Overcoming Adverse Selection through Performance Pay
Evidence from a Field Experiment in Pakistani Private Schools *

Tahir Andrabi  Christina Brown†

February 20, 2020

(Preliminary and Incomplete Draft - Do Not Cite or Circulate)

Abstract

Good teachers provide enormous short and long term benefits for their students. Attracting good teachers into the profession is a key area of policy interest for which there is relatively little evidence. This study uses teachers’ contract choices and randomized controlled trial with 2500 teachers in 252 private schools in Pakistan to study whether performance pay can induce positive sorting of teachers. Consistent with models of adverse selection, we find that performance pay can induce positive selection into schools. High value-added teachers and teachers who respond more strongly to incentives significantly prefer performance pay and sort into these schools in response to these preferences. Using information treatments and additional contract treatment arms, we are able to isolate that high-quality teachers sort because they have accurate information about their quality, rather than due to differences in risk or competition preferences. Teachers also have significantly more information about their quality than their principal both at the time of hiring and throughout their tenure at the school. Taking into account the sorting effects from performance pay suggests that we have significantly underestimated the benefits of performance pay contracts on teaching quality.

*I am extremely thankful for support and advice from Supreet Kaur, Chris Walters, and Ted Miguel. David Card, Jishnu Das, Stefano DellaVigna, Fred Finan, Anne Karing, Asim Khwaja, Sam Leone, Pat Kline, Jeremy Magruder, Gautam Rao, Jesse Rothstein, Heather Schofield, and seminar audiences at UC Berkeley, CERP and Stanford provided helpful feedback. I gratefully acknowledge generous funding and support by DFID’s RISE Programme, JPAL’s Post-Primary Initiative, the Weiss Family Fund, CEGA, the Institute for Research on Labor and Employment and the Strandberg Fund. Our wonderful team at the Center for Economic Research in Pakistan, Haya Mubasher, Anam Tariq, Attefaq Ahmed, Zahra Niazi, Mujahid Murtaza, and Zohaib Hassan, provided excellent research assistance. All remaining errors are our own. We received IRB approval from Pomona College. AEA Registry-0003835.

†Brown (corresponding author): University of California, Berkeley (christinabrown@berkeley.edu); Andrabi: Lahore University of Management Sciences and Pomona College
1 Introduction

Teachers are the most important input in the education production function. Having a better teacher for one year is associated with countless short term (literacy, numeracy, social skills, test scores, etc) and long term (income, crime, health, etc) benefits. However, it is notoriously challenging to predict who will be an effective teacher. Standard markers of productivity, like credentials, or subjective measures, like interview performance, are poor predictors of eventual teacher quality. Perhaps, as a result, most policy focuses on how to improve the existing stock of teachers through training, incentives, and support, rather than changing who enters or exists the profession. On average these interventions have had relatively small impacts on teacher quality.\(^1\)

Performance pay has been theorized to help overcome this selection problem and has been shown to improve selection in other occupations. Whether performance pay could improve the selection of workers in this setting depends on three underlying features of the labor market. First, (1) higher-quality teachers must prefer performance pay\(^2\). Second, (2) their preferences over pay structure must be large enough relative to their other preferences regarding occupation, location, non-wage amenities, etc to induce positive sorting. Third, (3) teachers’ preferences must provide additional predictive power beyond what the employer knows about teacher quality or because the employer is not empowered to act on any quality information (for example, due to firing constraints, principal-agent problems, etc).

This study conducts a contract choice experiment and randomized controlled trial with 2500 teachers in 252 private schools in Pakistan to study whether performance pay can induce positive sorting of teachers. First, we offer teachers the opportunity to choose what contract they would like for the following year, selecting their preferred contract from pairwise comparisons of flat raise versus performance-based raise. Teachers’ choices implemented in a randomly selected subset of schools to ensure incentive compatibility of responses. We also elicit the distribution of teachers’ beliefs about their value-added and their risk preferences through an incentivized activity. A random subset of teachers are also provided historical information about their value-added, moving individuals’ priors and certainty, and allowing us to identify directly the effects of their beliefs on contract choice.

We then randomize schools to one of four contract types: i). flat raise, ii). raise based

---

\(^1\)Teacher performance pay has increased student test performance by 0.1sd on average and training programs generally have almost no effect.

\(^2\)This could be either because higher-quality teachers are aware of their quality and know they would benefit financially from the contract or because quality is correlated with competitiveness or other features which may draw people to performance pay contracts.
on percentile value-added, iii). raise based on their principals’ rating and iv). providing the contract type they selected at baseline from among these three types. Teachers are informed that the contract type is associated with the school itself, which is important in this setting as 15% of teachers switch schools during the summer break.

