Discretionary Hiring and Inequality: Evidence from School Principals

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

March 21, 2025

Abstract

We exploit rich administrative data on the centralized teacher recruitment drive in Peru to quantify how school principals' hiring decisions shape the distribution of student achievement. Discretionary hiring selects slightly higher value-added teachers than test-based hiring, on average, but remains largely misaligned with the objective of maximizing student achievement. We find that this average hides substantial heterogeneity that we quantify using rich data on principals' ranking behavior. Heterogeneity in hiring practices generates a standard deviation of 0.16σ in the value-added of hired teachers, indicating that discretionary hiring largely contributes to inequality in student achievement.

(Preliminary and incomplete)

*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, gianmarco.leon@upf.edu
§Yale University & NBER, christopher.neilson@yale.edu
¶University of Cagliari & CRENoS, mgnieddu@unica.it

1 Introduction

The ability of managers to identify and hire high-performing workers is a fundamental challenge in labor markets, particularly when worker productivity is unobserved (Akerlof, 1970; Spence, 1973). This challenge becomes even more relevant when wage rigidities limit an organization's ability to leverage pay differentials in order to attract highly productive workers (Gibbons and Waldman, 1999; Lazear, 1995; Oyer and Schaefer, 2011). In the education sector, where the quality of teachers is a key determinant of student achievement (Rockoff, 2004; Chetty et al., 2014b), understanding how school principals select teachers is a critical issue.

In this paper, we investigate how school principals, as hiring managers, assess candidates for teaching positions and whether their selection decisions lead to the hiring of effective teachers. We leverage rich administrative data on the centralized teacher recruitment drive in Peru to gather detailed information on the universe of applicants and job postings and on every step of the hiring process. We document that principals' hiring decisions select slightly higher value-added teachers than test-based hiring, on average, but remains largely misaligned with the objective of maximizing student achievement. We further find that this average hides substantial heterogeneity as some principals tend to systematically select more effective teachers than others. We quantify that this heterogeneity generates a standard deviation of 0.16σ in the value-added of the hired teachers, thereby largely contributing to inequality in student achievement.

Our analysis draws upon a rich array of administrative datasets containing detailed information on the national recruitment drive for permanent public teacher positions over the period 2015-2019. We observe the universe of applicants and job postings, the set of short listed candidates evaluated by each school principal, the final candidate ranking in each school and students' academic performance. We complement these data with detailed information on applicant, principal, school and student characteristics. We leverage teacher-student linked data to estimate a flexible teacher value-added model where the treatment effect of a given teacher is allowed to be heterogeneous and vary with students' prior achievement measures and demographics.

We first document that principals use their discretion to select candidates in a way that, on average, performs slightly better than test-based ranking but is overall poorly aligned with

the objective of maximizing student achievement. School principals select teachers that have a value-added 0.02σ higher than the average applicant. On the other hand, ranking applicants based on value-added would select teachers that have a value-added 0.20σ higher than the average applicant. We further show that principals positively select internal candidates, i.e. candidates that they already know, irrespective of whether they are high or low value-added. This suggests that principals principals may have other motives beyond maximizing student achievement when selecting candidates explaining this misalignment. However, we find substantial heterogeneity in hiring practices. School principals in urban schools tend to positively select candidates based on their value added by 0.04σ while principals in rural schools tend to negatively select high value-added teachers by -0.05σ.

To fully exploit data on the ranking behavior of principals to quantify and characterize the extent of this heterogeneity, we build an empirical model of principals' hiring decisions. This approach also allows us to quantify how this heterogeneity feeds into inequality in teacher value-added and student achievement. Our findings reveal that, on average, principals put a low weight on teacher value-added relative to other directly observed measures of teacher quality. However, we also report substantial heterogeneity in how principals weight teacher characteristics relative to each other. A one standard deviation increase in the value-added coefficient more than triples its importance relative to other characteristics indicating that some principals might be more closely aligned to ranking candidates based on value-added than what is suggested by the mean.

Finally, we quantify how heterogeneity in principals' hiring practices contributes to the variation in the characteristics of the teachers ultimately hired. We find that variation in hiring behavior generates a standard deviation of 0.16σ in the value-added of the hired teachers. Overall, these findings suggest that giving discretion in hiring might perform better than rule-based hiring but at the cost of increased inequality caused by heterogeneity in hiring practices.

Our findings contribute to two strands of literature. First, we contribute to the organizational economics literature on managerial decision-making and hiring practices, particularly in public-sector settings (Bertrand and Schoar, 2003; Bloom et al., 2015). Existing research has examined how managers grapple with unobserved worker productivity and the balance between discretion and formal rules. This paper leverages features of the Peruvian education sector, where administrative data on multiple stages of the hiring process are available

and where we can assess the performance of hiring decisions using estimated measures of value-added, to provide novel insights regarding the role of hiring discretion in generating inefficiencies and inequality.

Second, we contribute to a literature studying the determinants of inequality in student achievement. While there exists a large body of work studying how teachers contribute to student achievement (Rivkin et al., 2005; Chetty et al., 2014a; Bau and Das, 2020; Ahn et al., 2023), our study highlights the critical role of worker selection in shaping student outcomes. By showing that principals exhibit substantial heterogeneity in their hiring behavior and that this leads to substantial variation in the value-added of hired teachers, we underscore a previously less emphasized mechanism through which disparities in teacher quality and student achievement across schools may arise.

