SCHOOL GOVERNANCE, TEACHER INCENTIVES, AND PUPIL-TEACHER RATIOS:
EXPERIMENTAL EVIDENCE FROM KENYAN PRIMARY SCHOOLS

Esther Duflo
Pascaline Dupas
Michael Kremer

School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence
from Kenyan Primary Schools
Esther Duflo, Pascaline Dupas, and Michael Kremer
NBER Working Paper No. 17939
March 2012, Revised April 2014
JEL No. D71,I21,M51,O15

ABSTRACT

Some education policymakers focus on bringing down pupil-teacher ratios. Others argue that resources
will have limited impact without systematic reforms to education governance, teacher incentives, and
pedagogy. We examine a program under which Kenyan Parent-Teacher Associations (PTAs) at randomly
selected schools were funded to hire an additional teacher on an annual contract renewable conditional
on performance, outside normal Ministry of Education civil-service channels, at one-quarter normal
compensation levels. For students randomly assigned to stay with existing classes, test scores did
not increase significantly, despite a reduction in class size from 82 to 44 on average. In contrast, scores
increased for students assigned to be taught by locally-hired contract teachers. One reason may be
that contract teachers had low absence rates, while centrally-hired civil-service teachers in schools
randomly assigned PTA contract teachers endogenously reduced their effort. Civil-service teachers
also captured rents for their families, with approximately 1/3 of contract teacher positions going to
relatives of existing teachers. A governance program that empowered parents within PTAs reduced
both forms of capture. The best contract teachers obtained civil service jobs over time, and we estimate
large potential dynamic benefits from supplementing a civil service system with locally-hired contract
teachers brought in on a probationary basis and granted tenure conditional on performance.

Esther Duflo
Department of Economics, E17-201A
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
eduflo@mit.edu

Pascaline Dupas
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305-6072
and NBER
pdupas@stanford.edu

Michael Kremer
Harvard University
Department of Economics
Littauer Center M20
Cambridge, MA 02138
and NBER
mkremer@fas.harvard.edu

1. Introduction

Social service delivery in developing countries often delivers very little. Patients frequently do not receive correct diagnosis and treatment (Amin, Das and Goldstein, 2007), and students score very low on internationally comparable tests (Hanushek and Woessman, 2010). Some policymakers seeking to address this problem focus on hiring more providers, while others focus on systemic reforms designed to improve governance and provider incentives.

In education, while several studies in developed country contexts have found that reducing class size can increase test scores,[1] the randomized trials that have been conducted in low-income countries consistently have failed to find an effect of pupil-teacher ratio on test scores, across a range of geographic contexts, starting with very different initial levels of class size.[2] One hypothesis is that weak governance institutions contribute to poor incentives (leading to high absenteeism rate, see Chaudhury et al., 2006) and reduce the impact of additional education expenditure.

Several different approaches have been advocated to improve governance and teacher incentives. Many countries have experimented with supplementing the civil-service teacher workforce with additional teachers hired and managed locally, on short-term contracts, rather than centrally, under civil-service rules.[3] Another, related approach is School-Based Management reforms to empower parents and other local decision makers.[4] While some applaud these programs, others are concerned that the abandonment of civil-service hiring rules may create the potential for capture by local elites and favoritism in hiring.

We examine the interaction of extra resources, incentives, and school governance reform, through a randomized evaluation of the Extra Teacher Program (ETP) in Kenya. Each Kenyan school has a school committee, and prior to Kenya's abolition of school fees, many of these committees hired extra teachers on short-term and low-pay contracts to supplement the civil-service teachers assigned to the school by central authorities at much higher salaries. After Kenya abolished school fees, school committees no longer had funds to hire these supplemental teachers. The ETP program provided 70

---

[1] For positive effects at the primary school level see Krueger and Whitmore (2002) in the US, Angrist and Lavy (1999) in Israel, Piketty (2004) in France. Hoxby (2000) finds no significant effect in the US. Bandiera et al. (2010) find evidence of non-linear class-size effects at the university level.

[2] See Banerjee, Jacob, and Kremer (2005), Banerjee et al. (2007), and, for a well-identified non-experimental study, Urquiola and Verhoogen (2009).

[3] Contract teachers have been used in Benin, Burkina Faso, Cambodia, Cameroon, Chad, Congo, Côte d'Ivoire, Guinea, India, Indonesia, Madagascar, Mali, Nicaragua, Niger, Senegal, and Togo. See Duthilleul (2005) for a review, and Kumar et al. (2005).

[4] School-Based Management programs have been implemented in Argentina, Australia, Bangladesh, Canada, Guatemala, Honduras, Hong Kong, India, Lebanon, Lesotho, Macedonia, Madagascar, Mexico, Nicaragua, the Philippines, Senegal, Serbia, Sri Lanka, The Gambia, the United Kingdom, and the United States. See Bruns et al. (2011) for a review.

randomly selected school committees with funds to hire supplemental teachers locally over a two year period. These teachers were fully qualified – typically recent graduates of teacher training colleges whose cohorts had not been able to obtain jobs as civil-service teachers due to freezes on civil-service hiring. They were placed on short-term contracts under the authority of school committees composed primarily of parents. These contracts were renewable after one year, upon approval by the school committee. Contract teachers followed the same curriculum as civil-service teachers, but rather than focusing on particular subjects, and rotating across grades, as many civil-service teachers do, the newly-hired teachers were assigned to work with a fixed group of first-grade students.

Several features of the experimental design allow us to shed light on the interaction between extra resources, incentives, and governance. First, within schools sampled for the program, students were randomly assigned to either the newly created class taught by the teacher hired and governed under the ETP program, or to a pre-existing class taught by civil-service teachers. Despite a dramatic reduction in the pupil teacher ratio for grade 1 from 82 to 44, the program generated little improvement in test scores for students who remained within the standard system. Our results suggest that the effort response of the teachers is at least part the explanation– civil-service teachers were 16 percentage points less likely to be found in class teaching if their school received funding to hire a contract teacher. The endogenous behavioral responses that we observe under weak governance institutions may help explain why studies in developing countries typically find little learning improvement from hiring additional teachers.[5]

In contrast, the program increased test scores by 29% of a standard deviation for students assigned to the contract teachers. These locally hired and monitored teachers appear to have much better incentives: they were 12 percentage points more likely to be found in class teaching during unannounced spot checks than civil-service teachers in comparison schools, and 28 percentage points more likely to be present than civil service teachers in the schools they were assigned to.[6]

In addition to reducing effort, civil-service teachers captured some of the benefits of the program for their extended family – about a third of the contract teachers hired through the program were relatives of existing civil-service teachers in the school, and students of these teachers appear to have learned less than students of other contract teachers.

---

[5] In contemporaneous work, Muralidharan and Sundaramanan (2013) also find that teachers reduce effort in response to hiring additional teachers. Das et al. (2004) and Das et al. (2013) find that parents respond endogenously to provision of non-teacher inputs by cutting back their own investments in these inputs.

[6] The incentive effect is potentially confounded both by the lower experience of the contract teacher, and by the fact that, as mentioned above, the program imposed that the PTA contract teacher be exclusively in charge of teaching all subjects to one specific grade 1 class, potentially allowing them to better understand those students' needs.

However, capture by teachers and their network can be reduced by promoting parental involvement and control at the local level. In a (randomly selected) subset of schools participating in the contract teacher program, members of the school committee and interested parents received a short School-Based Management (SBM) empowerment training (90 minutes) on how to interview and select job applicants, monitor and assess teachers' effort and performance, and perform a formal review of the contract teacher's performance to decide whether to renew her contract. Point estimates suggest that SBM empowerment program cut by half both the reduction in the regular teacher effort in response to the program by half and the fraction of contract teachers who were relatives of regular teachers. Moreover, it eliminated the test score differential associated with relatives of civil-service teachers.

Finally, contract teachers whose students scored well were more likely to be hired as civil servants, creating a dynamic that could potentially improve the quality of the teacher workforce (as in Gordon, Kane and Staiger, 2006). A calibration exercise suggests that if teachers started their careers by working for three years on contract for schools whose committee received SBM training and then were hired into the civil-service according to the process observed over the duration of our study, the steady state gain in test scores for the Kenyan education system as a whole would be up to 18% of a standard deviation.