We find evidence that performance pay induces positive selection into schools. (1) We find higher value-added teachers significantly prefer performance pay when allowed to choose their contract. Teachers who choose performance pay had 0.16 sd higher student test scores in the year before the intervention. (2) In the year after we introduce the contract, we find better teachers move into performance pay schools. Those schools see a gain of 0.12 sd in teacher value-added (VA) relative to flat schools during the intervention year. Most of these gains are derived from good teachers moving into performance pay schools, rather than poor performers leaving. Finally (3), when we control for information principals have about teacher quality, there is still a significant amount of residual variation explained by teacher’s contract choice and sorting decision. This suggests that teachers have more information about their quality than their principal. Combined, these results show performance pay can be effective in inducing positive selection into the teaching profession.

Our contract choice exercise also allows us to tease out what predicts contract choice. In the cross-section, teachers’ beliefs about their performance under the contract, the certainty of their beliefs, and their risk preferences all are related to their contract choice. In addition, we experimentally vary both the level and precision of their beliefs about their VA by introducing an information treatment. Increasing their believed level and the preciseness of their beliefs makes teachers more likely to choose performance pay. We also find that teachers who choose performance pay contracts at baseline have much larger direct behavioral responses to performance pay. This suggests teachers may have private information about how motivating they find performance pay.

This project speaks to the developing literature on adverse selection and teacher sorting (Biasi, 2018; Leaver et al, 2019). Current theory suggests that sorting effects may be larger than the direct effects of performance pay (Lazear, 2000) but this would require teachers to be able to predict their quality which there is weak evidence for (Springer, 2010; Rothstein, 2015). This project will allow us to tease out the extent of sorting in response to these incentives.

The remaining sections are organized as follows. Section 2 provides information about the teacher labor market in this context. Section 3 presents a model of job choice extending from Roy (1951). Section 4 details the contract choice, randomized controlled trial and data collection procedures. Section 5 presents the results and concludes.
2 Context

2.1 Teacher Performance Pay

In the last decade, there has been a significant push to tie teacher salaries to student outcomes in developed and developing countries (Goodman and Turner, 2013; Pham, et al 2017; Stetcher et al, 2018). The U. S. Department of Education has provided $2.4 billion for performance pay from 2010-2016, and Arne Duncan described the policy as the Department’s “highest priority” when he was Secretary. In a meta-analysis of teacher performance pay studies, there was substantial variation in effectiveness with an average of 0.1 sd increase in student test scores (Pham, et al 2017).

Recent work has tried to isolate the effect of different contract types on teacher selection. Biasi (2018) finds that schools who offer flexible salaries, rather than the standard tenure-based salary schedule, attract higher quality teachers. She finds that this appears to be due to schools using flexible salaries to reward high performers. Leaver et al. (2019) show that when given the option, high-quality new entrants select into performance pay schools, but they do not see evidence of existing teachers sorting in response to the contracts.

2.2 Study Context

This study partners with a large private school system in Pakistan. School fees are high and tend to attract upper-middle class and upper-class students. Teachers have significant pre-employment training and credentials. Teacher attendance and time on task are high. The context of the school in terms of resources, infrastructure, and instructional quality is similar to that of a low-income school in the US.

3 A Model of Job Choice

3.1 Employment Choice

We set up a simple Roy-like model of job choice. First, we restrict the employee’s decision to just include the wage type: performance pay or flat pay. We think of this as similar to the decision they make in the first part of the study, where they are allowed to select their contract. Second, we add on other non-wage preferences employees have over different jobs. This decision problem models the second phase of the study where teachers decide what
school to work at, which is a function of both their idiosyncratic preferences for location, colleagues, etc, plus the contract type.