2 Context and Data

2.1 Teacher Recruitment and the Role of School Principals

Permanent public teaching positions in Peru are allocated through a two-sided matching mechanism that combines centralized testing with decentralized school-level evaluation. Each applicant first takes the Prueba Única Nacional (PUN), a standardized national test assessing reading comprehension, logical reasoning, and pedagogical knowledge. Only candidates scoring above 60% in each component are eligible for permanent positions, while lower-scoring applicants may pursue short-term contract positions. Once, the Ministry of Education makes the list of available permanent positions publicly available, applicants eligible for a permanent position choose a province and submit a rank order list of schools within this province.[1] Each school then receives a shortlist of up to ten candidates that they evaluate in a decentralized stage.[2] A variant of serial dictatorship is used to assign applicants to these shortlist slots: applicants are ranked according to their test scores and are assigned sequentially to their three top ranked schools, among those that still have available slots in their short list.

The decentralized stage has three distinct assessment components administered by each school's evaluation committee. This committee is chaired by the school principal and includes the deputy principal (or a designated teacher) alongside another appointed teacher from the

[1] The maximum length of the list was limited to 5 in 2015 but became unrestricted from 2017 onward.
[2] The maximum length of the shortlist was 20 in 2015 and reduced to 10 in 2017.

same educational level or modality as the candidate.[3] The evaluation committee employs three assessment instruments: (1) a professional background verification, where candidates' submitted documents are authenticated against their online qualifications and experience declarations, i.e. an evaluation of the candidates' CV; (2) a personal interview assessing alignment with the school's educational mission and teaching vocation; and (3) a classroom observation requiring candidates to deliver a demonstration lesson. The first two components are worth up to 25 points each, while the classroom observation carries greater weight at 50 points maximum. The final candidate ranking integrates scores from both the centralized test score (maximum 200 points) and the decentralized evaluation components, with positions offered to the highest-ranked candidates.[4]

School principals play a key role in the final selection of teachers, as they are the most senior members of the evaluation committee. Principals must have at least five years of experience working as a teacher in a public school, and are appointed either directly by the school districts (prior to 2014), or through a centralized recruitment drive, which is conducted in parallel to the teacher recruitment drive.

2.2 Data

We focus our analysis on the assignment of primary school positions. We combine several administrative datasets containing detailed information on the national recruitment drive for permanent public teacher positions over the period 2015-2019. We observe the universe of applicants and job postings, the set of short listed candidates evaluated by each school principal, the final candidate ranking in each school and students' academic performance. We complement these data with detailed information on applicant, principal, school and student characteristics.

More specifically, we use the *MoE's administrative records* payroll dataset to track all teachers and principals that were in the Peruvian public school system from 2014 to 2020.

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

[4] About 24% of the candidates failed to submit the required documents at this stage, and a few others (less than 1%) did not comply with the minimum requirements. The average candidate obtains 7.02 points (s.d. 4.05) in this evaluation. 97% of candidates who go through the first stage of the process show up for the interview, and the average candidate obtained 19.54 points (s.d. 4.05). Finally, 96% of the candidates who did the interview also showed up to the classroom evaluation. Among these applicants, the average grade obtained is 16.94 points (s.d. 4.35).

This allows us to gather information about the career trajectories of both teachers and principals and study their influence on principals' hiring decisions. We combine these data with the *centralized assignment data* collecting information on all stages of the recruitment drive in each year. We observe the list of all vacancies, all applicants who participated in the centralized test, and their scores for each test module. This corresponds to a total of 31,700 vacancies and 51,500 applicants. We also observe the list of candidates shortlisted for each position and the scores assigned at each stage of the evaluation. This allows us to track how candidate rankings evolve throughout the selection process. Additional statistics on the teachers who apply to the assignment mechanism, the school that post vacancies are shown in Appendix Table A.1.

We enhance our primary dataset with several auxiliary sources containing additional information on schools and students. The *school census* offers detailed information on teacher composition, infrastructure (including libraries, computers, classrooms, and sports facilities), and access to essential amenities such as electricity, water, internet, banking, and public libraries. Student academic performance data is derived from the 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, with a 97% participation rate among eligible students. Through an administrative teacher-classroom dataset maintained by the Ministry of Education (SIAGIE - Sistema de Información de Apoyo a la Gestión de la Institución Educativa) we are able to recover the teacher identifier corresponding to each classroom and link this information to the centralized assignment dataset.

2.3 Measuring Teacher Value-Added

Throughout the rest of the analysis, we use two measures of teacher effectiveness: (i) scores from the national competency test and (ii) estimates of teacher value-added derived from the matched teacher-student dataset. We denote Y_{ij} the potential value of student i's 4th grade math test score if taught by teacher $j \in \{1, ..., J\}$ such that

$$Y_{ij} = \alpha_j + X'_{i1}\beta_j + X'_{i2}\gamma + \epsilon_{ij},$$

where $\mathbb{E}[\epsilon_{ij}] = 0$, $\mathbb{E}[X_{i1}\epsilon_{ij}] = 0$ and $\mathbb{E}[X_{i2}\epsilon_{ij}] = 0$ by construction. We normalize X_{i1} to be mean zero such that α_j can be interpreted as the average treatment effect (ATE) of teacher j and β_j captures the returns to X_{i1} when being exposed to teacher j. This model for potential outcomes allows teachers to have comparative advantages in their ability to teach students with different demographics or background.

