Signal Extraction in Hiring: Evidence from Teacher Recruiting*

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

September 29, 2026

Abstract

We quantify the value of discretionary hiring through signal extraction in teacher recruiting. Using rich data on schools' evaluations of candidates in Peru's nationwide teacher allocation platform, we find that discretion improves expected student achievement on average relative to test-based hiring, but gains vary across schools. We estimate a model separating heterogeneity in schools' objectives from heterogeneity in signal-extraction ability. Both margins explain one quarter of the variance in these gains, with the latter alone accounting for 11 percent. Disclosing teacher value-added estimates eliminates signal-extraction heterogeneity and more than doubles the status quo gains, but leaves cross-school inequality unchanged as heterogeneity in school objectives becomes more consequential.

*We thank the Ministry of Education of Perú for supporting rigorous policy evaluation efforts and continued collaboration. Matteo Bobba acknowledges financial support from the ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program) and grant 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

Hiring under uncertainty creates a fundamental organizational trade-off. Centralized screening can discipline selection by relying on standardized and observable information, but local decision makers may possess or acquire additional information about worker productivity (Aghion and Tirole, 1997; Dessein, 2002). Granting them discretion can therefore improve selection, but its benefits depend both on the quality of the information they extract and on the private objectives they are pursuing (Hoffman et al., 2018; Frankel, 2021). When decision makers differ along either dimension, the returns to discretion may vary substantially across firms. Understanding the sources of such heterogeneity is central to the design of screening and delegation in hiring.

Teacher recruiting provides a particularly important setting in which to study this trade-off. 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 are able to identify and select effective teachers among the applicants they face (Bates et al., 2025). We therefore ask when discretion improves teacher selection, whether its value differs across schools, and how screening policy should be designed.

We study these questions in Peru's recruitment system for public-school teachers, which combines centralized screening with school-level discretion. We observe how schools evaluate candidates at different stages of the hiring process and link recruited teachers to subsequent student outcomes. 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 a model of hiring under uncertainty and learning that separates schools' objective function from the information they acquire during screening. The estimates imply substantial heterogeneity along both dimensions and show that both explain one quarter of the variance in estimated gains from discretion across schools. Finally, counterfactual simulations show that disclosing stan-

dardized information on teacher effectiveness can substantially improve aggregate selection gains by removing heterogeneity in signal extraction, but does little to reduce cross-school inequality in screening performance, as differences in school objectives become increasingly salient.

Permanent teachers in Peru are hired through a two-stage process combining centralized screening with decentralized school evaluation. Applicants first take a national test and submit rank-ordered lists of schools. Applicants are then shortlisted in their preferred schools by order of their test score rank to be evaluated by each school in a second stage. School committees 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' prior 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. Even after conditioning on observable proxies of quality, average teacher effectiveness and match-effect coefficients remain highly heterogeneous: the residual standard deviation of average effectiveness is 0.383, while the residual standard deviations of the slope coefficient with respect to prior achievement and gender are 0.151 and 0.108.

We use these estimates to quantify how alternative screening rules affect expected student achievement, holding fixed the candidate set faced by each school. Ranking candidates by the centralized test score raises expected achievement by 0.010 standard deviations relative to random selection from the same shortlist. The pre-interview school ranking generates a similar gain of 0.011, but rises to 0.025 after the interview. The additional information acquired during screening therefore raises expected achievement by 0.014 standard deviations on average. This average effect masks substantial heterogeneity: the standard deviation of these gains across schools is about 0.19. Discretionary evaluation also changes selection along

other dimensions, most notably in favor of male applicants and incumbents. School rankings differ systematically from a rule based on predicted teacher effectiveness. The data therefore point to two distinct 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 allow the precision of schools prior to depend on whether an applicant is an incumbent or not. We also allow schools to value candidate attributes other than expected teacher effectiveness, including traits that may themselves be productive but are not captured by student test-score value added. We model preferences over these other traits as a function of observable candidate attributes whose effects may differ before and after the interview, together with unobservable match-specific attributes containing both a persistent and transitory component. This allows rankings to reflect learning about productivity as well as information acquisition about other traits. 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 call signal extraction ability.

Identification relies primarily on observing complete rankings of the same candidate set before and after schools acquire additional information. Pre-interview rankings identify how schools trade off expected effectiveness against other observed and unobserved candidate attributes. Post-interview rank changes identify signal precision from the extent to which candidates move according to posterior mean shifts in teacher value added relative to prior expectations. Observing complete rankings of candidates as well as schools repeatedly participating in the recruitment process over several years provides additional variation to recover the distribution of latent school heterogeneity in preferences and signal extraction ability.

Our estimates imply 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. The average posterior weight on the private signal schools receive during the interview is 0.23, with a standard deviation of 0.26. Among schools in the top decile of the signal-precision distribution,

post-interview beliefs correlate with underlying teacher effectiveness by as much as 0.76. We also find essentially no additional pre-interview productivity information about internal candidates, despite their substantially higher probability of selection after discretionary evaluation. 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.

Using the estimated model, we show that selection gains are concentrated among schools that both place greater weight on teacher effectiveness and receive more precise productivity signals while schools below the median on both dimensions generate essentially no gain from discretionary evaluation relative to test-based hiring. Heterogeneity in school objectives and signal extraction jointly account for 24.4 percent of the variance in selection gains across schools. Varying the weight on effectiveness alone accounts for 17 percent and varying signal precision alone for 11.4 percent. Strikingly, observable school and principal characteristics explain virtually none of the variation in either dimension. Identifying ex-ante which schools will use discretion effectively is therefore challenging, motivating policies that address information frictions directly rather than screening school principals or targeting delegation.

Finally, we compare counterfactual screening regimes that vary the information available to schools and the amount of discretion they possess. Replacing the interview with direct disclosure of teacher value-added estimates generates an average gain of 0.070 standard deviations in expected student achievement relative to random selection, more than doubling the status quo gains. Eliminating all uncertainty about teacher effectiveness raises it further to 0.105. These gains nevertheless remain well below the 0.265 achieved by assigning the position to the candidate with the highest posterior expected effectiveness. Heterogeneity in schools' objective functions thus impose a tight upper bound on the potential aggregate gains from improving information acquisition through discretionary screening. Disclosing standardized information is also ineffective at reducing cross-school inequality in screening performance. As information about candidates' productivity becomes increasingly accurate, differences in schools' objectives become more consequential: the share of the variance of selection gains attributable to school heterogeneity rises from 24.4 percent under the status quo to 29.7 percent under standardized disclosure and to 40 percent under full information, while the dispersion in expected selected teacher effectiveness remains essentially unchanged. Better information thus removes one source of heterogeneity while making another more

salient, suggesting that effective screening policy design should account for the strong complementarity between information and agency frictions.

Our paper first contributes 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 (Bobba et al., 2021; Biasi et al., 2021; Bates et al., 2025; Ederer, 2023). 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). At the same time, previous work suggests that discretionary evaluations can contain valuable information about teacher effectiveness beyond observable characteristics (Jacob and Lefgren, 2008; Rockoff and Speroni, 2010; Jacob et al., 2018). We contribute to this literature by studying the efficiency and equity consequences of allowing for discretion in teacher hiring. We find positive average gains from school-level discretion relative to test-based hiring, but substantial heterogeneity across schools. Providing standardized information has large aggregate effects and eliminates the role of heterogeneity in signal-extraction ability. However, hiring outcomes remain highly unequal, as differences in school objectives become more consequential. Unpacking the effect of both sources of heterogeneity is thus crucial for screening policy design.

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 (Jovanovic, 1979; Farber and Gibbons, 1996; Altonji and Pierret, 2001; Kahn and Lange, 2014; Pallais, 2014; Carranza et al., 2022; Lessem and Miller, 2026) while other work studies the role of managers and discretion in hiring (Autor and Scarborough, 2008; Hoffman et al., 2018; Estrada, 2019; Colonnelli et al., 2020; Weaver, 2021). We allow schools to differ both in their ability to extract productivity-relevant information and in the weight they place on productivity. We identify these margins separately by leveraging rich data on complete rankings of the same candidate sets before and after schools acquire additional information, which we can benchmark against a direct measure of worker productivity. Heterogeneity in signal-extraction ability accounts for 11 percent of the variance in gains from discretion across schools, while heterogeneity in schools' objectives accounts for 17 percent. Separating the two channels is therefore important for understanding why discretion performs differently across organizations and for designing screening and delegation policies.

2 Context and Data

2.1 The Allocation Mechanism for Permanent Teachers

There are two types of teaching positions in Perú's public schools. Short-term contracts last for up to two years and pay a flat wage. Permanent positions are of indefinite length, carry 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. Bobba et al. (2021) provide a description of the assignment mechanism, the structure of compensation, and the associated sorting implications for teachers who apply for short-term positions. In this paper, we focus instead on the recruitment process of permanent teachers, which combines a centralized, pre-screening stage with a school-based selection stage.

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 a school district and submit a rank-ordered list (ROL) of schools within that school district. The maximum length of the rank-ordered lists submitted by applicants was five schools in 2015 but became unrestricted from 2017 onward.

Teacher candidates are then ranked by test score and assigned sequentially to their three top-ranked schools among those with available interview slots, generating school-specific shortlists of at most ten candidates.[1] An applicant may therefore be evaluated by multiple schools in the same recruitment drive, and we refer to each school-year shortlist as a recruitment contest. Each school evaluates its shortlisted candidates. School committees are led by the school principal, and include the deputy principal (or a designated teacher) alongside another appointed teacher. The committee sequentially deploys three screening devices: (CV) an evaluation of the applicant's curriculum vitae, where candidates' submitted documents are authenticated against their online qualifications and experience declarations; (Inter) a personal interview assessing the candidate's overall fit with the school; and (Class)

[1]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.

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 test-based component and the school-based screening components.

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 school-based evaluation. Our analysis centers on public primary education, which offers the broadest geographic coverage of any schooling level in Perú.

