THE MISALLOCATION OF PAY AND PRODUCTIVITY IN THE PUBLIC SECTOR: EVIDENCE FROM THE LABOR MARKET FOR TEACHERS*

NATALIE BAU† JISHNU DAS‡

March 14, 2017

Abstract

We use a unique dataset of both public and private sector primary school teachers and their students to present estimates of (a) teacher value added (TVA) and its correlates in a low income country and (b) the link between TVA and teacher wages. Moving a student from a teacher in the 5th to the 95th percentile of the public school TVA distribution leads to a 0.54 standard deviation increase in test scores, relative to a 0.39-0.55 increase in the United States. Moving a student from the 5th to the 95th percentile in the overall distribution would increase mean test scores by 0.69 standard deviations. Although the first two years of experience, as well as content knowledge, are associated with TVA, all observed teacher characteristics explain no more than 5 percent of the variation in TVA. Finally, there is no correlation between TVA and wages in the public sector (although there *is* in the private sector), and a policy change that shifted public hiring from permanent to temporary contracts, reducing wages by 35%, had no adverse impact on TVA, either immediately or after 4 years. The study confirms the importance of teachers in low income countries, extends previous experimental results on teacher contracts to a large-scale policy change, and provides striking evidence of significant misallocation between pay and productivity in the public sector.

*Natalie Bau gratefully acknowledges the support of the National Science Foundation Graduate Research Fellowship and the Harvard Inequality and Social Policy Fellowship. Jishnu Das acknowledges funding from RISE. We are also grateful to Christopher Avery, Deon Filmer, Asim Khwaja, Michael Kremer, Nathan Nunn, Roland Fryer, Owen Ozier, Faisal Bari and seminar participants at the World Bank, CERP, NEUDC, IADB, the University of Auckland, and the University of Delaware for helpful comments. The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments they represent.

†University of Toronto. (email: natalie.bau@utoronto.ca)

‡World Bank. (email: jdas1@worldbank.org)

1 Introduction

How to recruit and reward teachers is one of the most contentious questions in education today. Some believe that the best way to improve the quality of teachers is to hire the brightest college graduates by offering high salaries.[1] Others argue that public school teachers are overpaid, and with increasing fiscal stress in many countries, teacher salaries are a natural target for retrenchment (Biggs and Richwine, 2011). Understanding the characteristics that make an individual a good teacher and whether the same characteristics are also highly rewarded by the outside labor market is key to this debate. If, for instance, the brightest college graduates can earn high salaries in other professions but are not better teachers, they may be the wrong population to target for recruitment. This question has received considerable attention in the United States, but sparse data has impeded similar investigations in low-income countries.

We examine both the question of what makes a good teacher and the link between wages and productivity using a unique dataset that we collected between 2003 and 2007 from the province of Punjab, Pakistan as part of the Learning and Educational Achievement in Pakistan Schools or LEAPS project. These data contain test-score information on matched teacher-child pairs, permitting teacher value-added (TVA) estimation for 1,533 public school teachers and 975 private school teachers from 574 public and 345 private primary schools, observed in grades 3, 4 and 5. These data allow us to estimate the association between teacher characteristics and teacher wages in both sectors. We are also able to combine these data with an unexpected regime change at the beginning of the data collection period, which led all new hires in the public sector to receive temporary contracts with lower wages. By contrasting the TVA of teachers hired under the old and the new regime, we assess whether the change adversely affected TVA, either for those hired immediately after the change or those hired several years later.

The main results are as follows. First, teachers matter, and the variation in teacher quality is greater when we include both the public and private sectors. A 1 standard deviation (sd) increase in public sector TVA leads to a 0.16sd increase in mean student test scores, and a 1sd increase in overall TVA leads to a 0.21sd incease in test scores. Moving a student from the 5th to the 95th percentile of the public sector TVA distribution increases test-scores by 0.54sd, which can be compared to an average annual test score gain of 0.41sd in our sample. Although observed teacher characteristics explain no more than 5% of the variation in TVA – a common result in the literature on teacher effectiveness – we do find that the first two years of teaching experience are associated with a significant increase in TVA. Since we are able to examine experience effects for the same

[1] An influential report by McKinsey, for instance discussed the need to hire top graduates (Auguste et al., 2010): "Given the real and perceived gaps between teachers' compensation and that of other careers open to top students, drawing the majority of new teachers from among top-third students would require substantial increases in compensation" (p. 7).

teacher over time, this result is arguably causal. Consistent with recent work by Bold et al. (2017) in Africa, we also find that higher content knowledge is associated with significantly higher TVA, a correlation that emerges clearly once we account for measurement error using multiple test scores from the same teacher.

Second, there is no link between TVA and wages in the public sector.[2] Public sector wages reward seniority and education – both of which have small associations with TVA – but not TVA directly. Although this "zero-gradient" result is widely believed to be true, to our knowledge this is the first direct empirical test in a low-income country. Using similar data from *private schools*, we compare the gradient in the "market" with that in the public sector. Rewards to seniority are one-fifth as high in the private sector and, strikingly, a 1sd increase in TVA increases wages by 11%. Even in the absence of a formal testing regime, it may be that TVA is somewhat observable and can be rewarded, but the public sector does not have a mechanism to do so.

Third, lowering the wages of public sector teachers has no effect on the TVA of new entrants.[3] We compare the TVA of teachers hired just before and after a change in the hiring regime that moved all new hires to temporary contracts with significantly lower wages. In our data, 93% of new hires in 1997 were permanent teachers and by 2002, 89% were contract teachers, who received salaries that were 35% *lower*, not counting further cost savings from benefits such as pensions.[4] Instead of a decline, we typically find a *positive*, though not consistently significant, impact of contract status on TVA. When we compare contract and permanent teachers with more similar levels of experience, we find larger and more significant positive effects. We also do not find evidence that the pool of new teachers worsened over time with similar TVA estimates for later and earlier hires.

We investigate two possibilities that can lead to this result: either (1) teachers hired prior to the policy change possessed attributes that were rewarded in the outside labor market, but these attributes are not correlated with TVA, or (2) there is a significant wage premium in the public sector even conditioning on teacher characteristics. The weight of the evidence favors the latter as the more plausible explanation. Average education levels and test scores of new hires, both of which are arguably rewarded in the non-teaching labor market, did not decline after the regime

[2]Here, we assume that TVA is a useful measure of productivity for teachers. We recognize that teachers have other functions that we do not capture, and how these are rewarded is an important agenda in its own right. Nevetheless, childrens' test scores remain a key component of educational performance in any system, especially since they are a strong predictor of adult outcomes (for example, see Chetty et al. (2014b)).

[3]Determining whether wages are "too high" is typically difficult since researchers must determine how teachers are compensated relative to their outside options. To do this, they must adjust for different schedules (summer vacation), education levels, cognitive ability and the type of teacher, and different adjustments lead to different conclusions about the relative size of teacher compensation. Weissman (2011) is a non-technical summary of the studies and these issues.

[4]We arrive at 35% by regressing log teacher salaries on teacher characteristics, including seniority, as well as an indicator variable for contract status. Thus, we report the difference between contract teacher and non-contract teacher wages after accounting for any differences in observable teacher characteristics.

shift. Moreover, the significant wage premium in public sector teaching jobs is also seen in a comparison of teachers' wages in public and private schools. In 2003, teacher salaries in the public sector were 5 times higher than those in the private sector (Andrabi et al., 2008) and by 2011, they were 8 times as high. When private sector wages are one-fifth as high, decreasing the public wage by one-third does not affect the quality of new entrants; the outside option is *never* more attractive than the public sector in our data.

Both our TVA estimates and the effect of the policy change we present are based on observational variation, and much of our sample is from small, rural schools with a single classroom per grade. Although this is a common scenario – in the census of primary public schools for the province, average class sizes for grades 3, 4 and 5 in 2005 were only 17, 16, and 13 – it does not permit grade-specific, within school-year TVA estimates since we typically do not observe multiple teachers teaching the same grade in the same school in the same year.

Given this extension of the TVA literature to small schools that provide the bulk of education in rural regions of low-income countries, it is particularly important to assess the robustness of our TVA estimates, In particular, although we observe multiple teachers teaching different classrooms through the years, our results are suspect if children sort systematically to teachers. Therefore, we assess whether our TVA estimates are biased in various ways, and in our most important test, we examine whether a teacher's TVA predicts test score gains among a sample of children who switch schools. We first show that the child's future teacher's TVA after she switches schools does not predict her current teacher's TVA, suggesting little persistence in the quality of children's teachers. We then show that the gain in test-scores for a child who switches schools is precisely predicted by the TVA of the teacher that she is matched to, suggesting that our estimated TVA indeed captures variation in teacher quality. This out of sample validation test, similar in spirit to that of Chetty et al. (2014a), who use teacher switching instead, suggests that we can extend TVA computations even to small schools with single grades. Such schools are the norm in low-income countries.

Additionally, we find no evidence that contract teachers were systematically allocated to better performing schools or students (which would be the main potential source of bias). In fact, contract teachers were typically assigned to schools with fewer facilities. Furthermore, there is no correlation between students' test score trends and future assigment to a contract teacher, either at the school or at the student level. Additionally, to account for unobserved selection of students or teachers to schools, in our main results, we compare contract and permanent teachers *within* the same school.

The TVA estimates that we present here are, to our knowledge, among the first such estimates from a large sample in a low-income country. In perhaps the first study of teacher and classroom effects outside of a high income country, Araujo et al. (2016) present estimates of the variance of teacher effects in Ecuador after randomly allocating students to teachers and also find that observed characteristics explain little of the variation in students' outcomes, even after including direct measures of classroom effort. Our almost precise replication of (a) the low explanatory power of observed characteristics and (b) the positive effects of the first two years of teaching experience that is also found in the U.S. literature (Rockoff, 2004; Chetty et al., 2014a Rivkin et al., 2005) is informative because the variation in observed characteristics in our data is arguably much greater. In our data, 49% of public school teachers and 74% of private school teachers do not have a bachelor's degree, and (self-reported) mean days absent per month range from 0.5 at the 10th percentile to 5 at the 90th percentile of the absentee distribution in the public sector and 0 at the 10th percentile and 4 at the 90th percentile in the private sector.[5] At the same time, content knowledge and TVA are highly correlated once we account for measurement error. It could be that, at the very low levels of learning that we see in our sample, both among children and among teachers, content knowledge becomes increasingly important.

Our results also add to a recent literature that highlights the misallocation in public resources between teachers' wages and other inputs in low-income countries (Pritchett and Filmer, 1999). In related work in Kenya, Duflo et al. (2011) and Duflo et al. (2014) show that contract teachers cost less, but the test-score gains of children randomly matched to contract teachers are higher. They also show that contract teachers with higher performance were more likely to be rewarded with tenure in later years, and thus career concerns provide additional incentives to exert effort. In India, Muralidharan and Sundararaman (2013) allocate a contract teacher to randomly chosen schools. They show that schools where contract teachers were assigned gained more in test-scores. Using observational data, they suggest that there is an independent contract-teacher effect beyond the reduction in student teacher ratios caused by the additional teacher. Finally, Bold et al. (2013) repeat the experiment in Duflo et al. (2014) with a NGO and the government. They are able to replicate Duflo et al.'s (2014) results when the NGO implements the policy, but not when the government implements. Our main contribution to this literature is to examine the effects of a policy change that affected the entire teacher labor market in a Pakistani province, allowing us to identify the broad labor market effects of replacing permanent teachers with contract teachers. Since contract teachers are substantially cheaper, even if these teachers have zero effect, a large-scale contract teacher policy would allow the government to reallocate spending toward other investments. Our estimates of the contract teacher effect that partially account for contract and permanent teachers' different levels of experience are positive and significant, consistent with Duflo et al. (2014) and Muralidharan and Sundararaman (2013), and across specifications, we never find evidence that a large-scale policy reducing teachers' salaries negatively affected teacher quality.[6] Additionally,

[5]For comparison, the absence rate of an average teacher in the U.S. is 5%, and is only 3.5 percentage points higher for schools whose proportion of African American students (a marker associated with disadvantage in the United States) is in the 90th percentile (Miller, 2012).

[6]Comparisons based on the performance of the average permanent and marginal contract hire could conflate

our results suggest that shorter experiments that compare inexperienced contract teachers and permanent teachers may *underestimate* the effectiveness of contract teachers due to the large TVA gains that we observe over the first two years of teaching.

More broadly, a large literature from the OECD typically finds public sector premia of 5 to 15%, with some portion of the gap explained by differential motivation, sector-specific productivity and the selection of workers (Disney and Gosling, 1998; Dustmann and Van Soest, 1998; and Lucifora and Meurs, 2006). In contrast, in our study, public sector wages are 5 times as large as private sector wages, and within the public sector, a decline in wages of (at least) 35% has no negative impact on productivity as measured by TVA.[7] The much higher wage premium in Pakistan is consistent with the findings of Finan et al. (forthcoming), who use household survey data from 32 countries to show that the wage premium in the public sector is higher in poorer countries. Taken as a group, these studies all point to large and significant misallocations in the pay of public sector teachers in low-income countries, which we are able to demonstrate in the context of a large-scale change in the public sector recruitment of teachers.

We must acknowledge that although we get close to the natural experiment of a wholesale reduction in wages, if contract teachers believed that their chances of obtaining a permanent contract would increase with their effort, the policy change affected both remuneration levels and career incentives. Separating the two effects would require a separate experiment.[8] If the career incentives channel is important, our results would show that temporary contracts induced a combination of teacher effort and quality that can yield the same learning at a lower cost, at least for some years, an issue we return to in Section 5.4 below.

The remainder of our paper is as follows. Section 2 describes the setting and context, and Section 3 discusses the data. Section 4 discusses TVA estimation, the results of regressions of TVA on teacher characteristics, and the robustness of the TVA measures. Section 5 discusses the link between teacher quality and teacher wages, and Section 6 concludes.

2 Setting and Context

The data are from rural Punjab, Pakistan, the largest province in the country with an estimated population of 100 million. The majority of children in the province can choose to attend free public schools, or they can pay to attend private school, and at the primary level, one-third of enrolled

differences due to experience or cohort effects with differences in contract status (for instance, an average permanent hire is older and more experienced in our data).

[7]Since we do not include future liabilities such as pensions in this accounting, the wage difference is a lower bound in our study.

[8]A related experiment in Indonesia doubled teachers salaries but again found no increase in learning; this is a conceptually separate experiment from ours since it focuses on the link between effort and wages for existing teachers rather than the link between wages and recruitment (De Ree et al., 2015).

children choose to do so.[9] Although funding for public schools has traditionally been small, in recent years, the government of Punjab has ratcheted up education budgets from 468 million dollars in 2001-2002 to 1.680 billion dollars in 2010-2011 (Ishtiaq, 2013). Much of this expenditure is on recurring budget items, and, similar to other low-income settings, teachers' salaries account for 80% of spending (UNESCO Islamabad, 2013).[10]

Whether public sector teachers wages in Pakistan are 'adequate' depends on the comparison. Comparisons with Indian states show that both Pakistani and Indian teachers earn, on average, 5-7 times GDP per-capita (Siniscalco, 2004 and Aslam, 2013). Comparisons with other professions again suggest similar levels of remuneration relative to other professionals. Each of these has obvious problems: Comparisons across countries require that teachers are efficiently compensated in the "benchmark" country. Comparisons across professions are subject both to selection concerns and differences in the job profiles across occupations.

Teacher salaries in the private sector provide an alternative benchmark. Andrabi et al. (2008) show that teachers' wages in private schools were one fifth of teacher salaries in public schools in 2003-2004, and public school salaries have only grown relative to private school salaries since then (Figure 1). Similar wage gaps have been documented in Colombia, the Dominican Republic, the Philippines, Tanzania, Thailand, and India (see Jimenez et al., 1991; Muralidharan and Kremer, 2008). These large wage premiums may reflect a lack of accountability and the strength of teachers' unions rather than greater productivity. Absenteeism is high in the public sector, and firing is rare since teachers are protected by permanent contracts (Chaudhury et al., 2006). In our sample, public school teachers self-reported absences of 2.6 days per month compared to 1.9 days per month for private school teachers. Recent research accounting for selection bias in both Pakistan (Andrabi et al., 2010) and India (Muralidharan and Sundararaman, 2015) shows that attending private schools, despite a lower per-student cost, improves student outcomes.