We denote $D_i \in \{1, ..., J\}$ as the realized assignment of student i to teachers and define $Y_i = \sum_j Y_{ij}\mathbb{1}\{D_i = j\}$ as the observed realized test score of student i. We assume selection on observables such that

$$\mathbb{E}[Y_{ij}|X_{i1}, X_{i2}, D_i] = \alpha_j + X_{i1}'\beta_j + X_{i2}'\gamma, \quad j = 1, ..., J,$$

where X_{i1} includes student i's lagged math test scores (2nd grade) and gender, and X_{i2} includes the variables in X_{i1} averaged at the school and classroom level, as well as squared lagged test scores, ethnicity and age at the student level, averaged at the classroom level and averaged at the school level. When measures of lagged outcomes are available, selection on observables has been shown to be a relatively plausible assumption yielding estimates of teacher and school value-added that feature little to no bias (Chetty et al., 2014a; Angrist et al., 2017). Under this assumption and provided that there is sufficient within-classroom variation in X_{i1}, a regression of Y_i on teacher fixed effects, X_{i1} interacted with teacher fixed effects, and X_{i2} yields unbiased estimates of $(\alpha_j, \beta_j, \gamma)$ denoted as $(\hat{\alpha}_j, \hat{\beta}_j, \hat{\gamma})$.

To avoid any attenuation bias resulting from using the noisy estimates $\hat{\alpha}_j$ and $\hat{\beta}_j$ as inputs in our analysis of principals' hiring decisions, we instead use the linear shrinkage estimator. We denote $\theta_j = (\alpha_j, \beta_j)$ and treat $\hat{\theta}_j$'s as normally distributed with variance estimates denoted as $\hat{\Omega}_j$. We assume that the θ_j's are drawn i.i.d from a common prior $\theta_j|\hat{\Omega}_j \sim \mathcal{N}(\mu_\theta, \Sigma_\theta)$. We estimate $(\mu_\theta, \Sigma_\theta)$ by MLE and construct the estimated posterior means for the teacher value-added coefficients θ_j

$$\hat{\theta}_j^* = (\hat{\Omega}_j^{-1} + \hat{\Sigma}_\theta^{-1})^{-1}(\hat{\Omega}_j^{-1}\hat{\theta}_j + \hat{\Sigma}_\theta^{-1}\hat{\mu}_\theta).$$

Throughout the rest of the analysis we use $\hat{\alpha}_j^*$ as a measure of teacher j's average effectiveness (ATE) and $X_{i1}'\hat{\beta}_j^*$ as a measure of teacher j's comparative advantage with students with characteristics X_{i1} (match effect). We refer to the the sum of the ATE and the match effect component as the value-added of teacher j for student i.

Table 1: Estimates Teacher Value-Added Distribution

			Correlation	
	Mean	SD	Lag Score	Female
	(1)	(2)	(3)	(4)
Average Effectiveness (ATE)	0	0.401 (0.006)	0.096 (0.027)	0.239 (0.075)
Lagged Score	0.504 (0.004)	0.156 (0.004)	-	0.068 (0.076)
Female	-0.077 (0.005)	0.095 (0.010)	-	-

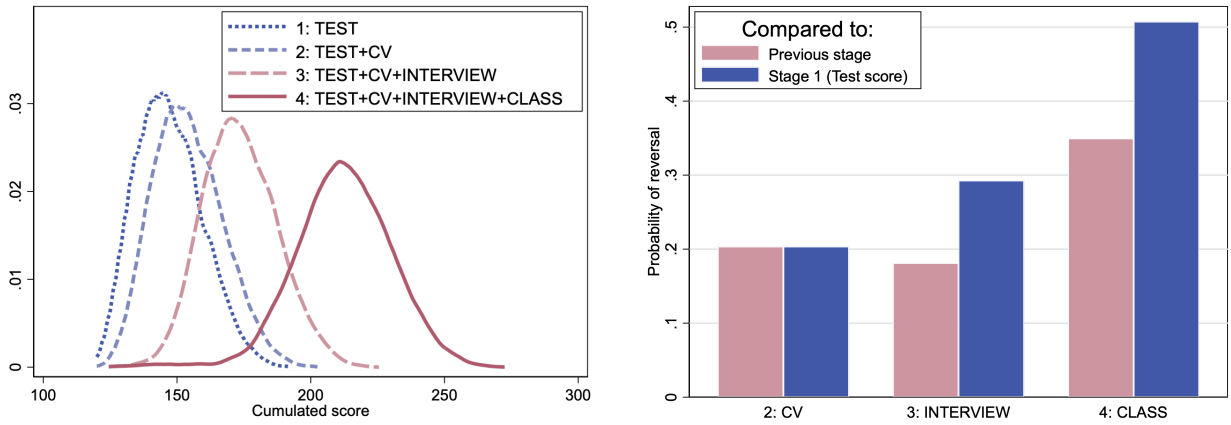
Table 1 displays estimates of the distribution of θ_j (see Table A.2 for estimates of γ). We find that the ATE varies substantially in the population of teachers (0.40σ standard deviation). Teachers also largely differ in their comparative advantages. The standard deviation of the slope coefficients of lagged scores and female are 0.16σ and 0.10σ, respectively. Teachers that are highly effective on average are also more likely to be effective with female students and with students having higher initial levels of achievement. Other measures of teacher quality, such as teacher test scores or experience, correlate with teacher value-added but explain very little of its overall variation (see Table A.5).