We link this rich dataset on the recruitment process of permanent teachers to additional data sources. Payroll registries track all teachers and school principals in the Peruvian public school system from 2014 to 2023. This information allows us to identify which teacher was selected for a position 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 allow us to construct the distance between each applicant's residence and each school. 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 (including libraries, computers, classrooms, and sports facilities), 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 (ECE - Evaluación Censal de Estudiantes) that measures proficiency in mathematics and Spanish language. This evaluation provides individual test scores for second- and fourth-grade students attending public primary schools in 2016 and 2018, which we can link to their respective classroom and teacher identifiers through an administrative teacher-classroom dataset (SIAGIE - Sistema de Información de Apoyo a la Gestión de la Institución Educativa).

We construct three samples corresponding to the different stages 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 effective-

ness. We retain applicants and vacancies for which the applicant, school, and geographic characteristics entering the choice model 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 with the achievement, demographic, classroom, and school characteristics required by the value-added specification and teachers with sufficient within-teacher variation in student characteristics to estimate the heterogeneous teacher effects described in Section 3. The resulting 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 schools' rankings and estimate the hiring model. We retain shortlisted applicant-school pairs for which candidates participated in the CV, interview, and classroom-observation stages and 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 therefore need not have the same composition. For instance, the applicant choice sample includes applicants and vacancies that ultimately remain unmatched, whereas the outcome sample necessarily requires a realized teacher–classroom match for which student achievement is observed. Similarly, the interview sample contains only schools for which a shortlist of evaluated candidates is observed; schools that receive no relevant applications or do not reach the school-based evaluation stage do not enter this sample. Differences in observable characteristics across the three samples therefore largely reflect these distinct sampling requirements rather than attrition from a common estimation sample. Appendix Table A.1 reports summary statistics for selected applicant and school characteristics across the three samples, together with the corresponding sample sizes.

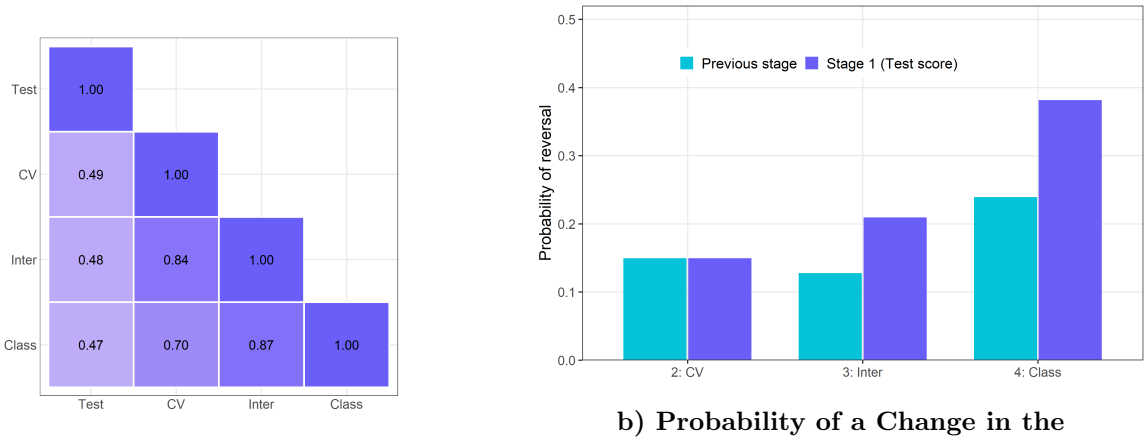
2.3 Discretion and the Interpretation of School Rankings

Our analysis below uses the rankings generated by school committees as revealed measures of 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 both schools' objectives and the additional information acquired during the evaluation process. This interpretation requires that the institutional scoring system leave schools meaningful latitude to reorder candidates, rather than mechanically determining rankings from centralized test scores or applicant characteristics. We therefore examine whether schools exercise discretion in practice and whether the resulting rankings appear to reflect the schools' own evaluations rather than mechanical or strategic features of the assignment process.

Figure 1 shows that schools make substantial use of this discretion. Panel (a) compares candidates' rankings across the successive stages of school evaluation. The pre-interview ranking is only moderately correlated with the centralized test-score ranking, and subsequent evaluation further changes candidates' relative positions. Thus, the school rankings used in the analysis contain substantial variation beyond the ordering generated by the centralized test. Panel (b) asks whether these changes are large enough to affect which candidate would ultimately be favored for appointment. The figure therefore 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 and subsequently included in our model. These characteristics explain part of the variation in CV scores, but leave substantial residual variation, indicating that committees retain considerable discretion even at the first stage of school evaluation. Appendix Figure A.1 provides two additional checks on how this discretion is exercised. Panel (a) shows that the scores assigned during the different stages of school evaluation are not systematically

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. Correlations are computed across candidate–contest observations without normalizing ranks by contest size or reweighting contests. 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.

related to candidates' initial positions in the centralized test-score ranking. Panel (b) similarly shows that, comparing the same applicant across schools, evaluation scores are largely unrelated to the applicant's position in competing shortlists. Thus, schools do not appear to systematically adjust their evaluations either 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 meaningful revealed rankings of candidates by schools under different information sets. We use these rankings in two steps. Section 3 first asks whether selecting candidates according to the observed rankings improves expected student achievement relative to alternative screening rules. Sections 4–5 then formalize how schools' ranking objectives and information acquisition shape the observed changes from the pre- to the post-interview ranking. Throughout the analysis, we hold the realized candidate sets fixed, so both estimates and counterfactuals capture changes in whom schools select from a given shortlist.

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 test score of student ℓ 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_ℓ contains student characteristics along which we allow teacher effectiveness to vary, while Z_ℓ contains additional student, classroom, and school controls. We normalize X_ℓ to have mean zero, so that α_j can be interpreted as the average treatment effect (ATE) of teacher j and β_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 average expected student achievement,

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

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

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

Thus, $Y_{ij} = q_{ij} + Z_i' \gamma$.

Consider now a school contest i with realized candidate set J_i. Throughout, we evaluate alternative hiring rules holding fixed this realized candidate set. Thus, our estimand captures the effect of selecting different candidates from a given shortlist rather than the effect of the recruitment mechanism on the composition of that shortlist.

For any selection rule r, let $j_i(r)$ denote the candidate ranked first 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 take random selection from the same realized candidate set as the benchmark. 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 turn to this measurement problem next.

3.2 Measuring Teacher Effectiveness

We denote $S_\ell \in \{1, \ldots, J\}$ as the realized assignment of student ℓ to teachers and define $Y_\ell = \sum_j Y_{\ell j}\mathbf{1}\{S_\ell = j\}$ as the observed realized test score of student ℓ. The vector X_ℓ in equation (1) includes the student's lagged math test score (2nd grade) and gender.[2] The vector Z_ℓ contains squared lagged math achievement, ethnolinguistic background, and age

[2]This specification for match effects mirrors recent work in the school/teacher value-added literature (Abdulkadiroğlu et al., 2020; Bobba et al., 2021; Ahn et al., 2025; Laverde et al., 2025).

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_ℓ. In particular, $\mathbb{E}[\epsilon_{\ell j} \mid X_\ell, Z_\ell, S_\ell = j]$ may be nonzero because the realized teacher–school matches are selected either because applicants may sort across schools on unobserved determinants of their preferences that are correlated with teacher–school productivity or because schools may rank candidates using private information about unobserved teacher–school match effects.

We first address selection generated by applicants' choices over schools. Following Abdulkadiroğlu et al. (2020), we use the observed rank-ordered lists R_j submitted by applicants to control for selection on unobserved teacher preferences for schools. 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 $u(x_i, z_j, d_{ij}, \tau_j)$ is a flexible function of observed applicant characteristics z_j, school characteristics x_i, distance between applicants and schools d_{ij}, an unobserved type $\tau_j = 1, ..., K$, and an $i.i.d.$ type-I extreme value unobserved shock η_{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 outside-option indicator whose coefficient varies with latent type and observed applicant characteristics, and additionally allow the outside-option intercept to vary across recruitment years.

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 ℓ is enrolled in and $\mu_\eta = \mathbb{E}[\eta_{i(\ell)j}]$. The parameter φ 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

$$\mathbb{E}[Y_{\ell j}|X_\ell, Z_\ell, S_\ell, R_j, z_j, d_j] = \alpha_j + X'_\ell\beta_j + Z'_\ell\gamma + \varphi\lambda_{i(\ell)}(R_j, z_j, d_j), \quad j = 1, ..., J, \tag{6}$$

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 φ requires variables that shift applicants' school choices, and therefore the control-function term, without directly affecting student achievement conditional on X_ℓ and Z_ℓ. We use two such excluded preference shifters: the distance between the applicant's residence and the school and the wage attached to the position. Distance to home is a standard preference shifter in school-choice and teacher-assignment settings (Walters, 2018; Abdulkadiroğlu et al., 2017). Wages are determined by a known compensation schedule based on predetermined school and locality characteristics. Both variables enter applicants' utility (5) but are excluded from the student-outcome equation (1).

Table A.6 assesses these exclusion restrictions using predetermined school outcomes that are not included in the TVA equation. We regress pre-sample student dropout rates, end-of-year examination failure rates, and year-to-year teacher retention on distance and wages, conditional on a rich set of predetermined school characteristics, including the school-level controls entering the outcome equation. The estimated associations are uniformly small in magnitude. In particular, neither distance nor wages display economically meaningful relationships with these predetermined outcomes, supporting the assumption that they shift applicants' preferences without directly capturing omitted determinants of student achievement.

Constructing the control function terms $\lambda(\cdot)$ requires estimating teachers' preferences $u(\cdot)$ in a first step. We assume that teachers rank schools by order of preferences such that we can write the likelihood of observing R_j conditional on $(z_j, \boldsymbol{d_j}, \tau_j)$ as the standard exploded logit formula:

$$\mathcal{L}(R_j | z_j, \boldsymbol{d_j}, \tau_j) = \prod_{i=1}^{N} \left(\frac{\exp\{u(x_i, z_j, d_{ij}, \tau_j)\}}{\sum_{l: iR_j l} \exp\{u(x_l, z_j, d_{lj}, \tau_j)\}} \right)^{1\{iR_j 0\}}.$$

where the denominator sums over the set of alternatives ranked below i denoted as $\{l : iR_j l\}$. Estimating $u(\cdot)$ requires maximizing the sample log-likelihood after integrating out the unobserved types τ_j.[3] We do so by using the EM algorithm with $K = 3$ and under the

[3] Although applicants formally choose a school district before submitting a rank-ordered list of schools within that district, we do not condition the choice set on the realized chosen district. Schools outside the chosen district enter the log-likelihood as unranked alternatives.

following specification for $u(\cdot)$

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

where the vector x_i contains the wage offered in school i, an amenity/infrastructure index,[4] indicators for whether school i belongs to specific categories that are eligible for specific 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 2^{nd} 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.