Unfortunately, a direct public-private comparison of the wage gap is also confounded by the large differences in observed teacher characteristics between the two sectors. Appendix Table A1 shows the large differences in training (90% versus 22% in the public relative to the private sector), education (51% hold a bachelor's degree versus 26%), gender (45% female versus 77%), and local residence (27% local versus 54%). Private school teachers also report 11 years less teaching experience on average. Using an Oaxaca-decomposition exercise, Andrabi et al. (2008) argue that controlling for observed characteristics explains little of the wage gap between public and private school teachers, but there is currently little direct evidence on the link between pay and productivity in the public sector.

[9]Religious schools, or madrassas, account for 1-1.5% of primary enrollment shares, and their market share has remained constant over the last two decades (Andrabi et al., 2006).

[10]Bruns and Rakotomalala (2003) show, in a study of 55 low-income countries, that teacher salaries account for 74 percent of recurring spending by the government on education.

2.1 Natural Experiment

Instead of relying on comparisons across countries or professions, we provide a direct test of the appopriateness of teacher remuneration: if remuneration is truly benchmarked to outside options, any decline should decrease teacher quality. If quality does not decline, two possibilities remain: either the wrong teachers were being hired (those with characteristics that are correlated with high wage outside options but not correlated with TVA), or public sector salaries are too high. Here 'too high' only means that a reduction in salaries does not affect teacher quality and has no implications for what constitutes a "fair wage."

The move to hiring teachers on temporary contracts with a decline in remuneration came about as follows. In the mid-1990s, the Government of Punjab started exploring changes in hiring practices, responding to both reports of low accountability and performance, and concerns about the budgetary implications of high wages and benefits for public sector employees. As these deliberations were gathering steam, unanticipated nuclear tests in 1998 led to international sanctions and a worsening of the budgetary position of the province, providing the final impetus for changes in public sector hiring practices and leading to a much wider use of contract teachers in public schools. Figure A1 shows that while the number of teachers hired each year varies, corresponding to the practice of "batch" hiring in the province (Bari et al., 2013), the period following the sanctions (1998-2001) is a uniquely long period of low hiring. After normal hiring resumed in 2002, almost all teachers in the province were hired on untenured, temporary contracts and received, as we will show empirically, 35% lower wages than permanent teachers with similar levels of experience.[11]

Separating the effects of tenure insecurity and lower compensation is challenging in our study, but some institutional details may shed light on the different mechanisms. Cyan (2009) notes that the institution of contract hiring was supported by a more centralized hiring process that relied on a point system based on employee qualifications, as well as interview performance. The policy also dictated that contract employees would undergo increased performance evaluation, and in surveys, 45% of contract teachers said that performance evaluations were linked to their contract renewal (Cyan, 2009). Performance evaluation may have increased teacher effort: 74% of surveyed contract teachers said that they were made to work more than regular teachers, and absenteeism and disciplinary infractions appeared to be lower among contract teachers. Importantly, as of 2009, there was no formal process for regularizing the contract teachers who were typically employed on 3-5 year contracts. Consistent with this, 71% of teachers said that they did not think their jobs offered them an opportunity for "professional growth," and 95% of teachers reported working on a temporary contract for more than three years. Therefore, in 2009, it seems that most contract

[11] Contract arrangements in Punjab became more common from 2000-2001 on (Hameed et al., 2014) and in 2004, the Government of Punjab announced its Contract Appointment Policy (Cyan, 2009).

teachers did not expect to be regularized in the future. This sets a time-frame for the wage savings from temporary contracts if all temporary contracts were to be converted to permanent in 4 to 7 years, the latter for those hired in 2002. In reality, it wasn't until 2012 that continued agitation by existing contract teachers led many to be converted to permanent status, receiving concomitantly higher wages thereafter.

Our natural experiment allows us to conduct a simple but important exercise to understand the effects of changes to teacher hiring policies. By examining what happens to teacher quality when the government decreases salaries by more than one-third for *all* incoming teachers, we can directly assess how large-scale contract teacher policies affect the quality of new entrants, as measured by their effect on student outcomes, with the caveat that the reform jointly affected incentives and remuneration, rather than remuneration alone.

3 Data

We use data collected across four rounds (2003 to 2007) of the Learning and Educational Achievement in Punjab Schools Survey (LEAPS). The original sample includes 823 schools (496 public) in 112 villages of 3 districts in the province of Punjab, with an additional 335 (111 public) schools entering the sample over the next four years.[12] The project was designed as part of a study of the rise of private schooling and, as a result, all the villages included in the study were randomly selected from a list frame of villages with at least one private school when the study began in 2003. As these villages tend to be larger and wealthier, the sample is representative of 60% of the rural population in the province of Punjab.

For our purposes, two parts of the data collection are key. First, a teacher roster was completed for all teachers within the school in each year of the survey. This roster included socio-demographic data on teachers (gender, age, educational attainment) and in the fourth round, month-level data on when the teacher began teaching. We use variables from the teacher roster to look at the differences between contract and permanent teachers in demographic characteristics, salaries, and subject knowledge. Appendix Table A1 provides summary statistics for these characteristics for both public and private school teachers across the four rounds of the survey.[13]

Our data collection reveals that most schools in our sample have 1 or fewer teachers per grade; we only observe multiple teachers teaching in a grade in 26% of public schools and 29% of private

[12]The three districts were chosen on the basis of an accepted stratification of the province into the better performing north and central regions and the poorly performing south.

[13]At times, we wish to compare teachers in terms of measures that were collected over multiple survey rounds, such as school facilities or teacher absences. To normalize these measures, we regress them on year fixed effects and teacher or school fixed effects, depending on the level at which the characteristic is observed. We then use the teacher or school fixed effect as the teacher-level measure. This process is analogous to how we combine test score data from multiple years to calculate teacher value-added measures.

schools. In Table 1, we report the number of teachers observed teaching each combination of grades 3, 4, and 5 in both public and private schools. Because teachers teach multiple grades, it is possible for a teacher to be observed teaching two or more grades even if they are observed once. For example, 8 public school teachers and 3 private school teachers are only observed once, but teach grades 3, 4, and 5 simultaneously. We report the teacher counts for both the full sample and a restricted sample that excludes teachers who ever appear to teach the same class in concurrent years (that is, more than 25% of their students in year t also were taught by them in year $t-1$). While many teachers are only observed once, which can happen if a teacher quits teaching (this occurs frequently after marriage in the private sector), transfers to another school, or starts teaching in 2007, a large number of teachers are observed two or more times, even when teachers who teach essentially the same students in subsequent years are excluded (615 total public school teachers observed teaching grade 3, 4, or 5 and 81 total private school teachers).

In the second part of our data collection, to assess learning outcomes, LEAPS tested children in the survey schools. English, Urdu, and mathematics tests were administered to children in grades 3, 4, and 5 between 2004 and 2007. In the first year of data collection, only classrooms with grade 3 students were tested. In subsequent years, those children were followed to new classrooms, and an additional cohort of 3rd graders was added in year 3 and followed in year 4. Appendix A discusses the implementation and scoring of these tests. Here we note that (a) the tests were low-stakes and designed by researchers to maximize precision over a range of abilities in each grade and (b) scores could be equated across years using a set of linked questions in each year together with Item Response Theory as in Das and Zajonc (2010). These test equating methods allow us to score all children in all years on the same scale in a comparable fashion. Appendix Table A1 documents test score gains by year over the four rounds of testing in the panels of public and private school students.

On the day that children were tested, we also asked teachers to take the same test as the children so that they could assess the test themselves. One worry is that these tests, designed to assess learning by third to fifth graders, are uninformative due to ceiling effects. Appendix Figures A2 and A3 show the histograms of teacher test-scores in public and private schools, and although there are ceiling effects, particularly in math, where 14% of public teachers achieve the maximum, there is also a great deal of variation. Equated to the child test score distribution, the mean public teacher test score is 3.04sd higher, but the 5th percentile of teachers was only 1.91sd higher than the average tested child. As the process of testing teachers was repeated each year, whenever we observe teachers multiple times, we have multiple observations of the teacher test score, which we will use to adjust for measurement error. Appendix Table A2 correlates teacher content knowledge with teacher characteristics for both public and private teachers.[14] Teacher characteristics only

[14]Running the regressions separately for public and private school teachers yields qualitatively similar results.

explain 7% of the variation in content knowledge, and reassuringly, having a bachelor's degree is robustly correlated with content knowledge. However, the correlation is relatively small (0.2 to 0.3sd), which could either reflect the quality of the degree or 'learning on the job' among those without a degree.

Teacher quality is identified following the TVA literature (Rockoff, 2004; Chetty et al., 2014a; and Kane and Staiger, 2008) by regressing student test scores on a function of their lagged test scores, round, grade, and teacher fixed effects. Teacher value-added is the estimated teacher fixed effect. The panel structure of the data, where both students and teachers are observed multiple times, is important for identification: to be included in the value-added calculations, students must be observed at least twice across consecutive years, since they require a lagged test score to control for selection. To separate correlation in student outcomes within years from TVA, at least some teachers must also be observed across years so that round fixed effects are identified. To estimate the variance of teacher value-added, we cannot simply take the variance of these value-added estimates, both because it would be biased by sampling bias and because we cannot separately identify classroom-level shocks from teacher value-added for teachers who are only observed once. While this does not bias our estimates when teacher value-added is an outcome variable, as long as classroom shocks are uncorrelated with the independent variables in the regression, it would bias estimates of the variance of teacher effects. Therefore, to estimate the variance of teacher value-added, we focus on the sample of teachers who are observed at least two times in the data.

We observe a total of 1,756 public school teachers in at least one round of the data linked to 22,857 unique public school students and 1,346 private school teachers linked to 9,741 unique students. Appendix Figures A4 and A5 document this variation in public and private schools. However, we are only able to estimate TVAs for 1,533 of the public teachers and 975 of the private teachers. This is primarily because the TVA estimation does not include students if they were not observed in the prior year. In particular, this means that we cannot calculate TVAs for teachers who were only observed in the first round of the data. Table 2 provides more information on the sources of variation for the TVA calculations in both public and private schools. When we correlate TVAs with teacher characteristics, our sample is further reduced to 1,383 public and 294 private teachers. This is because detailed data on when a teacher started teaching, which allows us to include our experience controls, was only collected in the fourth round of the LEAPs study.

To account for unobservable variables, we can also demean TVA estimates at the school level.[15]

[15] In practice, we present our estimates for both TVAs that are and are not demeaned at the school level. De-meaning at the school level means (1) that we can only consider teachers who teach in multi-teacher schools and (2) that our measures of teacher quality are essentially within-school rankings of teacher quality. If schools try to equalize marginal products across teachers (e.g. by reallocating resources that substitute for quality toward less effective teachers or giving more effective teachers larger class sizes), estimating teacher quality within schools may lead us to under-estimate the true variation in teacher quality.

Since we do not observe teachers in more than one school, we cannot separately identify pure school effects as opposed to a school simply having better teachers on average. Therefore, the demeaned TVAs should be interpreted as a within school ranking of teacher quality. Demeaning at the school level requires that more than one teacher was observed in the school over the course of the study. TVAs for teachers in the 158 public schools and 86 private schools where only one teacher was ever observed with tested students are left out of the within-school TVA sample; these teachers account for 2,357 child-year observations (1,771 unique children) in public schools and 936 child-year observations (414 unique children) in private schools.

4 Teacher Value-Added

4.1 Teacher Effectiveness: Estimating the Variance of Teacher and Classroom Effects

To measure the variance of teacher and classroom effects in our data, we closely follow Araujo et al. (2016), extending their methodology to produce the first estimates, to our knowledge, from a low-income country. These are policy-relevant measures since the variances of classroom and teacher effects are indicative of how much test scores would increase if a student moved to a 1sd better classroom or teacher. To measure the variance of these effects, we first generate estimates of classroom effects (in our sample, equivalent to a teacher-year effect) by estimating the regression

$$y_{it} = \sum_{a} \beta_a y_{i,t-1} I(grade = a) + \delta_{jt} + \alpha_t + \mu_g + \upsilon_{it}, \tag{1}$$

where y_{it} are test scores in math, Urdu, and English for a student i in year t, β_a is the grade-specific effect of lagged test scores, α_t is a year fixed effect, μ_g is a fixed effect for grade g, and δ_{jt} is the classroom effect, which includes both idiosyncratic classroom-level shocks, such as having a more disruptive student in the classroom in a given year, and the teacher effect. Since teachers do not change schools in our sample, we cannot separately identify school-specific effects. Instead, to estimate the variance of the classroom effects, following Araujo et al. (2016), we de-mean our estimates of δ_{jt} at the school-level. While demeaning at the school-level may be less attractive than demeaning at the school-year-grade level, it is a necessity in many low-income countries, like Pakistan, where there is often only 1 teacher and 1 class per grade in the modal school. Appendix Figure A6, which plots the distribution of grade sizes in Punjab in 2005, illustrates this fact. To estimate the variance of δ_{jt}, we cannot simply taking the variance of our estimates of δ_{jt}, since these estimates have sampling error, upwardly biasing the empirical variance. Instead, we derive the sampling bias, accounting for the de-meaning procedure, and subtract it from the empirical variance of the classroom effects.[16] This provides us with an unbiased estimate of the variance of the classroom effects.

To measure the variance of teacher effects, we can exploit the fact that we observe teachers in multiple classrooms over time. Following Araujo et al. (2016), Hanushek and Rivkin (2012), and McCaffrey et al. (2009), we can take advantage of the fact that the variance of the teacher effects is $Cov(\lambda_{jt}, \lambda_{j,t+1})$, where λ_{jt} is the true de-meaned classroom effect for teacher j in year t. However, the facts that we observe $\widehat{\lambda_{jt}}$, rather than λ_{jt} and that our estimates of school-level means include multiple observations of the same teacher complicate this approach. If a teacher is associated with an abnormally high estimated class-effect in one year, this will lead us to infer that her effect is lower in another year, mechanically inducing a negative correlation between our de-meaned estimates. To ensure that this does not affect our results, when we estimate the variance of the teacher effects, we de-mean $\widehat{\delta_{jt}}$ by subtracting the school-level mean of $\widehat{\delta_{-jt}}$. Even with this correction, $Cov(\widehat{\lambda_{jt}}, \widehat{\lambda_{j,t+1}})$ is an overestimate of the variance of the teacher effects since correlation in the estimation error in the school means will positively bias the covariance estimate. In Appendix B, we derive this bias. Then, we arrive at our estimates of the variance of the teacher effects by substracting the bias term from $Cov(\widehat{\lambda_{jt}}, \widehat{\lambda_{j,t+1}})$.[17]

Table 3 reports our estimates of the effect of moving to a 1sd better classroom or teacher on test scores for math, English, and Urdu. In the first row, we include the entire sample in the estimation proceedure. In the second row, we restrict the sample to teachers where less than 25% of students in their class are the same as the students they taught in the previous year to ensure that our results are not biased by teachers teaching the same students repeatedly. In the final row, we restrict our sample to public school teachers. In the full sample, moving to a 1sd better classroom increases mean test scores by 0.31sd and moving to a 1sd better teacher increases test scores by 0.21sd. Restricting the sample to teachers who teach different students from year to year leads to similar estimates (Row 2). In Row 3, when we restrict the sample to teachers in public schools, we find that a 1sd better teacher increases test scores by 0.16sd. This estimate suggests that moving a student from a teacher in the fifth to the ninety-fifth percentile of the public school TVA distribution would lead to a 0.54sd increase in mean test scores. Our estimates of the variance of the classroom and teacher effects are larger than the estimates in Ecuador (Araujo et al., 2016), and our estimates of the variance of teacher effects are on the high end of estimates in the United States (see Chetty et al. (2014a) and Rockoff (2004)). Given these estimates, teacher quality appears to be at least an important determinant of student outcomes in low-income countries like Pakistan as it is in the United States, if not more so.

[16] Our de-biasing procedure exactly follows the one described in Appendix D of Araujo et al. (2016).

[17] The variances of the teacher effects we estimate are almost identical to the variances of the teacher effects that we estimate when follow the methodology of Chetty et al. (2014a).