The remainder of this paper is organized as follows: Section 2 provides background on central government and local community institutions for school governance in Kenya. Section 3 describes the Extra Teacher Program, the experimental design, and the data. Section 4 presents the impact of the program on test scores and teacher behavior, as well as evidence on how school committees select contract teachers. Section 5 discusses the potential dynamic impact of contract teacher hiring on the teacher workforce, based on evidence about how contract teachers' performance affects their likelihood of being hired as civil-service teachers. Section 6 draws conclusions and discusses evidence on related programs.

2. Education Governance in Kenya

Many developing countries adopted highly centralized education systems at much lower levels of development than did some of today's developed countries. Newly independent states adopted central control in part to rapidly expand education levels from a low base and in part as an element in nation-building, designed to unify diverse segments of society. Civil-service rules were seen as an important bulwark against politicization, ethnic favoritism, and nepotism in hiring and in favor of professionalism.

Kenya has had an evolving mixture of local, informal education governance and governance by the central state, but over time the role of the central state has grown, while that of local institutions has shrunk. During the colonial period, churches and local communities started many independent schools. While the Harambee movement in the late 60s and 70s encouraged local communities to start schools and retained representation of parents, the local community, and founding churches on school committees, overall the post-independence government adopted a fairly centralized education system. Civil-service headmasters and teachers were sent to take over successful Harambee schools and school committees were *de facto* relegated to fundraising rather than teacher governance. The Ministry of Education sought to professionalize and standardize teacher training, hiring, and discipline. A strong teachers union strengthened the emphasis on formal educational qualifications and on accountability to professional norms rather than to local parents.

2.1 TSC and PTA teachers

Historically, Kenyan schools have had two types of teachers: those hired as civil servants through the Teachers Service Commission (TSC) of the Ministry of Education, and Parent-Teacher Association (PTA) teachers hired locally and informally by local school committees. TSC teachers, have long constituted the vast majority of teachers, but graduates of teacher training colleges typically have to queue for TSC jobs, often undergoing years of unemployment or short-term employment as PTA teachers before they are hired by the TSC. Once they are hired, promotions, transfers, and disciplinary measures are decided through the TSC and based heavily on formal, objective criteria, such as educational qualifications and experience. TSC teachers are represented by a strong union, have civil-service protection, and receive wages and benefits considerably above market-clearing levels. Disciplinary issues are addressed by the TSC through formal legalistic procedures with adjudication by panels dominated by professional educators. The national political system in Kenya may not have produced particularly strong TSC teacher accountability. Indeed, constituency elections for members

of parliament in Kenya have historically been competitive, even under single-party rule, and much of national politics has been based on constructing alliances between particular ethnic and regional groups. Thus there were arguably stronger political incentives for Ministers of Education to devote effort to delivering resources to their home areas or to areas of allies in national politics than to press for increased teacher accountability at the national level.

PTA teachers are hired by school committees at each school. School committees are primarily composed of parent representatives, elected for each class. School committees also include some additional *ex officio* members, including the headmaster or a member appointed by the founding church. There is considerable variance in the *de facto* relative strength of the headmaster and parents in the PTA. Some school committees are dominated by the headmaster, but in others parents or other community members play a very influential role. Hiring and supervision of PTA teachers is quite informal, with much left to the discretion of the school committee or headmaster. Parents presumably have an incentive to select good teachers for their children and to incentivize them to teach, although school committee members may also have other incentives (such as hiring relatives), and Kremer et al. (2003) argue that the system by which school committee members are chosen led school committees to prefer higher school fees and school committee budgets than the typical parent would have preferred.

PTA teachers are typically paid much less than their TSC counterparts. In the area of study, in 2004, PTA teachers received compensation in the range of 2,000 Kenyan shillings, or US\$ 25 per month. In comparison, the average civil service teacher received around US\$ 120 per month plus benefits, including housing allowances, provisions for retirement, and medical coverage (Glewwe et al., 2010). PTA teachers are effectively at will employees of the school committee. Despite the low pay and lack of job security, PTA positions are actively sought after by unemployed teachers, in part because teaching experience helps them obtain a formal civil-service teaching position.

2.2   Free Primary Education and Local Governance

Kenya's historical pattern of school committees supplementing TSC teachers with locally-hired teachers paid from school fees was fundamentally transformed by the introduction of free primary education, itself likely a consequence of the introduction of a competitive multi-party competition for national political leadership.

After primary school fees were abolished in 2003, enrollment rose nearly 30 percent. At the same time, since parents were no longer required to pay fees, local school committees were generally unable

to raise the funds necessary to hire PTA teachers. They received grants from the central government, but the grants had to be spent on physical inputs, not hiring teachers locally, so *de facto* the introduction of free primary education strengthened the role of the central ministries in teacher governance at the expense of parents and local elites. A survey we conducted with over 300 primary schools in 2004 showed that 80 percent had no locally-hired teachers. As a result of increased enrollment and decreased number teachers, average class size in first grade in 2005 was 83, and median class size was 74; 28 percent of grade 1 classes had more than 100 pupils.[7]

3. The Extra Teacher Program and Study Design

3.1 Program Description

Extra Teacher Program

The Extra Teacher Program (ETP), implemented in 2005 and 2006 in Kenya's Western Province, provided funds to school committees to hire one PTA teacher, thereafter called the "ETP contract teacher", to supplement TSC teachers teaching in lower grades. The program was implemented by the non-governmental organization International Child Support (ICS), with funding from the World Bank, in a random subset of schools in three districts of Western Kenya. ICS staff met with the headmaster, the civil-service teachers assigned to the lower grades, and the parents of students in grade 1 to explain how the program worked and ask if they wanted to participate. To qualify for ICS funding, the contract teacher hired under the ETP contract had to have the same academic qualifications as civil-service teachers, and schools participating in the program had to create an additional section in first grade. Crucially, students and the contract teacher were then randomly allocated across sections. (The randomization was carried out by ICS staff in the office, with support from the research team and was adhered to in the field.) Other inputs, such as classroom facilities, were also supposed to be equally allocated across sections.[8]

If the contract teacher hired under ETP left the school (to take on a TSC position, or for any other reason, including dismissal by the school committee), the school was encouraged to replace her as quickly as possible.

---

[7] Many parents exited the public system, turning to informal private schools, which have grown dramatically since the introduction of free primary education (Lucas and Mbiti, 2011).

[8] In practice, established civil-service teachers may have pulled rank and obtained better physical classroom infrastructure. We do not have data to estimate how common this was, but in any case, it would tend to bias downwards estimates of the contract teacher effect and bias upwards estimates of the effectiveness of class-size reductions.

Contract teachers followed the same curriculum as TSC teachers, but they focused on a specific class (one of the sections in first grade), in contrast to TSC teachers who most often focus on a particular subject and rotate across sections and grades. In addition to its potential pedagogical benefits, this setup mirrored the organization of school committees and made it easier for parents to hold teachers accountable. For example, if a parent learned that their child spent a lot of the day unsupervised, they would know who is responsible if a single teacher is assigned to the class. When the program continued the following school year, school committees were free to replace or keep the contract teacher, but they were asked to assign the teacher to the second grade with the students from that same group who had been promoted.[9]

The monthly allowance for contract teachers was 2,500 Kenyan shillings (around US$35), putting it at the top of the range of what is typically paid to locally-hired contract teachers by school committees in Kenya, but at only one quarter of the typical TSC teachers' salaries, and a smaller fraction of their compensation, since civil-service teachers also receive benefits including pensions and housing allowances.

The Extra Teacher Program represents a shift in the balance between local parent control and formal and centralized professional control but it should be seen as augmenting, rather than replacing, the role of the central state. Likewise, it does not replace the current system where teachers have tenure, but rather augments it with a probationary period. Eligibility for the program was restricted to graduates of teacher training colleges, and contract teachers reported to centrally appointed headmasters. As mentioned above, headmasters typically have considerable influence within school committees. Contract teachers hired under ETP were likely motivated in part by the prospect that they would eventually obtain civil-service jobs, and performance in previous teaching positions is one of the criteria used by the TSC when making appointments.