### 3.1.1 Employee Utility

Individuals choose between two jobs, $j_F$, which pays a fixed wage, or $j_P$ which pays a wage dependent on the worker’s output, $y$. Output is a function of the worker’s innate ability, $\theta$, their effort response to a performance pay contract, $\beta$, and noise, $\mu$.

\[
\begin{pmatrix}
\theta \\
\beta \\
\mu
\end{pmatrix}
\sim
N
\begin{bmatrix}
\begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix},
\begin{pmatrix}
\sigma^2_{\theta} & \rho_{\theta\beta} & 0 \\
\rho_{\theta\beta} & \sigma^2_{\beta} & 0 \\
0 & 0 & \sigma^2_{\mu}
\end{pmatrix}
\end{bmatrix}
\]

The wage received from each contract type is:

\[
w = \begin{cases} 
0 & \text{if } j = j_F \\
=f(\theta, \beta) + \mu & \text{if } j = j_P 
\end{cases}
\]

Wage in the performance pay job is an increasing function of $\theta$ and $\beta$. The individual chooses a performance pay job if $EU[f(\theta, \beta) + \mu] \geq 0$. This implies the likelihood of selecting a performance pay job is:

1. Increasing in workers’ ability, $\theta$
2. Increasing in workers’ response to performance pay, $\beta$
3. Decreasing in risk aversion
4. Decreasing in the noisiness of the contract, $\sigma^2_{\mu}$, for those who are risk-averse and increasing for those who are risk-loving

### 3.1.2 Adding non-wage utility from jobs

Assume now that jobs also carry non-wage utility, $\epsilon_{ij}$, that is employee-specific. These are the distance between the individuals’ home and the job and non-wage amenities (such as colleague and principal quality).
\[ w = \begin{cases} \epsilon_{iF} & \text{if } j = j_F \\ \theta + \beta + \mu + \epsilon_{iP} & \text{if } j = j_P \end{cases} \] (2)

These utilities are drawn from the same distribution, \( \epsilon_{iP}, \epsilon_{iF} \sim \mathcal{N}(0, \sigma_{\epsilon}) \), for flat and performance pay jobs in our study due to the randomization of contracts.

This implies

\[ 5. \frac{\partial(E(\theta | j = j_p) - E(\theta | j = j_f))}{\partial \sigma_{\epsilon}^2} < 0: \text{ Larger heterogeneity in distance costs and amenity value reduces the extent of positive sorting} \]

4 Experimental Design

4.1 Timeline and Treatments

In partnership with a network of private schools, we implemented a contract choice elicitation exercise and then a randomized controlled trial implemented between October 2017 through January 2019. Figure 1 details the study timeline.

*Contract Choice Exercise* – The contract choice exercise was conducted in October 2017 with teachers as part of the baseline data collection exercise. During the exercise, teachers were told the school system would be testing out several new raise policies in the following year. In some randomly selected schools, teachers would receive the raise type they had selected in the survey. The surveyor explained the randomization process and teachers also watched a video explaining the raise options, the randomization and that their choice would actually be implemented if their school was selected. Screen captures from the video are shown in Appendix figures 5 and 6.

One challenge previous performance pay interventions have faced is that teachers have a hard time understanding how their effort translates into pay. If teachers do not understand the incentive scheme or do not understand how the metrics are calculated then it is unlikely they would induce any positive selection. Several features of the contract design itself, along with procedures used during the elicitation, help assuage those concerns here. First, teachers are familiar with the exams used, as they are annual exams administered by the school system. Second, teachers have previously received raises based on other metrics of performance, so they are familiar with the raise bins used in this experiment. Third, understanding checks
during the choice exercise restricted teachers from moving on with the questionnaire if they did not have a sufficiently strong understanding of the incentive scheme. Finally, even with all of these assurances, a new performance metric will always have some uncertainty associated with it. Teachers have historically received performance ratings from their principal so they are familiar with this type of rating scheme. Therefore, we further compare preferences for a high (VA) versus low (subjective) noise contract to isolate the role of performance uncertainty in contract preferences.

Finally, during the choice exercise, teachers were randomized to information about their historical performance. Teachers were randomized to receive a precise signal, a “high” performance signal, a “low” performance signal or no signal.3

*Contract Randomization* Contracts were randomized at the school level and the contract applied to all core teachers (those teaching Math, Science, English, and Urdu) in grades 4-13. Elective teachers and those teaching younger grades received the status quo contract.

The treatment varies how the teachers’ yearly raise is determined but maintains the same budget implication for the school. Schools are randomized to the following contract options:

- Control: **Flat Raise** - All teachers receive a 5% raise at the end of the calendar year. This is fairly standard for professional sector jobs and is equal to about twice the cost of living inflation.

- Treatment 1: **Objective Performance Raise** - All teachers receive a raise based on percentile value-added (Barlevy and Neal, 2016)4 – their students’ average percentile within their lagged percentile comparison group on the standardized exam at the end of each semester. Teachers are ranked within the school based on this measure of value-added and receive a raise varying from 0-10%.