4.2 Estimating Teacher Effects

Following Rockoff (2004), we estimate TVA as a teacher fixed effect. This method is similar to the methods of Kane and Staiger (2008), Sass et al. (2014), Chetty et al. (2014a), and Chetty et al. (2014b), but unlike some of these approaches, the fixed effect approach allows us to estimate TVA for teachers who are only observed once in the data. To compute TVA, we estimate the following regression on our full set of teachers (public and private), including all child-year test score observations:

$$y_{it} = \beta_0 + \sum_a \beta_a y_{i,t-1} I(grade = a) + \gamma_j + \alpha_t + \mu_g + \varepsilon_{it},$$

where y_{it} is student i's test score in year t, γ_j is the teacher fixed effect, α_t is the round fixed effect, and μ_g is the grade fixed effect. Then, γ_j is the TVA, equivalent to the underlying unexplained variance in test score gains associated with students having the same teacher. As is conventional in the TVA literature, we control for year-specific and grade-specific shocks, as well as lagged test scores, which are allowed to affect students in different grades differently. These account for students' prior human capital attainment and the selection of students to teachers.

Like Chetty et al. (2014a) and Kane and Staiger (2008), we do not include child fixed-effects to account for additional unobservable selection of students to teachers (Sass et al., 2014). Identification with child fixed-effects would be based on a smaller sample of 7,696 unique children (one-third of the sample) who are observed with multiple teachers over time. More worryingly, measurement error in how teacher codes are entered into the data will lead to false switchers – students who appear to be switching teachers but actually are not. Even with a small number of false switchers, this could lead to large biases in the estimation of TVA by inducing spurious correlations between the TVA of teachers with similar ID numbers.[18]

We also do not use empirical Bayesian methods to estimate TVA. The empirical Bayes approach proposed by Kane and Staiger (2008) relies on the assumption that TVA is time invariant.[19] However, past work shows that teacher effectiveness non-linearly increases with experience in the first 1-2 years of teaching (Rockoff, 2004; Chetty et al., 2011), and we will confirm that this is the case in Pakistan as well. This violation of the assumptions of time invariance or stationarity is likely to be particularly problematic among Pakistan's inexperienced teacher labor force. Among

[18] Suppose that 1% of teacher IDs are randomly entered incorrectly. This will have little impact on TVA estimates that utilize the full sample. But if only 10% percent of students change teachers each year, when identifying variation comes only from the test scores of students who change teachers, these incorrect entries account for 9 percent of the variation. In other words, when we restrict the sample to students who change teachers, we always include incorrect ID entries, but we shrink the number of correct ID entries, increasing the percentage of the variation that is driven by students with incorrectly entered IDs. For more details, see Appendix C.

[19] Chetty et al. (2014a) relaxes time invariance and allows for drift in TVA but still assumes stationarity.

teachers observed in 2007 (the only year teachers were asked the year they started teaching), 13% of public school teachers and 54% of private school teachers had less than 3 years of experience. In our context, controlling for teacher experience would likely be necessary for the assumption of time invariance or stationarity to be valid. Unfortunately, experience is collinear with year hired, which in our setting, is highly correlated with contract status due to the sharp change in the hiring regime. Virtually all teachers with 0-5 years of experience are contract teachers, and virtually all teachers with more than 5 years of experience are permanent teachers. Since we cannot flexibly control for experience without subsuming the temporary contract effect, our estimates of γ_j utilize the full sample of students and teachers, averaging over teacher effectiveness at different experience levels.

Nevertheless, while we cannot fully non-parametrically control for experience effects, later in this paper, we control for whether teachers had less than 3 years of experience in 2007, implying that they had low levels of experience throughout the data collection period, to separate experience effects from the effects of other teacher characteristics on the TVA estimates. This method exploits non-linearity in the experience effect, as the literature shows that after the first two years of teaching, experience's effect on student outcomes plateaus (Rockoff, 2004). Using within-teacher observations of student test-scores in different years, we also verify that this is the case in Pakistan in the following section.[20]

Our fixed effect approach allows us to identify TVA for teachers who are only observed once in the data, as long as we observe their students at least twice. While we cannot separately identify teacher and classroom effects for these teachers, as long as the classroom-level shock is exogeneous to our treatments of interest, we can still use this combined measure of teacher and classroom effects to study the effects of the contract teacher policy on TVA. Additionally, since we focus on TVA as an outcome variable, decreasing classical measurement error in our TVA estimates (as methods like empirical bayes do) should not affect our estimates of interest.

Implementing the procedure described in this section results in TVA estimates for 1,533 public sector teachers and 975 private school teachers. In the next sections, we first examine TVA's correlations with teacher characteristics. We then use the TVA estimates to assess the link between productivity and wages, both in terms of the gradient (do higher TVA teachers earn more?) and the intercept (does lowering wages reduce average TVA?).

[20]Our TVA estimates do no capture teachers' heterogeneous effects on different students. In reality, such heterogeneity may be important for students' outcomes. For instance, Bau (2015) shows that schools in Pakistan can have different effects on the outcomes of more and less advantaged students. Relatedly, Aucejo (2011) shows that teachers responded to the incentive structure of No Child Left Behind in the United States by increasing the outcomes of their lower ability students at the expense of higher ability students. In other examples, Muralidharan and Sheth (2016), Antecol et al. (2015), Dee (2007), and Hoffmann and Oreopoulos (2009) measure the effects of the teacher-student gender match. If teachers have heterogeneous effects on students, our TVA measures can be thought of as capturing the effect of a teacher on the *average* student.

4.3 Teacher Value-Added Results

4.3.1 Teacher Characteristics and TVA

Using our TVA estimates, we estimate the association between TVA and student performance and the link between TVA and observed teacher characteristics for public school teachers using the following specification:

$$TVA_j = \beta_0 + \Gamma X_j + \alpha_d + \varepsilon_j,$$

where TVA_j is a teacher j's average value-added over math, Urdu, and English; X_j consists of teacher characteristics, including an indicator variable for some training, an indicator variable for having a bachelor's degree or greater, an indicator variable for having 3 or more years of experience in 2007, an indicator variable for female, an indicator variable for whether a teacher is local, an indicator variable for whether a teacher has a temporary contract, controls for age and age squared, and in some specifications, controls for a teacher content knowledge; and α_d is a district fixed effect. In some specifications, we also include a school fixed effect. Table 4 presents the results from this specification. Column 1 reports the means of the covariates of interest. Columns 2 and 3 report regression results without controlling for teachers' own test scores, and Columns 4-9 include different measures of teachers' test scores with the goal of reducing measurement error in teacher content knowledge.

Like in the United States, we are never able to explain more than 5% of the variation in mean TVA.[21] Presaging our discussion of wages and productivity in later sections of the paper, we find no significant, positive correlation between TVA and two key characteristics – education (measured as whether a teacher has a Bachelor's degree) and whether the teacher has some training. In fact, for training, the point estimate is negative and significant when we also include content knowledge. Nevertheless, two correlations are statistical significant and are of particular interest.

First, content knowledge has a significant correlation with estimated TVA. When we use subject-specific measures of the teacher test-score, the effects are generally small and insignificant (Columns 4 and 5). However, when we reduce measurement error by averaging over all subjects (Columns 6 and 7), the association becomes positive and significant. Columns 8 and 9 further reduce measurement error by instrumenting for the teacher's first cross-subject average test score with her second.[22] Our preferred IV specification suggests that a 1sd increase in teacher test scores increases TVA by 0.30-0.33sd, which is higher than the effects estimated by Metzler and Woess-

[21] We arrive at 5% by first regressing mean TVA on district or school fixed effects, then regressing mean TVA on the fixed effects and the teacher characteristics, and then calculating the difference in the adjusted R^2's.

[22] The inclusion of teacher test scores causes the sample size to fall since not all teachers took the test. Instrumenting for teachers' first test scores with their second test scores further reduces the sample since even fewer teachers took the test twice.

mann (2012) (0.1sd in Peru) but similar in magnitude to the effects estimated by Bold et al. (2017) (0.23sd-0.54sd in Africa). At first, it may seem surprising that teacher knowledge is statistically significantly correlated with mean TVA estimates while having a BA or better is not. However, recall from Appendix Table A2 that having a bachelor's degree is only associated with a 0.22sd (see Column 8) increase in a teacher's average test score. Based on the IV results, the implied effect of a bachelor's degree would be about 0.06sd, which is not significantly different from the estimated effect of 0.04sd (see Column 3 of Table 4). Our effect size is also substantially larger than the effect of IQ (0.04sd) estimated by Araujo et al. (2016). We can speculate that this is driven in part by the large variation in teacher knowledge levels in Pakistan, where a standard deviation in the teacher test score distribution in math is nearly as large as a standard deviation in the student test score distribution (0.87 vs. 1) and a teacher at the 5th percentile scores below a student at the 95th percentile in math on a test designed for primary school students.

Second, experience in the first two years of teaching also increases TVA. Columns 2-9 treat each teacher as a single observation, but unlike the other characteristics in these regressions, experience is not time invariant. Since we observe teachers multiple times at different levels of experience, we can use our panel data to better identify experience effects for public school teachers. In Column 10, an observation is a student-year, and we regress mean test scores on teacher experience and lagged student test scores, controlling for teacher fixed effects, which capture any time invariant teacher characteristics, using the following specification:

$$mean\,test\,score_{ijt} = \beta_0 + \beta_1 I(exp_j \leq 1) + \beta_2 mean\,test\,score_{ij,t-1} + \gamma_j + \varepsilon_{ijt},$$

where i denotes a student, j denotes a teacher, and t denotes a year. $I(exp_j \leq 1)$ is an indicator variable equal to 1 if a teacher j has 0 or 1 years of experience in time t, and γ_j is a teacher fixed effect. Then, we report β_1, our coefficient of interest. The results in Column 10 suggest that the experience effect is large: the outcomes of students of teachers with only 0 or 1 years of experience are 0.3sd worse than those of other students. In Appendix Table A3, we expand this specification to include controls for 0-1, 2, 3, and 4 years of experience. We find very large experience effects in the first year: students of a teacher with 0-1 years of experience have test scores between 0.55sd (in English) and 0.72sd (in math) lower than students of teachers with 5 or more years of experience. In the second year, the penalty is lower, ranging from -0.47sd for Urdu to (a marginally statistically significant) -0.26sd for English, and by the third year, there are no further significant experience effects.[23] Note that because we are comparing the outcomes of the students of the same teacher over time, these effects are more likely to have a causal interpretation relative to the associations between TVA and time invariant teacher characteristics.

[23]Estimates of the effects of 1, 2, 3 and 4 years of experience are qualitatively similar for private school teachers, and follow the same pattern, with large, negative non-linear effects from 0 or 1 years of experience.

Appendix Table A4 replicates the regressions of TVA on teacher characteristics for private school teachers. We find almost exactly the same pattern of effects, with only teacher content knowlege (after correcting for measurement error) and teacher experience consistently affecting the mean TVAs. As in the public sector, teacher characteristics explain little of the variation in mean TVAs in the private sector.

4.4 TVA Robustness

Our TVA estimates control for lagged student test scores and allow the effect of lagged student test scores to depend on the student's grade to account for the non-random assignment of students to different teachers. However, if these lagged tests scores do not sufficiently account for selection, our TVA estimates may be biased. We first assess the extent of the bias by checking whether adding additional controls for student and school characteristics alters our TVA estimates. In the new estimates, we control for the student's household assets index,[24] gender, and parental schooling, as well as two indices of school facilities that vary over time and time-varying school-level student-teacher ratios. Even though the new TVAs use substantially more information about students' socioeconomic status and school-level inputs (jointly, the controls explain 10% of the variation in mean test scores), they are highly correlated with the old TVAs with correlations of .98 in English, .97 in math, and .97 in Urdu (see Appendix Table A5).

Our next test relies on an "out-of-sample prediction" test following Chetty et al. (2014a). We focus on children who switched schools and test whether the TVA of a switcher's new teacher predicts her test score gains after switching schools, controlling for lagged test score. If TVA is a meaningful measure of teacher quality, a teacher's TVA should predict her student's learning gains. The specific regression specification is:

$$test\,score_{ijt} = \beta_0 + \beta_1 TVA_j + \beta_2 test\,score_{ij,t-1} + \alpha_s + \varepsilon_{ijt}, \tag{2}$$

where $test\,score_{ijt}$ is the test score of a student i with a teacher j in year t, which can be in math, Urdu, or English or the the average across all three, TVA_j is the value-added of a student's teacher in the relevant subject, and α_s is a school fixed effect. The sample consists of students who are in a new school in period t. Because we limit the sample to school-switchers, β_1 will not be influnced by common shocks at the school-level that are correlated over time.[25]

[24]We create an asset index by predicting the first factor of a principal components analysis of indicator variables for ownership of beds, a radio, a television, a refrigerator, a bicycle, a plow, agricultural tools, tables, fans, a tractor, cattle, goats, chicken, watches, a motor rickshaw, a scooter, a car, a telephone, and a tubewell following methods discussed by Filmer and Pritchett (2001).

[25]Focusing on children who switched schools ensures that our test will not find spurious correlations between future and current TVAs due to the fact that school-grade level shocks to the current teachers' students' outcomes will effect the lagged test scores used to calculate the future teachers' TVAs as described by Chetty et al. (2015).

Our test proceeds in two steps. We first ensure that β_1 is not biased by selection between students and teachers. If students who learn quickly are more likely to sort to certain teachers, even when they switch schools, then these teachers will appear to have a higher TVA and these high TVAs will also be related to students' outcomes. We follow Rothstein (2010) and test whether the TVA of a student's future teacher after a school switch predicts the TVA of her current teacher (Table 5). Across all subjects, there is no evidence of a correlation between current and future TVAs, suggesting that when children switch schools, the allocation to specific teachers mimics random sorting.

Table 6 reports the estimates from equation 2 and confirms that TVA in a subject is highly predictive of the students' test-score gains in that subject. For average test scores, a 1sd increase in TVA increases student test score gains by 0.852sd, and this coefficient is significant at the 1% level. In fact, we cannot reject that the true coefficient is 1, as would be the case if we had estimated the "true" TVA. These results suggest that our TVA estimates are very predictive of real student gains from teacher quality.

5 Teacher Productivity and Teacher Wages

Before moving to the link between TVA and wages, it is useful to have a a simple framework to interpret the results. Appendix D provides the formalization of a model where each teacher has a productivity θ_j drawn from a bounded distribution F, with a maximum of θ_{max} and a minimum of θ_{min}. In the public sector, teachers receive a wage w_{pub} set exogenously by the government, and the public sector hires randomly from its applicants. In the private sector, due to free entry, private schools make zero profits and a teacher receives her productivity, θ_j. A decline in w_{pub} can then have one of three potential effects. If w_{pub} is very low relative to the private sector, no teachers enter the public sector either before or after the wage change. If $w_{pub} > \theta_{max}$ both before and after the decline, then all teachers apply to the public sector in both cases, and jobs in the public sector are rationed. If w_{pub} is between θ_{min} and θ_{max} after the decline, more productive teachers sort into the private sector relative to before. In this case, lowering public wages will decrease the quality of new entrants. The final two cases are illustrated in Figures A7 and A8.

Since our empirical strategy allows us to identify the average productivity of teachers hired under higher and lower wages, we can directly test which of the figures is more applicable to the Pakistani educational system. If the results are in line with the predictions of Figure A7, it suggests that the government can reduce teachers' salaries without reducing student learning and without fear of causing shortages. However, we caution that the regime change we evaluate is more complex than the model presented here, since it may have also increased the returns to effort through career concerns.

5.1 Teacher Productivity and Teacher Wages

We now examine the association betweeen wages and TVA. In Table 7, we regress log salaries separately on public and private teacher characteristics (Column 1) and then include mean TVA in the regression (Columns 2-4) using the specification:

$$log(salary_j) = \beta_0 + \Gamma X_j + \alpha_d + \varepsilon_j,$$

where $log(salary_j)$ is the log of the mean salary of teacher j, and X_j consists of the same teacher characteristics as in equation 2. As before, α_d is a district fixed effect, and some specifications (Columns 3, 4, and 6) also include school fixed effects.

In public schools, receiving some training is associated with a 52% increase in teacher salaries and having a bachelor's degree is associated with a 26% increase.[26] In addition, seniority is heavily rewarded in the public sector, with every additional year of age resulting in a 5.8% (no school fixed-effects) to 6.3% increase in wages (with school fixed effects). Recall that the first two years of teaching experience have a large effect on TVA. While we cannot include both experience and age non-parametrically in this Micerian regression, we can include an indicator variable for whether the teacher had more than 3 years of experience in the final round for which data was collected. We find no additional effect of experience beyond the seniority effect. Similarly, teacher content knowledge (Column 4) has small and insignificant effects on teacher salaries. Unsurprisingly, teachers with temporary contracts make 35% less than teachers with permanent contracts.