3 Descriptive Analysis

School principals have a substantial amount of discretion in the hiring process of permanent teachers, as described in Section 2. Panel a) of Figure 1 shows the evolution of applicants' total scores throughout each step of the decentralized evaluation process. The distribution of cumulative scores mechanically moves to the right from one stage to the other. The dispersion of the scores also increases, which suggests that the evaluation committee uses the grades given at each stage to screen candidates and likely reshuffle the prior rankings. The stage of the decentralized evaluation involving the largest amount of screening appears to be the classroom evaluation, which significantly increases the variance of the score distribution.

Panel b) of Figure 1 displays the probability (in red) and the cumulative probability (in blue) of a change in the first-ranked candidate throughout the different steps of the

Figure 1: The Role of School Principals over the Recruitment Process



a) Teachers' Score **b) Probability of a Change in Top-Rank**

NOTES: Panel a) displays the empirical distributions of the applicants' score across the different stages of the recruitment process. Each line corresponds to the PDF of the cumulated score obtained by permanent teachers at each stage. The light blue dotted line is the competency test (TEST); the dashed blue line is the score at stage two, obtained as TEST plus the score from the professional background evaluation (CV); the dashed purple line is the overall score at stage three, obtained as TEST + CV + the score from the personal interview (INTERVIEW); finally, the solid line indicates the final score at stage four, which includes also the score from the mock class observation (CLASS). Panel b) shows the probability of a ranking reversal at each stage of the recruitment process. A ranking reversal occurs when there is a change in one of the top positions that, based on the current ranking, would guarantee appointment. Light bars indicate the probability of a reversal compared to the previous stage, while darker bars show the probability relative to the first stage of the recruitment process, where school-specific rankings are based solely on centralized test scores.

recruitment process. During the first step of the decentralized evaluation, in which candidates are screened based on their professional experience as recorded in the Curriculum Vitae (CV), the probability of a change in the first-ranked candidate when compared to the ranking based on standardized score (PUN) is 19 percentage points. This probability increases by another 16 percentage points after the interview stage and finally goes up by another 33 percentage points after the classroom observation. The cumulative probability of a switch in the first-ranked candidates across vacancies from the centralized stage to the end of the decentralized evaluation is close to 50 percentage points, suggesting that principals have a substantial amount of discretion in the hiring process.

We next document that principals use their discretion to positively or negatively select candidates on various traits. Column (1) of Table 2 displays the characteristics of the average applicant, including test scores, value-added, gender, experience and whether the candidate is internal (i.e. was working in the same school as a contract teacher before). Columns (2)-(4) report the difference between the characteristics of the first-ranked candidates according to different ranking rules and the average applicant. The second column shows the performance

Table 2: Principals Ranking Behavior

	All Applicants	Δ with Top Applicants Ranked by ...		
		Test Score	Principals	Value-Added
	(1)	(2)	(3)	(4)
ATE	0.134	-0.002	0.018	0.189
Match Effect	0.003	0.000	0.002	0.012
Test Score (std.)	1.863	0.243	0.124	0.006
Male	0.186	0.028	0.027	0.016
Experience > 5	0.242	0.005	0.033	0.019
Internal Candidate	0.111	-0.007	0.033	-0.012
$\times$ High Value-Added	0.056	-0.008	0.022	0.028
$\times$ Low Value-Added	0.055	0.001	0.011	-0.039

of the ranking rule based on test scores, i.e. simulating the scenario where principals did not change the original centralized ranking. Top scorers differ little from the average applicant in terms of value-added and experience and are 3 percentage points more likely to be male. In contrast, principals' ranking behavior performs slightly better and positively selects higher-value added teachers. As shown in column (3), first-ranked candidates after the decentralized evaluation have an average effectiveness (ATE) that is 0.02σ higher than the average applicant (column 1). The top-ranked candidates by principals are also positively selected on their test scores and experience and more likely to be male. Finally, Column (4) shows the performance of a ranking rule that would aim at selecting the most productive teachers (i.e. highest value-added). Principals' behavior massively under-performs this benchmarks which would attain gains of 0.20σ in value-added relative to the average applicant. It also shows that the test-score based rule, and principals' hiring rule over-select male and high-scoring applicants (consistently with Table A.5). Overall, this evidence shows that principals use their discretion to positively select better quality teachers and outperform the test-score based hiring rule, but massively outperforms a benchmark where they would rank teachers based on value-added.

These findings are suggestive of a systematic misalignment between principals ranking rule and the objective of hiring the most productive workers. This could be explained by mainly two reasons. First, principals might have personal motives unrelated to the objective of maximizing their students' achievement that get reflected in their hiring decisions. They might have a particular distaste for working with a certain type of people, or might want

