidates' ability to generate favorable

**Table 2:** Effects of Alternative Screening Rules on Expected Student Achievement

| | Test score | Pre-interview | Post-interview | Prior TVA |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Mean gain in expected student 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) |

Notes. This table reports the mean and cross-contest standard deviation of the treatment effects $\tau_i(r)$ defined in equation (4) for alternative screening rules, relative to random selection from the same realized candidate set. For each contest and screening rule, the treatment effect is the expected effectiveness of the top-ranked candidate minus the average expected effectiveness of candidates in the same contest. The sample consists of school × year contests with at least two candidates for whom teacher value added can be estimated; rankings are recomputed within this sample. Standard errors are reported in parentheses and are clustered at the school level. Standard errors for the cross-contest standard deviations are computed analytically using the influence function for the standard deviation, with the same school-level clustering.

signals conditional on true latent quality, or information about other latent traits valued by schools.

The selection effects in Table 3 also show that schools' pre-interview rankings differ systematically from a rule that ranks candidates according to expected teacher effectiveness. Relative to actual pre-interview rankings, ranking candidates based on prior expected teacher effectiveness in Column (5) places substantially greater weight on female candidates, high-scoring candidates, and candidates with private-sector experience, and less weight on public-sector experience and internal-candidate status.

These patterns suggest that schools' objective functions may be misaligned with the objective of maximizing teacher effectiveness, but they may also reflect the fact that schools' beliefs about teacher effectiveness are not aligned with our estimated prior. However, Table 2 documents no evidence that pre-interview rankings (Column 2) select higher value-added candidates than rankings based on the estimated prior (Column 4), suggesting that deviations from the prior do not reflect superior productivity-relevant information. Across alternative priors constructed from increasingly coarse subsets of applicant characteristics, Table A.8 shows that pre-interview rankings remain weakly aligned with predicted teacher effectiveness. We also find similar misalignment among more experienced school principals, for whom biased or naive priors are less plausible (see Table A.9). Finally, the estimates reported in Table A.10 suggest that schools are not more likely to select high value-added internal candidates than low value-added internal candidates, even though internal candi-

**Table 3:** Selection Effects on Observed Candidate Characteristics

| | 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 |
| | | (0.005) | (0.005) | (0.005) | (0.005) |
| Test score (std.) | 1.653 | 0.334 | 0.017 | 0.012 | 0.158 |
| | | (0.004) | (0.005) | (0.005) | (0.005) |
| Quechua | 0.089 | 0.003 | 0.007 | 0.004 | 0.009 |
| | | (0.003) | (0.003) | (0.003) | (0.003) |
| Exp public $> 0$ | 0.900 | 0.013 | 0.047 | 0.039 | 0.032 |
| | | (0.004) | (0.003) | (0.003) | (0.004) |
| Exp private $> 0$ | 0.712 | 0.000 | -0.027 | -0.014 | 0.104 |
| | | (0.006) | (0.006) | (0.006) | (0.005) |
| Internal candidate | 0.112 | 0.012 | 0.012 | 0.040 | 0.006 |
| | | (0.004) | (0.004) | (0.004) | (0.004) |
| N. observations | 3,642 | 3,642 | 3,642 | 3,642 | 3,642 |
| N. schools | 2,874 | 2,874 | 2,874 | 2,874 | 2,874 |

Notes. This table reports how alternative screening rules affect the observed characteristics of the top-ranked candidate. Column (1) reports the expected characteristic under random selection from the realized candidate set, computed as the average characteristic among candidates in each school × year contest and then averaged across contests. Columns (2)–(5) report the characteristic of the top-ranked candidate under the indicated rule minus the corresponding within-contest random-selection benchmark, averaged across contests. Each contest receives equal weight. Standard errors in parentheses are clustered at the school level. The Prior TVA rule ranks candidates according to the estimated prior expectation of teacher effectiveness defined in equation (11).

dates are precisely the applicants about whom recruiters may have more information. Taken together, this evidence is difficult to reconcile with a purely information-driven explanation and suggests that schools likely pursue objectives other than maximizing our measured teacher effectiveness.

These findings establish that decentralized screening generates valuable information about teacher effectiveness, but does not consistently lead schools to select the most effective candidates. The returns to discretion vary substantially across schools, raising the possibility of allocating screening opportunities toward those that benefit most from decentralized evaluations. Assessing the potential gains from such policies requires understanding the sources of this heterogeneity. In particular, we need to distinguish differences in schools' ability to acquire productivity-relevant information from differences in how much they value teacher

effectiveness. We next develop a model of discretionary hiring and learning that separates these two channels.

# 4 A Model of Discretionary Hiring and Learning

## 4.1 Beliefs and Learning

We assume that schools start the hiring process by having rational expectations about the teacher effectiveness parameters $\theta_j = (\alpha_j, \beta_j)$ conditional on observable characteristics, $z_j$, and applicant $j$'s submitted ranking, $R_j$. This implies that schools face a common prior on the distribution of $q_{ij} := \alpha_j + \beta_j' X_i$. Given the distributional assumptions on $\theta_j$ posited in Section 3.3, we write the prior distribution of $q_{ij}$ as

$$q_{ij} \mid z_j, R_j, d_j, X_i \sim \sum_{k=1}^{K} P(\tau_j = k \mid z_j, d_j, R_j) \mathcal{N}\left(\mu_\alpha(z_j, \tau_j = k) + \mu_\beta(z_j, \tau_j = k)' X_i, \sigma_q^2(X_i)\right),$$

where the prior mean is

$$\mu_q(z_j, R_j, d_j, X_i) = \sum_{k=1}^{K} P(\tau_j = k \mid z_j, d_j, R_j) \left(\mu_\alpha(z_j, \tau_j = k) + \mu_\beta(z_j, \tau_j = k)' X_i\right)$$

and the within-type variance is

$$\sigma_q^2(X_i) = \sigma_\alpha^2 + 2\Sigma_{\alpha\beta} X_i + X_i' \Sigma_\beta X_i.$$

We account for the fact that schools may have additional information about internal candidates. Let $I_{ij}$ denote an indicator for candidate $j$ being internal to school $i$. We assume that schools update their priors about internal candidates towards our shrunken estimates of teacher effectiveness $q_{ij}^*$ by a factor of $\rho \in [0, 1]$. Prior to the interview, schools' beliefs are therefore

$$\begin{aligned} &\mathbb{E}[q_{ij} \mid z_j, R_j, d_j, X_i, I_{ij}] \\ &\quad = \mu_q(z_j, R_j, d_j, X_i) + \rho I_{ij} \left[q_{ij}^* - \mu_q(z_j, R_j, d_j, X_i)\right], \end{aligned}$$

with within-type variance

$$(1 - \rho I_{ij})\sigma_q^2(X_i).$$

During the interview and classroom demonstration, we assume that school committees receive a private signal

$$\nu_{ij}|\theta_j, X_i \sim \mathcal{N}(\alpha_j + X_i'\beta_j, \exp(\gamma_i))$$

about $q_{ij}$ and update their beliefs. We assume that $\gamma_i \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2)$ such that the parameter $\sigma_\gamma^2$ captures heterogeneity in schools' ability to extract informative signals.

We assume that schools update their beliefs about teacher effectiveness conditional on applicant types, keeping fixed the posterior probabilities of these types inferred from applicants' submitted rankings. The posterior mean of $q_{ij}$ can then be expressed as

$$\begin{aligned}
&\mathbb{E}[q_{ij} \mid z_j, R_j, d_j, X_i, I_{ij}, \nu_{ij}] \\
&\quad = \mathbb{E}[q_{ij} \mid z_j, R_j, d_j, X_i, I_{ij}] + \frac{(1 - \rho I_{ij})\sigma_q^2(X_i)}{(1 - \rho I_{ij})\sigma_q^2(X_i) + \exp(\gamma_i)} \left(\nu_{ij} - \mathbb{E}[q_{ij} \mid z_j, R_j, d_j, X_i, I_{ij}]\right).
\end{aligned}$$

## 4.2 Hiring Objectives

We define the expected value that school $i$ gets from hiring teacher $j$ given the information acquired until stage $g = 1, 2$ of the interviewing process as

$$\begin{aligned}
V_{ij}^g &= \exp(\delta_i)\mathbb{E}[q_{ij} \mid S_{ij}^g] + v^g(z_j, I_{ij}) + \xi_{ij} + \varepsilon_{ij}^g, \qquad (13) \\
S_{ij}^1 &= (z_j, d_j, R_j, X_i, I_{ij}), \qquad S_{ij}^2 = (z_j, d_j, R_j, X_i, I_{ij}, \nu_{ij}).
\end{aligned}$$

The first term captures the value that schools place on expected teacher effectiveness. The remaining terms capture alternative determinants of candidate rankings. We allow their observable component, $v^g(\cdot)$, to differ between the pre- and post-interview stages. This stage-specific variation may capture information revealed during the discretionary evaluation about candidate attributes other than teacher effectiveness, systematic differences in interview or classroom-demonstration performance across observable groups, biased evaluation associated with observable characteristics, or changes in how schools weight observable candidate characteristics after the interview. We do not attempt to distinguish among these mechanisms. Together with the stage-specific shocks $\epsilon_{ij}^g$, this allows rankings to flexibly change for reasons unrelated to learning about $q_{ij}$.

We specify

$$v^g(z_j, I_{ij}) = z_j'\psi_v^g + \psi_I^g I_{ij}, \qquad g = 1, 2,$$

and impose that standardized applicant test score does not enter $v^g(\cdot)$ at either stage.[10] We assess the validity of this exclusion restriction directly using the scores assigned by school committees during the CV, interview, and classroom-observation stages. Table A.11 shows that residual test-score variation has very little explanatory power for these evaluations and is quantitatively too small to affect strict pairwise candidate orderings.[11]

The parameter $\delta_i$ captures the importance of expected teacher effectiveness relative to these other factors. We assume that $\delta_i \sim N(\mu_\delta, \sigma_\delta^2)$, so that $\exp(\delta_i)$ is the positive weight placed on expected effectiveness. We assume that $\delta_i$ and $\gamma_i$ are independent in the population. However, they might be dependent in their posterior distribution once we condition on the observed rankings. To preserve this dependence, all simulation exercises that follow will draw $\delta_i$ and $\gamma_i$ from their posterior distributions.

Additional unobserved determinants of candidate rankings are captured through a persistent unobserved component $\xi_{ij}$ and a stage-specific type-I extreme-value shock $\epsilon_{ij}^g$, i.i.d. across $(i, j, g)$. The persistent component captures applicant–school attributes that are unobserved to the econometrician, while $\epsilon_{ij}^g$ captures transitory stage-specific heterogeneity. We allow the dispersion of the persistent component to differ between internal and external candidates:

$$\xi_{ij} \sim N\left(0, \sigma_{\xi, I_{ij}}^2\right).$$

This allows schools to have additional persistent information or preferences about internal candidates along dimensions that are not captured by measured teacher effectiveness.

[10] With a slight abuse of notation, we write $v^g(z_j, I_{ij})$ although the empirical specification also includes $K - 1$ posterior type probabilities $P(\tau_j = k \mid z_j, d_j, R_j)$ constructed in Section 3.3. We suppress these additional arguments to simplify notations.

[11] We residualize both applicant test scores and school evaluation scores with respect to prior expected teacher effectiveness, the remaining observables entering $v^g(\cdot)$, and school-year fixed effects. A one-standard-deviation increase in residualized test score is associated with a 0.020 standard-deviation increase in the CV score, 0.022 in the interview score, and 0.009 in the classroom-observation score. The corresponding partial $R^2$s are below 0.1 percent, and removing the fitted test-score component does not reverse any strict within-contest pairwise ordering.

## 4.3 Identification and Estimation

At each stage $g = 1, 2$, schools form rankings $R_i^g$ of the set of shortlisted candidates $J_i$. We assume that principals are not strategic and rank candidates in order of expected value $V_{ij}^g$.[12] We also assume that the realized set of candidates is exogenous conditional on the observables entering the model, meaning that the distribution of $(\epsilon_{ij}^g, \xi_{ij}, \delta_i, \gamma_i, \nu_{ij})$ conditional on $J_i$ coincide with their marginal distributions. The shortlist formation process is solely a function of applicants' reported preferences and test scores, which attenuates these concerns.[13] Violations of this assumption would require applicants' reported preferences to depend on private information about persistent unobserved determinants of schools' hiring behavior.