School-Based Management Training (SBM)

The School-Based Management (SBM) training, implemented by a trained ICS officer in a random subset of ETP schools, was designed to empower school committee members (in particular, the parents, as opposed to only the headmaster) to monitor teacher performance. The training lasted about 90 minutes, and immediately followed the meeting that ICS held with parents to inform them of the Extra Teacher Program. All parents were welcome to stay for the SBM training if they so

---

[9] Students enrolled in grade 2 in 2005 and who repeated grade 2 in 2006 were randomly assigned to either the contract teacher or the civil service teachers in 2006.

wished.[10] During the training, school committee members were encouraged to supervise the recruiting of the contract teacher, taught how to conduct interviews, and taught techniques for soliciting input from other parents and checking teacher attendance. Two parents of grade 1 students were asked to volunteer to perform attendance checks on the teachers on a regular basis, and were given a notebook to record attendance. A formal sub-committee of first grade parents was formed to evaluate the contract teacher and deliver a performance report at the end of the first year. Finally, the school committees were asked to hold a formal review meeting at the end of the first school year of the program to assess the contract teacher's performance and decide whether to renew the teacher's contract or replace the teacher.

Although the SBM training focused on empowering parents to actively monitor the contract teacher, the encouragement and advice it provided on how to measure teacher performance may have helped parents to notice, and act on, the greater issue of absenteeism among civil-service teachers, and in particular to make sure the contract teacher was effectively used to increase resources available to grade 1 pupils, rather than to reduce workload for the other teachers.

Tracking

A separate set of schools were chosen for an alternative Extra Teacher Program. In those schools, the first-grade class was divided into two sections by initial achievement ("tracking"), and the ETP contract teacher was randomly assigned to one of these sections. This program is discussed in a companion paper (Duflo et al. 2011) and, since the pedagogy and classroom organizations changed significantly in those schools, data for the schools participating in that program is not part of the sample analyzed in this paper.

3.2 Experimental Design

Background data on enrollment, pupil-teacher ratios, and number of grade 1 sections was collected in 2004 from 210 primary schools in Western Province. Of these, 70 were randomly assigned to the tracking program and are not the focus of this paper. The remaining 140 schools were stratified by administrative division and number of grade 1 sections and randomly divided into a comparison group, and an Extra Teacher Program (ETP) group, each with 70 schools. Thirty four of the 70 ETP schools were randomly chosen to be offered the opportunity to participate in SBM training. ICS held

---

[10] This training took place before students had been randomized into the class taught by the contract teacher or a regular class, and all grade 1 parents were invited.

its school meetings to explain the program to parents and teachers in March 2005. All schools offered the opportunity to participate in the ETP and SBM programs chose to do so and agreed to respect the randomized assignment of students to sections, and of the contract teacher to a section. Schools selected for the Extra Teacher Program generally had a contract teacher in place by early May 2005, the start of the second of the three school terms in Kenya's academic year.

Panels A and B of Table 1 present summary statistics at the school level. Unsurprisingly given the randomized assignment of schools to treatment arms, Panel A shows no significant differences in observable characteristics at baseline. Panel B suggests that the program was successful at reducing pupil-teacher ratios over two years. While average class size in grade 1 in 2005 in the comparison schools was 82, it was only 44 in ETP schools.[11] This class size gap was somewhat attenuated the following year in second grade: in 2006, average class size in second grade was 68 in comparison schools and 42 in ETP schools. The class size gap fell over time because students assigned to the contract teacher in ETP schools were much more likely to be promoted to grade 2 than students in comparison schools (consistent with their better performance, which we discuss in section 4).[12]

Interestingly, a number of ETP schools decided that the students not assigned to the contract teacher would be assigned to one exclusive TSC teacher: the likelihood of class-based rather than subject-based assignment of TSC teachers was 18% in ETP schools compared to only 5% in control schools. We see no statistically significant difference in test scores of students taught by a single TSC teacher and students taught by rotating TSC teachers, however.

3.3 Study Sample

Summary statistics on the students and teachers in the study are shown in Panel C of Table 1. The student sampling frame is the cohort of approximately 13,500 children enrolled in first grade at the end of the first school term of 2005. Students averaged 7.7 years old at the outset of the program (with a standard deviation of 0.5 years), but ages ranged from 5 to 14. Baseline data on students' initial achievement was collected from the school records at the onset of the study. This data is not based on a common scale across schools, and thus can only be used for within-school comparisons. Random assignment of students to teachers in ETP schools led to sections with comparable average baseline scores.

---

[11] Class size did not fall by exactly 50% since some schools already had two grade 1 sections at baseline. In these schools, the introduction of the Extra Teacher Program reduced class size by only one third.

[12] The repetition rate is relatively high in Kenya. In year 2 of the program, the population in grade 2 thus included both those sampled students who had been promoted and did not transfer to another school, and grade 2 repeaters.

A total of 754 teachers taught lessons in first grade in 2005, in second grade in 2006, or both. Of those, 653 were centrally hired civil-service teachers, while 101 contract teachers were hired through the program by the 70 ETP schools over the five school terms (corresponding to 15 months of instruction) it operated. Average age was 42 for civil-service teachers and 27 for contract teachers, and contract teachers had 14 fewer years of experience than their TSC counterparts on average. Contract teachers were also much less likely to be females than TSC teachers in our sample (49% versus 66%). This is in part due to the fact that TSC teachers in our sample are those assigned to lower grades, and female teachers are disproportionately assigned to lower grades. The share of females in the entire TSC teacher body is on average 55% in our sample of schools, statistically indistinguishable from the percentage among contract teachers.

3.4 Data

Test Scores

Standardized tests covering math and literacy questions ranging from identifying numbers and letters to subtracting two-digit numbers and writing words were administered in all schools after 5 school terms (19 months total, but only 15 months of instruction), just before the program ended (November 2006). The same tests were administered again during a follow-up one year after the program ended (November 2007). Tests were administered by trained enumerators and graded blindly by a separate team of enumerators. In each school, 60 students were randomly drawn from the baseline sample to participate in the tests, for a total of around 8,000 students across all schools.

To limit attrition, enumerators were instructed to go to the homes of students who had dropped out or were absent on the day of the test and to bring them to school for the test. Not all students were found and tested, however. Overall, the attrition rate was 18 percent for the endline test, and 23 percent for the long-run follow-up test.[13]

Appendix Table A1 presents estimates of attrition by treatment groups. At endline, attrition was lower in ETP schools, especially for students assigned to the contract teachers (which is in itself of interest, and consistent with the higher performance level we observe among contract teachers). Low achievement students were significantly less likely to attrit from SBM schools, and less likely to attrit if they had been assigned to the contract teacher (although, surprisingly, if they were assigned to the contract teacher in an SBM school, they were not significantly less likely to attrit). This suggests that

---

[13] School choice is not a concern in our context. We observe quasi inexistent levels of student sorting across treatment and control schools during the experiment. In any case we conduct an intention-to-treat analysis, assigning treatment status based on the school the students were enrolled in prior to the program being announced.

the differential attrition will bias downwards the estimates of the impact of SBM on students assigned to regular teachers, and bias downwards the estimates of the short-run impact of contract teachers in non-SBM schools. The interaction between being in an SBM school and being taught by the contract teacher led to significantly lower attrition among higher-achieving students, however. Estimates of the effect of the combination of SBM and contract teacher may thus be upward biased. The patterns of attrition are roughly similar in the long-run follow-up. Given the substantial evidence of differential attrition, when estimating impacts on test scores below we estimate lower bounds, correcting for attrition à la Lee.

Teacher Effort
Over the course of the program, four unannounced school visits were made by the research team to measure teacher effort. Teacher effort was measured by the teacher's presence in school and by whether she was in class teaching when the observers entered the school compound. Data on pedagogy was collected through classroom observations and structured interviews with teachers.

Hiring, Retention, and Regularization of Contract Teachers
In the school term that followed the start of the program, data was collected on the contract-teachers hired through the program, including their demographic characteristics, past experience, relationship to the school (i.e., whether they are related to a teacher or a pupil), and the hiring process. Demographic data was also collected on civil-service teachers. In the school term that followed the end of the program, headmasters of program schools were interviewed about the status of the contract-teacher, and whether the school committee had taken over the extra teacher program after the ICS subsidy ended.

4. Results

This section first estimates the program impacts on pupil test scores (Subsection 4.1). To understand the mechanisms behind these results, we study impacts on teacher behavior (Subsection 4.2), and then study how the SBM program affected teacher selection and behavior (Subsection 4.3).