Table 3: Principals Ranking Behavior: Heterogeneity

	All Applicants	Δ with Top Applicants Ranked by ...		
		Test Score	Principals	Value-Added
	(1)	(2)	(3)	(4)
Panel A: Urban Schools				
ATE	0.147	0.011	0.039	0.202
Match Effect	0.003	0.001	0.001	0.009
Test Score (std.)	1.897	0.286	0.174	0.048
Male	0.154	-0.006	-0.006	-0.020
Experience > 5	0.219	-0.025	0.007	-0.014
Internal Candidate	0.122	-0.005	0.049	0.002
Panel B: Rural Schools				
ATE	0.088	-0.044	-0.049	0.148
Match Effect	0.003	-0.001	0.003	0.019
Test Score	1.744	0.104	-0.038	-0.129
Male	0.301	0.137	0.131	0.132
Experience > 5	0.324	0.101	0.117	0.127
Internal Candidate	0.072	-0.012	-0.018	-0.055

to hire candidates they already know. Second, teacher value-added might be unobserved to the principal. To provide evidence supporting the first hypothesis, we investigate whether principals positively select high value-added candidates when they are internal, i.e. when the principal is more informed about the performance of this candidate. Table 2 shows that principals positively select internal candidates irrespective of whether they are high or low value-added. This suggests that principals principals may have other motives beyond maximizing student achievement when selecting candidates.

Finally, Table 3 shows that principals hiring behavior displays substantial level of heterogeneity beyond the average patterns described in Table 2. We investigate whether principals working rural and urban schools differ in their ranking behavior. Such disparities could be particularly salient as sizable urban-rural disparities in teaching quality and student achievement have been documented in this context (Bobba et al., 2022). School principals in urban schools tend to positively select candidates based on their value added by 0.04σ while principals in rural schools tend to negatively select high value-added teachers by -0.05σ. This

could reflect differences in principals' characteristics between urban and rural schools and is likely to play an important role in explaining urban rural disparities in student achievement.[5]

Overall, we find that principals use their discretion to select candidates in a way that is, on average, poorly aligned with the objective of maximizing student achievement. However, we find substantial heterogeneity in hiring practices that is likely to explain a large fraction of inequality in student achievement. In the next section, we build an empirical framework that allows us to fully exploit data on the ranking behavior of principals to quantify and characterize the extent of this heterogeneity.

4 An Empirical Model of Hiring

Consider a set of principals $i = 1, ..., n$ and a set of teachers $j = 1, ..., J$. Each principal i faces a set of applicants in year t denoted $\mathcal{J}_{it}$, provided that the school in which they are employed has at least one vacant position in the recruitment drive for permanent teachers. We define the value that principal i gets from hiring teacher j in period t as:

$$U_{ijt} = z'_{ijt}\alpha_i + \varepsilon_{ijt},$$

where z_{ijt} is a vector of teacher characteristics including an intercept, national test scores, value-added, gender, experience, and an indicator for whether teacher j is an internal candidate. The vector α_i are weights pinning down the relative importance of each teacher characteristic in principal i's ranking decision. We specify α_i as a vector of normally distributed correlated random coefficients $\alpha_i \sim \mathcal{N}(\mu_\alpha, \Sigma_\alpha)$ measuring how those weights vary across the population of principals. Finally, ε_{ijt} is an i.i.d. type-I extreme value unobserved shock explaining further unobserved heterogeneity in principal i's ranking behavior. We normalize the value of leaving a position vacant as

$$U_{i0t} = 0 + \varepsilon_{i0t}.$$

We impose the following assumptions on the set of applicants $\mathcal{J}_{it}$:

[5] Indeed, Table A.3 in the Appendix shows large spatial gaps in the standardized scores that are used to assign school principals across schools.

Assumption 1 *For all $i = 1, ..., n$ and $t = 1, ..., T$:*
(i). $g_{\varepsilon|\mathcal{J}}(\varepsilon|\mathcal{J}_{it}) = g_{\varepsilon}(\varepsilon)$
(ii). $\alpha_i|\mathcal{J}_{it} \sim \mathcal{N}(\mu_\alpha, \Sigma_\alpha)$

Assumption 1 states that the distribution of random coefficients and taste shocks does not depend on the set of applicants $\mathcal{J}_{it}$ faced by the principal. This rules out scenarios where applicants would have private information about principals' idiosyncratic hiring criteria and would strategically apply to schools using this information. We denote the rank-order list (ROL) submitted by principal i in period t as R_{it} where $jR_{it}j'$ indicates that applicant j is ranked above j'. We impose the following assumptions on R_{it}:

Assumption 2 *For all $i = 1, ..., n$, $t = 1, ..., T$ and for $j \in \mathcal{J}_{it}$ such that $jR_{it}0$:*

$$jR_{it}j' \implies U_{ijt} > U_{ij't}, \text{ for all } j' \in \mathcal{J}_{it} \cup \{0\}$$

Assumption 2 implies that the observed ranking submitted by each principal reflects the hiring model described above, except for applicants ranked below the outside option. Under Assumptions 1 and 2, we can write the probability of observing the ROLs formed by principal i as:

$$\begin{aligned}\mathbb{P}(\{R_{it}\}_{t=1}^{T}|(z_{ijt})_{j\in\mathcal{J}_{it}}, \mathcal{J}_{it}) &= \prod_t \prod_{j\in\mathcal{J}_{it}} \mathbb{P}(U_{ijt} > U_{ikt}, \forall k \text{ s.t. } jR_{it}k|(z_{ijt})_{j\in\mathcal{J}_{it}}, \mathcal{J}_{it})^{1\{jR_{it}0\}} \\ &= \prod_t \prod_{j\in\mathcal{J}_{it}} \left(\int \frac{\exp\{z'_{ijt}\alpha_i\}}{\sum_{k:jR_{it}k} \exp\{z'_{ikt}\alpha_i\}} \phi(\alpha_i|\mu_\alpha, \Sigma_\alpha) d\alpha_i \right)^{1\{jR_{it}0\}}.\end{aligned}$$

Identification of μ_α and Σ_α in random coefficients logit models follows directly from Fox et al. (2012). Rank-ordered choice data as well as repeated observations of the same principal over time generate substantial within principal variation in choices and choice sets allowing to identify the distribution of α_i flexibly (Berry et al., 2004). We approximate the integral by Monte Carlo integration using 5000 draws from a Halton sequence and estimate μ_α and Σ_α by MLE.