In addition, to test additional mechanisms and to ensure the contract choice exercise was incentivized, there are two additional treatment schools could be randomized to:

---

3Teachers were randomized to one of four treatments (i). No information, (ii). Receiving their exact percentile from the previous year. (iii). Receiving a range they were in, in which their actual score was at the top of the range and (iv). Receiving a range in which their actual score is near the bottom. Traditional information treatments, tend to move both peoples’ priors and the certainty about those priors. This design allows us to separately identify the effect of changes to the mean and certainty of their beliefs. Comparing treatment (iii) and (iv) allows us to isolate the effect of moving mean beliefs but not certainty. The remaining variation between (i) and (ii) allows us to isolate the effect of increased certainty.

4The benefit of using this method is it only relies on ordinal information, allowing us to change the exams year-to-year and is much easier and more transparent to explain to teachers than many other value-added models. The method is becoming increasingly common to rank workers.
• Treatment 2: **Subjective Performance Raise** - All teachers are rated by their direct supervisor (usually a principal or vice principal) on the teacher’s “effort in improving students’ academic performance” on a scale from 0-100. Principals gave a short justification for their rating. Within the school, teachers are ranked according to this subjective score and given a raise from 0-10% based on their ranking.

This treatment allows us to isolate the channel through which positive sorting occurs. This contract has the same overall incentive structure but varies the metric in which teachers are evaluated.

• Treatment 3: **Teachers’ Contract Choice** - Teachers receive the contract they stated they preferred at baseline from the three options randomized

This treatment was implemented solely to ensure the contract choice exercise was incentivized. The fraction of schools randomized to this treatment is smaller than the three other arms and not intended to be used to look at outcomes from this treatment.

Randomization for the incentive treatment was conducted at the school level as there are possibilities for within-school spillovers. Due to implementation constraints from the school system, rather than have an even split for randomization, 42 schools are in the flat treatment, 42 schools are in the objective treatment and 80 are in the subjective treatment.

**Teacher Sorting** – Within this school system, transfers across schools are common. There are two types of transfers. Many schools operate on a larger campus. For example, there may be a primary school, middle school and high school all on the same larger campus, and a teacher applies to transfer from the primary school to the middle school. The other type is across campuses. For example, transferring from a middle school teacher at a school in Lahore to a different branch of the school system in Karachi. 6% of teachers make a within campus transfer and 11% of teachers make an across campus transfer each year.

Transfers are initiated by the teacher and need to be accepted by the receiving school. Transfers are recorded in the administrative data and we can observe rejected transfer applications. Transfers are nearly always accepted by the receiving school. This is because incumbent teachers have hiring priority and there is high turnover within the system, virtually guaranteeing open positions at the school of interest each summer. The vast majority of transfers and resignations happen over the summer break between school years. Therefore

---

5Pairwise randomization by baseline test performance was used, which generally performs better than stratification for smaller samples (Bruhn and Mckenzie, 2009).
it is appropriate to think of this setting as a one-sided choice problem, as the schools have little say in who within the transfer applicants is hired.

During the information campaign which explained the treatment status of their school, teachers were also told that if they switched schools they would be under the contract of that given school. Teachers were provided information over email and through their employee portal about the treatment status of all other schools. This was important for transparency. If an employee changes to a different site within a firm they should know the incentive scheme they will receive before moving. It also ensured full information for all study participants, allowing the possibility of positive selection. Teachers were also reminded of the treatment status of their school and other schools during the summer break via email and their employee portal as that is the time most transfers take place. Administrative data records the date of transfer application and acceptance, allowing us to see who transfers and when.

4.2 Data

Data used comes from four main sources: i). historical administrative data from the school system ii). teacher and principal survey conducted at baseline and endline and iii). student test and survey data conducted at endline iv). classroom observation data conducted during the intervention year.

Administrative data – The administrative data documents the teachers’ demographics, classes, subjects, school assignment, salary and performance rating from their supervisor from June 2014 to June 2019. It also provides information on their entry date into the school system and their leave date if they resigned during the intervention.

Student test scores from the school system serve as pre-interention data for four years, midline and follow up data. Test exams are designed by the school systems’ central office and administered to all students in a given grade. However, tests are graded locally by the school, often times, by the students’ teacher. Due to concerns of grade manipulation, for the midline tests, grading was audited by the research team\(^6\), and the endline test was conducted exclusively by the research team (described in Student test and survey section below).