4.1 Learning Outcomes

Recall that we have two layers of randomization: schools were randomly assigned to programs, and within schools students were randomly assigned to teachers. To simplify the analysis, we first analyze the overall program effect at the school level, comparing averages across school types. Specifically, we use the following reduced form equation to estimate the effects of each school-level program on student outcomes:

$$Y_{ij} = a_1 \, BasicETP_j + a_2 ETPwSBM_j + X_{ij}'a_5 + \varepsilon_{ij1} \qquad (1)$$

where Y_{ij} is the endline test score of student i in school j (expressed in standard deviations of the distribution of scores in the comparison schools),[14] $BasicETP_j$ is a dummy equal to 1 if school j participated in the Basic Extra Teacher Program, and 0 otherwise (including in schools that had the ETP program re-informed with SBM), $ETPwSBM_j$ is a dummy equal to 1 if school j participated in the ETP program augmented with School-Based Management training, and 0 otherwise, and X_{ij} is an (optional) vector of student control variables.

We then allow for the program effects to depend on the type of teacher a student was assigned to. Namely, we estimate the following equation:

$$Y_{ij} = b_1 \, BasicETP\_TSC_{ij} + b_2 \, BasicETP\_CT_{ij} + b_3 ETPwSBM\_TSC_{ij} + b_4 ETPwSBM\_CT_{ij} + X_{ij}'a_5 + \varepsilon_{ij2}$$

where $BasicETP\_TSC_{ij}$ is a dummy equal to 1 if student i was assigned to a class taught by a teacher from the TSC in a school that received basic ETP, and the other dummies are similarly defined.

The results are presented in Table 2. We present the results of specifications that control only for a region dummy and a linear trend in the endline test date to account for longer teaching time in school schools than in others depending on when the test was administered.[15] (The results are robust to including school-level and student-level controls, see Appendix Table A2).

In equation (1) estimated in columns 1-3 of Table 2, a_1 is the average effect on test scores of being in a school sampled for the basic ETP program. The effect is large, an increase of 0.189 standard deviation, significant at the 5% level. The coefficient estimate a_2 is the average effect of being in a school sampled for the ETP with SBM program. The effect is somewhat larger, an increase of 0.236

---

[14] An alternative specification of the endline test score for math, using item response theory, yields similar results (available from the authors). The format of the language score was not appropriate for this exercise.

[15] The data collection was phased in by region. On average the endline test was administered 31 days earlier in the Bungoma region. We also have some significant differences in the timing of the endline test by group: the endline test was done on average 6 days earlier in the basic ETP schools, and 4 days earlier in the ETP with SBM schools, compared to the comparison schools. If we do not control for a trend in the date of the test the program effects thus appear somewhat lower. (We use a linear trend, the results are robust to a higher order polynomial).

standard deviation, also significant at the 5% level. We cannot reject the null that the impacts of the two programs was the same, however (see the p-value for the test that $a_1=a_2$ at the bottom of the table, at 0.656). The gains were slightly stronger in math than literacy (columns 2 and 3).

Figure 1 and Columns 4 to 6 of Table 2 break down the program effects by teacher type. Focusing first on the TSC vs. contract teachers split, Figure 1 shows the distribution of the endline test scores for three groups: students in the comparison schools, students assigned to TSC teachers in ETP schools (whether Basic or SBM) and students assigned to PTA contract teachers in ETP Schools. The three distributions are clearly distinct from each other. The performance of students assigned to contract teachers strictly dominates that of students assigned to TSC teachers within ETP schools, and these latter students themselves perform strictly better than those in control schools.

Turning now to the effect by teacher *and* program type, coefficient b_1 in columns 4-6 of Table 2 is the effect of the basic ETP program for students assigned to the TSC teachers. Absent any endogenous response of teachers, this would mainly capture the effect of a reduced class size in the schools that were provided additional teaching staff. The effect is small (0.087 of a standard deviation) and insignificantly different from zero (standard error: 0.098), despite a decline in class size of almost 50%. In contrast, the effect of the basic ETP program on students assigned to a contract teacher is very large, at 0.294 standard deviation (coefficient b_2). We can reject at the 1% level equality of the effect of the basic ETP program between those assigned to TSC teachers and those assigned to contract teachers (see p-value for the test $b_1=b_2$ at the bottom of the table).

The gap between students of TSC vs. contract teachers is considerably smaller when the ETP program is implemented jointly with SBM. In SBM schools, coefficient estimates b_3 (for students assigned to TSC teachers) and b_4 (for students assigned to PTA teachers) are both large and significant, and we cannot reject that they are equal to each other. Ultimately, the ETP with SBM program appears to yield large gains in test scores for all students, while the Basic ETP program only benefitted students assigned to the contract teachers. Thus the very cheap SBM intervention somewhat increases the effectiveness of the Extra Teacher Program.[16]

---

[16] It should also be noted that, while we find positive and at times statistically significant effects of SBM training on test scores in the context of the main Extra Teacher Program, the estimated effects were smaller and not statistically significant in the tracking program (results are available upon request). This may merely reflect the more limited scope for improvement under that program: the tracking program already increases teacher effort (at least in the top track), probably leaving little scope to further increase effort and test scores.

Appendix Table A3 shows the estimated lower bounds for these specifications accounting for attrition, using the approach first proposed in Lee (2008).[17] The relative performance of the programs and their effects by teacher type appear relatively robust to controlling for attrition, though most of the conservative lower bounds we estimate are insignificant at conventional levels. The one result that remains consistently significant at conventional levels is the finding that contract teachers led to large and significant improvements in math scores, both when compared to control school students and to students assigned to TSC teachers within basic ETP schools.

While we see relatively large learning effects in the years contract teachers are present, these effects appear to fade out once students are reassigned to regular classes. One year after the program ended and students had gone back to being taught by rotating civil service teachers in large classes, the effects were no longer statistically significant (Appendix Table A4). Such fade-out is not out of line with the decay in test scores observed for other early interventions, including India (e.g. Banerjee et al, 2007) and the US (see Cascio and Staiger, 2011, for a review), and with rates of fade out estimated in developing countries (Andrabi et al. (2011) find that only a fifth to a half of learning persists between grades), but contrasts with the more persistent effects of tracking that we observed in the same context (Duflo et al, 2011). Of course, the decay in test score effects does not necessarily imply a lack of long-term impacts (Chetty et al. (2010)).

4.2 Teacher Effort and Pedagogy

To understand the mechanisms behind these results, table 3 presents evidence on teacher effort and on pedagogy using data from unannounced spot checks, classroom observations conducted with a subset of teachers, and surveys administered to teachers. For this analysis we consider five types of teachers: TSC teachers in comparison schools; TSC teachers in Basic ETP program schools; TSC teachers in ETP with SBM schools; contract teachers in Basic ETP program schools; and contract teachers in ETP with SBM schools. We regress a battery of teacher effort and pedagogy variables on dummies for each teacher type – the omitted category being the first, TSC teachers in comparison schools. Thus the coefficient estimates shown in Table 3 presents the average performance of a teacher of a given type compared to TSC teachers in comparison schools.

---

[17] The method amounts to the following: for each treatment group, we identify the "attrition deficit" (e.g., the excess proportion of students who did not attrit compared to the comparison group), and then "trimming" the upper tail of the test score distribution by this proportion, yielding a worst-case scenario bound. To refine the bound, we use two baseline covariates (gender and age above/below median) and, within each treatment group, use different trimming proportions for each gender-age category. The trimming proportions are shown in tables A3.

We find striking differences across school and teacher types. Focusing first on effort (Panel A), we find that contract teachers in Basic ETP schools were 11.7 percentage points more likely to be found in a classroom and teaching during a random visit than TSC teachers in comparison schools (the mean for the TSC teachers in the comparison group was 57.9%, see bottom of Panel A, column 1). In contrast, TSC teachers in Basic ETP schools were 15.7 percentage points *less* likely to be found in class teaching during a school visit than their counterparts in comparison schools. This effect corresponds to a 27 percent decrease in teacher presence in class compared to comparison schools. This suggests that TSC teachers took advantage of the presence of the contract teachers to work less. Contract teachers might not have been in a position to refuse teaching combined classes when TSC teachers were absent.

However, there is evidence that this reduction in effort by TSC teachers was mitigated through community empowerment. In ETP with SBM schools, the decrease in attendance among TSC teachers is only around half that observed in Basic ETP schools (-8.3 pp vs. -15.7 pp, p-value = 0.081). It seems likely that SBM directly affected TSC teachers, rather than influencing them by only reducing the likelihood contract teachers would cover classes for them, since we find no effect of SBM training on the likelihood that contract teachers were covering classes for TSC teachers during unscheduled school visits (results available upon request).