Table 4: Principals Preferences

	Mean	SD
	(1)	(2)
Intercept	-1.646	1.682
	(0.105)	(0.134)
Test Score (std.)	2.877	1.987
	(0.063)	(0.063)
Value-Added	0.221	0.530
	(0.079)	(0.224)
Male	0.261	0.447
	(0.031)	(0.059)
Experience = 0	-0.385	0.314
	(0.034)	(0.079)
Experience > 5	0.327	0.393
	(0.027)	(0.057)
VA Missing	0.109	0.399
	(0.034)	(0.072)
Internal Candidate	0.582	0.579
	(0.051)	(0.083)

5 Results

5.1 Estimates

We report in this section the estimates of μ_α and Σ_α. Column (1) of Table 4 displays the estimated mean of the random coefficient distribution. The average principal values observed measures of quality such as experience and test scores. Principals put a low weight on teacher value-added relative to other directly observed measures of teacher quality. An increase in value-added of one standard deviation is valued less on average than a 0.1 standard deviation increase in test scores and has less than half the value of hiring an internal candidate. Consistently with Table 2, principals hiring decision rule is poorly correlated with a decision rule that would rank teachers based on value-added.

Departing from the average principal yields different conclusions. Column (2) reveals the presence of substantial heterogeneity in how principals weight teacher characteristics relative to each other. A one standard deviation increase in the value-added coefficient more than triples its importance relative to other characteristics. While the estimated standard deviations of the other coefficients are also large, they are relatively smaller compared to

Table 5: Principals Preferences: Correlation Matrix

	Test Score	VA	Male	Exp = 0	Exp > 5	VA Missing	Internal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	-0.970	0.215	-0.083	0.290	-0.365	-0.311	0.081
	(0.029)	(0.296)	(0.169)	(0.240)	(0.157)	(0.170)	(0.174)
Test Score		-0.311	0.034	-0.362	0.425	0.090	-0.137
		(0.231)	(0.097)	(0.151)	(0.099)	(0.116)	(0.106)
Value Added			-0.113	0.632	-0.433	0.596	-0.408
			(0.463)	(0.709)	(0.481)	(0.307)	(0.419)
Male				0.075	0.077	0.022	-0.386
				(0.309)	(0.194)	(0.252)	(0.244)
Experience = 0					-0.397	0.392	-0.097
					(0.343)	(0.382)	(0.328)
Experience > 5						-0.374	-0.369
						(0.220)	(0.229)
VA Missing							0.124
							(0.259)

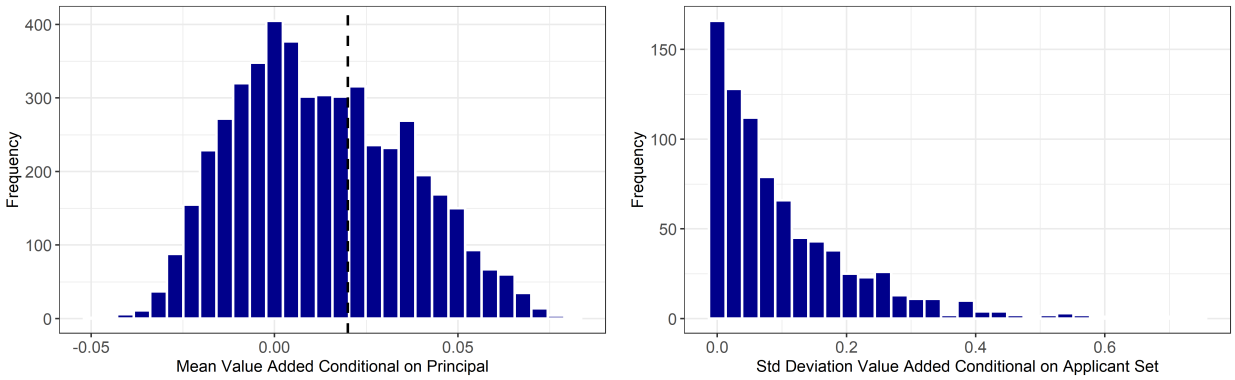
their mean. This evidence suggests the presence of substantial heterogeneity in the use of discretion in principals' hiring behavior. Some principals might be more closely aligned to the decision rule ranking teachers based on value-added than what is suggested by the mean.

5.2 Variance Decomposition

Conditional a fixed set of applicants $\mathcal{J}_{it}$, the heterogeneity in principals' ranking behavior documented in Table 4 is very likely to generate substantial variation in the characteristics of the first-ranked teacher. To quantify this variation, we perform the following exercise. Given $\mathcal{J}_{it}$, we generate one draw of ε_i for all $j \in \mathcal{J}_{it}$. We then generate 2000 draws of α_i from its estimated distribution. For each draw, we compute the corresponding counterfactual ROL $\tilde{R}_{it}$ and store the characteristics of the applicant ranked first. We then repeat this exercise for each set of applicant $\mathcal{J}_{it}$.