We construct the control function terms $\lambda(\cdot)$ in a second step extending Dubin and McFadden (1984) to a setting with unobserved discrete types. In particular, for ranked alternatives $\{k : kR_j0\}$, one can write:

$$\mathbb{E}[\eta_{ik} - \mu_\eta | R_j, z_j, \boldsymbol{d_j}, \tau_j] = \log\left(\sum_{l:kR_jl} \exp\{u(x_l, z_j, d_{lj}, \tau_j)\}\right) - u(x_k, z_j, d_{kj}, \tau_j).$$

We then integrate out the unobserved types τ_j as follows:

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

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

$$P(\tau_j = k | z_j, \boldsymbol{d_j}, R_j) = \frac{\pi_k \mathcal{L}(R_j | z_j, \boldsymbol{d_j}, \tau_j = k)}{\sum_{l=1}^{K} \pi_l \mathcal{L}(R_j | z_j, \boldsymbol{d_j}, \tau_j = l)}. \tag{7}$$

We provide additional details as well as formulas for the control functions terms corresponding

[4]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.

to unranked alternatives in Appendix B.1.

Provided that there is sufficient within-teacher variation in X_ℓ, a regression of Y_ℓ on teacher fixed effects, X_ℓ interacted with teacher fixed effects, Z_ℓ and the control function terms yields consistent estimates of $(\alpha_j, \beta_j, \gamma, \varphi)$ denoted as $(\hat{\alpha}_j, \hat{\beta}_j, \hat{\gamma}, \hat{\varphi})$ under these assumptions.[5] The estimated control-function coefficient is small and statistically insignificant: $\widehat{\varphi} = 0.018$ with a standard error of 0.014 in the baseline truth-telling specification (Table A.5). Thus, after conditioning on the rich set of student, classroom, and school controls, selection on unobserved determinants of applicants' school preferences appears to have little empirical effect on the TVA estimates.[6]

We also consider the second source of selection discussed above: school committees may possess private information about match-specific productivity that is not captured by q_{ij}. As a robustness check, 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. If committees possess economically important private information about the residual component of productivity, candidates receiving higher screening scores should subsequently generate higher residual student achievement. Table A.7 shows that the estimated coefficients are uniformly small in magnitude and not systematically positive, providing limited evidence that school evaluations select candidates based on a teacher–school match component omitted from our TVA measure.

We then use these estimates to estimate the prior distribution of the teacher-specific average effectiveness and comparative advantage parameters $\theta_j = (\alpha_j, \beta_j)$. Assuming that the value-added estimates are approximately normal,

$$\hat{\theta}_j | \theta_j, \hat{\Omega}_j \sim \mathcal{N}(\theta_j, \hat{\Omega}_j),$$

[5] 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.

[6] This finding is consistent with Bobba et al. (2021), who obtain a similarly small control-function correction when estimating teacher value added in the same institutional setting. We treat the estimated first-stage preference parameters as fixed when computing uncertainty in the TVA estimates. Given that the estimated control-function coefficient is very close to zero, accounting for uncertainty in the estimated control-function terms should be quantitatively immaterial here; Bobba et al. (2021) find the corresponding adjustment to be negligible in a closely related implementation.

we specify a normal joint prior on θ_j conditional on (z_j, τ_j)

$$\theta_j | z_j, \tau_j \sim \mathcal{N}(\mu_\theta(z_j, \tau_j), \Sigma_\theta). \tag{8}$$

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 θ_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 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 θ_j denoted as θ_j^* corresponding to posterior beliefs about θ_j after observing the estimates $\hat{\theta}_j$

$$\hat{\theta}_j^* = \sum_{k=1}^{K} P(\tau_j = k | z_j, d_j, R_j)(\hat{\Omega}_j^{-1} + \hat{\Sigma}_\theta^{-1})^{-1}(\hat{\Omega}_j^{-1}\hat{\theta}_j + \hat{\Sigma}_\theta^{-1}\hat{\mu}_\theta(z_j, \tau_j = k)). \tag{9}$$

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

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

Our baseline estimates above treat submitted rankings as truthful. Because applicants may instead strategically omit schools they regard as infeasible, we assess an alternative identification strategy following Fack et al. (2019).[7] Under this approach, applicants are assumed to be shortlisted at their preferred schools among those feasible for them, so that

[7] 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, and there is little systematic gradient in application behavior conditional on being feasible. Thus, any strategic behavior appears quantitatively limited and, if anything, takes the form of omitting schools that applicants are unlikely to reach—a behavior explicitly accommodated by the stability-based estimator described here.

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 3.2. 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 Σ_θ. Standard errors are reported in parentheses.

observed assignments reveal choices from personalized feasible sets. Preference estimates under this stability-based specification, reported in Table A.4, differ non-trivially from the truth-telling estimates in Table A.3. These differences in the first-stage preference model, however, have almost no effect on the resulting TVA estimates. Table A.5 shows that the common outcome coefficients and the control-function coefficient are virtually unchanged:

the stability-based estimate of φ is 0.022 with a standard error of 0.021. 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 truth-telling-based estimates throughout the remainder of the paper.

Panel A of Table 1 displays the estimates of the prior mean of θ_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 applicants' preferences over job attributes capture unobserved traits that are correlated with TVA. Panel B reports the estimated prior variance-covariance matrix. Average effectiveness varies substantially in our analysis sample, 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.

3.3 Results

Table 2 reports the average treatment effects defined above 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. Both rankings reflect discretionary school choices, but are made under different information sets. To simplify exposition, and consistently with the empirical framework that we will develop in Section 4, we consider jointly the interview and classroom-demonstration stages and refer to the resulting ranking as the post-interview ranking. The gain relative to random selection increases from 0.011 standard deviations before the interview to 0.025 afterwards. 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 direct evidence that schools learn about candidate characteristics that are predictive of teacher effectiveness. While the average productivity gain may seem small, there is substantial heterogeneity across school contests. The standard deviation of the contest-

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)
N. observations	526	526	526	526
N. schools	492	492	492	492

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. Each contest receives equal weight. 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.

specific gains is 0.189 under the pre-interview ranking and 0.193 under the post-interview ranking, both of which are substantial given the corresponding sample averages at 0.011 and 0.025 standard deviations, respectively.

Table 3 reports the corresponding selection effects on observed candidate characteristics. Test-score ranking generates a large gain in applicants' standardized test scores, 0.334 standard deviations, 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 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, the prior-based rule in Column (5) places sub-

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 (8).

stantially 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.

These findings 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 candidates are precisely the applicants about whom recruiters should have more information. Taken together, this evidence is difficult to reconcile with a purely information-based explanation. It suggests that pre-interview rankings are more likely to reflect schools' objectives over candidate attributes that extend beyond measured teacher effectiveness rather than imperfect, or coarse, beliefs about their teaching effectiveness.

This evidence establishes three facts. First, discretionary screening produces positive average gains in measured teacher effectiveness through learning. Second, the effects of discretionary screening are highly heterogeneous across school contests, with discretion improving selection in some contests but worsening it in others. This heterogeneity is central for policy design: whether and where discretion should be expanded, constrained, or targeted depends on why it works well in some schools but poorly in others. Third, the discretionary evaluation stage changes selection along dimensions such as gender and internal-candidate status that are consistent with objectives other than maximizing teacher effectiveness.

The evidence shown above does not reveal whether these differences arise because schools differ in how closely their objectives are aligned with maximizing teacher effectiveness or in the information they extract from discretionary evaluation. We next develop a model of discretionary hiring and learning that separates these two sources of heterogeneity, while flexibly allowing other determinants of candidate rankings to change across the stages of the hiring process.

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 θ_j posited in Section 3.2, we approximate the type mixture by a normal distribution with posterior-

probability-weighted mean and common within-type variance:

$$q_{ij} \mid z_j, R_j, d_j, X_i \sim N\left(\mu_q(z_j, R_j, d_j, X_i), \sigma_q^2(X_i)\right),$$

where

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

and

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

We account for the fact that schools may have additional information about candidates who are already working at the school. Let I_{ij} denote an indicator for candidate j being internal to school i, and let $q_{ij}^* = \alpha_j^* + \beta_j^{*\prime} X_i$ denote the school-specific effectiveness implied by the shrunken value-added estimates constructed in Section 3.2. We parameterize the additional information available about internal candidates by $\rho \in [0, 1]$. Prior to the interview, schools' beliefs are therefore

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

with conditional variance

$$\mathrm{Var}(q_{ij} \mid z_j, R_j, d_j, X_i, I_{ij}) = (1 - \rho I_{ij})\sigma_q^2(X_i).$$

For external candidates, schools therefore start from the common prior. For internal candidates, the prior mean is shifted toward the candidate-specific value-added estimate and prior uncertainty is reduced. The parameter ρ captures the amount of additional information that schools possess about candidates they have previously employed.

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 σ_γ^2 captures heterogeneity in schools' ability to extract informative signals. The posterior

mean of q_{ij} can be expressed as

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

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

$$\begin{aligned} V_{ij}^g &= \exp(\delta_i)\mathbb{E}[q_{ij} \mid S_{ij}^g] + v^g(z_j, I_{ij}) + \xi_{ij} + \varepsilon_{ij}^g, \\ S_{ij}^1 &= (z_j, d_j, R_j, X_i, I_{ij}), \qquad S_{ij}^2 = (z_j, d_j, R_j, \nu_{ij}, X_i, I_{ij}). \end{aligned} \tag{10}$$

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 not measured by 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 ϵ_{ij}^g, this allows rankings to 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.[8] Standardized applicant test score is therefore excluded as a direct regressor from $v^g(\cdot)$ at either stage. 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

[8] 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 3.2. We suppress these additional arguments to simplify notations.

evaluations and is quantitatively too small to affect strict pairwise candidate orderings.[9] In the empirical specification, $v^g(\cdot)$ includes gender, ethnolinguistic background, public- and private-sector experience, the posterior probabilities of the applicant preference types, and internal-candidate status.