Teacher and principal survey – The teacher survey was conducted with a sub-sample at baseline and the full sample at endline. The survey contained incentivized elicitation of

\(^6\)10% of all teachers’ exams were regraded. If the teachers’ grade and the auditor’s grade were off by more than 5%, another 10% of their tests were audited. If the average was still off by more than 5%, all of the teachers exams were regraded. Overall, grade manipulation was small and was generally driven by cases where teachers bumped up grades from failing to passing. There was no heterogeneity in grading accuracy by treatment.
the teachers’ belief distribution of their historical performance\textsuperscript{7}. It also measured their risk aversion using a high and low-stakes coin flip game, equivalent to half a day’s wage and a week’s wage. The endline survey also included measures of intrinsic motivation (Ashraf, 2015) and efficacy (Burrell, 1994).

The principal survey was conducted at baseline and endline and includes measures of beliefs about teacher VA and a measure of management quality (from the World Management Survey questionnaire).

\textit{Student test and survey} – The research team conducted an endline test and student survey in January 2019. The test was conducted in Reading (English and Urdu), Math and Science. The items were written in partnership with the school systems curriculum and testing department and the research organization and grading was conducted by the research organization. Items from international standardized tests (TIMSS and PERL) and a locally used standardized test (LEAPS) were also used to benchmark students.

The student survey includes measures of satisfaction and socio-emotional development. Question items were selected using measures of satisfaction with school (LASSI, National Student Survey), empathy (Bryant, 1982), social preferences (Afrobarometer), grit (LASSI, Litman and Spielberger) and global competence. These areas were selected as they were of importance to the school system, as they are components of their mission statement for developing well-rounded students. They are also areas that teachers’ could potentially have an impact on (Jackson, 2018).

\textit{Classroom observation} – Finally, classroom observations were conducted for a random sub-sample of schools. Two sets of observation tools were used. First, the research team used the CLASS observation rubric (Pianta et al, 2008). Second, local school administrators (principals and vice principals) used a simplified version of the CLASS tool.

\textsuperscript{7}The distribution of teachers’ beliefs were measured in two ways. The first was more complicated but allowed us to incentivize teachers. Teachers were asked what they believed their VA was in the previous year. They were then asked to put weight on how likely they thought it was they were in the bottom 20\% of teachers, second quantile, etc. Their payoff would be the weight they placed on the correct answer divided by the total weight they placed across all five options. Teachers watched a video explaining the payoff structure and were required to correctly answer understanding checks before moving on to the allocation. Screen captures from the video are shown in Appendix, Fig. 4. The second way we measure belief distribution was simply asking teachers beliefs of their percentile ranking and the highest and lowest percentile they believed they could have been ranked.
4.3 Sample and Implementation Checks

On average, teachers in our sample have 8 years of experience and have been teaching at this school system for 5 years. 76% of teachers are female and they are 37 years old on average. 67% have some post-BA credentials/degrees. Yearly turnover is 29%. There are a mix of career teachers and those who are less attached to their school. 70% and 36% expect to still be teaching at their current school in 1 year and 10 years, respectively. On average these characteristics are balanced across treatment. Table 1 presents balance checks between the three main treatment arms. Of 10 tests, 0 are significant at the 10% level. Administrative data exists for the entirety of the sample. The endline teacher survey was conducted with 65% of teachers who were in the system at the announcement of treatment. 29% of those who attrited were teachers who left the school system during the intervention year. The remaining were teachers who opted out of taking the survey. Attrition is balanced by treatment.

Understanding of their own appraisal system is strong. One year after the treatment announcement, 80% of teachers correctly identifying the key aspects of the appraisal system they were assigned to. Knowledge of appraisal systems in other schools is relatively low, though, which could impede sorting across schools. 15% of teachers could name off the top of their head a school which was assigned to given treatment arm.

5 Contract Choice

The following two sections present the main results of the paper. I first present the results of the contract choice exercise and then observed sorting during the RCT.

5.1 Contract Choice and Baseline VA

To understand whether schools that offer performance pay are likely to attract higher value-added teachers (Prediction 1 of the model), we first compare the value-added of teachers who choose for their raise to be based on their percentile value-added versus a flat (non-performance-based) raise. Table 2, Column 1, presents the results of the following specification:

\[ VA_{i,t-1} = \beta_0 + \beta_1 \text{ChosePP}_i + \epsilon_i \]  

where \( VA_{i,t-1} \) is a teachers’ pre-intervention value-added and \( \text{ChosePP}_i \) is the contract the teacher chose at baseline. Here we think of \( VA_{i,t-1} \) as teachers’ quality in the absence of
performance incentives, as this is the year prior to the intervention.