Strikingly, every attribute that the public sector appears to reward has no significant effect on TVA. When we add mean TVA to the regressions (Columns 2-4), the coefficient is small, insignificant, and negative. Moreover, adding mean TVA has no effect on the adjusted R^2, suggesting that mean TVA does not explain any of the variation in salaries. We infer that higher quality teachers do not appear to be rewarded with higher salaries in the public sector, consistent with our theoretical framework.

Perhaps TVA cannot be rewarded because it is difficult to observe or verify. Using our data on private schools, we replicate the specifications in Columns 2 and 3 for private school teachers in Columns 5 and 6. The differences in compensation schemes in Column 5 are striking. As has been noted before (Andrabi et al., 2008), the private sector pays teachers according to their outside option, penalizing women and teachers who are locally resident. The private sector also rewards training and education (in similar ways for education, but less so for training). However, the premium on seniority is much lower and TVA is highly correlated with salaries. A 1sd increase in TVA is associated with a 11% increase in wages, and this coefficient is statistically significant

[26] Almost all public school teachers have at least some training. Therefore, the large association between training and salaries relies on 44 individuals (3% of the sample) who have no training.

at the 5% level.

In Column 6, we replicate the regression in Column 5 including school fixed effects. The inclusion of school fixed effects eliminates the positive relationship between mean TVA and salaries. There are two possible reasons for this: (1) across-school TVA estimates are biased because unobservably better students attend more expensive schools or (2) the link between teacher pay and TVA is driven by the fact that better teachers are sorting into higher-paying, better private schools. While either explanation is possible, given that we find little evidence of bias in our across-school TVA estimates in Appendix Table A5 or Tables 5 and 6, we believe that the second explanation is more likely, indicating that high-performing private schools pay higher wages to attract better teachers.

Our results suggest that teacher compensation in the public sector *does not* reward more productive teachers, but it is unlikely that this is because teacher productivity is impossible to observe. In the private sector, more productive teachers work at schools where they earn substantially more than less productive teachers. Our framework points to the next natural question: would a decline in public sector wages lower the average quality of public school teachers? To answer this question, we now estimate the effects of the contract teacher policy, which lowered wages by 35%, on the characteristics of individuals entering the teaching profession and on TVAs.

5.2 Methodology

While our TVA measures do not appear to be biased, we cannot simply regress TVA or other teacher characteristics on a teacher's contract status to estimate the effect of a contract teacher policy since contract status is not randomly assigned. The 2% of teachers hired and retained on temporary contracts prior to 1998, as well as the 17% hired on permanent contracts after 1998, are likely to be highly selected. Instead, the hiring regime change in 1998 allows us to instrument for contract status using the budgetary shock. Moreover, because the shock changed contract status for much of the labor pool, our natural experiment allows us to understand the effect of a large-scale contract teacher policy on teacher labor supply.

To estimate the effect of the contract regime on what types of individuals become teachers, on which schools and students those individuals are assigned, and on teacher productivity, we first estimate the ordinary least squares regression:

$$\begin{aligned} y_j = {} & \beta_0 + \beta_1 TempContract_j + \beta_2 month\_hired_j \\ & + \beta_3 month\_hired_j \times Post_j + \alpha_d + \varepsilon_j, \end{aligned} \tag{3}$$

where y_j are the characteristics of teacher j, including her TVA, her students, and the school to which she is assigned. $TempContract_j$ is an indicator variable equal to 1 if a teacher has a temporary contract and 0 otherwise, $month\_hired_j$ is the time a teacher was hired, measured at the month level, $Post_j$ is an indicator variable equal to 1 if a teacher was hired in or after 1998, and α_d is a district fixed effect. We include time trends in teacher quality to account for the fact that most of the variation in contract status is driven by whether teachers were hired before or after the budgetary shock. Moreover, we exclude the small number of (likely highly selected) teachers hired during the hiring freeze from our sample. For some outcomes, like number of teachers in a school, student-teacher ratios, and school facilities, we have extreme outliers, likely due to data entry error and misreporting, that lead to very skewed distributions. To ensure that our results are not sensitive to these outliers, we exclude the top and bottom 1% of observations for these variables.

Even so, the estimates of β_1 from this OLS regression are likely to be biased for several reasons. First, as we discussed before, it does not account for the selection in contract status for the teachers who were hired on temporary contracts before 1998 and the teachers who were hired on permanent contracts after 2002. Second, we typically observe teachers hired on a temporary contract with fewer years of experience in our data since these teachers are hired later. Since the effects of experience on student learning are highly non-linear, linear time trend controls that span the entire sample are unlikely to fully account for these experience effects. As we decrease the bandwidth of the sample, including fewer hiring years around the policy change, linear trends will better account for the experience effect.[27] Alternatively, we can also directly exploit the non-linearity in experience effects by only calculating TVAs using years when contract teachers hired in 2002 and 2003 had 2 years of experience or more (2005-2007) and then comparing the experienced permanent teachers to the experienced contract teachers. Restricting the sample to a two year bandwidth does just this since 2004 (round 1) is not included in the TVA estimation (lagged test scores are not available) and the two year bandwith sample is restricted to teachers hired in 2002 and 2003.

In our main specification, we account for the sources of bias using a fuzzy regression discontinuity design comparing teachers hired right before and after the budgetary shock. This approach is analogous to an instrumental variables regression that incorporates time trends and includes a subset of the sample around the budgetary shock. Therefore, to estimate β_1 without selection on contract status, we instrument for $TempContract_j$ with the indicator variable $Post_j$. The first stage of this two stage least squares strategy is then:

[27]To see this, write our TVA estimates as $\widehat{TVA}_j = \theta_j + g(month\,hired_j)$ where g, the experience effect, is an arbitrary function of the month a teacher was hired and θ_j, teacher productivity, is our measure of interest. As the bandwith we include in the fuzzy regression discontinuity specification approaches zero, the linear trend in month hired will approach g and account for the experience effect.

$$TempContract_j = \delta_0 + \delta_1 Post_j + \delta_2 month\_hired_j + \delta_3 month\_hired_j \times Post_j + \alpha_d + \mu_j. \tag{4}$$

Following Lee and Lemieux (2010), who discuss regression discontinuities with discrete data, such as time, we cluster our standard errors at the month hired level.

We report these regressions in the following sequence. We first check for potential biases that could arise from the systematic allocation of contract teachers to schools and parents; our main concern is that contract teachers may have been assigned to children who learned faster. We then estimate the effect of temporary contracts on TVA to examine the effect of the policy change on teacher productivity. Finally, we examine whether the policy affected the composition of teachers, by looking at the diferences in teacher characteristics before and after the regime change.

5.3 Results

5.3.1 *Existence of a First Stage*

The first panel of Figure 2 shows the discontinuous effect of being hired after 1998 on contract status. Being hired after 1998 is associated with an 80 percentage point increase in the probability that a teacher is hired on a temporary contract. Each point in the figure is the average of the outcome variable for teachers hired that month (ranging from 1 to 182 tachers). The second panel shows the similar discontinuity in salaries, with regression equivalents in Table 8. Each coefficient in the table is the result from separate regressions of the form specified either in equation 3 (OLS) or equation 4 (fuzzy RD with a four-year bandwidth or three-year bandwidth such that the sample includes 1994-1997 and 2002-2005 or 1995-1997 and 2002-2004). In the OLS regression (Row 1, Column 1), temporary contract status is associated with a 28% decline in a teacher's salary, which is somewhat less dramatic than the fuzzy RD estimates of 44-54% (Row 1, Column 3 and 5). These effect sizes are also consistent with the effect of temporary contracts in the Mincerian regression (35%), which accounts for observable characteristics of teachers but not unobservable characteristics that may be related to contract status. For brevity, we only present two RD bandwidths here (3 and 4 years). However, the negative effect of temporary contracts on log teacher salaries is significant at the 1% level for all remaning bandwidths that we tested (5-10 years) with effect sizes ranging from 69-83%.

5.3.2 *Effect of the Policy on Allocation of Teachers to Schools*

Figures A9 plots school facilities (as indices), student-teacher ratios and the number of teachers against the year a teacher started teaching with regression equivalents in Table 9. Both the figure and the regression results show that contract teachers were assigned schools with fewer extra facilties. Rows 3 to 9 present the coefficients for the components of the extra facilities index separately, and the effect of contract status on the extra facilities index is driven by schools with fewer libraries, who are less likely to have computers or electricity. In addition, Figure A9 and the final three rows of Table 9 suggest that parental education (particularly father's education) was lower for children assigned to contract teachers. The index of assets appears to be somewhat higher, but this is statistically insignificant. These results are consistent in both across- and within-school regressions. Given the large number of outcome variables we consider and the fact that we do not find consistent effects for most outcomes, we only focus on two bandwidths for the RD of the allocation of teachers to schools and students (3 and 4 years). However, when we turn to our key outcome of interest – our measure of teacher productivity – we present results for a wider range of bandwidths.

5.3.3 *Were Contract Teachers Assigned to Lower Ability Children?*

The fact that contract teachers were assigned to smaller schools with fewer extra facilities and lower levels of parental education could suggest that they were teaching children whose learning was systematically lower. If this is indeed the case, our TVA estimates for contract teachers may be negatively biased. Fortunately, the panel structure of our data set allows us to directly test whether higher or lower ability students were selectively matched to contract teachers, which is ultimately the main plausible source of bias in the TVA estimates. We therefore test directly whether contract teachers were assigned to schools with higher ability students by testing whether student test score trends predict a school being assigned a contract teacher. In Table 10, we first test whether time trends are the same for schools that never received contract teachers and schools that eventually received them. We estimate

$$\begin{aligned} y_{it} = \beta_0 + \beta_1 year_{it} + \beta_2 I(Received\,Contract\,Teacher)_s + \\ \beta_3 I(Received\,Contract\,Teacher)_s \times year_{it} + \Gamma X_{it} + \varepsilon_{it}, \end{aligned}$$

where y_{it} is the outcome variable, mean student test scores, $I(Received\,Contract\,Teacher)_s$ is an indicator variable equal to 1 if a school ever hired a contract teacher and 0 otherwise, $year_{it}$ is the survey year, and X_{it} is a vector of controls consisting of district fixed effects and lagged student test scores. The sample does not include any student-year observations from schools that have contract teachers in the survey year (or received one in a past year). As Column 1 shows, test scores are not different on average between schools that did and did not receive contract teachers. Moreover, there is little difference in pre-trends in test scores between public schools that do and do not receive contract teachers. If anything, the pre-trends for schools that later received contract teachers are negative.

Next, we assess whether test-scores gains predict receiving a contract teacher at the student instead of the school level. Column 2 shows that, within schools, there is no significant difference between the test scores or test score trends of students who will and will not eventually receive contract teachers. Finally, Column 3 tests whether yearly test score gains predict having a contract teacher. It shows that, across schools, a student's average test score gains before receiving a contract teacher are not predictive of whether he or she later receives a contract teacher. In summary, despite the fact that contract teachers were assigned to children with less educated fathers and schools that were smaller, there is no evidence to suggest that learning among children who received a contract teachers was systematically different.

5.3.4 *Effect of Contract Status on Student Test Scores*

Table 11 now presents the results of the OLS and RD specifications of mean TVA on contract teacher status, and the final two panels of Figure 2 are the graphical representations of the reduced form of the regression discontinuity specifications. Since TVA is our main outcome of interest, we report the instrumented results for the full sample and for bandwidths of 2-7 years around the policy change. The OLS effect of teacher contract status on mean TVA both across and within schools is small and the sign varies (-0.004 and 0.024). However, this estimate may be downwardly biased by selection of teachers hired before 1998 into contract teacher status and the relative inexperience of contract teachers during the years students were tested. To account for these effects, we estimate the fuzzy RD in Rows 2 to 8 of Table 11. The smaller bandwidths better accomodate non-linear experience effects since they do not assume that the effects of month hired are linear throughout the entire sample and they also do not include teachers (such as those hired in 2006 and 2007) who are only observed with 0 or 1 years of experience. Consistent with this, estimated effect sizes for contract teachers are larger when the bandwidth is 2, 3, or 4 years (although the estimates for the 2 year bandwidth, which includes only 227 teachers, are imprecise). The effect sizes are generally similar within and across schools, with positive effect sizes around 0.2 standard deviations for the smaller bandwidths. Larger bandwidths include contract teachers observed with lower experience levels, which is likely to negatively bias our estimates, and assume linear time trends across a larger sample of teachers. While temporary contract teachers no longer have significant positive effects for these larger bandwidths, we never find evidence of a strong negative contract teacher effect. Indeed, when we compare teachers in the same schools, regardless of bandwidth, we never estimate a negative contract teacher effect. Therefore, while there is not conclusive evidence that the contract policy raised teacher quality, it is very unlikely that the policy lowered teacher productivity.

We address two additional issues. First, to adjust the analytical standard errors for estimation error in TVA, we also estimate the RD p-values with a clustered bootstrap procedure (see Appendix Table A6). The pattern of significance for the estimates are similar to those obtained using analytic standard errors. Second, we test if contract teachers are more effective because they have smaller class sizes (though student-teacher ratios are not significantly different in the RD). In Appendix Table A7, we repeat the RD analysis after re-estimating the TVAs controlling for average school-year student-teacher ratios. The results are qualitatively and quantitatively similar. Overall, we find no evidence of a decline in TVA following the regime change and some evidence that the TVA of contract teachers was higher than the TVA of permanent teachers hired prior to the regime change.

5.3.5 *Effect of the Policy on Teacher Characteristics*

Interestingly, and consistent with the idea that salaries may be higher than is necessary to incentivize high quality teachers to enter the teaching profession, we find no evidence of a change in key characteristics of the teacher pool. Figure 3 shows broad trends since 1970 towards greater feminization, higher education and a greater proportion of younger teachers, but despite yearly variation in teacher characteristics, there is little evidence of a large trend break following the policy change.

The remaining rows in Table 8 formally compare the characteristics of contract and permanent teachers. OLS specifications containing the full sample appear to reflect the general but non-linear trend that teachers hired later are more educated; having a temporary contract increases the probability of having a bachelor's degree by 32 percentage points. However, in the RD design, the effect of having a temporary contract on bachelor's degree is no longer significant and is only 0.3-10%. In fact, there are no robustly significant differences between the characteristics of teachers hired on permanent and temporary contracts, with the exception of test scores in English for the RD with the 3-year bandwidth. This positive effect of temporary status on English scores is also inconsistent with the idea that lower wages led higher quality teachers to exit the teaching profession. The fact that the change in regime did not lead to a decline in the fraction of teachers with a bachelor's degree suggests that the outside options for these teachers remain below the considerably lower contract teacher wages. Interestingly, the RD estimates suggest that the fraction of female teachers increased following the establishment of the policy (and continued to increase in Figure 3), although the coefficient is insignificant and imprecisely estimated. We discuss the potential implications for teacher recruitment below in Section 5.4.

5.3.6 *Quality of the Teaching Pool Over Time*

While applicants hired right after the budgetary shock appear to be similar to applicants hired previously, the quality of applicants may still have changed over time. In Figure 3, there does seem to be a reduction in teacher training and an increase in workforce feminization after 2002 (although both may continue pre-existing trends and neither characteristic is strongly associated with TVA). To assess whether the quality of new teachers is decreasing over time, we would like to compare the test scores of the students of contract teachers hired earlier to those hired later. This poses several problems. First, on average, we observe more recently hired teachers with fewer years of experience. Thus, we will only compare the outcomes of the students when we see them with inexperienced contract teachers (teachers with 0 or 1 year of experience) to mitigate the effects of different levels of teacher experience. A second challenge is that later hires are only observed with students in later testing rounds. If student test scores are improving over time for unrelated reasons and if we do not control for year of testing, the effect of being a later hire will be upwardly biased. Therefore, we also include a control group of permanent teachers hired before 1998 in our regression sample, so that we can include testing round fixed effects. We estimate the regression:

$$
\begin{aligned}
y_{it} &= \beta_0 + \beta_1 month\_hired_j + \beta_2 Post_j + \beta_3 Post_j \times month\_hired_j \\
&\quad + \sum_g \beta_g y_{i,t-1} I(grade = g) + \alpha_t + \varepsilon_{it},
\end{aligned}
$$

where y_{it} is the test score of a student i in year t, $month\_hired_j$ is the time a teacher j is hired, with data available at the month-level, $Post_j$ is an indicator variable equal to 1 if a teacher is hired after 1998 and 0 otherwise, $y_{i,t-1}$ is a student's lagged test score, g is her grade, and α_t is a round fixed effect. β_3 then captures the effect of the month a teacher was hired on student outcomes for teachers hired after 2002. The coefficient β_2 does not have a clear interpretation. Because the sample is limited to inexperienced contract teachers, β_2 here is not analogous to the reduced form contract teacher effect in the fuzzy regression discontinuity. Instead, it captures a combination of the contract teacher effect and the inexperience effect. When we estimate this equation, β_3 is a small and insignificant -0.007 (0.024). Accordingly, there is little reason to believe that over time teacher quality decreased in response to decreased teacher salaries and teacher tenure.