The SBM program did not affect TSC teachers' presence on the school compound, but increased the likelihood that teachers were in class teaching if present (Table 3, Columns 2 and 3). While teachers who are present but not teaching might sometimes be grading or performing some other duty for the school, in the great majority of the cases we observed teachers were drinking tea with other teachers or reading a newspaper, and therefore we are confident that the increase in classroom time we observe among TSC teachers in the SBM schools corresponds to an increase in effort, not to a reallocation of total effort towards grade 1 specific effort. Note that it is easier for parents to observe and enforce along this margin (and perhaps also less costly for teachers). That is because teachers who are absent from the school compound might claim that their absence was work-related (e.g., in-service training, proctoring, committee work), and parents would not be able to verify such claims without the assistance of the headmaster. [18]

---

[18] The estimates in section 4.1 and 4.2 could potentially be combined to disentangle the possible gains of decreasing class size (keeping effort constant) and the impact of effort (proxied by teacher's time in class). In an earlier draft of the paper, we considered two different IV estimates, with different exclusion restrictions. First, we assumed that, conditional on teacher's sex and experience (which are observed), the only difference between contract teachers and civil-service teachers is that contract teachers have stronger incentives. Under this assumption, the experimental design provides four instruments for class size and effort: whether the student is in an ETP school, whether the student is in an SBM school,

Besides teacher effort, the ETP program could have led to changes in pedagogy. First, the class was smaller. Moreover, the pedagogy adopted by contract teachers could systematically differ from that of TSC teachers, given their focus on a single class, or the fact that contract teachers were typically younger (more energetic and trained more recently, but also less experienced). To test this, Panels B and C of Table 3 present estimates of program impacts, by school and teacher type, on pedagogy, measured through direct observation of lessons and structured interviews with teachers about classroom processes. Overall, we find very little evidence that the ETP program changed pedagogy. Students in both types of ETP schools appear somewhat more likely to be called to the blackboard and to be called on to answer questions, and teachers in those schools (whether civil servants or contract teachers) also report having more time to provide individual attention to children. Compared to the massive differences in effort levels observed in Panel A, however, within ETP schools, Panels B and C provide little evidence of differences in pedagogy across teacher or ETP program variants.[19]

Our results should be interpreted as the impact of supplementing a civil service system with locally-hired contract teachers brought in on a probationary basis, with the expectation of eventually obtaining a civil-service position, not as the impact of replacing a civil service teacher system with a system based entirely on local hiring of contract teachers. The contract teachers we study likely worked hard in part because they believed this would help them obtain a civil-service position: as discussed in the next section, about half of the contract teachers in our study eventually obtained civil-service positions by the end of our study period, with better-performing teachers more likely to obtain these positions.[20]

---

whether the student was assigned to a contract teacher, and the interaction between being assigned to a contract teacher and being in an SBM school. The second IV estimate did not use being assigned to a contract teacher as an excluded instrument, thus allowing for an effect of assigning a single teacher to a class. The results suggested that a negative but modest (significant) effect of class size on test scores, and a positive and large significant effect of teacher effort on test scores. Given that the exclusion restriction for SBM as an instrument for effort is questionable (in particular, because SBM could have had a direct effort on parental involvement), those IV results should be considered illustrative rather than definitive.

[19] One might expect a bigger impact of moving from rotation to a system in which students are taught by a single teacher in schools where more teachers share the first-grade class. We therefore estimated how the contract teacher effect varied with the number of civil-service teachers rotating in the non-ETP section. This test has low power, but we do not find any indication that the contract teacher effect is greater in schools where more teachers share classes.

[20] Given these incentives, it is impossible to say what would happen if TSC teachers were phased out and replaced by contract teachers without any prospect of tenure.

4.3   Selection and Retention of Contract Teachers: Teacher Rent Seeking and Parents' Empowerment

The results above suggest that TSC teachers responded to the ETP program by reducing effort, thus capturing some of the benefits of the program for themselves. In this section, we show that TSC teachers also captured some of the benefits of the program for their extended family, by securing the contract teacher positions for relatives.[21] We provide evidence that the SBM program, just as it mitigated the effort response, mitigated such nepotism. This is another channel through which empowering parents increased the effectiveness of providing additional resources to the school. As background, in Kenya, those with formal sector jobs often face strong pressure from members of their extended families to help them obtain such jobs as well. This pressure may be particularly strong for males, because Kenya is a patrilocal society and contract teachers are overwhelmingly hired from the local area, so a greater proportion of male teachers will be local to the area of the school and will have relatives from the area interested in applying for the job.[22] Moreover, women are considered to have fewer obligations towards their birth family when they marry into a new family.

To understand the factors influencing hiring and retention of contract teachers it is useful to consider the following framework. Suppose PTA hiring committees can be motivated both by a desire to obtain rents for relatives and a desire to improve student performance. We assume that among the members of the PTA, parents put relatively more weight on maximizing student performance, while teachers (particularly male teachers) put more weight on delivering rents to their relatives. Insofar as SBM training empowers parents within the PTA committee, it will reduce the weight PTA committees place on delivering rents to relatives of teachers.

Suppose that PTA hiring committees receive applications and can obtain observable information such as sex, education, and experience along with an imperfect signal of teacher quality. Suppose also that teachers can supply additional information on their relatives creating a force leading to a potentially positive impact of hiring relatives. Once teachers have been hired, school committees obtain additional information on their performance and decide whether to renew their contract and, more generally, how to manage them, and TSC teachers may continue to deliver rents to relatives. Relatives choose effort levels knowing this.

---

[21] This finding relates to Durante et al. (2011), who find that a 1998 reform that increased autonomy by local university officials in Italy resulted, in some areas, in a significant increase in the incidence of nepotism towards relatives.
[22] While TSC teachers are initially posted outside of their home area, they become eligible to choose their location as they gain seniority, and it is very common for seasoned TSC teachers to be in posts in their home area.

The combined effects of information and rent seeking on aggregate hiring of relatives and on the relative test scores of students of relatives of existing TSC teachers are ambiguous, but this framework suggests that SBM should unambiguously reduce hiring of teacher relatives, particularly those hired due to rent-seeking motives as opposed to those hired based on positive information provided by teachers working at the school. Moreover, if it is politically more difficult for teachers on PTA committees to fire teacher relatives, then in equilibrium we might see lower effort among those relatives (due to lower incentives), but may increase equilibrium effort among teacher relatives working as contract teachers. Test scores of students of relatives should be higher under SBM than without it. We have limited power to test these predictions due to the small sample size, but the evidence we do have is consistent with the predictions.

Hiring of Relatives

In ETP schools without SBM, 31% of contract teachers hired at the onset of the program were relatives of existing TSC teachers, while in SBM schools the share of relatives among contract teachers was about half that, at 16% (Table 4, column 1; the t-statistic on the SBM effect is only 1.5 due to the limited sample size, but a similar SBM effect was observed in tracking schools, suggesting this is not coincidence.)

One frequent concern with local empowerment programs is the risk of local capture: empowering parents within PTA might merely lead to a bias towards locals, perhaps with a strong ethnic component. However, there is no evidence that the SBM program led to a shift in favoritism from relatives of teachers to locals. On the contrary, point estimates suggest SBM schools were 12 percentage points less likely to hire contract teachers from the local area, although the difference is not significant given the small sample size (see Column 2, Table 4.) More generally, it seems that SBM training increased the transparency of the contract-teacher recruiting process. SBM schools advertised for the position more broadly and interviewed more candidates (these results are shown in Appendix Table A5).

Another concern is that teachers may be more influenced by professional qualifications, while parents may be influenced by cultural factors, such as preferences for men. Point estimates suggest the possibility that empowering local parents reduced hiring of women as contract teachers, but here again effects are far from significant in our small sample (Table 4, column 3). Students of female teachers score better (see Table A2, column 2, row 8), and thus the impact of SBM on contract teacher gender seems to have counteracted the impact of SBM on hiring relatives. Overall, we cannot reject

the hypothesis that students of contract teachers perform equally well with and without SBM (see bottom of Table 2, test of equality between coefficients b_2 and b_4).