Under these assumptions, we can write the likelihood of observing $R_i^g$ conditional on $(\delta_i, \gamma_i, \xi_i, S_i^g)$ as

$$\mathcal{L}(R_i^g \mid \delta_i, \gamma_i, \xi_i, S_i^g) = \prod_{j \in J_i} \left( \frac{\exp\left\{\exp(\delta_i) E[q_{ij} \mid S_{ij}^g] + v^g(z_j, I_{ij}) + \xi_{ij}\right\}}{\sum_{k: j R_i^g k} \exp\left\{\exp(\delta_i) E[q_{ik} \mid S_{ik}^g] + v^g(z_k, I_{ik}) + \xi_{ik}\right\}} \right).$$

Identification of the hiring model relies on observing the same shortlist of candidates ranked before and after the interview. Pre-interview rankings reveal how schools trade off expected teacher effectiveness against other candidate attributes. The exclusion of test scores from $v^1(\cdot)$ provides variation in prior expected effectiveness that does not directly affect other components of preferences. The rankings of internal candidates relative to their estimated posterior value-added identifies $\rho$, while their average ranking advantage is captured by $\psi_{I1}$ and the dispersion of their persistent heterogeneity by $\sigma_{\xi,1}$.

Changes in rankings after the interview identify the learning technology. Conditional on prior expectations, how candidate rankings move relative to their measured teacher effectiveness identifies the precision of private signals. Allowing $v^2(\cdot)$ to differ from $v^1(\cdot)$ absorbs systematic changes associated with other observed attributes, while persistence in unexplained rankings across stages identifies the dispersion of $\xi_{ij}$. Appendix B.2 formalizes this

[12] As discussed in Section 2.3, this assumption is supported by the evidence shown in Figure A.1, which shows that school-assigned screening scores are largely insensitive to candidates' implied rankings in other shortlists.

[13] Applicants submit their school rankings before school-level screening takes place, and shortlists are then generated centrally based on these rankings and applicants' test scores. Applicants therefore cannot condition their applications on the subsequently realized interview signals or stage-specific ranking shocks.

argument for homogeneous $(\delta, \gamma)$, showing that a single cross section of rankings identifies $\delta$ up to utility scale and $\gamma$, without parametric restrictions on the distributions of $\xi_{ij}$ and $\varepsilon_{ij}^g$. Repeated observations of schools across recruitment contests then provide additional variation to identify the distributions of $\delta_i$ and $\gamma_i$ under our random-coefficients specification.

To simplify notation, we suppress the recruitment-year index. When a school is observed in multiple recruitment contests, we assume that its $(\delta_i, \gamma_i)$ are persistent across contests and construct the school-level likelihood by combining the ranking likelihoods from all of its observed contests.

We estimate the parameters

$$\left(\mu_\delta, \sigma_\delta^2, \mu_\gamma, \sigma_\gamma^2, \psi_v^1, \psi_v^2, \sigma_{\xi,0}, \sigma_{\xi,1}, \rho, \psi_I^1, \psi_I^2\right)$$

by maximizing the sample joint unconditional log-likelihood

$$\sum_{i=1}^{N} \log \int \prod_{g=1}^{2} \mathcal{L}\left(R_i^g \mid \delta_i, \gamma_i, \xi_i, S_i^g\right) \, dF(\delta_i, \gamma_i, \xi_i, \nu_i \mid \hat{\theta}).$$

We approximate this integral by simulation using 25,000 Monte Carlo draws. For each draw, we sample the school's weight on effectiveness $(\delta_i)$, signal-noise parameter $(\gamma_i)$, persistent unobserved component $(\xi_{ij})$, and interview signals $(\nu_{ij})$ from their respective distributions. Signals are drawn from their posterior distribution conditional on the estimated teacher value-added parameters $(\hat{\theta}_j)$ accounting for estimation noise.

# 5 Estimation Results

## 5.1 Model Estimates

Panel A of Table 4 reports the estimated parameters governing heterogeneity in schools' preferences and signal-extraction ability, while Panel B reports implied moments of these distributions. Additional estimates of the hiring model are reported in Appendix Table A.12.

The implied mean of $\exp(\delta_i)$ is 2.38 and its standard deviation is 3.78, indicating substantial cross-school heterogeneity in the weight placed on expected teacher effectiveness. Panel (a) of Figure 2 reports, by decile of $\delta_i$, the within-contest rank correlation between the expected teacher effectiveness component of utility and overall pre-interview hiring utility.

**Table 4:** Preferences for Teacher Effectiveness and Signal Extraction

| | Estimate | Std. Error |
|---|---|---|
| *Panel A. Estimated distribution parameters* | | |
| Mean log weight on effectiveness, $\mu_\delta$ | 0.237 | (0.243) |
| SD log weight on effectiveness, $\sigma_\delta$ | 1.122 | (0.139) |
| Mean log signal variance, $\mu_\gamma$ | -0.059 | (0.524) |
| SD log signal variance, $\sigma_\gamma$ | 2.081 | (0.878) |
| *Panel B. Implied moments* | | |
| Mean weight on effectiveness, $E[\exp(\delta_i)]$ | 2.378 | (0.350) |
| SD weight on effectiveness, SD$[\exp(\delta_i)]$ | 3.774 | (0.847) |
| Mean weight on interview signal | 0.234 | (0.067) |
| SD weight on interview signal | 0.259 | (0.078) |

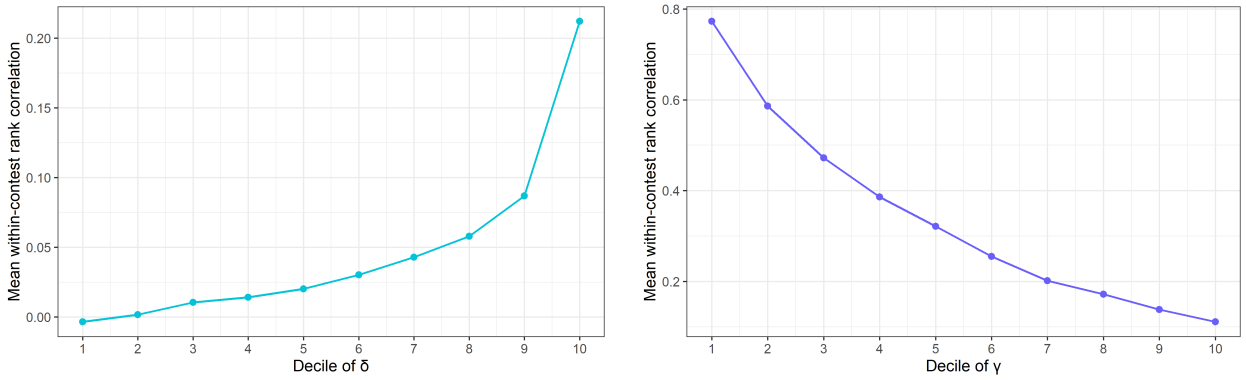
NOTES. Panel A reports the estimated parameters governing the population distributions of schools' log weight on expected teacher effectiveness $\delta_i$ and log interview-signal variance $\gamma_i$. Panel B reports moments implied by these distributions. The interview-signal weight for candidate $j$ at school $i$ is $\sigma^2_{q,ij}/[\sigma^2_{q,ij} + \exp(\gamma_i)]$, where $\sigma^2_{q,ij}$ is the remaining uncertainty about teacher effectiveness before the interview. Signal weights are averaged across candidates within schools before computing population moments. Standard errors for Panel B are obtained by the delta method; for signal weights, they account for estimation uncertainty in $(\mu_\gamma, \sigma_\gamma)$, holding the prior-variance distribution and other parameters fixed.

These correlations are uniformly small but highly heterogeneous. Among schools in the top decile of $\delta_i$, the rank correlation reaches 0.21, while it is close to zero among schools placing the least weight on teacher effectiveness.

The estimates also point to limited but highly heterogeneous signal extraction ability. The average weight placed on the interview signal in the posterior mean of teacher effectiveness is 0.23, with a standard deviation of 0.26. Panel (b) of Figure 2 translates the underlying heterogeneity in $\gamma_i$ into the implied accuracy of schools' posterior beliefs by reporting the rank correlation between post-interview expected teacher effectiveness and underlying teacher effectiveness by decile of $\gamma_i$. Schools receiving the most precise signals substantially improve their information about candidate productivity, with correlations as high as 0.76 while schools with the least precise signals show correlations as low as 0.11. Thus, the modest average gains from discretion documented in Table 2 mask substantial heterogeneity in schools' ability to extract productivity-relevant information during the screening process.

The additional estimates in Appendix Table A.12 show that schools favor internal candidates more strongly after the interview, despite no evidence that they possess additional pre-interview information about these candidates' teacher effectiveness. Female applicants also receive a negative weight post-interview. Persistent unobserved applicant–school at-

**Figure 2:** Rank Correlations and Heterogeneity in Preferences and Signal Extraction




**a) Expected Effectiveness and Pre-Interview Utility**

**b) Post-Interview Expected Effectiveness and underlying teacher effectiveness $q_{ij}$**

NOTES: Panel A reports, by decile of $\delta_i$, the Spearman rank correlation between the expected-effectiveness component of pre-interview utility and total pre-interview utility. Panel B reports, by decile of $\gamma_i$, the Spearman rank correlation between post-interview expected teacher effectiveness and underlying teacher effectiveness. Correlations are first calculated within school-year contests and then averaged within schools, giving each school equal weight. Schools are assigned to deciles separately for each draw from the posterior distribution, and the plotted values average the resulting decile-specific correlations across draws. Higher $\delta_i$ corresponds to a greater weight on expected teacher effectiveness; lower $\gamma_i$ corresponds to a more precise interview signal.

tributes also account for substantial variation in rankings.

Taken together, the estimates suggest that the modest average effectiveness gains from discretion documented in Section 3 reflect both limited alignment of school rankings with expected teacher effectiveness and limited productivity information extracted during screening. At the same time, substantial heterogeneity in both dimensions implies that these averages conceal large differences across schools in how effectively discretion is used.

## 5.2 Model Fit

We next assess whether the estimated model reproduces the empirical patterns associated with the two channels at the center of our analysis: imperfect and heterogeneous learning about teacher effectiveness, and heterogeneous hiring motives. We first ask whether the model captures the persistence and reshuffling of candidate rankings across stages, and whether the resulting changes in selection are related to teacher effectiveness consistently with the data. We then examine whether it reproduces systematic changes in selection along observable candidate characteristics, which help distinguish productivity-related selection from changes along other dimensions valued by schools. All model predictions are averaged across simulations of $(\epsilon^g_{ij}, \xi_{ij}, \delta_i, \gamma_i, \nu_{ij})$, with $(\delta_i, \gamma_i)$ drawn jointly from their posterior

distribution.

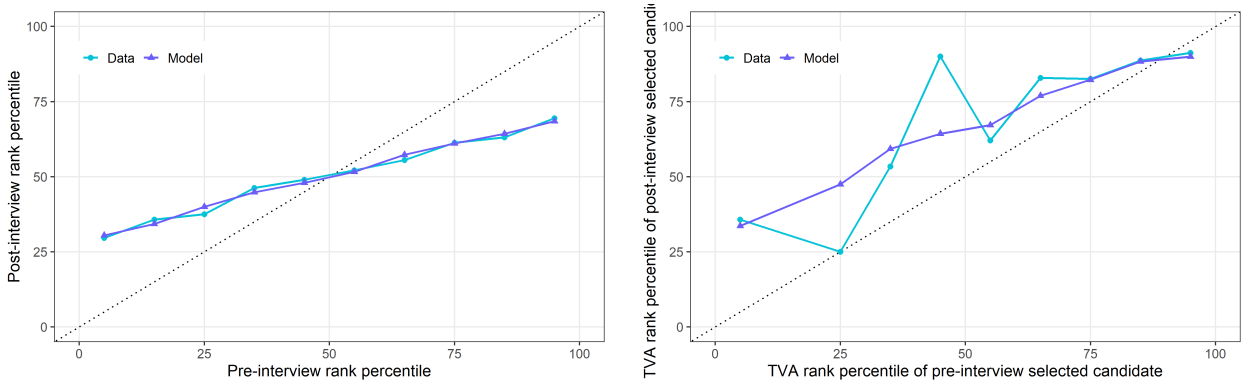
Panel A of Figure 3 compares the observed and simulated evolution of candidate rankings between the pre- and post-interview stages. The model closely replicates the observed relationship between average post-interview rank percentiles against pre-interview rank percentiles. Additionally, Table A.13 shows that the model closely matches the fraction of contests in which rankings change, the share of candidates who change rank, and the correlation between pre- and post-interview rankings.