5.4 Natural Experiment Robustness

In this section, we assess threats to identification and our ability to extrapolate the results beyond the specific context studied here.

Student Selection. The robustness tests in Section 4.3 indicate that our TVA estimates are not biased by selection of students to teachers. Therefore, it is unlikely that differences in student quality between contract and non-contract teachers are driving our results. Nonetheless, we can more formally test whether either observed or unobserved student quality drives the differences in student outcomes between contract and non-contract teachers. We follow Altonji and Mansfield (2014), who argue that the classroom level means of observable student characteristics can proxy for unobservable characteristics related to student outcomes. In revised TVA estimates (Appendix Table A8), we control for the classroom-level means of two of the most likely determinants of student-to-classroom and student-to-school sorting – lagged test scores and wealth. The point estimates are qualitatively and quantitatively similar to those reported previously in Table 11.

Selective Attrition: Assessing Student and Teacher Attrition. Our estimates of the contract teacher effect may also be biased if the students of contract teachers are differentially more likely to exit the sample or lower quality contract teachers are more likely to leave schools. 7% of contract teachers and 8% of permanent public teachers who appear in the data are not observed in the fourth and final round of data collection, while 73% of students observed appear in the fourth round (73% of those taught by a permanent teacher, as well as 73% of those taught by a contract teacher). Typically 80% of students observed in one year are observed in the next year.[28] Table A9 tests for both the types of attrition that could bias our estimates. In Column 1, we show that conditional on the year a student first appeared in the panel, the percent of times a student had a contract teacher has no significant effect on whether she appears in the fourth round of the panel. Column 2 shows that the percent of rounds a student was observed with a contract teacher does not predict the total number of years she is observed. In the remaining columns, we test for teacher attrition. Column 3 shows that the mean TVA of a contract teacher does not significantly predict whether she was present in the fourth round of the panel, and Column 4 shows that it does not predict the number of years a teacher was observed (conditional on the year she started teaching). Overall, we find no evidence of either differential attrition of the students of contract teachers or differential attrition of contract teachers by quality.

Permanent vs. Temporary Income. While the majority of contract teachers did not expect to be normalized in 2009 (Cyan, 2009), contract teachers did win a court case in 2012 which led many teachers to be tenured and receive salaries comensurate with permanent teachers. Therefore, teachers may have entered contract teacher teaching with the expectation that the salary reductions were temporary. If this is the case, to interpret our results, we must determine how much intial

[28]Students do not appear in the sample in a given year for a variety of reasons: they may be absent on the day of the test, they may have dropped out of primary school, or they may have moved on to a secondary school and are therefore no longer tested. For example, a student who was in 3rd grade in round (2004) will be in 6th grade and may no longer be in a primary school in round 4 (2007). In cases where students were absent on the day of the test, they typically reappear in the sample in later years; the probability that a student who was observed in year 1 is observed in year 3 is 74% (as opposed to the 64% we would expect if students never returned to the sample).

salary reductions reduced *permanent* incomes for contract teachers. This exercise requires several additional assumptions. We assume teachers had rational expectations and that the discount rate is a conservative 3%. Furthermore, we assume that a teacher expects to work for 40 years. For teachers hired in 2002, temporary contracts reduced their salaries by 35% for 10 of those 40 years. Even with this very low discount rate, the contract policy reduces permanent wages for teachers hired in 2002 by 13%, suggesting that there is still substantial room to lower wages without negatively affecting teacher quality.

External Validity. In Pakistan, the contract teacher policy was instituted in response to an economic crisis, which may have also negatively affected teachers' outside options. Thus, we should be cautious in applying these results to other contexts where teachers' outside options are unchanged. However, there is reason to believe that lowering teacher salaries, even in the absence of an economic crisis, would not result in a decline in productivity. We observe teachers who are hired as late as 2007, 9 years after the nuclear tests. The results in Section 5.3.6 indicate that these teachers are no worse than those hired in 2002. The recession in 1998 did not last 10 years; in fact, according to the World Bank, Pakistan experienced a period of relatively high per capita GDP growth from 2003-2007 (2.7-5.5% per year). Similarly, in 1997, the unemployment rate according to the World Bank was 5.8% and by 2007, it was 5.1%. Moreover, in our own data, we do not find that salaries fell for private school teachers hired after 1998. Taken together, these facts suggest that even if teachers' outside options fell after the nuclear tests, they had likely recovered by 2007.

6 Conclusion

This paper makes two important contributions to our understanding of the educational production function in low income countries. First, we provide among the first estimates of the correlations between teacher observable characteristics and teacher quality from a low income country, although the effect of good teachers on student test scores is large, and likely larger than in the U.S. This raises obvious questions of how teachers should be recruited and the relative benefits of systems with a probationary period followed by tenure for high performers. Our one result that adds to this discussion is that tests of content knowledge could potentially improve the quality of new entrants.

Second, this paper builds on work by Duflo et al. (2014) and Muralidharan and Sundararaman (2013) on the quality of contract teachers. Like these papers, which provide clean experimental estimates of the contract teacher effect, we find that contract teachers have as great as and perhaps moderately higher TVAs than permanent teachers. The large-scale poliy change that we study allows us to assess equilibrium effects, and we are able to demontrate the robustness of the previous experiments to such a change in an entirely different context. In fact, given the large experience effects in the first two years of teaching, it is likely that the previous papers actually underestimated the positive effects of experienced contract teachers.

The effect that we find may be linked to the greater hiring of female teachers after the regime change. One story consistent with our results is that when salaries go down, the number of male applicants decrease. If female applicants feel that they will not be fairly treated in a recruitment process (or, indeed, if they are not), the teaching pool will not include educated and high TVA female teachers. It could be that the lower salaries induced more women to apply and be hired for these jobs, leaving the quality of new entrants unchanged after the hiring regime change. We cannot directly test this since applicant data are not kept, but it would be consistent with a number of patterns that we find, and in particular, the lack of any effect of teacher salaries on observable characteristics, including education levels and test scores.

Our results also suggest that, at least in low income countries, policies that increase wage levels to attract higher-skilled teachers, like those advocated by Auguste et al. (2010), would be costly and ineffective. Since higher levels of education are not correlated with TVA and only weakly correlated with teacher test scores, the best teachers are not those with the greatest education, and government outlays increase by paying these teachers their outside options in low-income countries. More remarkably, wages are already so high that even a 35% decline has no impact on the education levels of new recruits. This paper suggests that public sector compensation could be signficantly redesigned to better account for the realities of low-income countries. Combining lower salaries (or salaries that are more strongly tied to teacher productivity) with greater investment in other school characteristics or student incentives could allow low-income countries to generate considerable fiscal savings.

References

Altonji, J. G. and R. K. Mansfield (2014). Group-average observables as controls for sorting on unobservables when estimating group treatment effects: The case of school and neighborhood effects. *NBER Working Paper #20781*. Cambridge, MA.

Andrabi, T., N. Bau, J. Das, and A. I. Khwaja (2010). Are bad public schools public “bads?” Test scores and civic values in public and private schools. *Working Paper*. Cambridge, MA.

Andrabi, T., J. Das, and A. I. Khwaja (2008). A dime a day: The possibilities and limits of private schooling in Pakistan. *Comparative Education Review 52*(3), 329–355.

Andrabi, T., J. Das, A. I. Khwaja, and T. Zajonc (2006). Religious school enrollment in Pakistan: A look at the data. *Comparative Education Review 50*(3), 446–477.

Antecol, H., O. Eren, and S. Ozbeklik (2015). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics 33*(1), 63–89.

Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady (2016). Teacher quality and learning outcomes in kindergarten. *Quarterly Journal of Economics 131*(3), 1415–1453.

Aslam, M. (2013). Focusing on teacher quality in Pakistan: Urgency for reform. *Right to Education*.

Aucejo, E. (2011). Assessing the role of teacher-student interactions. *Working Paper*. London, UK.

Auguste, B. G., P. Kihn, and M. Miller (2010). Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching: An international and market research-based perspective. *McKinsey & Company*.

Bari, F., R. Raza, M. Aslam, B. Khan, and N. Maqsood (2013). An investigation into teacher retention and recruitment in Punjab. *Institute of Development and Economic Alternatives*.

Bau, N. (2015). School competition and product differentiation. *Working Paper*. Toronto, ON.

Biggs, A. G. and J. Richwine (2011). Assessing the compensation of public-school teachers. *The Heritage Foundation*.

Bold, T., D. Filmer, M. Gayle, M. Ezequiel, R. Christophe, B. Stacy, J. Svensson, and W. Wane (2017). What do teachers know? does it matter?: Evidence from primary school teachers in africa. *World Bank Research Working Paper #7956*.

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a, and J. Sandefur (2013). Scaling up what works: Experimental evidence on external validity in Kenyan education. *Center for Global Development Working Paper #321*. Washington, DC.

Bruns, B. and R. Rakotomalala (2003). *Achieving universal primary education by 2015: A chance for every child*, Volume 1. World Bank Publications.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers (2006). Missing in action: Teacher and health worker absence in developing countries. *Journal of Economic Perspectives 20*(1), 91–116.

Chetty, R., J. Friedman, and J. Rockoff (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

Chetty, R., J. Friedman, and J. Rockoff (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 2633–2679.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from project star. *Quarterly Journal of Economics 126*(4).

Chetty, R., J. N. Friedman, and J. E. Rockoff (2015). Response to Rothstein (2014) 'revisiting the impacts of teachers'. *CEPR Discussion Paper #10768*. London, UK.

Cyan, M. (2009). Contract employment policy review. *Punjab Government Efficiency Improvement Program*.

Das, J. and T. Zajonc (2010). India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics 92*(2), 175–187.

De Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2015). Double for nothing? The effects of unconditional teacher salary increases on student performance. *NBER Working Paper #21806*. Cambridge, MA.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources 42*(3), 528–554.

Disney, R. and A. Gosling (1998). Does it pay to work in the public sector? *Fiscal Studies 19*(4), 347–374.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review 101*(5), 1739–74.

Duflo, E., P. Dupas, and M. Kremer (2014). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics 123*, 92–110.

Dustmann, C. and A. Van Soest (1998). Public and private sector wages of male workers in Germany. *European Economic Review 42*(8), 1417–1441.

Filmer, D. and L. Pritchett (2001). Estimating wealth effects without expenditure data or tears: An application to educational enrollments in states of India. *Demography 38*(1), 115–132.

Finan, F., B. A. Olken, and R. Pande (forthcoming). The personnel economics of the state. *Handbook of Field Experiments*.

Hameed, Y., R. Dilshad, M. Malik, and H. Batool (2014). Comparison of academic performance of regular and contract teachers at elementary schools. *Asian Journal of Management Sciences & Education 3*(1), 89–95.

Hanushek, E. A. and S. G. Rivkin (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics 4*(1), 131–157.

Hoffmann, F. and P. Oreopoulos (2009). A professor like me: The influence of instructor gender on college achievement. *Journal of Human Resources 44*(2), 479–494.

Ishtiaq, N. (2013). Understanding Punjab education budget 2012-2013: A brief for standing committee on education, provincial assembly of the Punjab.

Jimenez, E., M. E. Lockheed, and V. Paqueo (1991). The relative efficiency of private and public schools in developing countries. *The World Bank Research Observer 6*(2), 205–218.

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *NBER Working Paper #14607*. Cambridge, MA.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*, 281–355.

Lucifora, C. and D. Meurs (2006). The public sector pay gap in France, Great Britain and Italy. *Review of Income and Wealth 52*(1), 43–59.

McCaffrey, D. F., T. R. Sass, J. Lockwood, and K. Mihaly (2009). The intertemporal variability of teacher effect estimates. *Education 4*(4), 572–606.

Metzler, J. and L. Woessmann (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics 99*(2), 486–496.

Miller, R. (2012). Teacher absence as a leading indicator of student achievement: New national data offer opportunity to examine cost of teacher absence relative to learning loss. *Center for American Progress*.

Muralidharan, K. and M. Kremer (2008). *School choice international*. MIT Press.

Muralidharan, K. and K. Sheth (2016). Bridging education gender gaps in developing countries: The role of female teachers. *Journal of Human Resources 51*(2), 269–297.

Muralidharan, K. and V. Sundararaman (2013). Contract teachers: Experimental evidence from India. *NBER Working Paper #19440*. Cambridge, MA.

Muralidharan, K. and V. Sundararaman (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in India. *Quarterly Journal of Economics 130*(3), 1011–1066.

Pritchett, L. and D. Filmer (1999). What education production functions really show: A positive theory of education expenditures. *Economics of Education review 18*(2), 223–239.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005). Teachers, schools, and academic achievement. *Econometrica*, 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review P & P*, 247–252.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics 125*(1).

Sass, T. R., A. Semykina, and D. N. Harris (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review 38*, 9–23.

Siniscalco, M. T. (2004). Teachers' salaries. *Education for All Global Monitoring Report*.

UNESCO Islamabad (2013). Education budgets: A study of selected districts of Pakistan.

Weissman, J. (2011). Are teachers paid too much: How 4 studies answered 1 big question. *The Atlantic*.

Tables

Table 1: Variation in Grades Taught by Teachers and Number of Times Teachers are Observed

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Public Sector				Private Sector			
	Observed Once	Observed Twice	Observed Three Times	Observed Four Times	Observed Once	Observed Twice	Observed Three Times	Observed Four Times
Only Grade 3	235	37	14	14	346	35	2	0
Restricted Sample	235	33	8	11	346	32	1	0
Only Grade 4	166	14	1	0	274	8	0	0
Restricted Sample	166	12	0	0	274	8	0	0
Only Grade 5	0	0	0	0	0	0	0	0
Restricted Sample	0	0	0	0	0	0	0	0
Grade 3 and 4	31	235	53	12	29	83	19	6
Restricted Sample	31	31	1	0	29	11	1	0
Grade 3 and 5	13	37	8	0	11	14	6	0
Restricted Sample	13	35	5	0	11	9	1	0
Grade 4 and 5	25	110	18	0	27	31	19	0
Restricted Sample	25	26	1	0	27	10	0	0
Grades 3, 4, 5	8	48	214	83	3	28	25	28
Restricted Sample	8	14	9	1	3	6	2	0

This table reports counts of the number of teachers who are observed teaching only grade 3, only grade 4, only grade 5, only grades 3 and 4, only grades 4 and 5, and grades 3, 4, and 5 by how many times they were observed. The restricted sample excludes teachers who are ever observed teaching two classes of students who appear to be the same in two consecutive years (90% or more of the students in year t were taught by the same teacher in year $t-1$).

Table 2: Sources of Variation in Teacher Value-Added Calculations

	(1) Number of Teachers	(2) Number of Students	(3) Teachers in Schools With > 1 Teacher With Tested Students	(4) Students in Schools With > 1 Teachers With Tested Students
Public, Rd 1	486	8,340	4	131
Private, Rd 1	303	3,617	0	0
Public, Rd 2	593	9,327	214	3,290
Private, Rd 2	336	3,340	97	846
Public, Rd 3	1007	16,946	884	15,320
Private, Rd 3	579	6,777	524	6,247
Public, Rd 4	1103	15,357	812	12,610
Private, Rd 4	599	5,911	478	5,020

This table presents the breakdown of the data used to calculate within and across school TVAs. Within school TVAs require teachers to teach in schools where more than one teacher has tested students (such that the mean school effect is not equal to the sole teacher's TVA). The sample of students driving variation in the within school TVAs are the students who attend schools where more than one teacher has tested students.

Table 3: Classroom and Teacher Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Math		**English**		**Urdu**		**Average**	
	Class	Teacher	Class	Teacher	Class	Teacher	Class	Teacher
Full Sample	0.321	0.258	0.300	0.190	0.312	0.184	0.311	0.211
Restricted Sample	0.308	0.252	0.277	0.285	0.275	0.270	0.287	0.269
Public Schools Only	0.356	0.199	0.337	0.134	0.351	0.152	0.348	0.162

This table reports the effect of moving to a 1sd better classroom and the effect of moving to a 1sd better teacher on students' test scores in math, English, and Urdu. Test scores have been standardized to have a mean of 0 and a standard deviation of 1. In the first row, the sample includes all students and teachers in public and private schools. In the next row, the sample only includes classrooms where at least 75 percent of students had not been taught by the same teacher in the previous year. In the final row, the sample is restricted to students and teachers in public schools.