Teacher effort and student performance

The framework above suggests that teacher relatives hired under SBM should perform better than teacher relatives hired in the absence of SBM, both because SBM should reduce rent seeking while continuing to allow information flows and thus should improve selection of teachers, and because SBM schools may be better able to monitor and incentivize those relatives of existing TSC teachers who are hired. Looking at student tests scores in column 1 of Table 5, we estimate both the main effect of being taught by a contract teacher who is a relative of an existing teacher, and an interaction between SBM and this variable. We do find that relatives perform less well than non-relatives: the point estimate of the "teacher relative" effect on test scores is – 0.339 in non-SBM schools (significant at the 10% level). This indicates that hiring relatives must to some extent be a way to share rents, and is inefficient from society's point of view. The SBM program entirely undoes this effect however: the interaction is +0.385 standard deviations (significant at the 5% level). Taken together, these numbers imply that in SBM schools the total "teacher relative effect" is statistically undistinguishable from zero, consistent with the hypothesis that SBM undoes the tendency either to select weak relatives or to let relatives get away with low effort.

One worry with the analysis above is that schools that hired a relative may systematically differ from schools that did not, and therefore the heterogeneity in contract teacher performance by "relative" status may be picking up some differences across schools rather than the specific effect of nepotism or the lack thereof. To check this, in column 2 of Table 5 we test whether the "relative" and "relative x SBM" variables have predictive power for the performance of students assigned to TSC teachers, and find similarly signed coefficients but of smaller magnitudes and insignificant, suggesting that nepotism is at least part of the story in column 1.

Looking at teacher effort, the point estimates in column 3 of Table 5 suggest that while contract teachers who are relatives of TSC teachers are less likely to be in class teaching, this tendency is mitigated by the SBM program. Coefficients are large in magnitude and are significant at the 10% level. Interestingly, the effect of the SBM program on the performance of TSC teachers was concentrated among schools that hired relatives (column 4).

The framework also suggests that if existing TSC teachers are female, rent seeking will be lower but information channels should still operate, leading to better performance of relatives of females.

The data is consistent with this: fewer teacher relatives were hired in schools with a higher fraction of females among TSC teachers in lower grades (Table 4, column 1), and students of relatives hired as contract teachers learned more in those schools (Table 5, column 1, significant at the 10% level). Point estimates suggest relatives hired in these schools were present and teaching in class more often (Table 5, column 3).

5.   Potential Dynamic Effects of Local Probationary Teacher Hiring

The results presented above suggest that reintroducing elements of Kenya's historic mix of local and national teacher hiring by moving from a system in which teachers are immediately given tenure and managed solely through centralized state-run institutions to a system in which teachers are initially hired locally on temporary contracts before becoming eligible for TSC positions and local PTAs are trained in how to recruit and monitor teachers, could not only save roughly 75% on teacher salaries during teacher's initial period as locally-hired contract employees but could also have the potential to significantly strengthen incentives for teachers and produce better learning outcomes during this period. However, the long-run impact of adding a new phase to the career of teachers in which they are hired locally on short-run contracts before obtaining civil-service positions depends not only on the relative performance of TSC teachers and contract teachers at a given point in time, but also on how local hiring of contract teachers affects the teacher workforce over time. In this section, we first show that the TSC was able to identify and hire better-performing contract teachers. We then briefly present an (admittedly speculative) calibration of the impact of a system in which qualified teachers must first be hired by their local community on renewable contract and only then get tenure within the TSC.

Performance-based Retention

Table 6 presents evidence on how contract teachers' performance affected their outcomes.[23] By the beginning of year 3, the year after the program ended, 47% of the initial contract teachers had been hired by the TSC.[24] Experience (a proxy for cohort) was an important driver in TSC hiring. But holding experience and school quality constant, contract teachers whose students had good scores were more

---

[23] Data on the contract teacher are missing for 2 of 70 program schools.
[24] We had no contact with the TSC during the study, but we conducted interviews with school headmasters at the end of the program, and those interviews suggest that headmasters can recommend contract teachers to the TSC hiring commission.

likely to be hired by the TSC. A one standard deviation increase in the performance of students assigned to the contract teacher increases the likelihood that the contract teacher was absorbed by the TSC by 35 percentage points, after controlling for performance among students of TSC teachers (Table 5, column 1). It seems likely that benchmark competition between contract teachers and existing civil service teachers was responsible for the tendency for teachers with better scores to obtain TSC positions. Indeed, there was no relationship between student test scores and teacher hiring by the TSC in the schools where students were tracked by initial performance, which presumably made it more difficult to measure contract teacher performance.

Overall, 19% of contract teachers left their school (for reasons other than having obtained a TSC position) before the end of the program and had to be replaced (Table 5, column 2). The point estimate of -0.11 for student endline performance suggests that teachers whose students' had poor performance were more likely to leave, but the low performance of the students may have been caused by the early departure, rather than the other way around.[25]

More telling is what happened at the end of the program. The funds for the program stopped, but PTAs could raise their own funds to continue it. Schools with the SBM program were more likely to retain contract teachers into year 3. While 25% of the non-SBM schools kept the contract teacher on board after the ETP program ended, this figure was 0.25+0.21=46% among SBM schools (see Table 5, column 3). Thus parents were close to twice as likely to pay to continue employment of the contract teacher when they had stronger governance rights. We cannot distinguish the direct channel of parents being more willing to raise funds for a program because they felt they had a greater role in governance from the indirect channel of governance improving program performance and this in turn increasing willingness to pay.

Finally, among teachers not hired by the TSC, we see a positive but insignificant relationship between student performance and contract teacher retention by the school (Table 5, column 4). Relatives are not less likely to be retained, despite worse performance, so conditional on performance, relatives are more likely to be retained.

Long run dynamic impacts

The evidence above suggests that the TSC was able to identify and give permanent contracts to better performing contract teachers. In the Appendix, we do a calibration exercise and estimate, for a reasonable set of assumptions, the potential dynamic impact of a local contract teacher hiring program

---

[25] School committees never explicitly voted against renewing a contract in year 2.

embedded in an "up or out" promotion system. We estimate that if new entrants to the profession worked three years as a contract teacher and half were subsequently hired as civil service teachers where they remained for 27 years, then 20% of the teaching force would be made up of contract teachers in steady state. The extra incentives for the 20% of the workforce on temporary contract (net of the 'novice' effect identified in the literature times the higher proportion of novice teachers in steady state under such a system)[26] would yield a gain of 0.05 standard deviations in test scores; the positive selection into the TSC would yield an increase in student scores of 0.13 standard deviations; and thus the total learning gain from the system would be 0.18 standard deviations. This calculation is obviously not meant to be definitive, since it relies on a number of assumptions (e.g. that teachers serve for 27 years), and since the extent of capture by teachers or local elites may differ if a probation system is formalized and scaled, and also since much would clearly depend on the details of how such a system was implemented; but it provides a useful benchmark of the potential gains from reintroducing elements of the historic system in which many teachers were initially hired on short-term contracts by school committees and later absorbed into the civil service.

6. Conclusion

Efforts to improve education in developing countries often focus either on providing additional resources, typically by hiring more teachers to bring down class size, or on governance reform. We examine the impact of additional resources, and of two popular approaches governance reform: decentralized hiring of contract teachers, meant to decrease pupil-teacher ratios at low cost while at the same time strengthening teacher incentives through local control; and School-Based Management training programs, designed to empower parents. We examine these programs in Kenya, and find that they can work as complements. In the absence of parents' empowerment, the additional resources brought about by a contract teacher program are partly captured by the existing civil-service teachers in two ways. First, civil-service teachers reduce effort, which undoes the positive impact of class size reduction for their students. Second, they seek to capture rents by hiring their relatives, whose students perform less well than students of other contract teachers. Empowering parents through a short training session mitigates both these negative effects. First, in schools with SBM training, civil-service teachers were more likely to be present in class and teaching; second, in those schools, relatives of civil-service teachers were less likely to be hired as contract teachers; third, those relatives who were

---

[26] Novice teachers typically have lower value added (Kane et al., 2008; Rockoff and Staiger, 2010).

hired anyway performed as well as non-relatives (which could come from better selection of the remaining relatives, or stronger incentives).

Our results suggest that in the presence of weak institutions, increases in resources may be undermined by the behavioral responses of existing providers. However, local governance offers the potential to translate increased resources into better outcomes. In particular, programs devolving authority to hire teachers on short-term contracts have potential both to improve test scores in the short run and to improve the quality of the teaching labor force in the long run. But details matter: small differences in program design (e.g., a few hours of SBM training) can substantially affect outcomes.