The parameter δ_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 δ_i and γ_i are independent in the population. They need not be independent in their posterior distribution once the observed rankings are taken into account.

Additional unobserved determinants of candidate rankings are captured through a persistent unobserved component ξ_{ij} and a stage-specific type-I extreme-value shock ϵ_{ij}^g, i.i.d. across (i, j, g). The persistent component captures applicant–school attributes or information that affect rankings at both stages but are unobserved to the econometrician, while ϵ_{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.

4.3 Identification and Estimation

At each stage $g = 1, 2$, schools form rankings R_i^g of candidates in J_i. We assume that principals are not strategic and rank candidates in order of expected value V_{ij}^g. 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.

We also assume that the realized set of candidates is exogenous conditional on the observables entering the model, such that $(\epsilon_{ij}^g, \xi_{ij}, \delta_i, \gamma_i, \nu_{ij})$ is orthogonal to J_i. The timing and

[9] 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^2s are below 0.1 percent, and removing the fitted test-score component does not reverse any strict within-contest pairwise ordering.

centralized construction of shortlists make the most direct forms of endogenous selection less salient.[10] Violations of this assumption would require applicants' school choices to depend on private information about persistent unobserved determinants of schools' hiring behavior that is not captured by the observables in the model.

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. Before the interview, schools have not yet received the private signal generated by the discretionary evaluation. Their ranking therefore reveals how they use the information available at that point: the estimated prior over teacher effectiveness, additional information about internal candidates, observable applicant characteristics, and other persistent traits unobserved to the econometrician. Conditional on the estimated prior and the composition of the shortlist, variation in expected teacher effectiveness identifies the weight schools place on expected effectiveness relative to other candidate attributes. The exclusion of applicant test score from $v^1(\cdot)$ provides an additional source of identifying variation because test score predicts expected teacher effectiveness while having no direct effect on rankings outside this channel. Cross-school variation in the relationship between expected teacher effectiveness and rankings identifies the distribution of δ_i. Comparing the ranking of internal candidates with their candidate-specific value-added information identifies ρ, while the average preference for internal candidates is captured by ψ_I^1 and the dispersion of their persistent heterogeneity by $\sigma_{\xi,1}$.

In contrast, changes in rankings after the interview identify the learning technology. If interviews generate informative signals about teacher value-added, candidates whose value-added is higher relative to their pre-interview expectation should tend to move up. The extent to which these ranking changes systematically line up with teacher effectiveness identifies the precision of the interview signal and the distribution of γ_i. At the same time, allowing $v^2(\cdot)$ to differ freely from $v^1(\cdot)$ absorbs systematic changes in rankings associated with observable

[10] 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.

characteristics that need not reflect information about measured teacher effectiveness. Persistence in otherwise unexplained rankings across the two stages identifies the dispersion of ξ_{ij}.

Appendix B.2 formalizes these identification arguments, showing how complete rankings observed before and after the interview, together with repeated contests for the same school, generate distinct variation that separately pins down the preference and learning parameters.

To simplify notation, we suppress the recruitment-year index. When a school is observed in multiple recruitment contests, we assume that its (δ_i, γ_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 (δ_i), signal-noise parameter (γ_i), persistent unobserved component (ξ_{ij}), and interview signals (ν_{ij}) from their respective distributions. Signals are drawn from their posterior distribution conditional on the estimated teacher value-added parameters $(\hat{\theta}_j)$. Given each draw, we compute schools' beliefs about candidate effectiveness at both stages, construct the implied utilities, evaluate the two rank-ordered-logit likelihoods, and average across simulation draws.

5 Estimation Results

5.1 Model Estimates

Tables 4 and 5 report the estimates of the hiring model. These estimates allow us to separate the two sources of heterogeneity left unresolved by the descriptive evidence in Section 3: differences across schools in the weight placed on teacher effectiveness and differences in their

ability to extract information about teacher effectiveness during the discretionary evaluation. We find substantial heterogeneity along both dimensions. At the same time, the estimates suggest that, for many schools, candidate rankings are only weakly aligned with expected teacher effectiveness and that the discretionary evaluation reveals relatively little additional information about productivity.

Table 5 reports functionals of the estimated distributions of δ_i and γ_i. The implied mean of $\exp(\delta_i)$ is 2.37 and its standard deviation is 3.78, indicating substantial cross-school heterogeneity in the weight placed on expected teacher effectiveness. Because the scale of utility is normalized, the behavioral implications of this dispersion are more informative than the magnitude of $\exp(\delta_i)$ itself. Panel (a) of Figure 2 therefore reports, by decile of δ_i, the within-contest rank correlation between the expected-teacher-effectiveness component of utility and overall pre-interview hiring utility. These correlations are generally low, indicating that candidate attributes other than expected teacher effectiveness play an important role in school rankings. The degree of alignment nevertheless varies substantially across schools: among schools in the top decile of δ_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 learning about teacher effectiveness during discretionary evaluation. The average posterior weight placed on the interview signal is 0.23, with a standard deviation of 0.26. Panel (b) of Figure 2 translates the underlying heterogeneity in γ_i into its implications for what schools learn: it reports, by decile of γ_i, the rank correlation between post-interview expected teacher effectiveness and underlying teacher effectiveness. Schools receiving the most precise signals substantially improve their information about candidate productivity, with correlations as high as 0.76, while this relationship becomes considerably weaker as signal noise increases. Thus, the small average gains from learning documented in Section 3 mask substantial heterogeneity in schools' ability to extract productivity-relevant information during the screening process.

The remaining estimates help distinguish these productivity channels from other determinants of school rankings. In particular, Panel C of Table 4 provides essentially no evidence that schools possess additional productivity information about internal candidates before the interview: the estimate of ρ is 0.002 with a standard error of 0.157. Yet schools place a positive direct weight on internal candidates before the interview, and this coefficient rises from 0.192 to 0.811 after discretionary evaluation. The increased selection of internal candidates

Table 4: Estimates of the Hiring Model

Panel A: Preferences for effectiveness and learning

	Estimate	Std. Error
Mean log weight on effectiveness, μ_δ	0.237	(0.243)
SD log weight on effectiveness, σ_δ	1.122	(0.139)
Mean log signal variance, μ_γ	-0.059	(0.524)
SD log signal variance, σ_γ	2.081	(0.878)

Panel B: Other determinants of candidate rankings

	v^1	v^2
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 C: Private information and persistent applicant–school heterogeneity

	Estimate	Std. Error
Private information for internal candidates, ρ	0.002	(0.157)
Persistent heterogeneity: external candidate	1.257	(0.025)
Persistent heterogeneity: internal candidate	1.279	(0.069)

NOTES. Standard errors are reported in parentheses. Panel A reports the parameters governing the population distributions of the school-specific log weight on expected teacher effectiveness, δ_i, and log signal variance, γ_i. Panel B reports the coefficients entering the reduced-form components $v^1(\cdot)$ and $v^2(\cdot)$ of candidate rankings before and after the interview, respectively. Applicant type-1 and type-2 probabilities are the posterior probabilities of the latent applicant preference types constructed in Section 3.2; type 3 is the omitted category. Panel C 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.

documented in Section 3 therefore does not appear to arise because schools have superior information about their measured teacher effectiveness. Instead, it is captured by determinants of rankings outside the productivity component of the model.

Table 5: Functionals of Preferences for Teacher Effectiveness and Signal Extraction

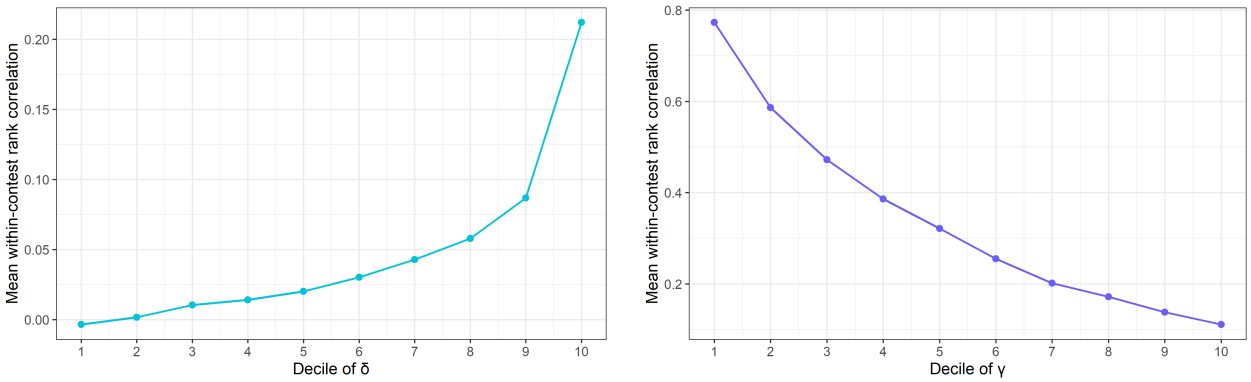
	Estimate	Std. Error
Mean weight on effectiveness, $E[\exp(\delta_i)]$	2.378	(0.350)
SD weight on effectiveness, $SD[\exp(\delta_i)]$	3.774	(0.847)
Mean weight on interview signal	0.234	(0.067)
SD weight on interview signal	0.259	(0.078)

NOTES. $\exp(\delta_i)$ is the weight placed on expected teacher effectiveness. The weight placed on the interview signal for candidate j in school i is $\frac{\sigma^2_{q,ij}}{\sigma^2_{q,ij}+\exp(\gamma_i)}$, where $\sigma^2_{q,ij}$ denotes the remaining uncertainty about teacher effectiveness immediately before the interview. For the signal-weight moments, weights are first averaged across candidates within each school, and the table then reports the mean and standard deviation of this school-level object across the population of schools. These moments integrate over the estimated distribution of γ_i and the observed distribution of prior variances. Standard errors are obtained by the delta method; for the signal-weight moments, they propagate the joint estimation uncertainty in $(\mu_\gamma, \sigma_\gamma)$, holding the prior-variance distribution and other structural parameters fixed at their point estimates.