Table 4: Relationship Between Teacher Characteristics and Mean Teacher Value-Added for Public School Teachers

	(1) Covariate Mean	(2) Mean TVA	(3) Mean TVA	(4) Mean TVA	(5) Mean TVA	(6) Mean TVA	(7) Mean TVA	(8) Mean TVA	(9) Mean TVA	(10) Mean Test Score
Female	0.445	0.070***	-0.036	0.080***	0.207	0.082***	0.213	0.086***	-0.382***	
		(0.026)	(0.134)	(0.026)	(0.225)	(0.026)	(0.224)	(0.032)	(0.056)	
Local	0.273	0.025	0.008	0.024	-0.004	0.023	-0.005	0.017	-0.113	
		(0.025)	(0.031)	(0.028)	(0.049)	(0.028)	(0.048)	(0.036)	(0.079)	
Some Teacher Training	0.904	-0.023	-0.101	-0.093	-0.213*	-0.093	-0.216*	-0.125	-0.598*	
		(0.055)	(0.072)	(0.075)	(0.126)	(0.075)	(0.126)	(0.119)	(0.329)	
Has BA or Better	0.514	0.054**	0.043	0.012	0.010	0.012	0.009	-0.066	-0.175	
		(0.025)	(0.031)	(0.033)	(0.059)	(0.033)	(0.058)	(0.046)	(0.108)	
Had >3 *Years of Exp in* 2007	0.868	0.060	0.076	0.037	0.163*	0.038	0.166*	-0.046	0.180	
		(0.038)	(0.052)	(0.047)	(0.097)	(0.047)	(0.097)	(0.072)	(0.276)	
Temporary Contract	0.229	-0.003	0.049	-0.020	0.051	-0.019	0.055	0.019	0.065	
		(0.036)	(0.048)	(0.043)	(0.083)	(0.043)	(0.082)	(0.057)	(0.105)	
Mean English Test Score	2.802			0.032**	0.015					
				(0.015)	(0.022)					
Mean Urdu Test Score	3.059			0.034	0.013					
				(0.023)	(0.037)					
Mean Math Test Score	3.262			0.023	-0.013					
				(0.022)	(0.034)					
Mean Teacher Knowledge	3.041					0.090***	0.018	0.298***	0.332*	
						(0.024)	(0.038)	(0.072)	(0.176)	
Have 0 or 1 Years Exp.	0.040									-0.305**
										(0.135)
Lagged Mean Score	0.324									0.717***
										(0.013)
Fixed Effects		District	School	District	School	District	School	District	School	Teacher
Number of Observations		1,383	1,383	919	919	919	919	622	622	27,089
Adjusted R Squared		0.224	0.450	0.228	0.415	3.247	0.727	0.068	0.038	0.721
Clusters		471	471	469	469	469	469	440	440	583
F		2.031	1.194	2.533	0.602	0.230	0.417	73.428	2.741	

This table reports estimates of the association between TVA and teacher characteristics. The first column reports the means of the covariates of interest. For columns 2-9, observations are at the teacher level and characteristics are time invariant. Column 10 identifies within-teacher experience effects, controlling for teacher fixed effects, and regressing student outcomes on whether a teacher had 0 or 1 year of experience. Observations for this column are at the student-year level. The F-statistic is for a F-test of all the covariates in columns 2-7. In columns 8 and 9, it is the F-statistic from the first stage of the instrumental variables regression. Standard errors are clustered at the school level.

Table 5: Does Future Teacher Value-Added Predict Current Teacher Value-Added When Students Change Schools?

	(1) Coefficient (se)	(2) N
Forward Lag of English	0.051	3,231
	(0.049)	
Forward Lag of English (Within School)	-0.023	1,976
	(0.056)	
Forward Lag of Math	-0.076	3,231
	(0.061)	
Forward Lag of Math (Within School)	-0.017	1,976
	(0.036)	
Forward Lag of Urdu	-0.081	3,231
	(0.072)	
Forward Lag of Urdu (Within School)	0.017	1,976
	(0.040)	
Forward Lag of Mean Score	-0.033	3,231
	(0.067)	
Forward Lag of Mean Score (Within School)	0.002	1,976
	(0.046)	

This table tests for bias in the teacher value-added calculations. The current teacher value-added of students who change schools in the next period is regressed on the value-added of their future teacher. Observations are at the child level, and standard errors are clustered at the teacher level.

Table 6: Out-of-Sample Validation of TVAs

	(1) Math Test Score	(2) English Test Score	(3) Urdu Test Score	(4) Mean Test Score
$Math\,TVA$	0.781***			
	(0.065)			
$English\,TVA$		0.857***		
		(0.068)		
$Urdu\,TVA$			0.845***	
			(0.077)	
$Mean\,TVA$				0.852***
				(0.078)
Lagged Score Control	Y	Y	Y	Y
Number of Observations	3,822	3,822	3,822	3,822
Adjusted R Squared	0.557	0.542	0.590	0.636
Clusters	1,090	1,090	1,090	1,090

This table tests if TVAs predict the test score gains of school changers who are allocated to the new teachers. If the TVA estimates perfectly predict the "true" teacher value-added, these coefficients should be 1. Standard errors are clustered at the teacher level.

Table 7: Relationship Between Mean TVA and Log Salary for Public and Private School Teachers

	(1) Log Salary Public	(2) Log Salary Public	(3) Log Salary Public	(4) Log Salary Public	(5) Log Salary **PRIVATE**	(6) Log Salary **PRIVATE**
Mean TVA		-0.007	-0.028	-0.044	0.111**	-0.011
		(0.014)	(0.025)	(0.036)	(0.046)	(0.049)
Female	-0.036***	-0.035***	0.154**	0.054	-0.413***	-0.287***
	(0.013)	(0.013)	(0.070)	(0.094)	(0.043)	(0.047)
Local	-0.052***	-0.051***	-0.049	-0.019	-0.178***	-0.043
	(0.019)	(0.019)	(0.032)	(0.043)	(0.029)	(0.035)
Some Teacher Training	0.518***	0.518***	0.392***	0.837***	0.165***	0.127***
	(0.141)	(0.141)	(0.140)	(0.316)	(0.045)	(0.040)
Has BA or Better	0.255***	0.255***	0.263***	0.211***	0.334***	0.282***
	(0.019)	(0.019)	(0.028)	(0.042)	(0.045)	(0.042)
$Had > 3\,Years\,of\,Exp\,in\,2007$	0.063	0.064	0.120*	0.122	0.020	0.058*
	(0.042)	(0.042)	(0.064)	(0.101)	(0.029)	(0.031)
Temporary Contract	-0.354***	-0.355***	-0.327***	-0.308***		
	(0.032)	(0.032)	(0.059)	(0.092)		
Age	0.058***	0.058***	0.063***	0.039	0.016**	0.022***
	(0.015)	(0.015)	(0.020)	(0.029)	(0.007)	(0.008)
Age^2	-0.000***	-0.000***	-0.001**	-0.000	-0.000**	-0.0002*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.0001)
Mean English Score				0.016		
				(0.017)		
Mean Urdu Score				-0.006		
				(0.029)		
Mean Math Score				0.020		
				(0.025)		
Mean Salary	6,987	6,987	6,987	6,745	1,403	1,403
Fixed Effects	District	District	School	School	District	School
Adjusted R Squared	0.616	0.615	0.662	0.707	0.459	0.768
Number of observations	1,383	1,383	1,383	919	807	807
F	108.304	96.471	35.025	12.496	38.522	16.157
Clusters	471	471	471	469	294	294

This table reports estimates from regressions of log mean teacher salaries in public (columns 1-4) and private (column 5 and 6) schools on teacher characteristics, including mean TVA (columns 2-6) and average teacher test scores in English, Urdu, and math (column 4). All regressions include either district or school fixed effects, and standard errors are clustered at the school level.

Table 8: Effect of the Discontinuity on Teacher Characteristics

	(1) OLS	(2) SE	(3) RD (3 Year)	(4) SE	(5) RD (4 Year)	(6) SE
Log(Salary)	-0.284***	0.062	-0.554*	0.273	-0.444**	0.205
Bachelor's	0.318***	0.032	0.003	0.186	0.109	0.140
Some Training	0.003	0.031	0.013	0.120	0.010	0.096
Local	-0.017	0.037	-0.006	0.178	-0.066	0.134
Age Started	0.072***	0.024	1.193	1.550	0.943	1.116
Single	0.148***	0.032	-0.006	0.176	0.053	0.136
Female	-0.005	0.044	0.288	0.254	0.273	0.190
Mean Teacher English Score	0.326***	0.080	0.570**	0.231	0.319	0.248
Mean Teacher Urdu Score	0.076	0.067	0.429	0.342	0.217	0.336
Mean Teacher Math Score	-0.013	0.080	0.604	0.427	-0.502	0.375

This table presents OLS and fuzzy regression discontinuity results for the effect of temporary contracts on teacher characteristics. The RD includes either teachers hired 4 years before 1998 and 4 years after 2001 or teachers hired 3 years before 1998 and 3 years after 2001. Standard errors are clustered at the month hired level for the regression discontinuity results and the school level for the OLS results. Log(salaries), which were observed multiple times over several years, were normalized by calculating the teacher fixed effect, controlling for year fixed effects and demeaned at the district-level. Each cell is a coefficient estimate (or standard error estimate) for the temporary contract teacher effect.

Table 9: Effect of the Discontinuity on Student Characteristics and School Characteristics

	(1) Mean	(2) OLS	(3) SE	(4) RD (3 Year)	(5) SE	(6) RD (Within School)	(7) SE	(8) RD (4 Year)	(9) SE	(10) RD (Within School)	(11) SE
School Attributes											
Basic School Facilities	-0.473	0.066	0.059	0.433	0.270			0.274	0.248		
Extra School Facilities	-0.607	-0.480***	0.115	-1.025*	0.582			-0.922**	0.434		
Library	-0.039	-0.118***	0.031	-0.350*	0.178			-0.295**	0.125		
Computer	-0.136	-0.040**	0.017	-0.145*	0.087			-0.170**	0.067		
Sports	-0.076	-0.117***	0.026	-0.296	0.188			-0.200	0.153		
Hall	-0.048	-0.030	0.019	-0.075	0.079			-0.087	0.076		
Wall	-0.117	-0.029	0.029	0.117	0.101			0.134	0.088		
Fans	-0.129	-0.085**	0.034	-0.118	0.126			-0.131	0.101		
Electricity	-0.102	-0.099***	0.035	-0.218*	0.113			-0.211**	0.087		
Number Teachers	0.294	0.438	0.786	-0.295	2.615			-0.772	2.012		
Student Teacher Ratio	6.671	1.701	1.127	-1.399	6.273			2.784	5.492		
Parental Attributes											
Student Household Assets	-0.236	0.044	0.091	0.693	0.473	0.154	0.316	0.453	0.372	0.236	0.263
Student Mother Education	-0.065	-0.077***	0.023	0.020	0.129	-0.073	0.093	-0.001	0.116	-0.121	0.073
Student Father Education	-0.057	-0.042	0.026	-0.130	0.094	-0.154*	0.082	-0.134	0.081	-0.166**	0.071

This table presents OLS and fuzzy regression discontinuity results for the effect of temporary contracts on student and school characteristics. The 4 year RD includes teachers hired 4 years before 1998 and 4 years after 2001. The 3 year RD includes teachers hired 3 years before 1998 and 3 years after 2001. Standard errors are clustered at the month hired level for the regression discontinuity results and the school level for the OLS results. Characteristics observed multiple times over several years were normalized by calculating the teacher or school fixed effect (depending on the level at which the characteristic is observed), controlling for year fixed effects. Each cell is a coefficient estimate (or standard error estimate) for the temporary contract teacher effect.

Table 10: Do Student Test Score Trends Predict Being Taught by a Contract Teacher?

	(1) Mean Test Scores	(2) Mean Test Scores	(3) Had a Contract Teacher
$Year$	0.134***	0.145***	
	(0.013)	(0.013)	
$I(Received\,Contract\,Teacher)$	0.048	0.069	
	(0.078)	(0.083)	
$Year \times I(Received\,Contract\,Teacher)$	-0.015	-0.011	
	(0.023)	(0.024)	
$Mean\,Test\,Score\,Gain$			-0.014
			(0.016)
District FE	Y	Y	Y
School FE	N	Y	N
Grade by Lagged Test Score Interactions	Y	Y	N
Number of Observations	25,296	25,296	15,956
Adjusted R Squared	0.637	0.677	0.037
Clusters	478	478	497

This table tests whether better students are allocated to contract teachers. The first column compares trends in student test scores before the receipt of a contract teacher in schools that did and did not receive contract teachers. The next column compares the test score gains of students within schools who did or did not receive contract teachers before the receipt of the contract teacher. The final regression regresses an indicator for whether a student ever had a contract teacher on their mean test score gains (residualized by testing round and grade) in the years prior to receiving a contract teacher. In this sample, each student is observed once. Standard errors are clustered at the school level.

Table 11: The Effect of Teacher Contract Status on TVA

	(1) Mean TVA	(2) SE	(3) One-Sided T-test	(4) N	(5) Within School Mean TVA	(6) SE	(7) One-Sided T-test	(8) N
OLS (Full Sample)	-0.004	0.042	0.541	1,337	0.024	0.026	0.181	1,278
RD (Full Sample)	-0.004	0.052	0.533	1,337	0.056	0.041	0.088	1,278
RD (2 Year)	0.840	0.550	0.068	227	0.360	0.322	0.137	201
RD (3 Year)	0.219	0.241	0.184	376	0.254**	0.123	0.022	336
RD (4 Year)	0.350	0.234	0.070	393	0.193*	0.097	0.026	350
RD (5 Year)	-0.074	0.120	0.732	661	0.035	0.057	0.268	604
RD (6 Year)	-0.026	0.106	0.598	690	0.040	0.053	0.225	631
RD (7 Year)	-0.036	0.106	0.632	692	0.035	0.052	0.250	632

This table regresses mean TVAs on whether a teacher has a temporary contract in all public schools using the ordinary least squares and 6 fuzzy regression discontinuity specifications. In the RD specifications, contract status is instrumented for with an indicator variable for whether a teacher was hired after 1998. All regressions contain linear time trends which are allowed to differ before and after the budgetary shock and district fixed effects. The table presents instrumental variables (RD) specifications for the full sample and for bandwidths of 2-7 years before and after the budgetary shock. Observations are at the teacher level, and standard errors are clustered at the school level in the OLS specification and the month hired level in the regression discontinuity specifications. Columns 3 and 7 report p-values for a one-sided t-test of whether the temporary contract effect is negative.

Figures

Figure 1: Teacher Salaries in Public and Private Schools

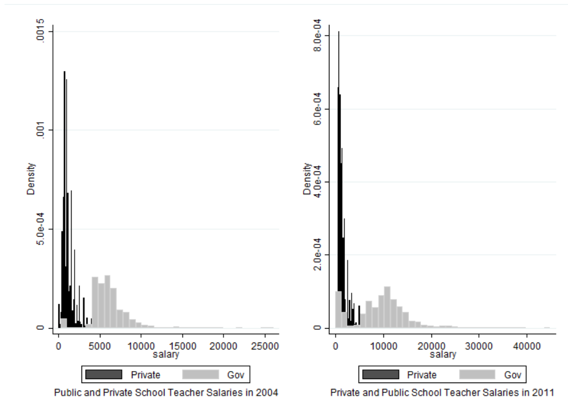
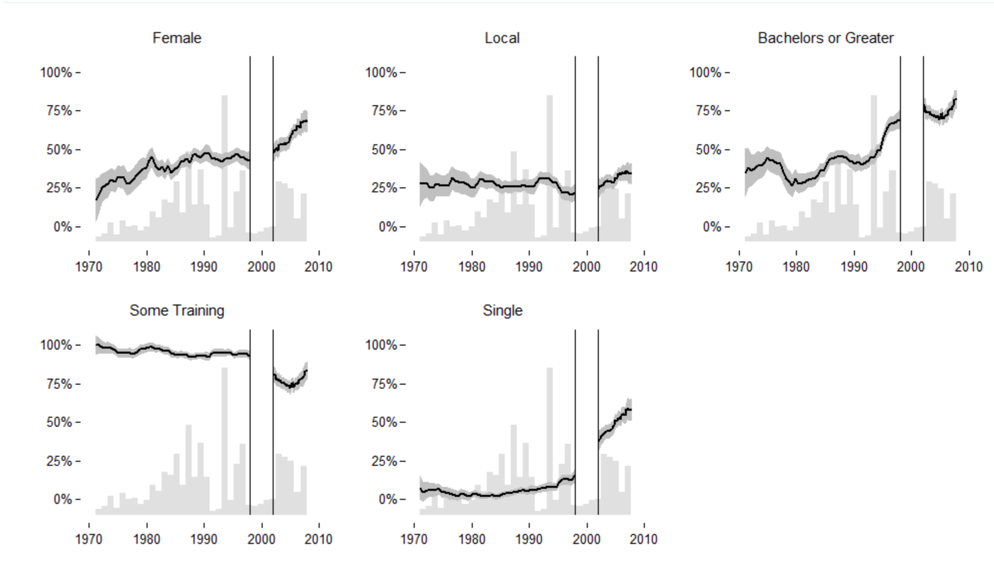


Figure 2: Teacher Contract Status, Salary, and Productivity by Year Hired

Figure 3: Trends in Teacher Characteristics by Month Hired



Appendix

Appendix A describes how our test score data was collected, and Appendix B details the derivation of the bias term in the estimate of the variance of the teacher effects. Appendix C documents how data entry errors in teacher ids could lead to greater bias in TVA estimates that control for child fixed effects. Appendix D provides a theoretical framework to understand the effects of public sector wage reductions on the quality of teachers entering the public sector.