In this light it is worth considering two related studies. A large-scale randomized study carried out across Andhra Pradesh (AP) State in India contemporary to ours (Muralidharan and Sundararaman, 2013) examines a program in which most contract teachers were not teacher training college graduates and thus were not eligible to become civil-service teachers. They find average test score gains very similar to those we observe, as well as reductions in civil-service teacher presence, suggesting that very similar forces may be at play in a different geographic and institutional context. Although the AP experiment did not include random assignment of students to contract and civil-service teachers so cannot estimate the causal impact of reductions in pupil-teacher ratios within the existing institutions, the authors observe that the correlation between class size and test score gains declines at higher grades, suggesting heterogeneous returns to pupil-teacher ratio reductions as the mechanism for the larger effect of the contract teacher program on lower grades.

The fact that similar effects were found in a context in which most contract teachers were not eligible to become civil-service teachers suggests that contract teachers may perform well even with less formal training and even without the incentive to perform well created by the possibility of a formal sector job. Of course, on the other hand, structuring a program without opportunity for contract teachers to become civil servants eliminates the potential discussed in Section 5 of this paper for positive selection of better-performing contract teachers to raise the quality of the civil-service workforce over time. The AP experiment did not include an interaction with SBM empowerment training.

Results from a study following up on ours (Bold et al., 2013), also in Kenya, suggest that the impact of contract teacher hiring is indeed sensitive to the institutional context and whether the program is executed as designed. A contract teacher program in which administration was contracted out to an NGO which made payments to school committees to hire teachers yielded learning gains similar to

those we found in a variety of contexts. But in a variant in which payments to schools were made through the district offices of the Ministry of Education (a separate institution from the TSC, which normally handles payments to civil-service teachers), contract teacher positions were less likely to be filled, monthly payments to teachers were often seriously late (the average delay was 2.33 months and 10% of teachers waited 10 months to be paid), and the program did not significantly improve test scores. Bold et al. (2013) show that part of the difference is accounted for by the unfilled vacancies and the late salary payments. Possibly more important, Bold et al. (2013) also show that contract teachers hired through the Ministry were much more likely than those hired by the NGO to believe that they would be automatically promoted to TSC teachers at the end of their contract, which as expected seriously undermined incentives.

A classic objection to decentralization and local control is that it may lead to capture by local elites. We indeed find evidence of such capture, but in our context, capture is not by local gentry or high caste landowners, but rather by the local representatives of the central state itself. Kenya has long had efforts to devolve control, for example through district-based planning. However, these have taken place within the context of a centralized state without local, elected legislatures. Our analysis suggests that unless such initiatives are combined with institutional reforms to allow democratic control at the local level, local agents of the state may use devolution to capture rents. Kenya's new constitution, overwhelmingly approved by voters, combines decentralization with the creation of elected legislatures at the local level. We do see some potential indication, although far from statistically significant, that empowerment of parents leads to selection of fewer female teachers, so it may be worth considering combining decentralization initiatives with centrally-enforced safeguards for women and other groups that may be discriminated against at a local level, such as minority communities.

Our results also have implications for the literature on corruption, governance, and ethnic conflict. Much of the recent literature in economics on corruption focuses on financial corruption – taking bribes or stealing state funds, for example. Yet we would argue that obligations to relatives often conflict with obligations to the state, or to employers more generally, and that favoring relatives is a widespread form of corruption.

It is also a form of corruption that interacts closely with ethnic politics. When civil servants receive rents in the form of above-market wages, enjoy civil-service protection and thus long-term employment, and can in turn influence the hiring of other civil servants, favoritism toward relatives can launch a dynamic process of growth of a network of coethnics within the state. Our analysis

suggests that altruism towards those family members (or greater ability to enforce reciprocity within the family) could be one reason why politics in much of Africa is often defined by ethnic cleavage, rather than any other arbitrary variable (such as religion or class). Since marriage in the society we study is primarily within ethnic group, relatives will overwhelmingly be coethnics, and since the society is patrilineal, relatives of males will be coethnics. In societies in which only a minority have the educational qualifications required for most civil-service jobs, networks of coethnics with sufficient education to qualify for these jobs may be tight enough that individuals benefit indirectly from hiring of unrelated coethnics.

What are the political prospects for local hiring of contract teachers? While teachers unions are likely to resist efforts to reduce their members' wages, they may be willing to compromise on a system in which existing members are protected, but new entrants to the teaching profession are first hired on short-term contracts at lower pay, and then are eventually brought in as civil-service employees. Under some standard models of union bargaining, in which unions can set or negotiate wages and benefits but the government controls total hiring of new civil-service teachers, some teacher training college graduates who are willing to work at less than the government's reservation value will not be hired. One possible solution would be for the government to hire these teachers outside union rules at a lower wage. Ex post, this would be statically efficient. One problem, from the union's point of view, is that the government may not be able to commit not to use the fact that it expects to hire teachers on this basis in its wage negotiations with the teachers union. Moreover, the government could simply replace TSC teachers with PTA teachers over time. To avoid this, the government would have to commit to limiting the number of contract teachers hired over time. One way to do this would be through an "up or out" system in which the workers outside the union bargaining agreement have to be either hired as civil servants or fired after a certain time. A key issue would be to ensure that the selection process maintains the positive correlation with student performance found in our sample.

 "Up or Out" systems are of course common in the higher education labor market in the United States. Short-term contracts prior to long-term jobs have also been a feature of some European labor market reforms. It is possible that such arrangements could be part of education reforms in developing countries with strong unions. Pritchett and Murgai (2006) proposed such a system for India.

Subsequent to our study, the Kenyan government, which had long had a freeze on hiring of new civil-service teachers, hired 18,000 contract teachers. Initial plans included no guarantee of civil-service employment afterwards. However, the Kenyan National Union of Teachers opposed the initial plans and under the eventual agreement, contract teachers were hired at much higher salaries than in the

program we study. Hiring was done under civil-service rules heavily weighting the cohort in which applicants graduated from teacher training college rather than the judgment of local school committees, and contract teachers hired under the program were promised civil-service positions. Our analysis suggests that these features could potentially undermine both the incentive and selection effects of local hiring of contract teachers by PTAs. Understanding the selection and incentive impacts of this program, and of potential program variants, is an important topic for future research.

References

Amin, Samia, Jishnu Das and Markus Goldstein (edited) (2007). *Are you Being Served? New Tools for Measuring Service Delivery.* World Bank, Washington DC.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc (2011). "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics" *American Economic Journal: Applied Economics*, 3(3): 29–54.

Angrist, Josh, and Victor Lavy (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114, 533-575.

Bandiera Oriana, Valentino Larcinese and Imran Rasul (2010). "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students," *Economic Journal* 120(549): 1365-1398.

Banerjee, Abhijit, Shawn Cole, Esther Duflo and Leigh Linden (2007). "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics*, 122(3): 1235-64.

Banerjee, Abhijit, Brian Jacob and Michael Kremer (2005), with J. Lanjouw and P. Lanjouw, "Moving to Universal Education! Costs and Tradeoffs," MIT mimeo.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a and Justin Sandefur (2013). "Pitfalls of Scaling-up Proven Education Interventions: Evidence from an RCT in Kenya". CGD Working paper 321.

Bruns, Barbara, Deon Filmer, and Harry A. Patrinos (2011). *Making Schools Work. New Evidence on Accountability Reforms.* The World Bank, Washington D.C.

Cascio, Elizabeth, and Douglas Staiger (2011). "Skill, Standardized Tests, and Fadeout in Educational Interventions". Mimeo.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2010). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126(4): 1593-1660, 2011

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Hasley Rogers (2006). "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives* 20 (1): 91-116

Das, Jishnu, Stephane Dercon, James Habyarimana, and Pramila Krishnan (2004). "When Can School Inputs Improve Test Scores?" Policy Research Working Paper 3217, World Bank, Washington DC.

Das, Jishnu, Stephane Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan and Venkatesh Sundararaman (2013). "School Inputs, Household Substitution, and Test Scores." *American Economic Journal: Applied Economics*, 5(2): 29-57.

Duflo, Esther, Pascaline Dupas and Michael 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.

Duthilleul, Yael (2005) *Lessons Learnt in the Use of "Contract" Teachers. Synthesis Report* International Institute for Educational Planning (IIEP) UNESCO; Paris.

Durante, Ruben, Giovanna Labartino and Roberto Perotti (2011). "Academic Dynasties: Decentralization and Familism in the Italian Academia," CEPR Discussion Paper 8645.