Panel B assesses whether the model replicates evidence of learning about teacher effectiveness. It computes the relationship between the TVA percentile rank of the post-interview top-ranked candidate and of the pre-interview top-ranked candidate and shows that the model reproduces the average trend of its observed counterpart, despite the data series looking noisier. Table A.14 provides a complementary comparison using the effects of the pre- and post- interview selection rules on expected student achievement reported in Table 2. The model predicts gains of 0.002 and 0.011 standard deviations under the pre- and post-interview rankings, compared with 0.011 and 0.025 in the data. While the model slightly under predicts the selection gains of each rule, it captures most of the additional gain associated with learning during the discretionary evaluation, as the observed gain rises by 0.014 standard deviations between the two stages, compared with 0.009 in the model. It also reproduces the dispersion across schools in these gains very closely.

Finally, Table 5 assesses whether the model reproduces the observed changes in selection across candidate characteristics. The model reproduces closely the decline in the selection of female candidates and the increase in the selection of internal candidates after the discretionary evaluation, as well as the changes in selection on public- and private-sector experience, test scores, and ethnolinguistic background.

Taken together, these exercises show that the estimated model closely reproduces the main empirical patterns generated by signal extraction and schools preferences. Specifically, it captures how candidate rankings change across stages, the productivity improvements associated with learning during the interview, as well as changes in selection along other teacher characteristics. We next use the estimated model to quantify how heterogeneity in schools' objectives and in their ability to extract productivity information contributes to the large dispersion in the gains from discretionary hiring documented in Section 3.

**Figure 3:** Model Fit: Ranking Changes and Selection on Teacher Effectiveness




**a) Pre- and Post-Interview Rank Percentiles** **b) TVA Rank of Selected Candidates**

NOTES: Panel A compares pre- and post-interview candidate rankings in the data and their model-simulated counterparts. Rankings are transformed into within-contest percentiles, with 100 denoting the top-ranked candidate and 0 the bottom-ranked candidate. The pre-interview percentile is partitioned ex ante into ten intervals, $[0, 10), [10, 20), \ldots, [90, 100]$, and each point reports the average post-interview percentile within the corresponding interval. Candidates receive weight equal to the inverse of contest size, so each contest receives equal total weight. Panel B plots the shrunken-TVA percentile rank of the post-interview selected candidate against that of the pre-interview selected candidate. It is restricted to contests with at least two candidates for whom TVA can be estimated. TVA ranks are constructed among candidates with observed TVA, while the percentile normalization uses the full number of candidates in the contest. Model curves average across model simulations. The dotted line is the 45-degree line.

**Table 5:** Model Fit: Characteristics of Top-Ranked Candidates

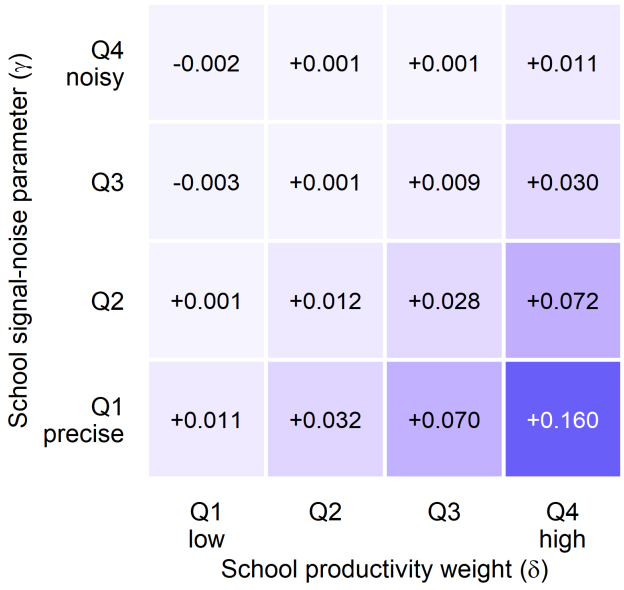
| | Random | Pre-interview | | Post-interview | |
|---|---|---|---|---|---|
| | | Data | Model | Data | Model |
| Female | 0.802 | 0.006 | 0.010 | -0.024 | -0.023 |
| Test score (std.) | 1.653 | 0.017 | 0.012 | 0.012 | 0.011 |
| Quechua | 0.089 | 0.007 | 0.006 | 0.004 | 0.007 |
| Exp public $> 0$ | 0.900 | 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 |

Notes. The table compares observed candidate selection with the corresponding model predictions. The Random column reports the expected characteristic under random selection from the realized candidate set. The pre- and post-interview columns report the characteristic of the top-ranked candidate minus the corresponding candidate-set average, separately for the data and the model. Each contest receives equal weight. All rows use the full structural-estimation sample. Model moments are averaged across model simulations.

## 5.3 Unpacking Aggregate Gains from Discretionary Hiring

We evaluate the gains from discretionary hiring using draws of underlying teacher effectiveness from its posterior distribution, rather than the shrunken empirical-Bayes estimates used in Section 3. For each draw, we compute the selection effects defined in (4). Figure 4

**Figure 4:** Heterogeneity in Gains from Discretionary Evaluation




NOTES: The figure reports the average gain in teacher value-added (TVA) from discretionary evaluation relative to random selection from the same candidate pool, by quartiles of schools' productivity weights ($\delta_i$) and signal-noise parameters ($\gamma_i$). Gains are evaluated using draws of underlying teacher effectiveness from its posterior distribution. Schools are assigned to quartiles separately for each draw of their latent $(\delta_i, \gamma_i)$. For schools observed in multiple contests, selection gains are first averaged across contests so that each school receives equal weight. Each cell reports the average across schools and posterior-predictive simulations. Higher ($\delta_i$) corresponds to a greater weight on teacher effectiveness, while higher ($\gamma_i$) corresponds to noisier private signals and lower signal-extraction ability.

averages these gains across draws, while the variance decomposition in Table 6 holds teacher effectiveness and the remaining shocks fixed to isolate variation generated by school-specific preferences and signal-extraction ability.

Our estimates imply that heterogeneity in signal extraction ability and schools preferences generate substantial heterogeneity in the gains from discretion. Figure 4 reports the average treatment effect $\tau_i(r)$ of the post-interview ranking rule, evaluated using draws of the underlying teacher effectiveness $q_{ij}$, by the joint quartiles of $\delta_i$ and $\gamma_i$. Selection gains are concentrated among schools that place greater weight on teacher effectiveness and extract more precise signals. At the intersection of the highest quartile of $\delta_i$ and the lowest quartile of $\gamma_i$, average gains reach up to 0.16 standard deviations, while more than half of the distribution generate worse or equal gains relative to test-based hiring.

Table 6 quantifies the extent to which this heterogeneity accounts for cross-sectional variation in screening performance. We cross draws from the estimated joint distribution of $(\delta_i, \gamma_i)$ with the observed applicant pools and apply the law of total variance. Differences in applicant pools account for approximately three-quarters of the total variance in selec-

**Table 6:** Variance Decomposition of Selection Gains

| Source of variation | Variance | Share of total |
|---|---|---|
| Total variance | 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 | 0.014 | 0.170 |
| Heterogeneity in $\gamma_i$ only | 0.010 | 0.114 |

Notes. The table reports a variance decomposition of the gains in teacher value-added from discretionary evaluation relative to random selection from the same candidate pool. To isolate variation generated by school objectives and signal-extraction ability, we fix one realization of underlying teacher effectiveness, persistent applicant–school heterogeneity, interview innovations, and post-interview preference shocks. We then draw joint school types $(\delta_i, \gamma_i)$ from the estimated distribution and evaluate each type against every observed school environment. The school-environment component is the variance across environments of average selection gains, while the joint school-type component is the average within-environment variance across types. The last two rows report the contributions from varying $(\delta_i)$ and $(\gamma_i)$ separately, holding the other parameter at its median. These contributions need not sum to the joint contribution because the two dimensions of heterogeneity interact in determining selection outcomes.

tion gains, while heterogeneity in $(\delta_i, \gamma_i)$ accounts for the remaining quarter. Varying $\delta_i$ alone accounts for 17 percent of the total variance, while varying $\gamma_i$ alone accounts for 11.4 percent.[14]

Importantly, this heterogeneity is difficult to target using observable school characteristics. Tables A.15 and A.16 show that student composition, urban status, principals' experience, and principals' standardized test scores explain almost none of the variation in the posterior mean of $\delta_i$ or $\gamma_i$. This limits the scope for efficiently allocating screening capacity across schools using observable characteristics. Section 6 first quantifies the potential gains from allocating discretion according to schools' signal-extraction ability, before examining whether centralized priorities can instead direct screening opportunities toward more productive applicants.

[14]These quantities do not sum to the joint contribution because the two dimensions interact nonlinearly in determining which candidate is selected and may also be dependent in their posterior distribution.

# 6 Counterfactuals

## 6.1 Targeting Screening Capacity

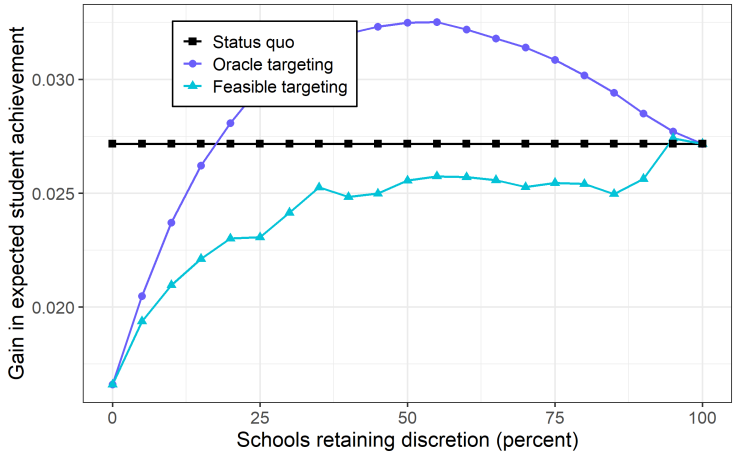
The heterogeneity in signal-extraction ability documented in Section 5 suggests that restricting discretion to schools that extract more precise signals can improve teacher selection. We consider policies that retain the status quo discretionary evaluation in a fraction of schools with signal extraction ability above a given threshold, while the remaining schools select the candidate with the highest centralized test score. We rank schools by their signal-noise parameter $\gamma_i$, with lower values corresponding to more precise signals, and vary the fraction retaining discretion from zero to one in increments of five percentage points.

We consider two targeting rules. Under *oracle targeting*, the policymaker observes each school's true $\gamma_i$. Under *feasible targeting*, schools are ranked by the posterior mean of $\gamma_i$ estimated from their observed hiring decisions. At each cutoff, both rules delegate hiring to the same number of schools. The endpoints correspond to test-score selection in every school and universal discretion, respectively.

Figure 5 shows that selective delegation can improve selection when schools' signal-extraction ability is perfectly observed, but the gains are modest. The hump shape is explained by the declining marginal value of discretion as delegation expands to schools with lower signal-extraction ability. For schools with more precise signals, the information acquired during screening generates selection gains relative to test-based selection. As signal precision declines, this informational advantage weakens. For these marginal schools, test-based hiring performs better, so average selection gains peak when discretion is retained in the half of schools with the highest signal-extraction ability, at 0.033 standard deviations compared with 0.027 under universal discretion.

These improvements largely disappear under feasible targeting, which performs worse than universal discretion for all delegation rates. Even under oracle targeting, the gains are modest. Thus, allocating decentralized screening across schools according to their signal-extraction ability offers limited scope for improving teacher selection.

**Figure 5:** Targeting Screening Capacity




NOTES: The figure reports average selection gains in expected student achievement relative to random selection from the same candidate pool, as a function of the percentage of schools retaining discretion. Schools without discretion select the candidate with the highest centralized test score. Under oracle targeting, schools are ranked by their true signal-noise parameter $\gamma_i$, while feasible targeting uses their estimated posterior mean. In both cases, schools with lower $\gamma_i$ retain discretion first. The horizontal line represents the status quo, where all schools retain discretion.