Appendix A: Test Data

In each round of the LEAPS data collection, we tested students in math, Urdu (the vernacular), and English. To avoid the possibility of cheating, project staff, with clear instructions not to interfere, administered the test directly to students. Test booklets were retrieved after class, so there was no missing testing material. Tests were scored and equated across the four rounds using Item Response Theory, yielding scores in each subject with a mean of 0 and a standard deviation of 1 (Das and Zajonc, 2010). Item response theory weights questions differently according to their difficulty and allows us to equate tests over years so that a standard deviation gain in year 1 is equivalent to a standard deviation gain in year 4. The tests could be equated because we included linking questions across any two years and for some questions, across multiple years.

Appendix Table 2 provides more information on the sources of variation for the TVA calculations. In year one, since only 3rd graders were tested, very few students were observed in schools where more than one classroom was tested. In future years, some students were held back, others were promoted, and another sample of 3rd graders was added in year 3, allowing students in a larger number of classrooms to be tested. Columns 1 and 2 describe the sample used to calculate the cross-school TVA estimates. Columns 3 and 4 describe the variation used to calculate the within school TVA measures.

Appendix B: Derivation of Bias

In this appendix, we derive the relationship between $Cov(\widehat{\lambda_{jt}}, \widehat{\lambda_{j,t+1}})$, where $\widehat{\lambda_{jt}}$ is our estimate of the de-meaned class effect associated with teacher j in year t, and our object of interest, $Var(\gamma_j)$, which is the variance of the teacher effects. To derive the bias, we assume that ε_{ijt} is i.i.d and homoskedastic with variance σ^2. Following from equation 1, and given the fact we do not include the estimates for teacher j in our estimate of the school's mean,

$$\widehat{\lambda_{jt}} = \lambda_{jt} + \Big(\frac{\sum_{i=1}^{N_{jt}} \varepsilon_{ijt}}{N_{jt}} - \frac{\sum_{kt}\sum_{i=1}^{N_{kt}} \varepsilon_{ikt}\mathbf{1}_{k\neq j}}{\sum_{kt} N_{kt}\mathbf{1}_{k\neq j}}\Big),$$

where i indexes a student, and N_{jt} is the number of students in the class of teacher j in year t. Now, assume that $\lambda_{jt} = \gamma_j + \theta_{jt}$, where γ_j is the time-invariance teacher effect and θ_{jt} is the idiosyncratic class-level effect, which is not correlated over time. Then,

$$Cov(\widehat{\lambda_{jt}}, \widehat{\lambda_{j,t+1}}) = Var(\gamma_j) + Cov\Big(\frac{\sum_{kt}\sum_{i=1}^{N_{kt}} \varepsilon_{ikt}\mathbf{1}_{k\neq j}}{\sum_{kt} N_{kt}\mathbf{1}_{k\neq j}}, \frac{\sum_{kt}\sum_{i=1}^{N_{kt}} \varepsilon_{ikt}\mathbf{1}_{k\neq j}}{\sum_{kt} N_{kt}\mathbf{1}_{k\neq j}}\Big).$$

The first term on the right-hand side is our term of interest. The second term is the bias, which we must estimate, and we define as Φ. To arrive at our estimable expression for the bias, we see that

$$E(\Phi) = E\Big(\frac{\sigma^2}{\sum_{kt} N_{kt}\mathbf{1}_{k\neq j}}\Big).$$

Appendix C: Incorrect Variation in Teacher Switching Due to Data Entry Errors

In this appendix, we show how a small amount of data misentry can lead to a large amount of bias when we include child fixed effects in the TVA estimation. Suppose that 1% of teacher IDs are randomly entered incorrectly. If 10% of students change teachers each year, when identifying variation comes only from the test scores of students who change teachers, these incorrect entries account for 9% of the variation. To arrive at this number, note that there are three cases where a student-year observation will provide identifying variation in a specification that includes child fixed effects: (1) the teacher ID was incorrectly entered, but no switch actually occurred (probability = $0.01 \times 0.9 = 0.009$), (2) the teacher ID was correctly entered and a switch occurred (probability = $0.99 \times 0.1 = .099$), and (3) the ID was incorrectly entered and a switch occurred (probability = $0.1 \times 0.01 = 0.001$). Then the probability that the teacher ID is mis-attributed in an observation that provides identifying variation is $\frac{0.01}{(0.009+0.099+0.001)} = 0.09$.

In order to assess potential bias more formally, consider a case where students are identical and TVA is randomly distributed, so there is no correlation between a student's future teacher's TVA and his current teacher's TVA as long as she changes teachers. Now, also assume that a student has a probability p of changing teachers each year, and an ID has a probability e of being incorrectly entered. Then, when the TVA of teacher is calculated for teacher j, it will be a weighted mean of the teacher's true TVA and the TVAs of teachers of any students with mis-attributed IDs. Therefore,

$$E(\widehat{TVA}_j) = \frac{p}{e(1-p)+p(1-e)+ep} TVA_j + \frac{e}{e(1-p)+p(1-e)+ep} \overline{TVA_j},$$

where $\overline{TVA_j}$ is the mean TVA in the teacher population and $\widehat{TVA}_j$ is the estimate of the TVA for teacher j. This expression formalizes the intuition that the bias decreases in the true probability of switching p and increases in the error rate e.

Appendix D: Theoretical Framework

Suppose there are N teachers and $M < N$ positions in the public sector. Each teacher has a productivity θ_j drawn from a bounded distribution F, with a maximum of θ_{max} and a minimum of θ_{min}. In the public sector, teachers receive a wage w_{pub} set exogenously by the government, and the public sector hires randomly from its applicants. As Section 5.1 shows, public wages are in fact unrelated to productivity, consistent with this assumption. In the private sector, due to free entry, private schools make zero profits and a teacher receives her productivity, θ_j. Note that this implies a link between productivity and wages in the private sector, consistent with our findings in Table 7. For simplicity, teachers can costlessly apply to a public sector job or go directly to the private sector. If she applies, a teacher will get a public sector job with probability $p = \frac{T(w_{pub})}{M}$, where T is the endogeneous number of teachers applying to public positions. If she does not get the public sector job, she enters the private sector and receives the private sector wage. Since applying to the public sector is costless, a teacher will always apply for a public sector job if $\theta_j < w_{pub}$.

Then, for a given w_{pub}, there are three possible outcomes:

1. $w_{pub} < \theta_{min}$: In this case, even the least productive teacher makes more in the private sector than they would in the public sector, so no teachers enter the public sector, and there is a shortage in the public sector.

2. $w_{pub} > \theta_{max}$: In this case, even the most productive teachers would make more in the public sector, so all teachers apply to the public sector, and the average productivity in the public sector is $\int_{\theta_{min}}^{\theta_{max}} \theta_j f(\theta_j) \partial \theta_j$. Therefore, there is no shortage since there are $N > M$ applicants.

3. $\theta_{min} < w_{pub} < \theta_{max}$: Then, there exists $\theta^* = w_{pub}$ such that all teachers with productivity greater than θ^* do not apply to the public sector and all teachers with productivity less than θ^* do. Thus, the average productivity of the public sector is $\int_{\theta_{min}}^{w_{pub}} \theta_j f(\theta_j) \partial \theta_j < \int_{\theta_{min}}^{\theta_{max}} \theta_j f(\theta_j) \partial \theta_j$. In this case, it is ambiguous whether there is a shortage of teachers since $T = N \times F(\theta^*)$ may be less than or greater than M.

The first corner case isn't relevant for our empirical context, since we observe that there are teachers entering the public sector before and after the wage change. Therefore, when we study the effect of a decline in w_{pub} to $w'_{pub} < w_{pub}$, we focus on the second two possible equilibria under w_{pub} and w'_{pub}, in which at least some teachers always enter the public sector. First, consider the case where $w_{pub} > \theta_{max}$. Then, there are two possibilities once wages decline to w'_{pub}:

1. $w'_{pub} > \theta_{max}$, and the average productivity in the public sector is still $\int_{\theta_{min}}^{\theta_{max}} \theta_j f(\theta_j) \partial \theta_j$. In this case, there is no shortage ($T > M$), since all teachers apply for public positions.

2. $w'_{pub} < \theta_{max}$, and $\exists\ \theta^{**} = w_{pub}$ such that all teachers with productivity greater than θ^{**} do not apply to the private sector and all teachers with productivity less than θ^{**} do. Then, the new average productivity of the public sector will be $\int_{\theta_{min}}^{w'_{pub}} \theta_j f(\theta_j)\partial\theta_j < \int_{\theta_{min}}^{\theta_{max}} \theta_j f(\theta_j)\partial\theta_j$. Therefore, under w'_{pub}, average productivity in the public sector declines. As before, it is ambiguous whether there is a shortage.

Now consider the case where $\theta_{min} < w_{pub} < \theta_{max}$. Then, under $\theta_{min} < w'_{pub} < w_{pub} < \theta_{max}$, there will be a new $\theta^{**} < \theta^{*}$, where all teachers with $\theta_j > \theta^{**}$ do not apply to the public sector. In this case, average productivity in the public sector declines from $\int_{\theta_{min}}^{w_{pub}} \theta_j f(\theta_j)\partial\theta_j$ to $\int_{\theta_{min}}^{w'_{pub}} \theta_j f(\theta_j)\partial\theta_j$. Again, it is ambiguous whether there is also a shortage.

Thus, when we study the effect of the wage decline in our data, there are two empirically relevant possibilities. The first is that the average productivity of entering public school teachers remains the same after the wage decline, suggesting that both w_{pub} and w'_{pub} are greater than θ_{max}, and there is no shortage under either wage. The second possibility is that average productivity of public school teachers declines, suggesting that $w'_{pub} < \theta_{max}$, while w_{pub} may be less than or greater than θ_{max}, and a shortage may occur. Figures A7 and A8 graph these two cases. Figure A7 shows the case where even after the salary reduction public school salaries are greater than private school salaries for all teachers, so all teachers continue to apply to public sector positions. Figure A8 shows the case where the salary reduction leads private salaries to be greater than public salaries for a subset of teachers. In this case, the most productive teachers no longer apply to the public sector.

Appendix Tables

Table A1: Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
	Government			Private		
	Mean	SD	N	Mean	SD	N
Female	0.449	0.497	3,829	0.768	0.422	4,733
Local	0.273	0.445	3,827	0.538	0.499	4,731
Some Training	0.904	0.294	3,829	0.221	0.415	4,731
BA Plus	0.514	0.500	3,829	0.255	0.436	4,734
Mean Salary	7671 ($129)	3746 ($63)	3,829	1407 ($24)	997 ($17)	4,731
Temporary Contract	0.229	0.420	3,824	0.838	0.368	4,646
Year Started	1,990.80	10.710	3,432	2,002.17	7.749	3,159
Mean Days Absent	2.644	3.297	3,825	1.936	3.368	4,728
Mean Teacher Test Score	3.041	0.569	1,175	2.861	0.606	1,046
Mean School Basic Facilities	-0.519	0.831	3,686	0.606	1.353	4,697
Mean School Extra Facilities	-0.607	1.401	3,686	0.716	1.033	4,697
Mean Student Household Assets	-0.236	0.242	1,699	0.484	1.022	1,311
Mean Student Mother Primary Education	0.298	0.212	1,699	0.378	0.276	1,311
Mean Student Father Primary Education	0.580	0.245	1,699	0.739	0.242	1,311
Mean Change in Math Scores	0.393	0.499	1,533	0.355	0.488	975
Year 2 - Year 1	0.206	0.647	557	0.226	0.546	322
Year 3- Year 2	0.438	0.463	662	0.511	0.403	316
Years 4 - Year 3	0.475	0.561	1,041	0.354	0.490	573
Mean Change in English Scores	0.393	0.474	1,533	0.388	0.461	975
Year 2 - Year 1	0.303	0.652	557	0.187	0.459	322
Year 3- Year 2	0.375	0.454	662	0.408	0.402	316
Years 4 - Year 3	0.462	0.530	1,041	0.389	0.490	573
Mean Change in Urdu Scores	0.444	0.453	1,533	0.423	0.434	975
Year 2 - Year 1	0.306	0.633	557	0.317	0.459	322
Year 3- Year 2	0.444	0.424	662	0.497	0.368	316
Years 4 - Year 3	0.533	0.502	1,041	0.445	0.451	573
Mean Change in Mean Scores	0.410	0.413	1,533	0.372	0.399	975
Year 2 - Year 1	0.272	0.575	557	0.243	0.411	322
Year 3- Year 2	0.419	0.372	662	0.472	0.327	316
Years 4 - Year 3	0.490	0.461	1,041	0.396	0.409	573

This table presents teacher-level summary statistics across 4 rounds of the LEAPS survey (2004-2007). Changes in test scores are calculated by averaging over the difference between a student's test scores in time t and time $t-1$. Household assets and school basic and extra facilities are predicted from a principal components analysis of indicator variables for the presence of different assets, and school facilities and are normalized by year observed.

Table A2: Correlations Between Teacher Test Scores and Teacher Characteristics

	(1) Math	(2) Math	(3) English	(4) English	(5) Urdu	(6) Urdu	(7) Mean	(8) Mean
Female	-0.252***	-0.188	-0.110*	0.232	-0.116***	-0.297	-0.159***	-0.085
	(0.042)	(0.236)	(0.057)	(0.467)	(0.039)	(0.338)	(0.037)	(0.270)
Local	0.021	-0.047	0.006	-0.089	-0.006	-0.026	0.007	-0.054
	(0.043)	(0.090)	(0.063)	(0.115)	(0.043)	(0.085)	(0.038)	(0.070)
Some Teacher Training	0.311	0.222	0.281	0.226	0.107	-0.037	0.233	0.137
	(0.211)	(0.300)	(0.214)	(0.288)	(0.134)	(0.180)	(0.150)	(0.197)
Has BA or Better	0.246***	0.225***	0.312***	0.268***	0.201***	0.154**	0.253***	0.216***
	(0.051)	(0.085)	(0.061)	(0.094)	(0.042)	(0.072)	(0.039)	(0.061)
Had $>$ 3 *Years of Exp in* 2007	-0.040	0.023	0.092	-0.002	0.071	0.206	0.041	0.075
	(0.099)	(0.184)	(0.091)	(0.190)	(0.069)	(0.130)	(0.072)	(0.134)
Temporary Contract	-0.111*	0.110	0.212***	0.369***	0.013	0.185	0.038	0.221**
	(0.064)	(0.135)	(0.068)	(0.140)	(0.056)	(0.116)	(0.049)	(0.102)
Fixed Effects	District	School	District	School	District	School	District	School
Number of Observations	1,105	1,105	1,105	1,105	1,105	1,105	1,105	1,105
Adjusted R Squared	0.070	0.042	0.062	0.167	0.049	0.114	0.085	0.200
Clusters	19.125	2.637	15.886	5.406	8.121	2.508	20.595	5.979
F	491	491	491	491	491	491	491	491

This table reports estimates of the association between teacher test scores and teacher characteristics. Observations are at the teacher level and characteristics are time invariant. In cases where a teacher was tested more than once, the outcome variables are the average across multiple test scores. Standard errors are clustered at the school level.