A similar distinction is useful for interpreting the decline in the selection of female candidates after the discretionary evaluation. The coefficient on female applicants changes from 0.128 before the interview to -0.361 afterwards. This shift is consistent with the descriptive pattern in Section 3, but our approach does not distinguish whether it reflects biased evaluation, systematic gender differences in performance during the school-based evaluation, information about other candidate attributes, or changes in the weights schools place on observable characteristics. Other candidate characteristics also receive sizable weights, including public- and private-sector experience and the latent applicant preference types, and several of these weights change following discretionary evaluation. Finally, the substantial dispersion of the persistent applicant-school component, ξ_{ij}, indicates that rankings also depend heavily on dimensions of applicant desirability that are not observed by the econometrician.

Taken together, the estimates suggest that the modest average effectiveness gains from discretionary screening 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. The model also separates these productivity channels from systematic changes in rankings along other dimensions, such as gender and internal-candidate status.

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}

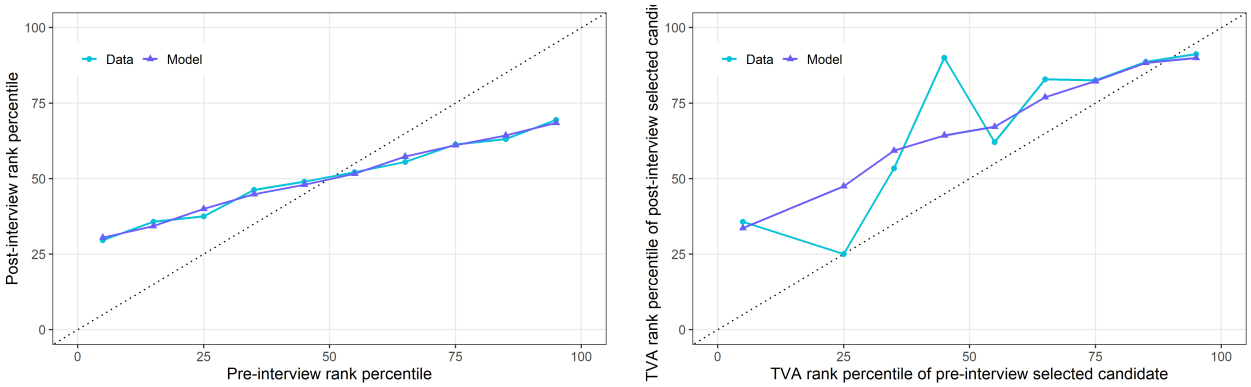
NOTES: Panel A reports, by decile of δ_i, the Spearman rank correlation between the expected-effectiveness component of pre-interview utility and total pre-interview utility. Panel B reports, by decile of γ_i, the Spearman rank correlation between post-interview expected teacher effectiveness and underlying 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 δ_i corresponds to a greater weight on expected teacher effectiveness; lower γ_i corresponds to a more precise interview signal.

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 that need not be aligned with maximizing teacher effectiveness. We therefore examine model fit along several complementary dimensions. 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 as they are in 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.

Panel A of Figure 3 compares the evolution of candidate rankings between the pre- and post-interview stages in the data and in the model. It plots average post-interview rank percentiles against pre-interview rank percentiles, and the two relationships are very similar. The model therefore reproduces both the substantial persistence in candidate rankings and the amount of reshuffling generated by discretionary evaluation. Table A.15 reaches the same conclusion using aggregate measures of ranking mobility: the model closely matches the fraction of contests in which rankings change, the share of candidates who change rank,

Figure 3: Model Fit



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 model. 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.

and the correlation between pre- and post-interview rankings.

Panel B turns specifically to the productivity-relevant component of this reshuffling. It compares the TVA percentile rank of the post-interview top-ranked candidate with that of the pre-interview top-ranked candidate and shows that the model broadly reproduces the relationship observed in the data. Table A.14 provides a complementary comparison using the effects on expected student achievement defined in Section 3. The model does not match the levels of the gains exactly: it 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. However, it captures roughly two-thirds 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 almost exactly the substantial cross-contest dispersion in these gains. Because the model is estimated from the complete rankings rather than by directly targeting these aggregate treatment-effect moments, an exact match in levels is not imposed. The fact that it captures most of the incremental productivity gain from learning as well as its large heterogeneity across contests is reassuring.

Finally, Table 6 assesses whether the model reproduces the observed changes in selection

Table 6: 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.

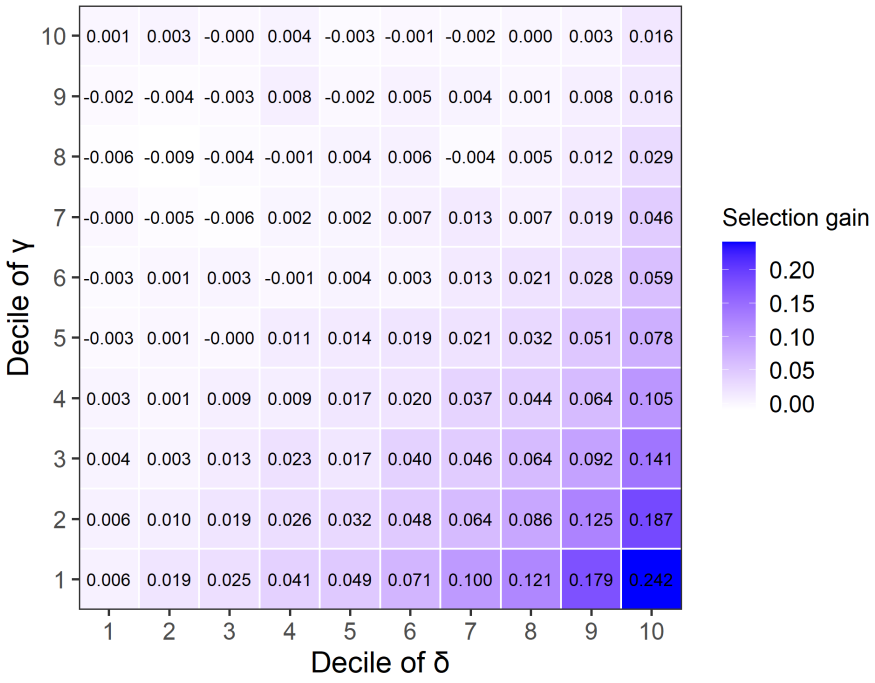
across candidate characteristics. The fit is particularly close. The model reproduces the decline in the selection of female candidates and the increase in the selection of internal candidates after the discretionary evaluation, as well as the changes in selection on public- and private-sector experience, test scores, and ethnolinguistic background.

Taken together, these exercises show that the estimated model closely reproduces the main empirical patterns generated by the two channels we seek to separate. It captures the persistence and reshuffling of candidate rankings, the productivity improvements and substantial heterogeneity associated with learning during discretionary evaluation, as well as systematic changes in selection along dimensions other than measured teacher effectiveness. We therefore use the model next 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.

5.3 Unpacking Aggregate Gains from Discretionary Hiring

We first clarify how we evaluate the gains generated by the estimated model. Our target remains the treatment effect $\tau_i(r)$ defined in equation (4). In Section 3, because the underlying teacher effectiveness q_{ij} is unobserved, we estimate these effects using the shrunken empirical-Bayes estimates $\widehat{q}^*_{ij}$. In the model-based exercises below, we instead draw the latent teacher-effectiveness parameters from their posterior distribution and construct the

Figure 4: Heterogeneity in Gains from Discretionary Evaluation



NOTES: The figure reports the average treatment effect $\tau_i(r)$ of the post-interview ranking rule relative to random selection from the same candidate pool, evaluated using draws of the underlying teacher effectiveness q_{ij}. For each draw, q_{ij} is constructed from a draw of the latent teacher-effectiveness parameters from their posterior distribution and is treated as the underlying effectiveness of candidate j in school i. For schools observed in multiple school-year contests, selection gains are first averaged across contests so that each school receives equal weight. Deciles of δ_i and γ_i are recomputed separately for each draw from their posterior distribution, and each cell averages the corresponding school-level gains across draws. Higher δ_i corresponds to a greater weight on teacher effectiveness, while higher γ_i corresponds to noisier private signals and therefore lower signal-extraction ability.

corresponding q_{ij}. Conditional on a given draw, we treat q_{ij} as the underlying teacher effectiveness and evaluate the selected candidate relative to random selection from the same candidate pool. Figure 4 averages these gains across draws, while the variance decomposition in Table 7 holds one realization of q_{ij}, and the remaining shocks, fixed in order to isolate variation generated by school environments and school types.

Our estimates imply that decentralized screening improves teacher selection on average, but that these gains vary substantially across schools. 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 deciles of δ_i and γ_i. Selection gains are concentrated among schools that place greater weight on expected teacher effectiveness (high δ_i) and receive more precise productivity signals from the interview (low γ_i). For schools with below-median weight on effectiveness and signal-extraction ability, discretionary evaluation generates essentially no gains in expected student achievement.

Table 7: Variance Decomposition of Gains from Discretionary Evaluation

Moment	Estimate	Share of total
Total variance of selection gain	0.083	1.000
School environment	0.063	0.756
Joint heterogeneity in δ_i and γ_i	0.020	0.244
Heterogeneity in δ_i only	0.014	0.170
Heterogeneity in γ_i only	0.010	0.114

Notes. The table reports a crossed school-environment-by-school-type variance decomposition of the treatment effect $\tau_i(r)$ defined in equation (4), where r is the post-interview ranking rule. To isolate variation generated by school objectives and signal extraction, we fix one realization of the underlying teacher effectiveness q_{ij}, persistent applicant–school heterogeneity, interview innovations, and post-interview type-I extreme-value shocks. The fixed q_{ij} is drawn from the posterior distribution of the latent teacher-effectiveness parameters and is treated as the underlying effectiveness of each candidate-school pair. Paired (δ_i, γ_i) types are then drawn from the equal-weight mixture of the school-specific posterior distributions and each type is evaluated against every observed school environment. The school-environment component is the variance across environments of the average $\tau_i(r)$ across school types; the joint school-type component is the average across environments of the variance of $\tau_i(r)$ across school types. The rows labeled δ_i only and γ_i only vary the indicated primitive over its estimated marginal distribution while fixing the other primitive at its median. These one-at-a-time contributions need not sum to the joint contribution because δ_i and γ_i interact nonlinearly in determining the selected candidate and may also be dependent in their posterior distribution.