## 6.2 Designing Prescreening

We next turn to a different margin of mechanism design: how the centralized platform allocates screening opportunities across applicants. We ask whether these priorities matter for teacher selection, and examine whether more informative measures of productivity should be used for centralized prescreening or disclosed directly to schools.[15]

We compare three centralized priority rules: random priorities, the existing test-score priorities, and priorities based on posterior average teacher effectiveness (ATE). For each rule, we vary the maximum shortlist length from the status quo to $L \in \{5, 3, 1\}$, allowing schools to shortlist at least as many candidates as they have vacancies. We also consider two information regimes. Without disclosure, schools evaluate their shortlisted applicants and acquire private productivity signals with school-specific precision ($\gamma_i$). With TVA disclosure, schools additionally observe estimates of teacher effectiveness before their interviews and update their beliefs using both these estimates and their private signals, retaining discretion over their rankings. For applicants without observed TVA estimates, we generate synthetic estimates of underlying effectiveness and construct the corresponding posterior beliefs. These

[15] We simulate final hiring outcomes within the restricted market consisting of the schools and applicants in our interview sample, rather than the full national market. Final hires are simulated using school-proposing deferred acceptance based on schools' post-interview rankings. We therefore abstract from competition with schools and applicants outside the estimation sample and from the institutional rule that incorporates national test scores in the final ranking.

**Table 7:** Centralized Prescreening and Information Disclosure

| Shortlist capacity | Interviews | | | TVA disclosed + interviews | | |
|---|---|---|---|---|---|---|
| | 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 |

Notes. The table reports outcomes under alternative centralized priority rules (test score, random, and teacher value added), maximum shortlist capacities, and information regimes. In both information regimes, schools retain discretionary interviews; under TVA disclosure, they additionally observe estimates of teacher effectiveness before evaluating candidates. Status quo refers to existing institutional shortlist capacities, while $L$ denotes the maximum shortlist length, subject to a minimum equal to the number of advertised vacancies. Panel A reports the mean value added (TVA) of assigned teachers, Panel B the share of unfilled vacancies, and Panel C the cross-school variance of average assigned TVA among schools with at least one hire. All outcomes are computed within the restricted simulation market, excluding competition from schools and applicants outside the interview sample. Final assignments use schools' discretionary rankings rather than the official composite ranking including test scores. The value-added statistics use relative teacher effectiveness, net of the common contribution of student composition. Results are averaged across posterior-predictive simulations.

productivity estimates also underpin the ATE-based priority rule.

We begin by establishing that centralized prescreening matters. Panel A of Table 7 compares random priorities with the existing test-score rule. Under the status quo shortlist capacities, the two rules lead to almost identical average teacher effectiveness, at 0.181 and 0.180 standard deviations, respectively. When screening opportunities become more restrictive, however, their outcomes diverge. At $L = 1$, average effectiveness falls to 0.163 under random priorities but remains at 0.182 under test-score priorities. Thus, the information used to allocate screening opportunities matters once capacity is scarce. At the same time, making test-score-based prescreening more restrictive yields virtually no improvement in the effectiveness of hires, even though test scores are positively correlated with teacher value added.

We next consider whether using more accurate productivity estimates for centralized

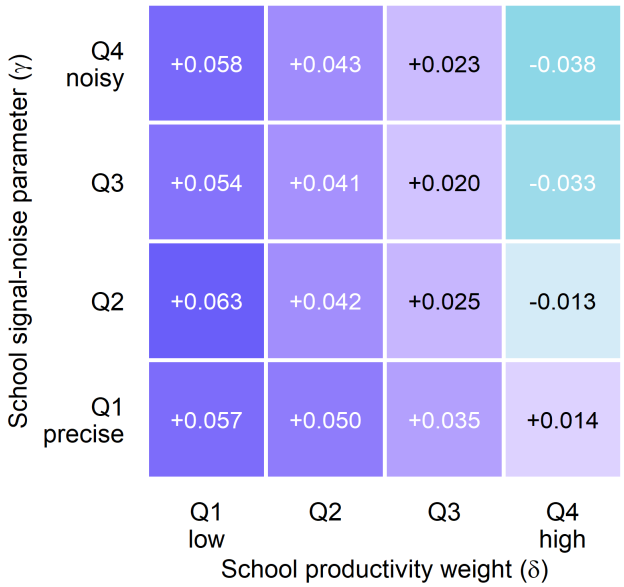
priorities can improve selection. Under existing shortlist capacities, replacing test-score priorities with ATE-based priorities has essentially no effect on average effectiveness. As capacity is restricted, however, the gains increase sharply. At $L = 5$, average effectiveness rises to 0.220 under TVA priorities, compared with 0.182 under test-score priorities; the differences reach 0.102 standard deviations at $L = 3$ and 0.158 at $L = 1$. Whereas test-score priorities leave the effectiveness of hires nearly unchanged as screening capacity contracts, more informative centralized priorities direct the remaining screening opportunities toward more effective applicants.

When more accurate productivity information becomes available, however, it is not obvious that the central platform should use it for prescreening. It could instead disclose that information directly to schools and let them combine it with what they learn during their decentralized evaluations. Under the status quo shortlist capacities and test-score priorities, TVA disclosure increases average effectiveness from 0.180 to 0.199 standard deviations, even though schools retain their private interview signals and discretionary rankings. Its returns diminish as screening capacity is restricted: disclosure raises effectiveness from 0.182 to 0.194 at $L = 5$, and from 0.181 to 0.187 at $L = 3$. In comparison, ATE-based prescreening with interviews raises effectiveness to 0.220 at $L = 5$, exceeding the gains from disclosure under the existing mechanism. Combining TVA priorities with disclosure produces only a modest additional improvement, from 0.220 to 0.228. Using information about teacher effectiveness for prescreening seems thus more effective than disclosing this information directly to schools.

The sources of heterogeneity identified in Section 5 help explain why the two interventions perform differently across schools. Figure 6 compares ATE-based priorities with $L = 5$ and decentralized interviews against TVA disclosure under the existing test-score priorities and shortlist capacities, retaining interviews in both regimes. ATE-based prescreening performs particularly well for schools that place relatively little weight on teacher effectiveness, regardless of how precisely they extract signals. Schools in the top quantile of the distribution of $\delta_i$ mostly benefit more from information disclosure, particularly when their signal extraction ability is low. Overall, the relative effectiveness of information disclosure and prescreening crucially depends on whether schools pursue alternative objectives and on the precision of the signals they extract through their own discretionary evaluations.

Importantly, the gains from using ATE-based priorities increase as the length of the short-

**Figure 6:** Heterogeneity in the Effects of Centralized Prescreening and Information Disclosure




NOTES: The figure compares (i) ATE-based centralized priorities with $L = 5$ and decentralized interviews, and (ii) standardized TVA disclosure before decentralized interviews under existing test-score priorities and shortlist capacities. Under both policies, schools retain discretionary evaluations and their private interview signals. Each cell reports the difference in average assigned teacher value added between (i) and (ii), by quartile of the school-specific weight on expected teacher effectiveness ($\delta_i$) and interview-signal noise ($\gamma_i$). Positive values favor ATE-based centralized prescreening. Quartiles are formed within each posterior-predictive draw using the latent school types governing hiring decisions in that draw. Policy effects use the same underlying teacher-productivity draws, school types, and private interview signals, and are computed for school-years with successful hires under both policies. Effects are averaged first across school-years within a school and draw, then across schools and draws. Higher $\delta_i$ denotes greater weight on effectiveness; higher $\gamma_i$ denotes noisier signals and lower signal-extraction ability.

list decreases, but this has adverse consequences on the number of matches formed. Panel B of Table 7 shows that ATE-based priorities with $L = 5$ raise average teacher effectiveness from 0.180 to 0.220 while reducing the share of unfilled vacancies from 37.5 to 35.9 percent. Further reducing $L$ improves the effectiveness of hired teachers but leaves substantially more vacancies unfilled. The share of unfilled vacancies rises to 43.0 percent at $L = 3$ and 56.9 percent at $L = 1$. With more aggressive shortlisting, schools increasingly compete for the same candidates. The resulting congestion limits the number of successful matches.

Finally, Panel C of Table 7 shows that improvements in average effectiveness do not translate into comparable reductions in disparities across schools. The cross-school variance of assigned TVA remains close to its status quo level of 0.117 across the counterfactual policies.

Taken together, these results show that using information available to the centralized platform to allocate screening opportunities can generate large gains and dominates revealing

this information directly to schools, particularly when school objectives are not aligned with productivity. More aggressive prescreening increases the effectiveness of the matches formed, but eventually reduces the flexibility needed to resolve competition for candidates and leaves more vacancies unfilled. The relevant design trade-off is therefore between directing scarce screening capacity toward better potential matches and preserving enough opportunities for those matches to be realized.

# 7 Conclusion

We study how centralized assignment mechanisms should organize decentralized information acquisition, using Peru's teacher recruitment system. We show that decentralized screening generates productivity-relevant information and improves teacher selection on average, but its returns vary substantially across schools. We estimate a model of hiring under uncertainty that separates schools' preferences from their ability to extract informative signals. Both dimensions contribute to the heterogeneity in returns to discretion, but are difficult to predict from observable school characteristics. Allocating screening capacity according to schools' signal-extraction ability therefore offers limited scope for improving selection.

We then investigate how centralized mechanisms can instead improve the allocation of screening opportunities across applicants. We show that centralized priorities matter for the effectiveness of hires, but that more restrictive prescreening based on imperfect proxies for productivity generates little improvement. When more informative productivity measures become available, using them to allocate screening opportunities can generate larger gains than disclosing them directly to schools, even when schools retain their ability to acquire private information. These gains are particularly large when schools' objectives are poorly aligned with productivity. However, concentrating screening opportunities too narrowly leaves schools with fewer alternatives when competing for the same applicants, reducing the number of successful matches.

Our findings extend beyond teacher recruitment to other centralized matching markets where acquiring information about match quality is valuable but costly, such as college admissions and public-sector hiring. More broadly, our results show that the allocation of scarce opportunities for decentralized information acquisition is itself an important dimension of market design.

# References


**Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A Pathak**, "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match," *American Economic Review*, 2017, *107* (12), 3635–89.

**_ , Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters**, "Do Parents Value School Effectiveness?," *American Economic Review*, 2020, *110* (5), 1502–1539.

**Altonji, Joseph G and Charles R Pierret**, "Employer learning and statistical discrimination," *The Quarterly Journal of Economics*, 2001, *116* (1), 313–350.

**Artemov, Georgy**, "Assignment mechanisms: Common preferences and information acquisition," *Journal of Economic Theory*, 2021, *198*, 105370.

**Autor, David H and David Scarborough**, "Does job testing harm minority workers? Evidence from retail establishments," *The Quarterly Journal of Economics*, 2008, *123* (1), 219–277.

**Bates, Michael, Michael Dinerstein, Andrew C Johnston, and Isaac Sorkin**, "Teacher Labor Market Policy and the Theory of the Second Best," *The Quarterly Journal of Economics*, 2025, *140* (2), 1417–1469.

**Biasi, Barbara, Chao Fu, and John Stromme**, "Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage," *NBER Working Paper 28530*, 2021.

**Bobba, Matteo, Tim Ederer, Gianmarco Leon-Ciliotta, Christopher Neilson, and Marco G Nieddu**, "Teacher Compensation and Structural Inequality: Evidence from Centralized Teacher School Choice in Perú," *NBER Working Paper 29068*, 2021.

**Carranza, Eliana, Robert Garlick, Kate Orkin, and Neil Rankin**, "Job search and hiring with limited information about workseekers' skills," *American Economic Review*, 2022, *112* (11), 3547–3583.

**Chetty, Raj, John N Friedman, and Jonah E Rockoff**, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 2014b, *104* (9), 2633–79.

**Colas, Mark and Chao Fu**, "Information Frictions and the Labor Market for Public School Teachers," Technical Report, National Bureau of Economic Research 2025.

**Colonnelli, Emanuele, Mounu Prem, and Edoardo Teso**, "Patronage and selection in public sector organizations," *American Economic Review*, 2020, *110* (10), 3071–3099.

**Dessein, Wouter**, "Authority and communication in organizations," *The Review of Economic Studies*, 2002, *69* (4), 811–838.

**Echenique, Federico, Ruy Gonzalez, Alistair J Wilson, and Leeat Yariv**, "Top of the Batch: Interviews and the Match," *American Economic Review: Insights*, 2022, *4* (2), 223–238.