We first consider a thought experiment where one would take a given principal α_i, apply their hiring rule to all sets of applicants $\mathcal{J}_{it}$, and measure the average characteristics of the selected applicants. One can then repeat this exercise for all draws α_i to study how different hiring rules would perform. Panel A of Figure 2 summarizes the results of this experiment by

Figure 2: Counterfactual Simulations



a) Mean VA Conditional on α_i **b) SD VA Conditional on $\mathcal{J}_{it}$**

plotting the distribution over α_i of the average difference in value-added between the first-ranked teacher and the average applicant. Consistently with Table 2, the average principal's hiring rule leads to selecting teachers that have a value-added 0.02σ higher than the average applicant (dashed vertical line). This average hides significant heterogeneity with more than 25% of principals selecting teachers that are worse than the average applicant. On the other hand, the upper decile of the distribution would select teachers that are 0.05σ better than the average applicant. Yet, this upper tail is still quite far from performing as well as ranking applicants based on value-added which would select teachers that are on average 0.20σ more effective than the average applicant.

We next consider an alternative thought experiment where one would instead fix the set of applicants $\mathcal{J}_{it}$, determine the characteristics of the highest ranked teacher for each draw of α_i, and measure the variance of those characteristics across α_i. One can then repeat this exercise for all set of applicants $\mathcal{J}_{it}$. Panel B of Figure 2 summarizes the results of this experiment by plotting the distribution over $\mathcal{J}_{it}$ of the standard deviation of the value-added of the selected applicant implied by the variance of α_i. We find that the variance in value-added generated by heterogeneity in principals' hiring rules is above 0.07σ for half of the set of applicants and above 0.25σ for the top decile.

To quantify the role of heterogeneity in α_i in generating variation in the characteristics of the first-ranked teacher, we consider the following variance decomposition

$$\text{Var}(z^*_{it}) = \mathbb{E}[\text{Var}(z^*_{it}|\mathcal{J}_{it})] + \text{Var}(\mathbb{E}[z^*_{it}|\mathcal{J}_{it}]),$$

Table 6: Variance Decomposition

	SD Explained by			
	α_i	$\mathcal{J}_{it}$	Total	Var % α_i
	(1)	(2)	(3)	(4)
Test Score (std.)	0.188	0.404	0.446	0.178
Value-Added	0.157	0.295	0.334	0.222
Male	0.185	0.371	0.414	0.199
Experience = 0	0.168	0.274	0.321	0.273
Experience > 5	0.218	0.388	0.445	0.241
Internal Candidate	0.167	0.290	0.334	0.249

where z_{it}^* denotes the characteristics of the first ranked teacher given α_i and $\mathcal{J}_{it}$. The first term corresponds to the share of the total variance in the characteristics of the first-ranked teacher explained by variation in α_i, while the second term corresponds to the share explained by variation in the set of applicants $\mathcal{J}_{it}$. We can simulate both terms by simulating draws of α_i following the procedure described above. Table 6 displays the results of this exercise. Heterogeneity in principals' hiring rules explains from 18 to 27% of this variance. Importantly, variation in hiring behavior generates a standard deviation of the value-added of first-ranked teachers of 0.157σ which corresponds to half of the total standard deviation. Overall, this suggests that heterogeneity in hiring practices largely contributes to inequality in student achievement.

6 Conclusion

This paper highlights three key results. First, school principals' hiring behavior is, on average, misaligned with the objective of hiring the most effective teachers but performs better than test-based rules. Second, this average hides substantial heterogeneity in hiring practices suggesting that some principals are much more likely to positively select high-value added teachers. These principals tend to be located in urban schools which has the potential to generate large inequities in teacher quality. Third, heterogeneity is hiring practices explain a large fraction of the variance in teacher value-added and student achievement. These findings suggest that giving discretion in hiring might perform better than rule-based hiring but at the cost of increased inequality caused by heterogeneity in hiring practices.

References

Ahn, Tom, Esteban Aucejo, and Jonathan James, "The Importance of Matching Effects for Labor Productivity: Evidence from Teacher-Student Interactions," Technical Report, California Polytechnic State University, Department of Economics 2023.

Akerlof, George A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters, "Leveraging Lotteries for School Value-Added: Testing and Estimation," *The Quarterly Journal of Economics*, 2017, *132* (2), 871–919.

Bau, Natalie and Jishnu Das, "Teacher value added in a low-income country," *American Economic Journal: Economic Policy*, 2020, *12* (1), 62–96.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 2004, *112* (1), 68–105.

Bertrand, Marianne and Antoinette Schoar, "Managing with style: The effect of managers on firm policies," *The Quarterly journal of economics*, 2003, *118* (4), 1169–1208.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen, "Do private equity owned firms have better management practices?," *American Economic Review*, 2015, *105* (5), 442–446.

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*, 2022.

Chetty, Raj, John Friedman, and Jonah Rockoff, "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 2014, *104(9)*, 2633–2679.

_ , _ , and _ , "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014, *104(9)*, 2593–2632.

_ , **John N Friedman, and Jonah E Rockoff**, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014a, *104* (9), 2593–2632.

Fox, Jeremy T, Kyoo il Kim, Stephen P Ryan, and Patrick Bajari, "The random coefficients logit model is identified," *Journal of Econometrics*, 2012, *166* (2), 204–212.