Table 7 quantifies the extent to which this heterogeneity accounts for cross-sectional variation in screening performance. We cross draws from the estimated joint distribution of (δ_i, γ_i) with the observed school environments and apply the law of total variance. Differences in school environments account for approximately three-quarters of the total variance in selection gains, while heterogeneity in (δ_i, γ_i) accounts for the remaining quarter. Varying δ_i alone accounts for 17 percent of the total variance, while varying γ_i alone accounts for 11.4 percent. These quantities need 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. Hence, both heterogeneity in schools' objectives and heterogeneity in signal-extraction ability are quantitatively important determinants of variation in the effectiveness of discretionary screening.

Importantly, this heterogeneity is difficult to target using observable school characteristics. Tables A.12 and A.13 show that student composition, urban status, principals' experience, and principals' standardized test scores explain almost none of the posterior variation in either δ_i or γ_i. The policy question, therefore, becomes how to neutralize the two wedges identified by the model directly rather than targeting delegation in hiring at pre-specified schools. We next use the estimated model to evaluate counterfactual screening policies that alter two key

margins of hiring discretion: imperfect and heterogeneous learning about teacher effectiveness and hiring objectives that need not be aligned with maximizing teacher effectiveness.

6 Counterfactuals

6.1 Counterfactual Screening Regimes

We first consider a policy that removes the interview and instead discloses standardized information about teacher value added. For applicants with observed TVA, schools receive the TVA estimate together with its estimated variance.[11] Schools incorporate this information into their existing information set and update their beliefs about teacher effectiveness according to the Bayesian structure of the model. The resulting posterior expectation of teacher effectiveness corresponds, for applicants with observed TVA, to the shrunken empirical-Bayes estimate constructed in Section 3.

This policy has a natural interpretation through the empirical-Bayes literature. If schools and the policymaker shared the objective of maximizing the expected effectiveness of the selected teacher, the Bayes-optimal decision would be to select the candidate with the highest posterior expected effectiveness, equivalently minimizing posterior expected regret from selecting a less effective candidate (Gu and Koenker, 2023). The result needs not extend to discretionary hiring. In our model, schools rank candidates according to their own objective, which places positive but heterogeneous weight on expected teacher effectiveness relative to other candidate attributes—see equation (10). The policymaker, instead, evaluates the resulting assignment according to the underlying effectiveness of the selected teacher. The policy therefore has the structure of an information-design problem in which the sender controls the information available to a receiver who takes a non-contractible action under a potentially different objective (Kamenica and Gentzkow, 2011). Additional productivity information can therefore change choices in ways that trade off productivity against other attributes valued by schools, so more information need not increase the policymaker's payoff.

We compare four regimes that differ in the information available to schools and in whether the hiring decision remains delegated. Under the *status quo*, the discretionary evaluation

[11] For applicants without observed TVA, we generate a synthetic TVA estimate centered on their simulated underlying effectiveness and choose its sampling variance so that the average posterior variance implied by Bayesian updating matches that observed among applicants with TVA. The synthetic estimate and corresponding standard error are then disclosed to schools.

takes place and schools' choices depend on both their signal-extraction parameter γ_i and their post-interview objective. Under *TVA disclosure*, the interview stage is removed and all schools receive the same standardized information about teacher effectiveness. Schools update their beliefs according to the Bayesian structure of the model and rank candidates using their pre-interview preferences. This regime removes the role of heterogeneous signal extraction, while preserving school discretion in hiring decisions through the role of δ_i and the other determinants of the pre-interview ranking. Because the discretionary evaluation itself is removed, the comparison with the status quo also eliminates the stage-specific changes in other determinants of candidate rankings captured by the difference between $v^1(\cdot)$ and $v^2(\cdot)$ in equation (10).

We consider two additional regimes aimed at separating these informational and preference margins. Under *full information*, schools observe candidates' underlying teacher effectiveness q_{ij} while retaining their pre-interview preferences. This provides a benchmark in which information frictions are entirely removed while the school objective remains unchanged. In this regime, we assess how much further selection can improve by increasing the precision of productivity information alone. Under the *optimal EB* benchmark, the same statistical information as under TVA disclosure is used to assign the position directly to the candidate with the highest posterior expected effectiveness. Holding the information available about teacher effectiveness fixed, this regime isolates the consequences of replacing the school-specific objective with an assignment rule that maximizes expected teacher effectiveness.

6.2 Results

To evaluate these regimes, we use the same treatment effect parameter $\tau_i(r)$ defined in equation (4). Unlike the empirical estimates in Section 3, which use shrunken estimates of teacher effectiveness on the restricted TVA sample, the model-based counterfactuals evaluate selection using draws of the underlying teacher effectiveness q_{ij} on the full sample. We first average contest-specific gains within school and then across schools, so that each school receives equal weight.

Panel A of Table 8 shows that replacing discretionary evaluation with standardized productivity information substantially improves teacher selection. Under the status quo, the selected teacher's underlying effectiveness is 0.027 standard deviations above random selec-

tion from the same candidate pool. TVA disclosure raises this gain to 0.070. Among the delegated regimes that use the same pre-interview objective, moving from TVA disclosure to full information raises the gain further, from 0.070 to 0.105, isolating the value of more precise productivity information. Conversely, holding fixed the statistical information available under TVA disclosure, the optimal EB benchmark raises the gain from 0.070 to 0.265 by assigning the position to the candidate with the highest posterior expected effectiveness. School-specific objectives therefore generate substantial losses even when compared at the same information set, and the optimal EB rule produces larger gains than decentralized hiring even under full information.

The counterfactuals also illustrate why removing one source of heterogeneity needs not reduce the overall dispersion of screening outcomes. Table 8 shows that the share of selection-gain variance generated by decision-maker heterogeneity rises from 24.4 percent under the status quo to 29.7 percent under TVA disclosure and 40.0 percent under full information. Figure 5 helps unpack this result. Panel B shows that, conditional on a given decile of δ_i, the dispersion of selection gains weakly falls as productivity information becomes more precise, particularly under full information. Standardizing information therefore reduces heterogeneity in screening outcomes among schools with similar objectives. At the same time, Panel A shows that mean selection gains become increasingly concentrated among schools with high δ_i. As the information friction is removed, differences in the weight that schools place on teacher effectiveness become more consequential for how they use the available information. The resulting increase in heterogeneity across δ_i types offsets the decline in dispersion within δ_i types. Under the optimal EB rule, this gradient largely disappears because δ_i no longer enters the assignment rule. Figure 5 therefore illustrates how removing heterogeneity in signal extraction can make heterogeneity in school objectives more salient rather than reducing overall heterogeneity in screening performance.

This mechanism also helps explain why the large efficiency gains from alternative screening policies translate into little change in the overall dispersion of selected teacher effectiveness. The cross-school standard deviation of expected selected teacher effectiveness is 0.309 under the status quo, 0.306 under TVA disclosure, 0.315 under optimal EB, and 0.308 under full information. Improvements in the information available for screening can therefore reduce some sources of heterogeneity while simultaneously making remaining differences in school objectives or environments more consequential. Cross-school dispersion in expected teacher

Table 8: Counterfactual Screening Regimes

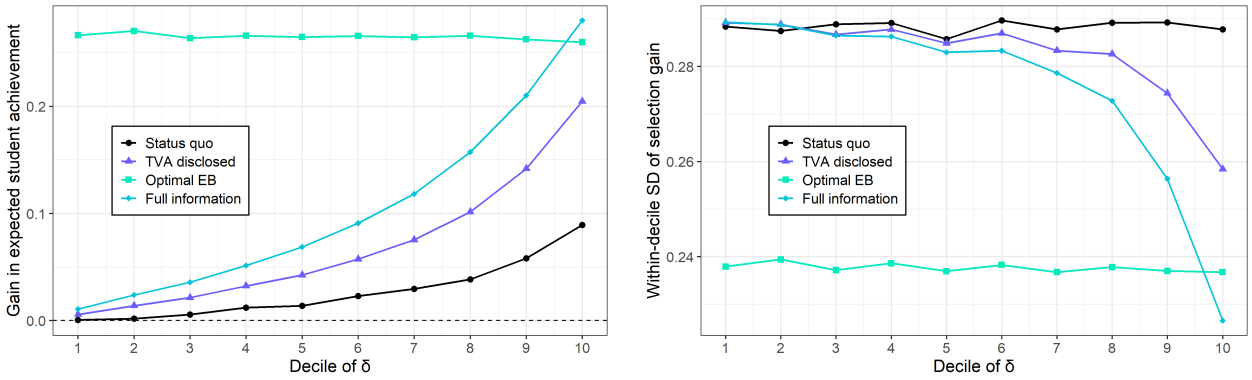
	Status quo	TVA disclosed	Optimal EB	Full information
Panel A. Efficiency and cross-school dispersion				
Mean gain in expected student achievement	0.027	0.070	0.265	0.105
SD of selection gains across schools	0.032	0.045	0.102	0.052
SD of expected selected teacher effectiveness	0.309	0.306	0.315	0.308
School-environment share of selection-gain variance	0.756	0.703	1.000	0.600
Decision-maker heterogeneity share of selection-gain variance	0.244	0.297	0.000	0.400
Panel B. Characteristics of selected candidates				
Female	-0.023	0.009	0.013	0.008
Test score (std.)	0.011	0.008	0.023	0.008
Quechua	0.007	0.006	0.001	0.005
Exp public > 0	0.035	0.055	0.006	0.053
Exp private > 0	-0.016	-0.026	0.021	-0.027
Internal candidate	0.038	0.011	-0.005	0.009
N. contests	3642	3642	3642	3642
N. schools	2874	2874	2874	2874

Notes. All selection gains correspond to the treatment effect $\tau_i(r)$ relative to random selection from the same full candidate set and are evaluated using draws of the underlying teacher effectiveness q_{ij}. Under the status quo, discretionary evaluation takes place and schools rank candidates using their estimated post-interview preferences and school-specific signal precision. Under TVA disclosure, discretionary evaluation does not take place and schools receive standardized empirical-Bayes TVA information while retaining their pre-interview preferences. The optimal EB rule selects the candidate with the highest disclosed posterior expected effectiveness and therefore does not depend on school-specific preferences. Under full information, schools observe underlying teacher effectiveness while retaining their pre-interview preferences. In Panel A, selection gains are first averaged across contests and model simulations within each school, and the table then reports their mean and standard deviation across schools. The variance shares use a crossed school-environment-by-decision-maker-type decomposition analogous to Table 7, holding one simulated realization of the underlying effectiveness and other shocks fixed. The decision-maker type is (δ_i, γ_i) under the status quo, δ_i under TVA disclosure and full information, and is absent under the optimal EB rule. In Panel B, entries report the characteristic of the selected candidate minus the average characteristic of candidates in the same contest. Model moments are averaged across model simulations and all regimes use the full candidate set.

effectiveness therefore remains essentially unchanged across the four regimes.