**Ederer, Tim**, "Labor market dynamics and teacher spatial sorting," *CESifo Working Paper*, 2026.

**Fack, Gabrielle, Julien Grenet, and Yinghua He**, "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions," *American Economic Review*, 2019, *109* (4), 1486–1529.

**Farber, Henry S and Robert Gibbons**, "Learning and wage dynamics," *The Quarterly Journal of Economics*, 1996, *111* (4), 1007–1047.

**Fernandez, Marcelo Ariel, Kirill Rudov, and Leeat Yariv**, "Centralized matching with incomplete information," *American Economic Review: Insights*, 2022, *4* (1), 18–33.

**Frankel, Alex**, "Selecting applicants," *Econometrica*, 2021, *89* (2), 615–645.

**Friedrich, Benjamin, Martin B Hackmann, Adam Kapor, Sofia J Moroni, and Anne Brink Nandrup**, "Interdependent values in matching markets: Evidence from medical school programs in denmark," Technical Report, National Bureau of Economic Research 2024.

**Hakimov, Rustamdjan, Dorothea Kübler, and Siqi Pan**, "Costly information acquisition in centralized matching markets," *Quantitative Economics*, 2023, *14* (4), 1447–1490.

**Hoffman, Mitchell, Lisa B Kahn, and Danielle Li**, "Discretion in hiring," *The Quarterly Journal of Economics*, 2018, *133* (2), 765–800.

**Kahn, Lisa B and Fabian Lange**, "Employer learning, productivity, and the earnings distribution: Evidence from performance measures," *The Review of Economic Studies*, 2014, *81* (4), 1575–1613.

**Laverde, Mariana, Elton Mykerezi, Aaron Sojourner, and Aradhya Sood**, "Match Effects and the Gains from Alternative Job Assignments: Evidence from a Teacher Labor Market," Working Papers 2025-001, Human Capital and Economic Opportunity Working Group Feb 2025.

**Lee, Robin S and Michael Schwarz**, "Interviewing in two-sided matching markets," *The RAND Journal of Economics*, 2017, *48* (3), 835–855.

**Pallais, Amanda**, "Inefficient hiring in entry-level labor markets," *American Economic Review*, 2014, *104* (11), 3565–3599.

**Rivkin, Steven G, Eric A Hanushek, and John F Kain**, "Teachers, Schools, and Academic Achievement," *Econometrica*, 2005, *73* (2), 417–458.

**Rockoff, Jonah E**, "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review*, 2004, *94* (2), 247–252.

**Walters, Christopher R**, "The demand for effective charter schools," *Journal of Political Economy*, 2018, *126* (6), 2179–2223.

**Weaver, Jeffrey**, "Jobs for sale: Corruption and misallocation in hiring," *American Economic Review*, 2021, *111* (10), 3093–3122.

# A Additional Tables and Figures

**Table A.1:** Analysis Samples: Summary Statistics and Sample Sizes

| | Applicant choice | Outcome | Interview |
|---|---|---|---|
| *Panel A. Applicant characteristics* | | | |
| Applicant test score (std.) | 1.528 (0.375) | 1.111 (0.595) | 1.649 (0.429) |
| Female | 0.786 (0.410) | 0.740 (0.439) | 0.812 (0.391) |
| Speaks Quechua | 0.129 (0.336) | 0.099 (0.299) | 0.092 (0.288) |
| Experience $> 0$ (public) | 0.832 (0.374) | 0.820 (0.384) | 0.880 (0.325) |
| Experience $> 0$ (private) | 0.697 (0.459) | 0.643 (0.479) | 0.722 (0.448) |
| *Panel B. School characteristics* | | | |
| Rural school | 0.604 (0.489) | 0.360 (0.480) | 0.370 (0.483) |
| Amenity/infrastructure index | -0.130 (1.009) | 0.230 (0.845) | 0.228 (0.837) |
| Travel time to provincial capital (hours) | 2.966 (5.030) | 0.944 (1.470) | 1.075 (1.677) |
| Log locality population | 8.383 (3.817) | 10.310 (4.010) | 10.238 (4.084) |
| Mean lagged math achievement | -0.119 (0.723) | 0.138 (0.558) | 0.096 (0.545) |
| Mean lagged Spanish achievement | -0.286 (0.754) | 0.020 (0.557) | -0.009 (0.554) |
| Female student share | 0.493 (0.150) | 0.488 (0.108) | 0.492 (0.130) |
| Quechua student share | 0.141 (0.317) | 0.098 (0.259) | 0.092 (0.246) |
| Mean student age (centered) | 0.102 (0.336) | 0.001 (0.252) | 0.013 (0.233) |
| *Panel C. Sample size* | | | |
| 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 |

NOTES. This table reports summary statistics for selected covariates in the three analysis samples defined in Section 2.2. Panels A and B report means with standard deviations in parentheses. Applicant characteristics are measured once per applicant-year in the applicant choice and interview samples and once per teacher-year in the outcome sample. School characteristics are measured once per school-year. Panel C reports sample sizes.

**Table A.2:** Determinants of the CV score

| | (1) |
|---|---|
| Test score (std.) | 0.650 |
| | (0.462) |
| Test score squared | -0.259 |
| | (0.263) |
| Test score cubed | 0.041 |
| | (0.048) |
| Experience > 0 (public) | 0.731 |
| | (0.026) |
| Experience > 0 (private) | -0.210 |
| | (0.019) |
| Female | 0.034 |
| | (0.020) |
| Quechua speaker | 0.269 |
| | (0.030) |
| Recruitment-year FE | Yes |
| R-squared | 0.218 |
| Observations | 14,479 |

NOTES. The dependent variable is the standardized professional-background (CV) score. The regression is estimated on the final interview sample used in the hiring model. The covariates are a third-order polynomial in the standardized centralized test score and indicators for prior public-sector teaching experience, prior private-sector teaching experience, female applicants, and Quechua-speaking applicants. Recruitment-year fixed effects are included. Standard errors in parentheses are clustered by school-year.

**Table A.3:** Applicant Preferences: Truth-Telling Specification

| | Baseline ($\tau = 1$) | Baseline ($\tau = 2$) | Baseline ($\tau = 3$) | × Female | × Std Score | × Quechua | × Exp Pub | × Exp Priv |
|---|---|---|---|---|---|---|---|---|
| Wage | 0.788 | 0.836 | 1.257 | -0.508 | -0.413 | -0.186 | 0.377 | -0.082 |
| | (0.157) | (0.173) | (0.154) | (0.063) | (0.076) | (0.081) | (0.086) | (0.060) |
| Amenity Index | 0.078 | 0.008 | 0.057 | 0.016 | 0.012 | 0.063 | -0.027 | 0.057 |
| | (0.028) | (0.032) | (0.028) | (0.013) | (0.013) | (0.018) | (0.013) | (0.012) |
| Border | -0.114 | -0.023 | -0.435 | -0.002 | 0.356 | -0.108 | 0.193 | 0.023 |
| | (0.105) | (0.109) | (0.108) | (0.045) | (0.051) | (0.048) | (0.061) | (0.037) |
| VRAEM | 0.258 | 0.425 | 0.695 | -0.241 | -0.105 | -0.139 | -0.198 | -0.191 |
| | (0.152) | (0.142) | (0.126) | (0.050) | (0.064) | (0.060) | (0.074) | (0.049) |
| Multigrade | -0.335 | -0.193 | -0.238 | -0.181 | -0.097 | 0.220 | 0.312 | -0.222 |
| | (0.080) | (0.093) | (0.080) | (0.033) | (0.038) | (0.043) | (0.045) | (0.030) |
| Single Teacher | -0.324 | 0.031 | -0.277 | -0.571 | -0.136 | 0.328 | 0.348 | -0.151 |
| | (0.331) | (0.374) | (0.327) | (0.109) | (0.154) | (0.162) | (0.206) | (0.107) |
| Bilingual | -1.795 | -1.450 | -2.424 | -0.196 | 0.047 | 1.178 | -0.155 | -0.358 |
| | (0.107) | (0.110) | (0.102) | (0.036) | (0.050) | (0.039) | (0.063) | (0.035) |
| Travel Time | 0.019 | 0.192 | -0.092 | -0.179 | -0.117 | 0.075 | -0.002 | -0.095 |
| | (0.042) | (0.046) | (0.043) | (0.017) | (0.019) | (0.020) | (0.026) | (0.015) |
| log Population | -0.245 | 0.062 | -0.203 | -0.016 | 0.004 | 0.029 | -0.002 | -0.001 |
| | (0.006) | (0.006) | (0.006) | (0.003) | (0.003) | (0.004) | (0.003) | (0.003) |
| Time × log Pop | 0.003 | -0.036 | 0.018 | 0.022 | 0.017 | -0.004 | 0.001 | 0.012 |
| | (0.006) | (0.006) | (0.006) | (0.002) | (0.003) | (0.003) | (0.004) | (0.002) |
| Lagged Math Score | 0.193 | -0.054 | 0.036 | 0.014 | -0.046 | -0.045 | -0.025 | -0.020 |
| | (0.087) | (0.095) | (0.087) | (0.038) | (0.042) | (0.047) | (0.045) | (0.033) |
| Lagged Spanish Score | -0.369 | 0.179 | -0.083 | -0.071 | 0.068 | 0.085 | 0.042 | -0.091 |
| | (0.092) | (0.102) | (0.093) | (0.040) | (0.044) | (0.050) | (0.049) | (0.035) |
| Share Female | 0.120 | -0.061 | 0.033 | 0.011 | -0.033 | 0.119 | -0.068 | -0.091 |
| | (0.158) | (0.188) | (0.162) | (0.074) | (0.077) | (0.104) | (0.077) | (0.068) |
| Share Quechua | 0.095 | 0.574 | -0.343 | 0.014 | -0.072 | 0.941 | 0.020 | 0.128 |
| | (0.095) | (0.093) | (0.089) | (0.035) | (0.042) | (0.034) | (0.046) | (0.033) |
| Student Age | -0.064 | -0.315 | -0.068 | -0.020 | -0.119 | -0.134 | 0.273 | 0.004 |
| | (0.100) | (0.115) | (0.104) | (0.043) | (0.048) | (0.061) | (0.056) | (0.039) |
| Outside Option | 0.484 | 8.040 | 3.918 | -1.627 | -1.286 | 0.545 | 0.931 | -0.541 |
| | (0.329) | (0.370) | (0.328) | (0.138) | (0.163) | (0.184) | (0.179) | (0.130) |
| Previous School | 3.493 | 6.651 | 3.561 | -0.322 | -0.035 | -0.724 | -0.207 | 0.337 |
| | (0.098) | (0.109) | (0.103) | (0.044) | (0.045) | (0.056) | (0.053) | (0.038) |
| Distance: 0–20km | -0.110 | -0.104 | 0.036 | -0.011 | -0.015 | 0.027 | 0.024 | -0.024 |
| | (0.003) | (0.006) | (0.003) | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) |
| Distance: 20–100km | -0.062 | 0.033 | -0.020 | -0.002 | -0.001 | 0.003 | -0.002 | 0.002 |
| | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Distance: 100km+ | -0.034 | -0.002 | -0.022 | -0.000 | -0.001 | -0.002 | 0.000 | 0.000 |
| | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |

NOTES. This table reports maximum-likelihood estimates of applicants' preferences over schools specified in equation (5), under the assumption that applicants truthfully rank schools in order of preference. The first three columns report the baseline coefficients for the three latent preference types $\tau_j = 1, 2, 3$. The remaining columns report interactions with applicant gender, standardized test score, Quechua-speaking status, public-sector experience, and private-sector experience. Distance enters as a linear spline with knots at 20 and 100 kilometers. Standard errors are reported in parentheses.