Using the contract choice exercise, we find that teachers who choose performance pay, relative to flat pay contracts, have higher pre-intervention value-added. On average their students’ had 0.16 sd higher test scores in the previous year. These effects are larger than most direct effects of performance pay. This stands in contrast to previous evidence in which teachers are asked to predict their VA rank (Springer, 2010). When asked straight out what they believe their ranking to be, teachers are highly overconfident, a result we replicate in our sample as well (Figure 3).

So, why do we see higher types now choosing performance pay contracts? There are two possible explanations for this result: (i). teachers do, in fact, have incorrect beliefs about their VA, but VA is correlated with other preferences (risk, etc) that make high types more likely to choose performance pay, or (ii). moving from an unincentivized/sensitive question to an incentivized/non-sensitive question allows people to reveal their true preferences.

We posit the second mechanism is at work. Overall, we do not find evidence for the first claim. Higher VA teachers do not have different risk preferences, preferences for competition or beliefs about the fairness of certain contracts. VA is also not correlated with any of the predictors of contract choice we discuss in the following section. We also find that when we provide teachers information about their historical VA, this trickles through to affect their contract preferences. However, from this data we cannot disentangle whether it is the stakes themselves of the contract or the fact that the question is now less personal which moves people to reveal their true preferences. However, this highlights the importance of using real decision-making outcomes, when it comes to outcomes individuals’ may exhibit overconfidence toward.

5.2 Other predictors of contract choice

We find evidence of predictions 2, 3, and 4 of the model as well. To test prediction 2, we compare the direct effects of performance pay for those who selected it pre-RCT.

\[ TestScores_i = \beta_0 + \beta_1 AssignedPP_j + \beta_2 ChosePP_i + \beta_3 AssignedPP_j \cdot ChosePP_i + \epsilon_i \]

for teacher \( i \) at school \( j \), where \( \beta_3 \) gives us the effect of performance pay for those who wanted it. Table 3, shows that the effects of performance pay are on average 0 for those who did not want it. Whereas, for those who chose performance pay, the effects are 0.18 sd, which are large relative to most performance pay interventions.
We also find that individuals’ risk preferences are predictive of their choice in contract, as is the certainty in the belief about their performance in the following year. These results are of course correlations but give us insight into what could be affecting their contract choice. Figure 2 shows the coefficients of a regression predicting contract choice.

While we are unable to randomly vary individuals’ risk preferences, as discussed in the previous section we update individuals’ beliefs about rank and see this feeds through to their contract preference. We can also compare individuals’ contract preferences on an aspect of performance they are well-acquainted with, their subjective performance rating by their manager, compared to VA, which they are less certain about. Subjective performance ratings have been conducted for the last decade in the school system, and teachers feel much more certain about their future performance in this dimension than in VA, which is being introduced to them for the first time during the exercise. Table 2, Column 2, shows the relationship between contract preference (performance raised based on their subjective rating versus flat raise) and their actual performance. Comparing columns 1 and 2, we see the relationship is much stronger for a measure of performance they are more certain about.

6 Performance Pay and Teacher Sorting

While the contract choice is crucial to help us better understand the extent of private information teachers may have, the policy-relevant outcome is the quality of teachers performance pay schools will attract relative to flat pay schools. Switching costs or other preferences for location may significantly outweigh contract preferences, swamping out any possibility of positive sorting.

To test this question, we estimate the quality of individuals who end up in performance pay schools after a year.

\[ VA_{i,t-1} = \beta_0 + \beta_1 TeachatPP_i + \beta_2 Post_i + \beta_3 TeachatPP_i \times Post_i + \epsilon_i \]

*TeachatPP* is a dummy for whether they teach at a school that was assigned performance pay, *Post* is a dummy which is 1 for December 2018, the end of the intervention, and 0 for December 2017, the month before the announcement of treatments. \( \beta_1 \) tells us the difference in quality between schools assigned performance raises versus flat raises at the time the contract was announced. This coefficient is a test of balance between the treatment and control groups, as there should be no difference in teacher quality just before the contract is announced. \( \beta_2 \) tells us the change in quality of teachers teaching at flat raise schools between
the beginning and end of the intervention year. $\beta_3$ is the key coefficient of interest. It tells us whether performance raise schools attracted better teachers over the year of the intervention.