Finally, Panel B of Table 8 connects the counterfactuals back to the changes in observable selection documented in Section 3. Under the status quo, the model implies that the selected candidate is 2.3 percentage points less likely to be female and 3.8 percentage points more likely to be internal than a randomly selected candidate from the same shortlist, closely reproducing the corresponding post-interview patterns in Table 3. When discretionary evaluation is replaced by standardized TVA information, these differences become +0.9 and +1.1

Figure 5: Counterfactuals



a) Mean selection gains by decile of δ_i **b) SD selection gains by decile of δ_i**

NOTES: Panel A reports the average treatment effect $\tau_i(r)$ relative to random selection from the same candidate pool, evaluated using draws of the underlying teacher effectiveness q_{ij}, by decile of the school-specific weight on teacher effectiveness, δ_i. Panel B reports the corresponding cross-school standard deviation of selection gains within each δ_i decile. For each draw from the posterior distribution, schools are assigned to deciles using that draw of δ_i, and the same decile assignment is used across all counterfactual regimes. The figure compares the status quo, TVA disclosure, the optimal empirical-Bayes benchmark, and full information. Under the status quo, schools rank candidates using post-interview beliefs and preferences. Under TVA disclosure, schools receive standardized TVA information and update beliefs according to the Bayesian structure of the model while retaining their pre-interview preferences. Under full information, schools observe underlying teacher effectiveness while retaining their pre-interview preferences. Under the optimal empirical-Bayes rule, the position is assigned to the candidate with the highest posterior expected effectiveness. Since δ_i does not enter the optimal empirical-Bayes assignment rule, differences across δ_i deciles under this benchmark reflect correlation between estimated school type and the underlying school environment rather than a direct effect of δ_i on selection.

percentage points, respectively, close to the pre-interview selection patterns in the data. Under optimal EB they are +1.3 and −0.5 percentage points, while under full information they are +0.8 and +0.9 percentage points. Together with the near-zero estimate of additional pre-interview productivity information about internal candidates (see Panel C in Table 4), these results suggest that the post-interview increase in the selection of internal applicants is not primarily explained by superior information about their productivity.

7 Conclusion

We study discretionary hiring in Peru's market for permanent public-school teachers using administrative data that link applicants' school choices, complete school rankings before and after discretionary evaluation, and subsequent student outcomes. We first estimate a flexible teacher value-added model that allows for heterogeneity in average effectiveness and comparative advantage across students, correcting for endogenous teacher sorting using applicants' rank-ordered school choices. Discretionary evaluation raises expected student achievement on average by revealing information about teacher effectiveness, but also shifts selection along

other dimensions and generates highly heterogeneous effects across schools. We then estimate a model in which schools differ both in the weight they place on student achievement and in their ability to extract productivity-relevant information during screening. The estimates imply substantial heterogeneity along both margins, which together account for about one quarter of the cross-school variance in the gains from discretion. Gains are concentrated among schools that both place greater weight on student achievement and extract more precise signals. Counterfactuals show that standardizing productivity information substantially improves average selection, but does little to reduce cross-school inequality because differences in schools' objectives become more consequential as information improves.

These results highlight demand-side hiring frictions as an important part of teacher labor-market design. Policies that improve the supply or allocation of effective teachers can only translate into better student outcomes if schools subsequently identify and select them from the applicant pool. Our counterfactuals also show that improving information alone need not eliminate disparities in screening performance: removing heterogeneity in signal extraction can make heterogeneity in objectives more consequential. Effective screening policy therefore requires accounting jointly for information and agency frictions rather than treating either in isolation.

A natural direction for future work is to study how organizations can improve these two margins directly—for example, whether signal-extraction ability can be measured or improved through screening protocols and training, and how the disclosure of standardized information can be combined with local discretion when decision makers pursue heterogeneous objectives. These issues extend well beyond the recruitment of teachers, especially in settings where worker productivity is difficult to observe and discretionary hiring remains central to selection.

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.

Aghion, Philippe and Jean Tirole, "Formal and real authority in organizations," *Journal of political economy*, 1997, *105* (1), 1–29.

Ahn, Tom, Esteban M Aucejo, and Jonathan James, "The Importance of Student-Teacher Matching: A Multidimensional Value-Added Approach," *Review of Economics and Statistics*, 2025, pp. 1–45.

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

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.

Dubin, Jeffrey A and Daniel L McFadden, "An econometric analysis of residential electric appliance holdings and consumption," *Econometrica*, 1984, pp. 345–362.

Ederer, Tim, "Labor Market Dynamics and Teacher Spatial Sorting," *Job Market Paper*, 2023.

Estrada, Ricardo, "Rules versus Discretion in Public Service: Teacher Hiring in Mexico," *Journal of Labor Economics*, 2019, *37* (2), 545–579.

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.

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

Gu, Jiaying and Roger Koenker, "Invidious comparisons: Ranking and selection as compound decisions," *Econometrica*, 2023, *91* (1), 1–41.

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

Jacob, Brian A and Lars Lefgren, "Can principals identify effective teachers? Evidence on subjective performance evaluation in education," *Journal of labor Economics*, 2008, *26* (1), 101–136.

_ , Jonah E Rockoff, Eric S Taylor, Benjamin Lindy, and Rachel Rosen, "Teacher applicant hiring and teacher performance: Evidence from DC public schools," *Journal of Public Economics*, 2018, *166*, 81–97.

Jovanovic, Boyan, "Job matching and the theory of turnover," *Journal of Political Economy*, 1979, *87* (5, Part 1), 972–990.

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.

Kamenica, Emir and Matthew Gentzkow, "Bayesian persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

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.

Lessem, Rebecca and Robert A Miller, "Matching Job Applicants to Vacancies: An Empirical Model of Multistage Hiring," 2026.

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.

_ **and Cecilia Speroni**, "Subjective and objective evaluations of teacher effectiveness," *American Economic Review*, 2010, *100* (2), 261–266.

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 in the natural units of each sample.

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 the applicant school-choice utility 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. The latent preference types are integrated out using the EM algorithm. 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: Estimates Teacher Value-Added: γ

	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, φ	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 φ 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 (α_j, β_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.6: 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.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: Observable Predictors of Estimated Preferences and Signal Extraction

	Posterior mean δ_i		Posterior mean γ_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 δ_i and γ_i. Higher δ_i corresponds to a greater weight on teacher effectiveness, while higher γ_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.13: Observable Characteristics by Estimated School Type

	Overall	Low δ, low γ	Low δ, high γ	High δ, low γ	High δ, high γ
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 δ_i and γ_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 γ_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.

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

	Pre-interview		Post-interview	
	Data	Model	Data	Model
Expected student achievement gain	0.011	0.002	0.025	0.011
SD of gains across contests	0.189	0.193	0.193	0.193
N. observations	526	526	526	526
N. schools	492	492	492	492

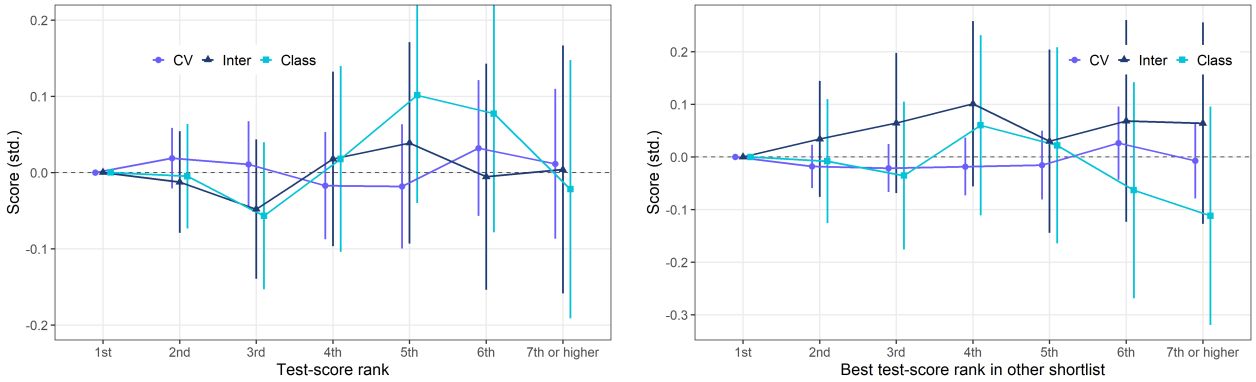
Notes. The table compares the treatment effects of the pre- and post-interview ranking rules with the corresponding model predictions. As in Table 2, treatment effects are evaluated using the shrunken empirical-Bayes estimates $\widehat{q}^*_{ij}$: for each contest, the effect is the shrunken effectiveness 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: 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.