**Table A.4:** Applicant Preferences: Stability-Based Specification

| | Baseline ($\tau = 1$) | Baseline ($\tau = 2$) | Baseline ($\tau = 3$) | × Female | × Std Score | × Quechua | × Exp Pub | × Exp Priv |
|---|---|---|---|---|---|---|---|---|
| Wage | 0.652 | 1.509 | 0.927 | -0.535 | -0.414 | -0.623 | 0.017 | -0.174 |
| | (0.174) | (0.224) | (0.175) | (0.138) | (0.067) | (0.167) | (0.133) | (0.129) |
| Amenity Index | 0.092 | 0.064 | 0.099 | 0.039 | -0.001 | 0.048 | -0.057 | 0.048 |
| | (0.027) | (0.039) | (0.029) | (0.023) | (0.010) | (0.032) | (0.022) | (0.021) |
| Border | 0.595 | 0.830 | 0.426 | -0.134 | 0.189 | -0.332 | 0.022 | -0.178 |
| | (0.135) | (0.121) | (0.134) | (0.101) | (0.045) | (0.116) | (0.092) | (0.084) |
| VRAEM | -0.576 | -0.337 | 0.030 | -0.161 | 0.078 | -0.070 | 0.635 | -0.027 |
| | (0.222) | (0.189) | (0.140) | (0.113) | (0.062) | (0.129) | (0.112) | (0.109) |
| Multigrade | -0.311 | -0.517 | -0.184 | -0.112 | 0.067 | 0.315 | 0.091 | -0.148 |
| | (0.076) | (0.112) | (0.078) | (0.062) | (0.030) | (0.076) | (0.060) | (0.057) |
| Single Teacher | -0.083 | -0.096 | -0.341 | -0.589 | 0.078 | 0.164 | 0.269 | 0.005 |
| | (0.200) | (0.281) | (0.218) | (0.164) | (0.094) | (0.253) | (0.169) | (0.162) |
| Bilingual | -1.995 | -1.643 | -2.663 | -0.139 | 0.011 | 1.315 | 0.079 | -0.344 |
| | (0.124) | (0.144) | (0.122) | (0.083) | (0.046) | (0.092) | (0.083) | (0.079) |
| Travel Time | -0.211 | -0.179 | -0.291 | -0.077 | -0.069 | 0.086 | 0.012 | -0.061 |
| | (0.045) | (0.051) | (0.050) | (0.037) | (0.019) | (0.034) | (0.035) | (0.035) |
| log Population | -0.249 | 0.026 | -0.205 | -0.009 | 0.017 | 0.013 | -0.010 | 0.012 |
| | (0.010) | (0.011) | (0.011) | (0.008) | (0.003) | (0.012) | (0.008) | (0.007) |
| Time × log Pop | 0.037 | 0.022 | 0.047 | 0.008 | 0.010 | -0.005 | -0.002 | 0.008 |
| | (0.006) | (0.008) | (0.007) | (0.005) | (0.003) | (0.006) | (0.005) | (0.005) |
| Lagged Math Score | 0.177 | -0.282 | -0.005 | 0.019 | 0.004 | -0.120 | -0.002 | 0.025 |
| | (0.077) | (0.105) | (0.083) | (0.064) | (0.030) | (0.077) | (0.060) | (0.059) |
| Lagged Spanish Score | -0.269 | 0.471 | -0.001 | -0.016 | -0.018 | 0.133 | 0.035 | -0.144 |
| | (0.082) | (0.112) | (0.088) | (0.067) | (0.032) | (0.082) | (0.064) | (0.062) |
| Share Female | 0.023 | -0.236 | 0.053 | 0.148 | -0.107 | -0.083 | -0.170 | -0.169 |
| | (0.141) | (0.223) | (0.157) | (0.125) | (0.056) | (0.164) | (0.118) | (0.115) |
| Share Quechua | -0.120 | 0.233 | -0.421 | 0.054 | -0.009 | 1.217 | -0.113 | -0.064 |
| | (0.121) | (0.119) | (0.103) | (0.083) | (0.041) | (0.079) | (0.078) | (0.075) |
| Student Age | -0.076 | -0.377 | -0.083 | 0.068 | -0.070 | -0.292 | 0.127 | -0.044 |
| | (0.090) | (0.134) | (0.097) | (0.076) | (0.036) | (0.106) | (0.072) | (0.070) |
| Outside Option | -3.590 | 4.619 | 0.270 | -1.256 | -1.155 | -0.523 | -0.255 | -0.530 |
| | (0.390) | (0.535) | (0.431) | (0.326) | (0.156) | (0.420) | (0.312) | (0.303) |
| Previous School | 3.429 | 7.208 | 3.713 | 0.150 | -0.053 | -0.681 | 0.121 | 0.351 |
| | (0.084) | (0.111) | (0.099) | (0.072) | (0.031) | (0.109) | (0.063) | (0.065) |
| Distance: 0–20km | -0.122 | -0.132 | 0.072 | -0.020 | -0.011 | 0.045 | 0.005 | -0.029 |
| | (0.004) | (0.013) | (0.005) | (0.004) | (0.002) | (0.005) | (0.003) | (0.003) |
| Distance: 20–100km | -0.075 | 0.026 | -0.028 | -0.000 | -0.001 | 0.005 | -0.004 | 0.002 |
| | (0.001) | (0.002) | (0.001) | (0.001) | (0.000) | (0.001) | (0.001) | (0.001) |
| Distance: 100km+ | -0.015 | -0.003 | -0.026 | 0.001 | -0.000 | -0.001 | 0.000 | 0.000 |
| | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |

NOTES. This table reports estimates of the applicant school-choice utility under the alternative stability-based identification strategy described in Section 3.2, which allows applicants to omit schools they regard as infeasible. The first three columns report the baseline coefficients for the three latent preference types $\tau_j = 1, 2, 3$. The remaining columns report interactions with applicant gender, standardized test score, Quechua-speaking status, public-sector experience, and private-sector experience. Distance enters as a linear spline with knots at 20 and 100 kilometers. Standard errors are reported in parentheses.

**Table A.5:** Balance Tests for Excluded Preference Shifters

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Dropout | Failing rate | Teacher retention | Dropout | Failing rate | Teacher retention |
| Intercept | 0.021 | 0.051 | 0.839 | 0.015 | 0.046 | 0.825 |
| | (0.004) | (0.005) | (0.024) | (0.004) | (0.005) | (0.026) |
| 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) |
| School Lagged Math Score | 0.001 | 0.006 | 0.010 | 0.000 | 0.005 | 0.014 |
| | (0.002) | (0.003) | (0.016) | (0.002) | (0.003) | (0.013) |
| School Lagged Spanish Score | -0.010 | -0.013 | 0.009 | -0.008 | -0.013 | 0.006 |
| | (0.002) | (0.003) | (0.017) | (0.002) | (0.003) | (0.014) |
| School Lagged Math Score$^2$ | 0.002 | -0.005 | 0.000 | 0.002 | -0.002 | -0.003 |
| | (0.001) | (0.002) | (0.010) | (0.001) | (0.002) | (0.009) |
| School Lagged Spanish Score$^2$ | -0.001 | 0.002 | 0.003 | -0.001 | 0.003 | -0.004 |
| | (0.001) | (0.002) | (0.008) | (0.001) | (0.001) | (0.007) |
| School Female | -0.006 | 0.001 | 0.087 | -0.004 | 0.002 | 0.101 |
| | (0.006) | (0.008) | (0.039) | (0.005) | (0.006) | (0.033) |
| School Quechua | -0.008 | 0.006 | -0.056 | -0.007 | 0.006 | -0.039 |
| | (0.002) | (0.002) | (0.011) | (0.001) | (0.002) | (0.009) |
| School Age | 0.032 | 0.101 | -0.058 | 0.031 | 0.098 | -0.048 |
| | (0.002) | (0.003) | (0.016) | (0.002) | (0.003) | (0.013) |
| Amenity Index | -0.002 | 0.001 | 0.006 | -0.002 | 0.000 | 0.006 |
| | (0.001) | (0.001) | (0.004) | (0.000) | (0.001) | (0.003) |
| Travel Time | -0.001 | 0.001 | -0.036 | -0.002 | 0.002 | -0.042 |
| | (0.001) | (0.002) | (0.008) | (0.001) | (0.001) | (0.007) |
| log(Population) | 0.000 | -0.000 | 0.002 | 0.000 | -0.000 | 0.004 |
| | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.001) |
| Travel Time × log(Pop) | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 | 0.005 |
| | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.001) |

NOTES. This table assesses the exclusion restrictions for the two variables used to shift applicants' school choices in the control-function correction. Columns (1)–(3) regress predetermined school dropout rates, examination failure rates, and teacher retention, respectively, on applicant–school distance. Columns (4)–(6) regress the same outcomes on the wage attached to the position. Distance is measured in thousands of kilometers and wages in thousands of soles. All regressions control for the predetermined school characteristics reported in the table. Standard errors are reported in parentheses.

**Table A.6:** Estimates Teacher Value-Added: Common Coefficients

| | Truth-telling | Stability |
|---|---|---|
| Lagged Score$^2$ | -0.001 (0.003) | -0.001 (0.003) |
| Quechua | -0.055 (0.017) | -0.055 (0.017) |
| Age | -0.130 (0.006) | -0.130 (0.006) |
| School Lagged Score | 0.304 (0.055) | 0.294 (0.055) |
| School Lagged Score$^2$ | 0.088 (0.050) | 0.093 (0.050) |
| School Female | -0.108 (0.255) | -0.131 (0.254) |
| School Quechua | 0.463 (0.291) | 0.452 (0.291) |
| School Age | -0.717 (0.136) | -0.706 (0.135) |
| Class Lagged Score | -0.158 (0.032) | -0.158 (0.032) |
| Class Lagged Score$^2$ | -0.066 (0.028) | -0.068 (0.028) |
| Class Female | 0.583 (0.136) | 0.574 (0.136) |
| Class Quechua | -0.497 (0.302) | -0.499 (0.302) |
| Class Age | 0.049 (0.075) | 0.041 (0.075) |
| Class Size | 0.002 (0.006) | 0.003 (0.006) |
| Class Size$^2$ | -0.000 (0.000) | -0.000 (0.000) |
| Control function, $\varphi$ | 0.018 (0.014) | 0.022 (0.021) |

NOTES. This table reports the common coefficients in the student-outcome equation used to estimate teacher effectiveness, together with the coefficient $\varphi$ on the applicant-preference control function. The Truth-telling column constructs the control function using the applicant preference estimates in Table A.3, while the Stability column uses the stability-based estimates in Table A.4. Teacher fixed effects and teacher-specific interactions with lagged achievement and gender, which recover the teacher-effectiveness parameters $(\alpha_j, \beta_j)$, are not reported. Standard errors are reported in parentheses. Inference treats the estimated first-stage applicant-preference parameters and the resulting control-function terms as fixed. We include an indicator for observations for which the control-function term is missing.

**Table A.7:** Selection on Private School Information

| | CV | Interview | Classroom observation |
|---|---|---|---|
| Screening score | -0.0028 | -0.0054 | 0.0029 |
| | (0.0049) | (0.0048) | (0.0052) |
| Observations | 380 | 372 | 384 |
| $R^2$ | 0.001 | 0.003 | 0.001 |

NOTES. This table tests whether school screening scores predict student-achievement variation not captured by the baseline teacher-value-added specification. The dependent variable is the teacher–school–year average residual from the baseline student-outcome equation. Each column reports a separate bivariate regression of this residual on the indicated screening score. CV, interview, and classroom-observation scores are standardized within the original school × year contest. Standard errors are reported in parentheses.

**Table A.8:** Candidate Selection under Coarser Prior Information Sets

| | Pre-interview | Prior TVA | Coarse prior: no $R_j$ | Coarse prior: no $R_j$, gender, ethnicity |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Female | 0.006 | 0.065 | 0.136 | 0.008 |
| | (0.005) | (0.005) | (0.004) | (0.005) |
| Test score (std.) | 0.017 | 0.158 | 0.239 | 0.268 |
| | (0.005) | (0.005) | (0.005) | (0.004) |
| Quechua | 0.007 | 0.009 | -0.027 | -0.004 |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| Exp public $> 0$ | 0.047 | 0.032 | -0.004 | -0.012 |
| | (0.003) | (0.004) | (0.004) | (0.004) |
| Exp private $> 0$ | -0.027 | 0.104 | 0.189 | 0.203 |
| | (0.006) | (0.005) | (0.004) | (0.004) |
| Internal candidate | 0.012 | 0.006 | 0.002 | 0.005 |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| Shrunken TVA | 0.011 | 0.016 | -0.002 | 0.004 |
| | (0.008) | (0.008) | (0.008) | (0.008) |
| N. observations | 3,642 | 3,642 | 3,642 | 3,642 |
| N. schools | 2,874 | 2,874 | 2,874 | 2,874 |

Notes. This table assesses whether the difference between schools' pre-interview rankings and the estimated prior ranking can be explained by schools conditioning on a coarser information set. All entries report the characteristic of the top-ranked candidate under the indicated rule minus its expected value under random selection from the same realized candidate set, averaged across school $\times$ year contests. Column (1) uses the observed pre-interview school ranking and Column (2) ranks candidates by the estimated prior expectation of teacher effectiveness. Column (3) ranks candidates using a coarse prior constructed without information from applicants' rank-ordered lists $R_j$, while Column (4) additionally removes gender and ethnicity from the information used to construct the coarse prior. The shrunken-TVA row is estimated on the TVA subsample used in Table 2; all other rows use the full analysis sample. Each contest receives equal weight. Standard errors in parentheses are clustered at the school level.