Table 4 presents the results. It appears there is some imbalance at baseline but the differences are not statistically significant. Overall, schools assigned performance raises see an increase in teacher quality by 0.12sd in teacher VA relative to flat raise schools during the intervention year. This translates into about a 0.03 sd increase in student test scores between performance and flat pay schools from sorting effects. This increase is driven almost entirely by high VA teachers moving into performance pay schools. We see very few low VA teachers moving out as a result of the intervention.

Is this positive sorting due to teachers’ knowing their own quality or could it be that higher VA teachers have other preferences that affect their choice to move into performance pay schools? First, as shown in Section 5.2, VA is not correlated with risk or competition preferences. Second, the subjective performance raise serves as an effective falsification test of this question. If high VA teachers were simply interested in the competitive nature of the contract or some other feature of performance pay, we would expect to see them sorting into subjective performance raise schools as well. And yet, table 5 shows that we do not see any sorting in response to this contract type. This is because a teachers’ subjective performance score does not tend to carry with them from school to school. It appears it is a principal-specific aspect of type, so teachers cannot sort based on it. Teachers are aware of this feature – stating that they expect to receive a 0.3 sd lower subjective score if they move schools.

While we do observe positive sorting, it is important to highlight the magnitude of the results. As prediction 5 suggests, there is much less positive selection comparing the contract choice (0.16sd) to the actual sorting we observe (0.03sd). This is as expected because non-salary preferences and switching costs will influence the extent of sorting. However, because those who prefer performance pay also have larger behavioral responses the combined effect of selection (both better teachers and those who are more motivated by the contract) results in substantial positive sorting.

Finally, teachers appear to have significantly more information about their own quality than their principal. Table 6 shows that when we control for principals’ beliefs about teacher VA, the contract decision is still highly predictive teacher quality.
References


Bloom, Nicholas, Christos Genakos, Raffaella Sadun, and John Van Reenen. “Management Practices Across Firms and Countries”, Academy of Management Perspectives 2012 26:1, 12-33


8 Figures

Figure 1: Study Timeline
Figure 2: Predictors of contract choice

Fraction of raise want based on outcome

Teacher Prediction (Mean)
Teacher Prediction (Variance)
Risk Lovingness

Value Added
Principal Evaluation

Bivariate
Multivariate
Figure 3: Teachers beliefs about their VA
## Tables

Table 1: Balance Table

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) Flat Mean/SE</th>
<th>(2) Objective Mean/SE</th>
<th>(3) Subjective Mean/SE</th>
<th>(1)-(2) T-test P-value</th>
<th>(1)-(3) T-test P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>37.929 (0.682)</td>
<td>37.259 (0.564)</td>
<td>37.770 (0.544)</td>
<td>0.448</td>
<td>0.855</td>
</tr>
<tr>
<td>First year teacher</td>
<td>0.193 (0.031)</td>
<td>0.228 (0.020)</td>
<td>0.178 (0.019)</td>
<td>0.351</td>
<td>0.683</td>
</tr>
<tr>
<td>Years of experience</td>
<td>4.851 (0.339)</td>
<td>4.748 (0.323)</td>
<td>5.147 (0.302)</td>
<td>0.824</td>
<td>0.515</td>
</tr>
<tr>
<td>Female</td>
<td>0.755 (0.038)</td>
<td>0.785 (0.034)</td>
<td>0.746 (0.048)</td>
<td>0.566</td>
<td>0.880</td>
</tr>
<tr>
<td>Value-added</td>
<td>-0.653 (0.363)</td>
<td>-0.501 (0.135)</td>
<td>-0.696 (0.112)</td>
<td>0.263</td>
<td>0.701</td>
</tr>
<tr>
<td>N</td>
<td>1108</td>
<td>711</td>
<td>3,348</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clusters</td>
<td>42</td>
<td>42</td>
<td>170</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses, * p<0.10, ** p<0.05, *** p<0.01
Table 2: Contract Choice and Baseline Performance

<table>
<thead>
<tr>
<th></th>
<th>Student Z-Score</th>
<th>Subjective Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choose PP Raise (vs. flat raise)</td>
<td>0.163***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0608)</td>
<td></td>
</tr>
<tr>
<td>Choose Subj. PP Raise (vs. flat raise)</td>
<td></td>
<td>0.374***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.124)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0238</td>
<td>-0.304***</td>
</tr>
<tr>
<td></td>
<td>(0.0352)</td>
<td>(0.0762)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.002</td>
<td>0.010</td>
</tr>
<tr>
<td>Observations</td>
<td>159195</td>
<td>133527</td>
</tr>
<tr>
<td>Clusters</td>
<td>1154</td>
<td>920</td>
</tr>
</tbody>
</table>
Table 3: Direct effects of performance contract by contract preference