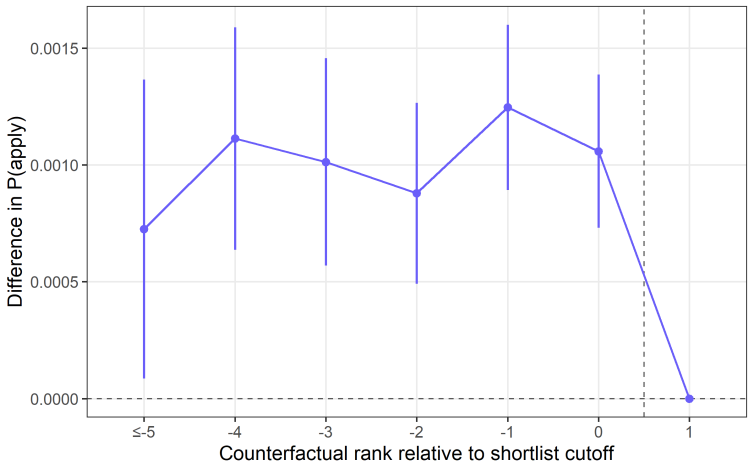
Figure A.1: School Evaluation Scores and Candidate Ranking Positions



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

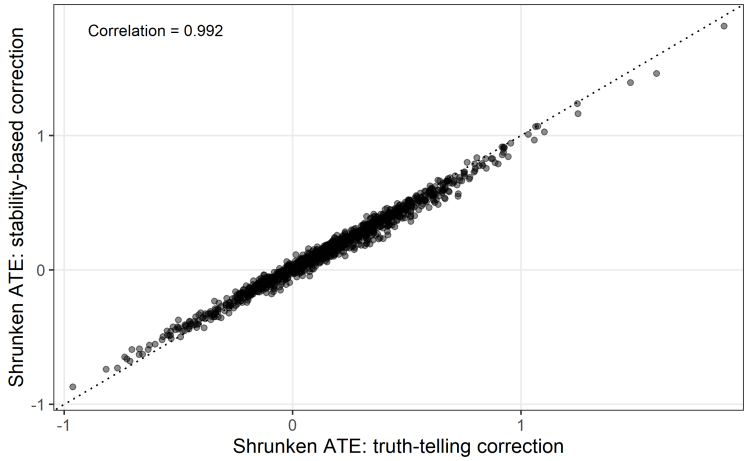
NOTES: Panel (a) relates standardized CV, interview, and classroom-observation scores to the applicant's centralized test-score rank in the focal school-year contest. Rank 1 is omitted and ranks 7 or worse are pooled. The regressions control flexibly for test score and observed applicant characteristics 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 used in Section 3.2. 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 η_{ij} is i.i.d. Type-I extreme value with mean μ_η.

We represent the observed rank-ordered list as a sequence of conditional choice problems. Let $\mathcal{A}_{jm}$ denote the set of alternatives remaining at ranking occasion m, after removing the alternatives ranked above m, and let r_{jm} denote the alternative chosen at that occasion. Conditional on type h, the probability that alternative $i \in \mathcal{A}_{jm}$ is chosen at occasion m is

$$p^m_{ijh} = \frac{\exp(u_{ijh})}{\sum_{\ell \in \mathcal{A}_{jm}} \exp(u_{\ell jh})}. \tag{11}$$

The probability of the observed ranking conditional on type h is therefore

$$\mathcal{L}_j(h) = \prod_{m=1}^{M_j} p^m_{r_{jm}jh}, \tag{12}$$

where the outside option is appended as the terminal chosen alternative. Thus, at the final ranking occasion, the applicant chooses the outside option over all schools that remain unranked.

For a standard logit choice problem, the Dubin–McFadden generalized residual for an alternative that is chosen from a choice set is

$$E\left[\eta_{ij} - \mu_\eta \mid i = r_{jm}, \mathcal{A}_{jm}, \tau_j = h\right] = -\log p^m_{ijh}. \tag{13}$$

Accordingly, if school i is ranked at position m in applicant j's submitted list, we construct the type-specific control-function term as

$$\lambda^R_{ijh} = -\log p^m_{ijh}. \tag{14}$$

We also require a control-function term for schools that applicant j does not rank. Let M_j denote the terminal choice occasion, at which the outside option is chosen, and let $p_{ijh}^{M_j}$ denote the conditional choice probability of an unranked school i in the choice set that remains at this occasion. Because $E[\eta_{ij} - \mu_\eta] = 0$, the law of total expectation implies

$$0 = p_{ijh}^{M_j} E[\eta_{ij} - \mu_\eta \mid i \text{ chosen}, \tau_j = h] + (1 - p_{ijh}^{M_j}) E[\eta_{ij} - \mu_\eta \mid i \text{ not chosen}, \tau_j = h].$$

Using equation (13), the generalized residual for an unranked alternative is therefore

$$\lambda_{ijh}^{U} = \frac{p_{ijh}^{M_j}}{1 - p_{ijh}^{M_j}} \log p_{ijh}^{M_j}. \tag{15}$$

Hence, ranked and unranked schools receive different generalized residuals because the former are observed as chosen alternatives at one of the exploded-logit choice occasions, whereas the latter are observed as not chosen when the applicant terminates the ranking by selecting the outside option.

The latent preference type τ_j is not observed by the econometrician. Let π_h denote the population probability of type h. After estimating the preference model, we use the applicant's complete observed ranking to form the posterior probability

$$\omega_{jh} \equiv P(\tau_j = h \mid R_j, z_j, d_j) = \frac{\pi_h \mathcal{L}_j(h)}{\sum_{s=1}^{K} \pi_s \mathcal{L}_j(s)}. \tag{16}$$

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

$$\lambda_{ij}(R_j, z_j, d_j) = \sum_{h=1}^{K} \omega_{jh} \lambda_{ijh}, \qquad \lambda_{ijh} = \begin{cases} \lambda_{ijh}^{R}, & \text{if school } i \text{ is ranked,} \\ \lambda_{ijh}^{U}, & \text{if school } i \text{ is unranked.} \end{cases} \tag{17}$$

In the baseline specification, $K = 3$. The posterior weights in equation (16) are applicant-specific and summarize what the submitted ranking reveals about the applicant's latent preference type. The same posterior weights are then used to integrate the school-specific generalized residuals in equation (17).

Under the alternative stability-based identification strategy described in Section 3.2, the construction is unchanged except that each exploded-logit choice occasion is defined over the

applicant's feasible choice set rather than the full set of schools implied by truth-telling. We recompute both the conditional choice probabilities and the posterior type probabilities under that alternative preference model before constructing the corresponding control-function terms.

B.2 Identification of the Hiring Model

To make the panel structure explicit, let $c \in \mathcal{C}_i$ index recruitment contests observed for school i. The school-specific parameters (δ_i, γ_i) are constant across contests, while candidate sets and candidate–school shocks vary across c. We condition throughout on the realized candidate sets.

Identification of δ_i. Before the interview,

$$V^1_{icj} = \exp(\delta_i) E[q_{icj} \mid S^1_{icj}] + v^1(z_j, I_{icj}) + \xi_{icj} + \epsilon^1_{icj}.$$

For two candidates j and k in the same contest, define

$$\Delta q^1_{icjk} = E[q_{icj} \mid S^1_{icj}] - E[q_{ick} \mid S^1_{ick}]$$

and

$$\Delta v^1_{icjk} = v^1(z_j, I_{icj}) - v^1(z_k, I_{ick}).$$

Conditional on the persistent shocks, the probability that j is ranked above k is strictly increasing in

$$\exp(\delta_i)\Delta q^1_{icjk} + \Delta v^1_{icjk}.$$

Applicant test score is excluded as a direct regressor from $v^1(\cdot)$ but shifts expected teacher effectiveness. Hence, conditional on the remaining determinants of $v^1(\cdot)$, variation in test scores generates excluded variation in Δq^1_{icjk}. If the support of this variation is nondegenerate, the mapping from δ_i into pre-interview ranking probabilities is injective.

Full rankings provide multiple such comparisons within each contest. Moreover, for schools observed in several contests, the same δ_i must rationalize rankings across different candidate sets and realizations of Δq^1_{icjk}.

Identification of γ_i. Let

$$s_{icj}^2 = (1 - \rho I_{icj})\sigma_q^2(X_{ic})$$

denote uncertainty about candidate effectiveness immediately before the interview. The updating equation implies

$$\begin{aligned} &E\left[E[q_{icj} \mid S_{icj}^2] - E[q_{icj} \mid S_{icj}^1] \mid q_{icj}, S_{icj}^1\right] \\ &\quad = \frac{s_{icj}^2}{s_{icj}^2 + \exp(\gamma_i)}\left(q_{icj} - E[q_{icj} \mid S_{icj}^1]\right). \end{aligned}$$

Thus the responsiveness of posterior beliefs to the productivity innovation

$$q_{icj} - E[q_{icj} \mid S_{icj}^1]$$

is

$$w_{icj}(\gamma_i) = \frac{s_{icj}^2}{s_{icj}^2 + \exp(\gamma_i)},$$

which is strictly decreasing in γ_i. In particular,

$$\exp(\gamma_i) = s_{icj}^2 \frac{1 - w_{icj}}{w_{icj}}.$$

Hence signal precision is uniquely determined by the degree to which post-interview beliefs respond to productivity innovations.

The ranking data reveal this response through changes in relative rankings between the two stages. Conditional on δ_i and on the stage-specific observable components $v^1(\cdot)$ and $v^2(\cdot)$, stronger alignment between post-interview rank changes and candidate-specific productivity innovations implies a lower γ_i. Because the same school is observed across multiple candidate comparisons, and potentially across multiple contests, the model observes repeated realizations of this relationship for a common γ_i.

Persistent heterogeneity and the distribution of school types. The persistent component ξ_{icj} enters both rankings within a contest, whereas the type-I extreme-value shocks are stage specific. Conditional on the systematic utility components, cross-stage rank persistence therefore identifies the dispersion of ξ_{icj}.

Let $\mathcal{R}_i$ collect all rankings observed for school i, and let $\mathcal{X}_i$ collect the corresponding

candidate sets and covariates. After integrating out contest-specific shocks and signals, write the school-level likelihood as

$$K_i(\mathcal{R}_i \mid \delta_i, \gamma_i, \mathcal{X}_i) = \prod_{c \in \mathcal{C}_i} K_{ic}(\mathcal{R}_{ic} \mid \delta_i, \gamma_i, \mathcal{X}_{ic}).$$

The population distribution $G(\delta, \gamma)$ then satisfies

$$P(\mathcal{R}_i \mid \mathcal{X}_i) = \int K_i(\mathcal{R}_i \mid \delta, \gamma, \mathcal{X}_i)\, dG(\delta, \gamma).$$

Identification of the mixing distribution requires the usual injectivity condition for this mixture representation. The relevant source of variation is particularly rich in our setting: each contest provides complete rankings under two information sets, while repeated contests provide different candidate pools for the same persistent school type. Our baseline normality assumptions on δ_i and γ_i provide a parsimonious parameterization of this mixing distribution rather than the source of the identifying variation.