**Table A.9:** Candidate Selection by Principal Experience and Managerial Skill

| | Random selection | Test score | Pre-interview | Post-interview | Prior TVA |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Panel A: Principal with at least 2 years of experience* | | | | | |
| Female | 0.808 | -0.009 | 0.008 | -0.022 | 0.062 |
| | | (0.007) | (0.007) | (0.007) | (0.006) |
| Test score (std.) | 1.686 | 0.328 | 0.006 | 0.005 | 0.139 |
| | | (0.005) | (0.006) | (0.006) | (0.006) |
| Quechua | 0.094 | 0.005 | 0.007 | 0.003 | 0.008 |
| | | (0.004) | (0.004) | (0.004) | (0.004) |
| Exp public $> 0$ | 0.909 | 0.011 | 0.030 | 0.032 | 0.021 |
| | | (0.005) | (0.004) | (0.004) | (0.005) |
| Exp private $> 0$ | 0.721 | 0.000 | -0.012 | -0.012 | 0.100 |
| | | (0.007) | (0.007) | (0.007) | (0.007) |
| Internal candidate | 0.124 | 0.014 | 0.011 | 0.046 | 0.004 |
| | | (0.006) | (0.006) | (0.006) | (0.006) |
| Shrunken TVA | 0.276 | 0.014 | 0.007 | 0.022 | 0.018 |
| | | (0.011) | (0.010) | (0.011) | (0.011) |
| N. observations | 2,130 | 2,130 | 2,130 | 2,130 | 2,130 |
| N. schools | 1,776 | 1,776 | 1,776 | 1,776 | 1,776 |
| *Panel B: Principal with managerial score above median* | | | | | |
| Female | 0.841 | -0.017 | 0.000 | -0.022 | 0.049 |
| | | (0.010) | (0.010) | (0.010) | (0.009) |
| Test score (std.) | 1.698 | 0.353 | 0.016 | 0.007 | 0.161 |
| | | (0.007) | (0.009) | (0.010) | (0.010) |
| Quechua | 0.076 | 0.005 | 0.004 | 0.005 | 0.006 |
| | | (0.005) | (0.006) | (0.005) | (0.006) |
| Exp public $> 0$ | 0.874 | 0.009 | 0.067 | 0.043 | 0.037 |
| | | (0.008) | (0.007) | (0.008) | (0.008) |
| Exp private $> 0$ | 0.765 | -0.010 | -0.030 | -0.001 | 0.083 |
| | | (0.010) | (0.011) | (0.011) | (0.009) |
| Internal candidate | 0.122 | 0.011 | 0.015 | 0.032 | 0.002 |
| | | (0.008) | (0.009) | (0.009) | (0.009) |
| Shrunken TVA | 0.321 | 0.019 | 0.008 | 0.036 | 0.045 |
| | | (0.014) | (0.014) | (0.015) | (0.014) |
| N. observations | 985 | 985 | 985 | 985 | 985 |
| N. schools | 749 | 749 | 749 | 749 | 749 |

Notes. This table reports the selection effects of alternative ranking rules for subsamples of principals for whom poorly informed or naive prior beliefs are less plausible. Panel A restricts the sample to principals with at least two years of experience. Panel B restricts the sample to principals whose managerial score is above the median among principals with an observed score. Within each panel, Column (1) reports the expected characteristic under random selection and Columns (2)–(5) report the characteristic of the top-ranked candidate under the indicated rule minus the corresponding random-selection benchmark. The shrunken-TVA row uses the corresponding TVA subsample; the remaining rows use all eligible contests in the panel. Each contest receives equal weight. Standard errors in parentheses are clustered at the school level.

**Table A.10:** Selection of Internal Candidates

| | Random selection | Pre-interview | Post-interview |
|---|---|---|---|
| | (1) | (2) | (3) |
| Internal candidate | 0.121 | 0.005 | 0.033 |
| | | (0.010) | (0.010) |
| Shrunken TVA | 0.272 | 0.011 | 0.025 |
| | | (0.008) | (0.008) |
| High-VA internal | 0.059 | 0.003 | 0.017 |
| | | (0.008) | (0.008) |
| Low-VA internal | 0.061 | 0.001 | 0.017 |
| | | (0.007) | (0.007) |
| N. observations | 526 | 526 | 526 |
| N. schools | 492 | 492 | 492 |

Notes. This table examines whether schools' preference for internal candidates is concentrated among internal candidates with higher measured teacher effectiveness. Column (1) reports the expected characteristic under random selection from the realized candidate set. Columns (2) and (3) report the characteristic of the top-ranked candidate under the pre- and post-interview rankings, respectively, minus the corresponding random-selection benchmark. High-VA and Low-VA internal candidates are internal candidates whose shrunken TVA lies above or below, respectively, the median shrunken TVA among internal candidates in the TVA analysis sample. The table uses the TVA subsample described in Table 2. Each contest receives equal weight. Standard errors in parentheses are clustered at the school level.

**Table A.11:** Test-Score Exclusion Restriction

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

Notes. This table assesses the exclusion of applicant test scores from the reduced-form components $v^g(\cdot)$ of school rankings. For each school-evaluation component—CV, interview, and classroom observation—we residualize both the standardized evaluation score and standardized applicant test score with respect to the estimated prior mean of teacher effectiveness, the other applicant characteristics entering $v^g(\cdot)$, and school $\times$ year fixed effects. Residualized test scores are then normalized to have unit standard deviation. The first row reports their coefficient in a regression of the residualized evaluation score on residualized test score; standard errors in parentheses are clustered at the school level. The second row reports the corresponding partial $R^2$. The third row reports the share of strict within-contest pairwise candidate orderings that would be reversed after removing the fitted test-score component from the evaluation score; tied candidate pairs are excluded.

**Table A.12:** Additional Estimates of the Hiring Model

| | Pre-interview, $v^1$ | Post-interview, $v^2$ |
|---|---|---|
| *Panel A. Other determinants of candidate rankings* | | |
| Female | 0.128 | -0.361 |
| | (0.053) | (0.053) |
| Quechua | 0.212 | 0.230 |
| | (0.087) | (0.092) |
| Public-sector experience | 1.593 | 0.794 |
| | (0.061) | (0.062) |
| Private-sector experience | -0.315 | -0.164 |
| | (0.048) | (0.050) |
| Applicant type-1 probability | -0.119 | -0.031 |
| | (0.061) | (0.061) |
| Applicant type-2 probability | -0.274 | -0.198 |
| | (0.093) | (0.097) |
| Internal candidate | 0.192 | 0.811 |
| | (0.067) | (0.069) |
| *Panel B. Private information and persistent heterogeneity* | | |
| | Estimate | Std. Error |
| Private information for internal candidates, $\rho$ | 0.002 | (0.157) |
| SD persistent heterogeneity, external candidates | 1.257 | (0.025) |
| SD persistent heterogeneity, internal candidates | 1.279 | (0.069) |

NOTES. Panel A reports coefficients entering the reduced-form components $v^1(\cdot)$ and $v^2(\cdot)$ of candidate rankings before and after the interview, respectively. Applicant type probabilities are posterior probabilities from Section 3.3, with type 3 omitted. Panel B reports the additional pre-interview information parameter for internal candidates and the standard deviation of the persistent applicant–school component, separately for external and internal candidates. Standard errors are reported in parentheses.

**Table A.13:** Model Fit: Changes in School Rankings

| Moment | Data | Model |
|---|---|---|
| Any pre-post rank change | 0.658 | 0.674 |
| Share of applicants changing rank | 0.514 | 0.528 |
| Pooled correlation of rank positions | 0.705 | 0.697 |

Notes. "Any pre–post rank change" is the share of contests in which at least one candidate changes rank between the pre- and post-interview stages. "Share of applicants changing rank" is first calculated within each contest and then averaged across contests, giving each contest equal weight. "Pooled correlation of rank positions" is the Pearson correlation between pre- and post-interview rank positions across candidate–contest observations. Model moments are averaged across model simulations.

**Table A.14:** Model Fit: Gains in Expected Student Achievement

| | Pre-interview | | Post-interview | |
|---|---|---|---|---|
| | Data | Model | Data | Model |
| Mean gain in expected student achievement | 0.011 | 0.002 | 0.025 | 0.011 |
| SD of gains across contests | 0.189 | 0.193 | 0.193 | 0.193 |

Notes. The table compares the treatment effects of the pre- and post-interview ranking rules with the corresponding model predictions. As in Table 2, treatment effects are evaluated using the shrunken empirical-Bayes estimates $\widehat{q}^{*}_{ij}$: for each contest, the effect is the shrunken effectiveness of the top-ranked candidate minus the average shrunken effectiveness of candidates in the same realized candidate set. The first row reports the average treatment effect across contests and the second row its cross-contest standard deviation. The sample is restricted to contests with at least two candidates for whom teacher value added can be estimated, and rankings are recomputed within this sample. Each contest receives equal weight. Model moments are averaged across model simulations.

**Table A.15:** Observable Predictors of Estimated Preferences and Signal Extraction

| | Posterior mean $\delta_i$ | | Posterior mean $\gamma_i$ | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Average 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 (years) | | 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 |

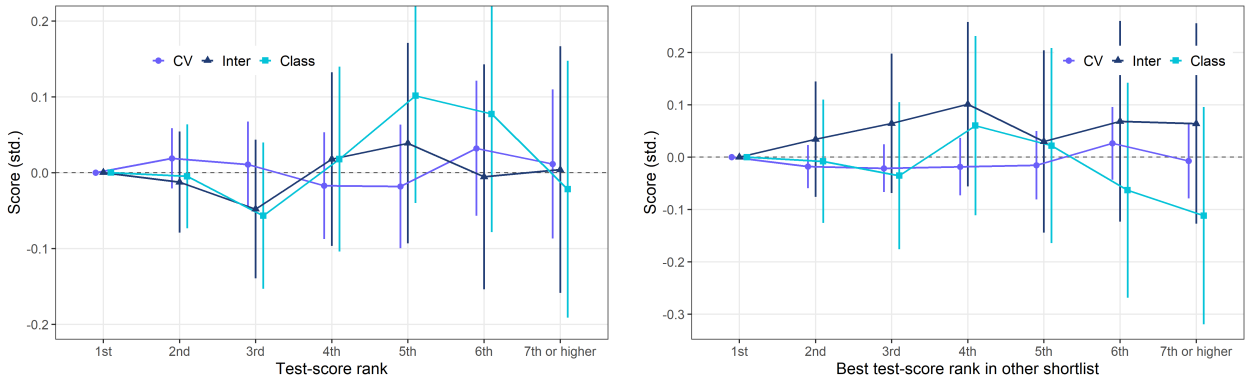
Notes. The dependent variables are the posterior means of the school-specific coefficients $\delta_i$ and $\gamma_i$. Higher $\delta_i$ corresponds to a greater weight on teacher effectiveness, while higher $\gamma_i$ corresponds to a noisier private signal and therefore lower signal-extraction ability. Columns (1) and (3) use the school characteristics available for the full estimation sample. Columns (2) and (4) additionally include principal experience and the principal's standardized test score and therefore use complete cases for these variables. School-year covariates are first collapsed to one observation per contest and then averaged within school. Heteroskedasticity-robust HC1 standard errors are reported in parentheses. The regressions are descriptive and treat posterior means as estimated school-level outcomes.