<table>
<thead>
<tr>
<th></th>
<th>(1) Endline Test Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assigned PP Raise</strong></td>
<td>-0.0128</td>
</tr>
<tr>
<td></td>
<td>(0.0679)</td>
</tr>
<tr>
<td><strong>Chose PP Raise (vs. flat)</strong></td>
<td>-0.00255</td>
</tr>
<tr>
<td></td>
<td>(.0636)</td>
</tr>
<tr>
<td><strong>Assigned PP Raise*Chose PP Raise (vs. flat)</strong></td>
<td>0.175**</td>
</tr>
<tr>
<td></td>
<td>(0.0821)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>-0.0299</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.063</td>
</tr>
<tr>
<td>Class and Subject Fixed Effects</td>
<td>Yes</td>
</tr>
<tr>
<td>School Baseline, Contract Pref</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment Clusters</td>
<td>143</td>
</tr>
<tr>
<td>Observations</td>
<td>52,005</td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01
Table 4: Sorting during Intervention

<table>
<thead>
<tr>
<th></th>
<th>(1) Baseline VA (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teach at PP School</td>
<td>0.152 (0.135)</td>
</tr>
<tr>
<td>Post</td>
<td>-0.0244 (0.0222)</td>
</tr>
<tr>
<td>Teach at PP School * Post</td>
<td>0.118** (0.0588)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.653* (0.363)</td>
</tr>
</tbody>
</table>

\[ R^2 \] 0.096

Class and Subject Fixed Effects, School Baseline: Yes
year Treatment Clusters: 128
Observations: 2112

Clustered standard errors in parentheses, * p<0.10, ** p<0.05, *** p<0.01
Table 5: Sorting during Intervention, Subjective Performance Raise

<table>
<thead>
<tr>
<th></th>
<th>(1) Baseline VA (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teach at PP School (subj)</td>
<td>-0.0431 (0.112)</td>
</tr>
<tr>
<td>Post</td>
<td>-0.0244 (0.0222)</td>
</tr>
<tr>
<td>Teach at PP School (subj) * Post</td>
<td>-0.00124 (0.0528)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.653* (0.363)</td>
</tr>
</tbody>
</table>

\[ R^2 \] 0.096
Class and Subject Fixed Effects, School Baseline Yes
year Treatment Clusters 128
Observations 2112

Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01
Table 6: Sorting during Intervention, Controlling for Principal beliefs

<table>
<thead>
<tr>
<th></th>
<th>(1) Baseline VA (z-score)</th>
<th>(2) Baseline VA (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teach at PP school</td>
<td>0.149</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td>(0.137)</td>
</tr>
<tr>
<td>Post</td>
<td>-0.0245</td>
<td>-0.0247</td>
</tr>
<tr>
<td></td>
<td>(0.0221)</td>
<td>(0.0221)</td>
</tr>
<tr>
<td>Teach at PP school * Post</td>
<td>0.115**</td>
<td>0.114**</td>
</tr>
<tr>
<td></td>
<td>(0.0575)</td>
<td>(0.0563)</td>
</tr>
<tr>
<td>Principal belief about VA</td>
<td></td>
<td>-0.00418</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.126)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.599</td>
<td>-0.546</td>
</tr>
<tr>
<td></td>
<td>(0.369)</td>
<td>(0.505)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.095</td>
<td>0.096</td>
</tr>
<tr>
<td>Class and Subject Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>School Baseline</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment Clusters</td>
<td>128</td>
<td>128</td>
</tr>
<tr>
<td>Observations</td>
<td>76974</td>
<td>76974</td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01
10 Appendix

Baseline Survey Features

Figure 4: Screen capture from baseline survey: Incentivized belief distribution elicitation

Example: Tahir thinks its likely he’ll get a B

Example: Tahir gets a B

Weight on B = 4
Weight on all grades = 6

$\frac{4}{6} \times 500 = 333$ winning
Figure 5: Screen capture from baseline survey: Contract randomization

What appraisal system will my branch receive?

Employees who liked 1 get 1. Employees who liked 3 get 3.
You get to pick for just yourself.

System 1
9 7
4 10

System 3
1 8
11 3

Your choice!
12 5
6 2

Figure 6: Screen capture from survey video: Calculation of percentile VA

Example: 5th grade math teacher Mrs. Qureshi

Ahmed performed better than 80% of students who began the year at the same place
Implementation

Figure 7: Timing of transfers
Figure 8: Treatment assignment map