Gibbons, Robert and Michael Waldman, "Careers in organizations: Theory and evidence," *Handbook of labor economics*, 1999, *3*, 2373–2437.

Lazear, Edward P, *Personnel economics*, Vol. 1993, MIT press Cambridge, MA, 1995.

Oyer, Paul and Scott Schaefer, "Personnel economics: Hiring and incentives," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 1769–1823.

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.

Spence, Michael, "Job Market Signaling," *The Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

A Additional Tables and Figures

Table A.1: Summary Statistics

	(1) N	(2) Mean	(3) Sd	(4) Min	(5) Max
Panel A: Applicants' Characteristics					
Test Score (std.)	27763	1.653	0.439	0.57	3.55
Male	27763	0.218	0.413	0.00	1.00
Missing VA	27763	0.866	0.341	0.00	1.00
ATE	3718	0.111	0.319	-1.19	1.57
Experience = 0	27763	0.142	0.350	0.00	1.00
Experience > 5	27763	0.304	0.460	0.00	1.00
Internal Candidate	27763	0.089	0.285	0.00	1.00
Panel B: School Characteristics					
N. of applicants	7279	3.814	3.240	1.00	20.00
N. of vacancies	7279	1.637	1.123	1.00	14.00
Applicants > Vacancies	7279	0.636	0.481	0.00	1.00
Rural school	7279	0.538	0.499	0.00	1.00

NOTES. This table presents summary statistics on applicants to recruitment drives for permanent teaching positions (Panel A) and on schools posting vacancies (Panel B).

Table A.2: Estimates Teacher Value-Added: γ

	Math
	(1)
Lagged Score^2	-0.000 (0.003)
Quechua	-0.041 (0.017)
Age	-0.133 (0.006)
School Lagged Score	0.328 (0.051)
School Lagged Score^2	0.079 (0.047)
School Female	-0.126 (0.240)
School Quechua	0.419 (0.285)
School Age	-0.676 (0.129)
Class Lagged Score	-0.172 (0.031)
Class Lagged Score^2	-0.053 (0.027)
Class Female	0.498 (0.131)
Class Quechua	-0.525 (0.301)
Class Age	0.051 (0.070)
Class Size	-0.002 (0.005)
Class Size^2	-0.000 (0.000)

NOTES. This table displays the estimates of γ. Standard errors are in parentheses.

Table A.3: Principal Characteristics by Rurality

	Rural Schools		Urban Schools	
	Mean	Sd	Mean	Sd
Male	0.658	(0.474)	0.471	(0.499)
Age	47.54	(7.211)	50.42	(5.736)
Experience (years)	1.995	(1.608)	2.074	(1.617)
Overall score (std.)	-0.0560	(0.970)	0.386	(0.942)
Reading score (std.)	-0.0403	(0.939)	0.291	(0.966)
Management score (std.)	0.130	(0.877)	0.363	(0.957)
Curricular score (std.)	-0.149	(0.979)	0.254	(0.926)

NOTES. This table presents descriptive statistics on the characteristics of school principals, separately for rural and urban schools.

Table A.4: Change in Top Ranked Candidates by Applicant Characteristics

	(1)	(2)	(3)	(4)
	Stage 1 leader	Cv vs (1)	Interview vs (1)	Classroom vs (1)
	mean/sd	b/se	b/se	b/se
Female	0.769	0.003	-0.003	-0.017*
	(0.421)	(0.007)	(0.007)	(0.007)
Speaks Quechua	0.169	0.003	0.002	0.004
	(0.374)	(0.006)	(0.006)	(0.006)
No experience	0.137	-0.021***	-0.026***	-0.027***
	(0.344)	(0.006)	(0.005)	(0.005)
Experience >5 years	0.302	0.031***	0.037***	0.040***
	(0.459)	(0.008)	(0.008)	(0.008)
In same school before	0.104	0.005	0.010*	0.027***
	(0.305)	(0.005)	(0.005)	(0.005)
Test score (std)	1.967	-0.016*	-0.035***	-0.130***
	(0.416)	(0.007)	(0.007)	(0.007)
Value added: ATE Math	0.096	0.002	0.013	0.033*
	(0.312)	(0.014)	(0.014)	(0.014)
Value added: Match effect	-0.001	0.008	0.007	-0.003
	(0.246)	(0.011)	(0.011)	(0.011)

NOTES. Standard deviation in parenthesis for Column 1. Standard errors in parenthesis for Columns 2 to 4.

Table A.5: Correlation ATE and Teacher Characteristics

	(1)	(2)	(3)	(4)
	Test Score	Male	No exp.	Exp >5
ATE	0.071***	-0.067***	-0.038**	-0.029
	(0.022)	(0.024)	(0.016)	(0.022)
Constant	1.672***	0.230***	0.113***	0.278***
	(0.007)	(0.007)	(0.006)	(0.008)
R-squared	0.003	0.003	0.002	0.000
N	3718	3718	3718	3718

NOTES. This table reports estimates from a set of regression on teacher characteristics on teacher value added in Math. In Column 1, the dependent variable is the overall teacher test score (standardized); in Column 2 it is a binary indicator for male teachers; in Column 3 and 4 it is a binary indicator for teachers with no previous teaching experience in public schools, and teachers with more than five years of experience, respectively. Robust standard error in parentheses.