**Table A.16:** Observable Characteristics by Estimated School Type

| | Overall | Low $\delta$, low $\gamma$ | Low $\delta$, high $\gamma$ | High $\delta$, low $\gamma$ | High $\delta$, high $\gamma$ |
|---|---|---|---|---|---|
| Average student test score | 0.220 | 0.141 | 0.242 | 0.226 | 0.261 |
| Female student share | 0.473 | 0.491 | 0.487 | 0.456 | 0.463 |
| Urban | 0.762 | 0.726 | 0.824 | 0.750 | 0.746 |
| Principal experience (years) | 1.621 | 1.355 | 1.665 | 1.897 | 1.499 |
| Principal test score (std.) | 0.550 | 0.434 | 0.582 | 0.570 | 0.606 |
| Share of classified schools | 1.000 | 0.221 | 0.242 | 0.292 | 0.246 |
| Number of schools | 281 | 62 | 68 | 82 | 69 |

Notes. For each draw from the posterior distribution, schools are classified according to whether $\delta_i$ and $\gamma_i$ lie above or below their respective cross-school medians. For each school, the posterior probability of each of the four joint types is the fraction of posterior draws in which the school is assigned to that type. A school is assigned to its modal type only if its posterior probability for that type is at least 0.40; remaining schools are left unclassified. The Overall column therefore refers to the classified-school sample. Low $\gamma_i$ corresponds to more precise signals and greater signal-extraction ability. School-year characteristics are first measured once per contest and then averaged within school, so schools are not weighted by the number of applicants or contests in which they appear. Principal characteristics use the years in which they are observed. "Share of classified schools" reports each type's share of the classified sample.

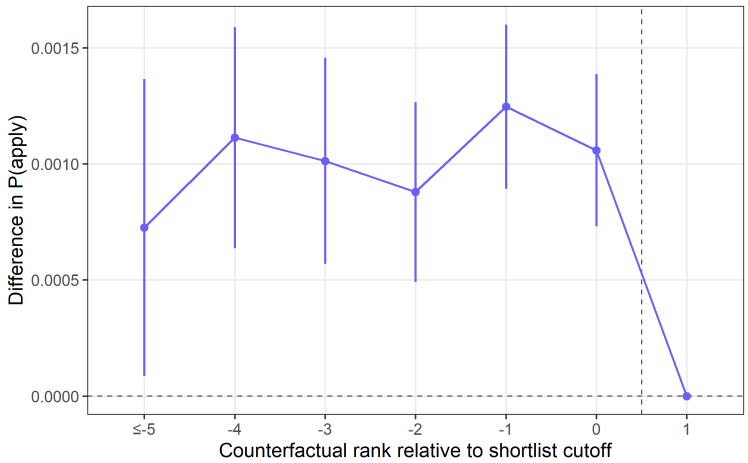
**Figure A.1:** School Evaluation Scores and Candidate Ranking Positions




**a) Initial Position in the Focal Shortlist** **b) Position in Competing Shortlists**

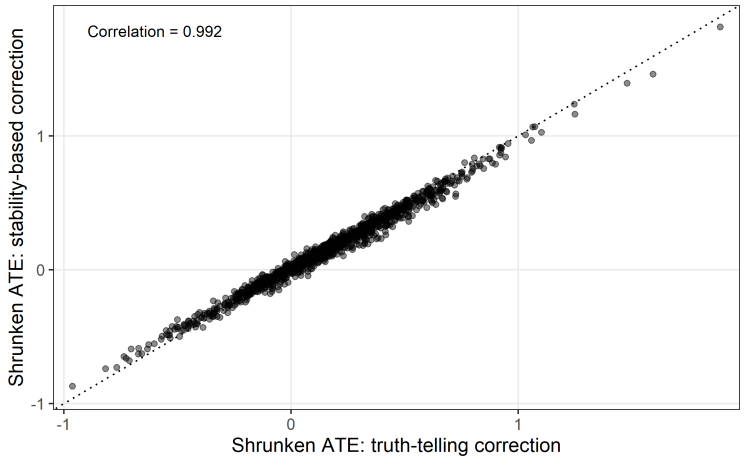
NOTES: Panel (a) relates standardized CV, interview, and classroom-observation scores to the applicant's centralized test-score rank in the focal school-year contest. Rank 1 is omitted and ranks 7 or worse are pooled. The regressions control flexibly for test score and observed applicant characteristics and include applicant and school-year fixed effects; standard errors are clustered by applicant. Panel (b) relates the same scores to the applicant's best test-score rank in another contemporaneous shortlist. These regressions include applicant-year and school-year fixed effects, with standard errors two-way clustered at those levels. Points are coefficient estimates and bars report 95% confidence intervals.

**Figure A.2:** Applicant Rankings and Shortlist Feasibility




NOTES: For each applicant–school pair involving a school-year with a reconstructable binding shortlist, we compute the applicant's counterfactual test-score rank relative to the institutional shortlist cutoff. Zero denotes the last feasible shortlist position, negative values positions inside the shortlist, and positive values positions outside it; positions five or more ranks inside the cutoff are pooled. The figure reports coefficients from a linear probability model for applying to the school, with the +1 position omitted. The specification includes applicant-year and school-year fixed effects, controls for distance and previous employment at the school, and two-way clustered standard errors. Categories above +1 are not displayed because support is sparse. Points are coefficient estimates and bars report 95% confidence intervals.

**Figure A.3:** Shrunken ATE under Truth-Telling and Stability-Based Preference Corrections




NOTES: Each point represents a teacher and compares the shrunken estimate of average teacher effectiveness obtained using the truth-telling applicant-preference specification on the horizontal axis with the corresponding estimate obtained using the stability-based specification on the vertical axis. The dashed line is the 45-degree line. The correlation between the two sets of estimates is 0.992.

# B Technical Appendix

## B.1 Construction of the Control-Function Terms

This appendix provides additional details on the construction of the control-function terms described in Section 3.3. We adapt the rank-ordered selection correction of Abdulkadiroğlu et al. (2020) to a setting with discrete unobserved applicant preference types. Applicant $j$'s utility from school $i$, conditional on latent preference type $\tau_j = h$, is

$$U_{ijh} = u_{ijh} + \eta_{ij}, \qquad u_{ijh} \equiv u(z_j, x_i, d_{ij}, \tau_j = h),$$

where $\eta_{ij}$ is i.i.d. Type-I extreme value with mean $\mu_\eta$.

Conditional on applicant preference type $h$, we construct the control functions $\lambda_{ijh} = E[\eta_{ij} - \mu_\eta \mid R_j, z_j, d_j, \tau_j = h]$ using the analytical expressions derived by Abdulkadiroğlu et al. (2020), which account for the applicant's entire rank-ordered list and apply to both ranked and unranked alternatives. Let $\pi_h$ denote the population probability of type $h$. After estimating the preference model, we then use the applicant's complete observed ranking to form the posterior probability

$$\omega_{jh} \equiv P(\tau_j = h \mid R_j, z_j, d_j) = \frac{\pi_h \mathcal{L}(R_j \mid z_j, d_j, \tau_j = h)}{\sum_{s=1}^{K} \pi_s \mathcal{L}(R_j \mid z_j, d_j, \tau_j = s)}. \tag{14}$$

The control-function term used in the student-outcome equation integrates the type-specific conditional expectations over these posterior probabilities:

$$\lambda_{ij}(R_j, z_j, d_j) = \sum_{h=1}^{K} \omega_{jh} \lambda_{ijh}, \qquad \lambda_{ijh} = \begin{cases} \lambda^R_{ijh}, & \text{if school } i \text{ is ranked,} \\ \lambda^U_{ijh}, & \text{if school } i \text{ is unranked.} \end{cases} \tag{15}$$

Under the alternative stability-based identification strategy described in Section 3.2, the same construction applies, with the choice sets defined by the applicant's feasible alternatives rather than the full set of schools implied by truth-telling. We recompute the conditional choice probabilities, type-specific control functions, and posterior type probabilities under that alternative preference model.

## B.2 Identification of Preferences and Signal Precision

We show that homogeneous $(\delta, \gamma)$ are identified from a single cross section of pre- and post-interview rankings and teacher value-added estimates, without imposing distributional assumptions on unobserved preferences beyond independence and a utility-scale normalization.

Consider two external candidates $j$ and $k$ with identical characteristics entering $v^g(\cdot)$ at both stages. Their direct contributions to utility through $v^g(\cdot)$ therefore cancel in pairwise comparisons. Crucially, because test scores are excluded from $v^g(\cdot)$ but affect prior expected teacher effectiveness, these candidates may still differ in their expected productivity. This provides the variation needed to separately identify preferences for teacher effectiveness from preferences over other candidate attributes. Define

$$\Delta\mu = \mathbb{E}[q_{ij} \mid S_{ij1}] - \mathbb{E}[q_{ik} \mid S_{ik1}], \qquad \Delta q = q_{ij} - q_{ik}.$$

Let

$$\kappa_i = \frac{\sigma_q^2(X_i)}{\sigma_q^2(X_i) + \exp(\gamma)}$$

denote the weight placed on the interview signal. Define the pairwise disturbances

$$\omega_{ijk}^1 = \xi_{ik} - \xi_{ij} + \varepsilon_{ik1} - \varepsilon_{ij1},$$

and

$$\omega_{ijk}^2 = \xi_{ik} - \xi_{ij} + \varepsilon_{ik2} - \varepsilon_{ij2} - \exp(\delta)\kappa_i \left[(\nu_{ij} - q_{ij}) - (\nu_{ik} - q_{ik})\right].$$

We assume that the preference shocks are jointly independent of $(\Delta\mu, \Delta q)$ and the interview-signal innovations, conditional on the remaining observables. We also assume that the interview-signal innovations are independent of $(\Delta\mu, \Delta q)$ and that $(\Delta\mu, \Delta q)$ has two-dimensional interior support, with zero belonging to the interior support of $\Delta\mu$. Let $F_g$ denote the conditional distribution of $\omega_{ijk}^g$ for $g = 1, 2$. We assume that $F_1$ and $F_2$ are continuously differentiable, with finite, strictly positive densities. Their distributions are otherwise unrestricted.

Although $q_{ij}$ is not directly observed, the econometrician observes $\widehat{\theta}_j$ for every applicant. Under the measurement-error structure described in Section 3.3, the estimation errors are independent of rankings conditional on true teacher effectiveness, with known variances

$\widehat{\Omega}_j$ and nonvanishing characteristic functions. Standard deconvolution arguments therefore identify the joint distribution of rankings and $q_{ij}$, and hence the conditional pairwise ranking probabilities.

Using the updating rule in Section 4.1, post-interview utility differences satisfy

$$\begin{aligned} V_{ij2} - V_{ik2} = \; & \exp(\delta)\big[(1-\kappa_i)\Delta\mu + \kappa_i \Delta q\big] \\ & + \xi_{ij} - \xi_{ik} + \varepsilon^2_{ij} - \varepsilon^2_{ik} \\ & + \exp(\delta)\kappa_i\big[(\nu_{ij} - q_{ij}) - (\nu_{ik} - q_{ik})\big]. \end{aligned}$$

By independence, the conditional probability that $j$ is ranked above $k$ can be written

$$\Pr(jR^2_i k \mid \Delta\mu, \Delta q) = F_2(\exp(\delta)[(1-\kappa_i)\Delta\mu + \kappa_i \Delta q])\,,$$

Differentiating and taking ratios eliminates both $F_2$ and the utility scale, yielding

$$\frac{\partial \Pr(jR^2_i k \mid \Delta\mu, \Delta q)/\partial \Delta q}{\partial \Pr(jR^2_i k \mid \Delta\mu, \Delta q)/\partial \Delta\mu} = \frac{\kappa_i}{1-\kappa_i} = \frac{\sigma^2_q(X_i)}{\exp(\gamma)}.$$

Since $\sigma^2_q(X_i)$ is identified from the teacher-effectiveness model, this identifies $\gamma$.

Before the interview, utility differences satisfy

$$V_{ij1} - V_{ik1} = \exp(\delta)\Delta\mu + \xi_{ij} - \xi_{ik} + \varepsilon^1_{ij} - \varepsilon^1_{ik}.$$

Consequently,

$$\Pr(jR^1_i k \mid \Delta\mu) = F_1\left(\exp(\delta)\Delta\mu\right),$$

Differentiating with respect to $\Delta\mu$ and evaluating at zero yields

$$\left.\frac{\partial \Pr(jR^1_i k \mid \Delta\mu)}{\partial \Delta\mu}\right|_{\Delta\mu=0} = \exp(\delta) f_1(0).$$

Provided that zero lies in the interior support of $\Delta\mu$ and $f_1(0) > 0$, this identifies $\delta$ up to the scale of the unobserved preference disturbance. In our parametric specification, this scale is normalized by imposing standard type-I extreme-value shocks. Thus, homogeneous $(\delta, \gamma)$ are identified up to a utility-scale normalization, without imposing parametric restrictions on the distributions of unobserved preferences.