Table A3: Non-Linearities in Within-Teacher Experience Effects

	(1) Math	(2) Urdu	(3) English	(4) Mean
Has 0 or 1 Years of Exp.	-0.722***	-0.677***	-0.548***	-0.589***
	(0.269)	(0.208)	(0.194)	(0.197)
Has 2 Years of Exp.	-0.425**	-0.470***	-0.256*	-0.346**
	(0.188)	(0.154)	(0.149)	(0.136)
Has 3 Years of Exp.	-0.112	-0.124	-0.033	-0.080
	(0.170)	(0.134)	(0.146)	(0.124)
Has 4 Years of Exp.	-0.076	-0.065	-0.024	-0.070
	(0.146)	(0.116)	(0.105)	(0.093)
Teacher FE	Y	Y	Y	Y
Lagged Test Scores	Y	Y	Y	Y
Number of Observations	26,508	26,508	26,508	26,508
Adjusted R Squared	0.624	0.644	0.647	0.721
Clusters	569	569	569	569

This table tests for non-linearities in the effect of teacher experience on student test scores. All regressions control for teacher fixed effects and lagged student test scores. The sample is restricted to public school teachers. Standard errors are clustered at the school level.

Table A4: Relationship Between Teacher Characteristics and Mean Teacher Value-Added for Private School Teachers

	(1) Covariate Mean	(2) Mean TVA	(3) Mean TVA	(4) Mean TVA	(5) Mean TVA	(6) Mean TVA	(7) Mean TVA	(8) Mean TVA	(9) Mean TVA	(10) Mean Test Score
Female	0.768	-0.006	-0.002	-0.005	-0.031	-0.008	-0.037	0.051	-0.496**	
		(0.029)	(0.043)	(0.031)	(0.061)	(0.032)	(0.061)	(0.052)	(0.234)	
Local	0.538	-0.038*	-0.047	-0.040	-0.056	-0.037	-0.050	-0.047	-0.234**	
		(0.023)	(0.030)	(0.025)	(0.051)	(0.025)	(0.048)	(0.037)	(0.103)	
Some Teacher Training	0.221	0.002	0.002	-0.022	-0.044	-0.017	-0.045	-0.018	-0.009	
		(0.026)	(0.042)	(0.029)	(0.067)	(0.029)	(0.065)	(0.045)	(0.235)	
Has BA or Better	0.255	0.055*	0.002	0.060*	0.068	0.062**	0.064	0.031	0.291**	
		(0.030)	(0.044)	(0.032)	(0.064)	(0.031)	(0.062)	(0.054)	(0.128)	
$Had > 3 Years of Exp in 2007$	0.467	0.040	0.034	0.031	0.040	0.034	0.039	0.053	-0.027	
		(0.024)	(0.031)	(0.027)	(0.046)	(0.027)	(0.045)	(0.056)	(0.186)	
Temporary Contract	0.838	0.019	-0.002	0.042	0.060	0.041	0.056	0.029	0.127	
		(0.032)	(0.045)	(0.033)	(0.067)	(0.033)	(0.066)	(0.050)	(0.181)	
Mean English Test Score	2.754			0.069**	0.005					
				(0.027)	(0.048)					
Mean Urdu Test Score	2.921			-0.009	-0.029					
				(0.022)	(0.041)					
Mean Math Test Score	2.907			0.007	0.018					
				(0.022)	(0.044)					
Mean Teacher Knowledge	2.861					0.066***	-0.001	0.232***	-0.084	
						(0.025)	(0.045)	(0.069)	(0.160)	
Have 0 or 1 Years Exp.	0.194									-0.301***
										(0.075)
Lagged Mean Score	0.275									0.682***
										(0.028)
Fixed Effects		District	School	District	School	District	School	District	School	Teacher
Number of Observations		808	808	561	561	561	561	198	198	27,089
Adjusted R Squared		0.104	0.295	0.128	0.237	0.123	0.239	0.099	-0.100	0.720
Clusters		294	294	289	289	289	289	174	174	347
F		1.707	0.687	2.583	0.579	2.960	0.697	37.653	0.419	317.594

This table reports estimates of the association between TVA and teacher characteristics for private school teachers. The first column reports the means of the covariates of interest. For columns 2-9, observations are at the teacher level and characteristics are time invariant. Column 10 identifies within-teacher experience effects, controlling for teacher fixed effects, and regressing student outcomes on whether a teacher had 0 or 1 year of experience. Observations for this column are at the student-year level. The F-statistic is for a F-test of all the covariates in columns 2-7. In columns 8 and 9, it is the F-statistic from the first stage of the instrumental variables regression. Standard errors are clustered at the school level.

Table A5: Correlation Between TVA Specifications

	(1) Across Schools	(3) Within Schools
English	0.977	0.955
Math	0.969	0.951
Urdu	0.963	0.944

English, math, and Urdu TVAs are calculated with and without individual level controls for gender, age, household assets, basic and extra facilities indices, and student-teacher ratios. Each cell of the table gives the correlation between the TVA estimated with the parsimonious specification and the TVA estimated with the detailed specification.

Table A6: Bootstrapped Regression Discontinuity Results

(1) Bandwidth (Years)	(2) Mean TVA	(3) P-value	(4) Within School, Mean TVA	(5) P-value
Full Sample	-0.004	0.213	0.056*	0.062
2	0.840**	0.030	0.360	0.186
3	0.219	0.248	0.254*	0.058
4	0.350*	0.076	0.193**	0.048
5	-0.074	0.809	0.035	0.347
6	-0.026	0.625	0.040	0.325
7	-0.036	0.668	0.035	0.347

This table replicates the fuzzy regression discontinuity estimates of the effect of temporary contract status on TVA using a bootstrap procedure to estimate the p-values to account for estimation error. The bootstrap was clustered at the month hired level. The estimates are based on 10,000 random samples of the data. The coefficients reported are the same coefficients as in Table 11.

Table A7: Regression Discontinuity Results Including Student-Teacher Ratio Controls

(1) Bandwidth (Years)	(2) Mean TVA	(3) Se	(4) Within School, Mean TVA	(5) Se
Full Sample	-0.016	0.056	0.054	0.041
2	0.716	0.526	0.354	0.311
3	0.267	0.276	0.279**	0.138
4	0.225	0.211	0.208*	0.106
5	-0.059	0.108	0.041	0.058
6	-0.009	0.098	0.045	0.054
7	-0.011	0.096	0.040	0.054

This table replicates the fuzzy regression discontinuity estimates of the effect of temporary contract status including controls for school-year student-teacher ratios. Standard errors are clustered at the month hired level.

Table A8: Regression Discontinuity Results Controlling for Classroom Mean Characteristics

(1) Bandwidth (Years)	(2) Mean TVA	(3) Standard Error	(4) Within School, Mean TVA	(5) Standard Error
Full Sample	0.013	0.0614	0.079**	0.039
2	1.215*	0.598	0.786*	0.423
3	0.272	0.261	0.263*	0.138
4	0.348	0.237	0.188*	0.108
5	-0.112	0.145	0.020	0.058
6	-0.051	0.125	0.035	0.056
7	-0.054	0.123	0.033	0.054

This table replicates the fuzzy regression discontinuity estimates of the effect of temporary contract status on TVA. TVA estimates include controls for the mean household asset index of a classroom and mean lagged student test scores in the classroom. Standard errors are clustered at the month hired level.

Table A9: Tests for Selective Attrition of Students and Teachers

	(1)	(2)	(3)	(4)
	Student Attrition		Teacher Attrition	
	Present In Year 4	Years Observed	Present In Year 4	Years Observed
Percent of Rounds Observed with a Contract Teacher	-0.015	-0.038		
	(0.019)	(0.028)		
Mean TVA			0.026	0.015
			(0.070)	(0.142)
Year Student Entered Panel FE	Y	Y	N	N
District FE	Y	Y	Y	Y
Year Teacher Started FE	N	N	Y	Y
Number of Observations	22,596	22,596	298	298
Adjusted R Squared	0.157	0.503	0.057	0.203
Clusters	512	512	200	200

This table examines whether the students of contract teachers selectively leave the sample and whether lower quality contract teachers selectively leave the sample. The outcome variables are an indicator variable for whether a student was in the sample in round 4, the number of rounds a student was observed, an indicator whether a teacher was in the sample in round 4, and the number of rounds a teacher was observed. In the first two columns, the sample is all tested public school students. In the second two, it is all contract teachers. Standard errors are clustered at the school level.

Appendix Figures

Figure A1: Number of Teachers Hired Per Year in the Sample

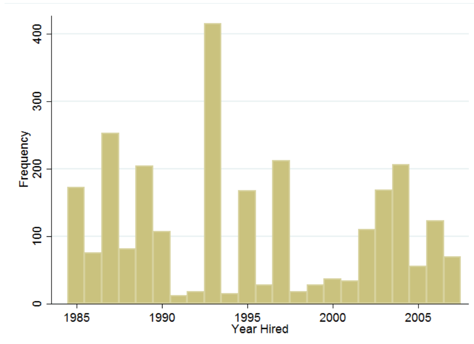


Figure A2: Teacher Test Scores in Public Schools

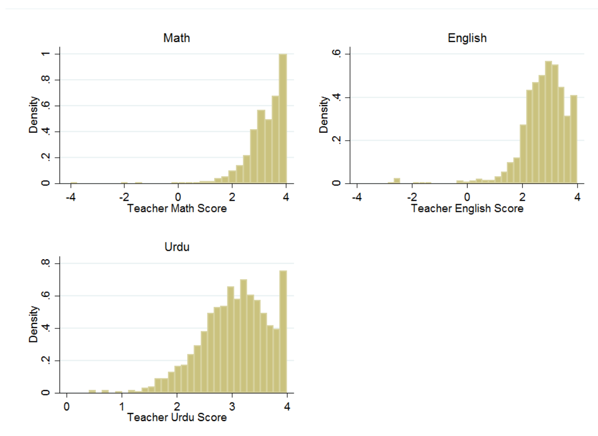


Figure A3: Teacher Test Scores in Private Schools

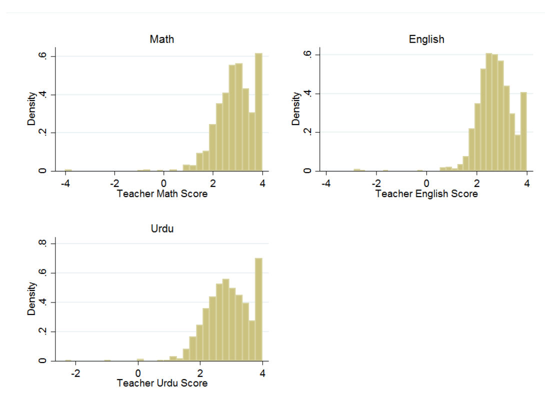


Figure A4: Number of Rounds Public School Teachers and Students are Observed

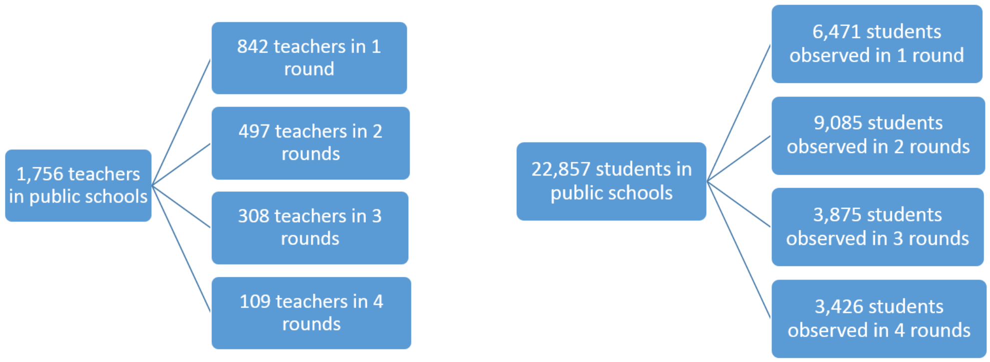


Figure A5: Number of Rounds Private School Teachers and Students are Observed

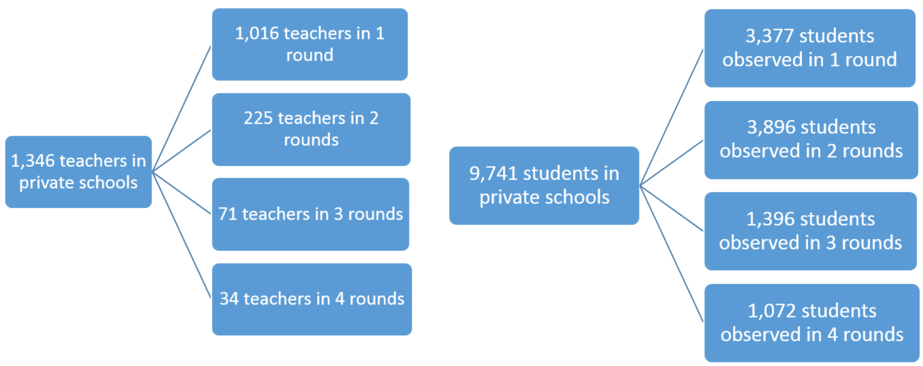


Figure A6: Third Grade Sizes in Public Schools in Punjab

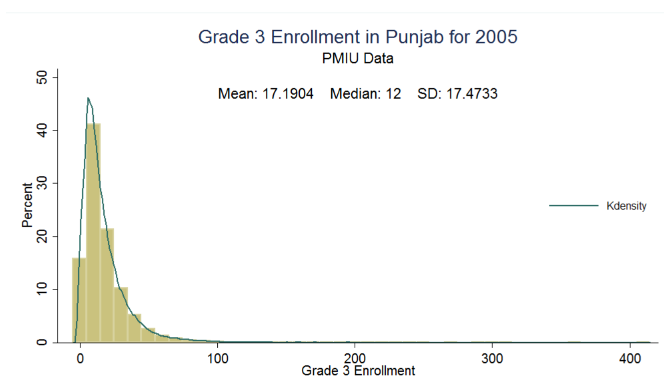


Figure A7: Case I: $w_{pub}, w'_{pub} \geq \theta_{max}$

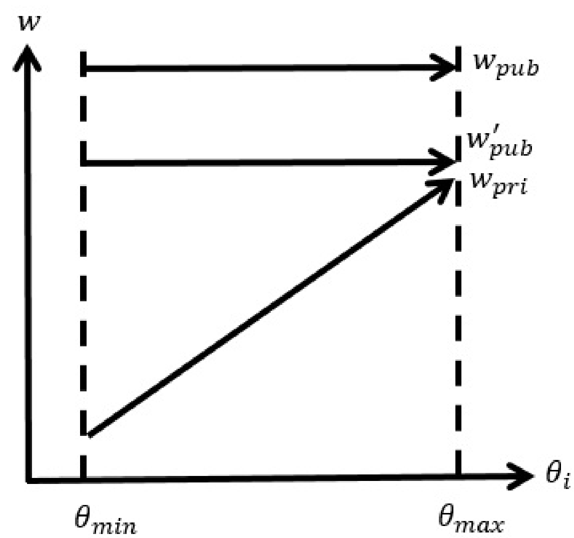


Figure A8: Case II: $w'_{pub} < \theta_{max}$

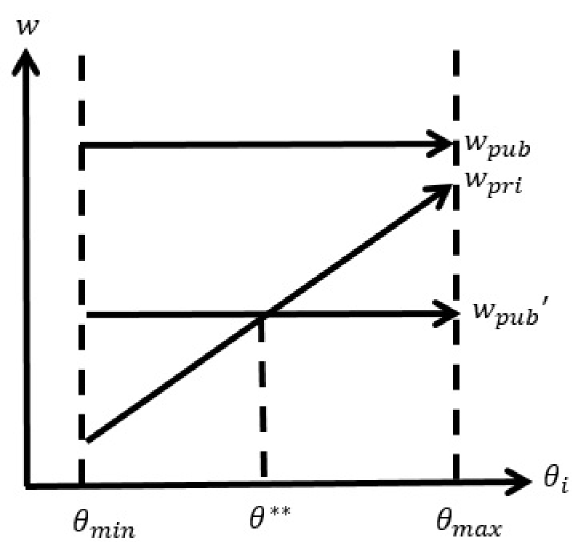


Figure A9: School and Parent Characteristics by Teacher Month Hired