Glewwe, Paul, Naumam Ilias and Michael Kremer (2010). "Teachers Incentives." *American Economic Journal: Applied Economics*, 2(3): 205-27

Goldhaber, D., and Hansen, M. (2013). "Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance". *Economica* 80(319): 589–612.

Gordon, Robert, Thomas J. Kane, Douglas O. Staiger (2006). "Identifying Effective Teachers Using Performance on the Job" The Hamilton Project, Washington, DC: Brookings Institute.

Hanushek, Eric A., and Ludger Woessmann (2010). "The Economics of International Differences in Educational Achievement", in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann (ed.), *Handbook of the Economics of Education*, Vol. 3. Amsterdam: North Holland.

Hoxby, Caroline M. (2000). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 115.4.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger (2008) "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27: 615–631.

Kremer, Michael, Moulin, Sylvie, and Robert Namunyu. (2003) "Decentralization: A Cautionary Tale," Working Paper Harvard University.

Krueger, Alan and Diane Whitmore (2002). "Would Smaller Classes Help Close the Black-White Achievement Gap?" In John E. Chubb and Tom Loveless, eds., *Bridging the Achievement Gap*. Washington: Brookings Institution Press.

Kumar, K., M. Priyam, And S. Saxena (2005): "The Trouble with Para-Teachers," *Frontline*, 18.

Lavy, Victor (2010). "Do Differences in School's Instruction Time Explain International Achievement Gaps in Math, Science, and Reading? Evidence from Developed and Developing Countries". NBER Working Paper w16227.

Lee, David (2008). "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", *Review of Economic Studies*, 76(3) 1071-1102.

Lucas, Adrienne and Isaac Mbiti (2011). "The Effect of Free Primary Education on Student participation, stratification and achievement: evidence from Kenya"

McCaffrey, D., Sass, T., Lockwood, J., & Mihaly, K. (2009). "The intertemporal variability of teacher effect estimates." *Education Finance and Policy*, 4(4), 572-606.

Muralidharan, Karthik, and Venkatesh Sundararaman (2013). "Contract Teachers: Experimental Evidence from India". Mimeo, UC San Diego.

Piketty, Thomas (2004). "L'Impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises: une estimation à partir du panel primaire 1997." Mimeo, PSE.

Pritchett, Lant, and Rinku Murgai (2006). "Teacher Compensation: Can Decentralization to Local Bodies Take India from the Perfect Storm Through Troubled Waters to Clear Sailing?" India Policy Forum, The Brookings Institution, vol. 3(1), pages 123-177.

Rockoff, Jonah E. and Douglas O. Staiger (2010). "Searching for the Effective Teachers with Imperfect Information." *Journal of Economic Perspectives*, 24(3): 97-118.

Schiff, Martin (1976). "The Educational Failure of Community Control in Inner-City New York". *The Phi Delta Kappan*, Vol. 57, No. 6, pp. 375-378 .

Urquiola, Miguel and Eric Verhoogen (2009). "Class size caps, sorting, and the regression discontinuity design." *American Economic Review*, 99(1), 179-215.

Appendix: Calibrating the Potential Dynamic Impact of Contract Teachers on Teacher Workforce

We follow an approach similar to that of Gordon, Kane, and Staiger (2006), who estimate that in Los Angeles, dropping the bottom quartile of teachers after their first year of teaching would increase the average value added of retained teachers by 1.5 percentile points. In their analysis, this his would be partially offset by the need to increase the flow of new teachers into the system and thus the increased proportion of novice teachers, so the net increase in student test score gains would be 1.2 percentile points per year.

In our context, the potential benefits of a system in which new teachers initially work as contract teachers are far greater, for two reasons. First, the negative novice teacher effect is counteracted by a positive contract teacher effect, as seen in Section 4. Second, the gap in teacher effectiveness appears larger–the gap we observe in our data between the students of the roughly 50% of contract teachers that were hired into the TSC and the students of those not hired is 0.32 standard deviations, even though performance was not the only criterion used by TSC when hiring.[27]

To get a sense of how big the total effect might be, suppose teachers work for 27 years.[28] If new entrants to the profession worked three years as a contract teacher and half were subsequently hired as civil service teachers, then 20% of the teaching force would be made up of contract teachers in steady state. The proportion of novices in the system would double from 1 in 30 to 2 in 30, bringing down average scores by 0.065/30 or 0.0022 standard deviations. Assuming that students of experienced contract teachers score 0.275 standard deviations more than students of their civil-service teacher counterparts (the average of coefficients b_2 and b_4 in Table 2), this effect would boost average test scores by 0.275 x 0.2 − 0.0022 = 0.052 standard deviations among students taught by contract teachers. (Note that this includes the impact of any class size reduction due to the additional contract teachers.) The teacher workforce made up of civil service teachers would have scored, while on contract, 0.32 standard deviations more than their counterparts who were not hired into the civil service. If we assume that only half of this effect persists once they become civil-service teachers,[29]

---

[27] This estimate was obtained through an OLS regression run on the subset of students assigned to the ETP contact teacher. We regress students' endline scores on a dummy equal to 1 if the contract teacher was hired by TSC within 2 years of the program start, controlling for the student and school level controls of Table A2. The coefficient is 0.32 and the standard error is 0.11.

[28] The standard retirement age for civil servants in Kenya is 55 and contract teachers in our sample are 27 years old on average; what's more firing is quasi-inexistent, and voluntary quits among teachers in Kenya are very rare.

[29] The literature on the inter-temporal stability of teacher effectiveness is mostly based on US data. Raw correlations are around 0.3-0.4, adjusted correlation are higher, at 0.6-0.7. See McCaffrey et al. (2009) and Goldhaber and Hansen (2013).

and since 80% of teachers would be TSC teachers, the gain in test scores among students overall from improved quality of the TSC workforce is 0.16 x 0.8 = 0.13 standard deviations. Overall, the gain would thus be approximately 0.052 + 0.13 = 0.182 standard deviations.

The analysis above assumes that the entire gain in student performance documented in section 4 for contract teachers is due to their exerting higher effort in response to stronger incentives and therefore does not persist once they become civil servants. However, it could also reflect positive selection among those initially hired as contract teachers compared to civil servants, which would mean some of the effect would persist, in which case our 0.182 standard deviation estimate above should be interpreted as a lower bound. To obtain an upper bound of the possible effect, if the entire test score gain in section 4 were due to selection rather than incentives, then in steady state rather than only roughly 20% of the teaching workforce generating the test score gain we observe among students of contract teachers, 100% would, increasing the estimated effect by an additional 0.22 standard deviations, for a total gain of approximately 0.4 standard deviation.[30]

Of course we cannot measure all potential channels of dynamic impact. Incentives to become a teacher could potentially either increase or decrease under a system in which teachers initially were hired on a contract basis by local PTAs. As discussed earlier, under the current system, civil-service teacher wages are held far above market clearing levels by the politically powerful teacher union. This motivates many to train as teachers. Jobs are rationed by queuing, so those entering teacher training college currently can expect several years of unemployment before getting a TSC position. Replacing a period of unemployment with a period of contract teaching at low wages could potentially increase the net present value of becoming a teacher. On the other hand, entering teacher training college will be less attractive to the extent prospective teacher training students fear that they will perform badly as contract teachers and therefore will not eventually obtain civil-service positions.[31]

It is also possible that there is complementarity or substitutability between teacher value added and the incentive system, so the gap in performance between teachers under civil-service contracts could be either smaller or larger than under short-term contracts.

Using multiple years of data to reduce the noise coming from variation in students, McCaffrey et al. (2009) estimate within-teacher correlation in value added ranging from 0.5 in elementary grades to 0.8 in middle grades.

[30] Another reason this figure should be considered an upper bound is that it is based on comparing contract teachers to civil-service teachers who teach the lower grades and thus may not be representative of civil-service teachers as a whole.

[31] While in general equilibrium if programs to fund PTAs to hire locally on temporary contracts were expanded to cover all of the schools in the area, the pool of potential contract teachers would be somewhat reduced, this effect is likely small, as two thirds of schools in the area were provided funds to hire contract teachers either through this program and contract teachers in the program were drawn from the local area.

Figure 1: Distribution of Endline test scores, by School and Teacher type

Kolmogorov-Smirnoff test of equality:
- p-value (Control=TSC teacher) = 0.007
- p-value (Control=Contract teacher) = 0.0001
- p-value (TSC teacher=Contract Teacher) = 0.0001
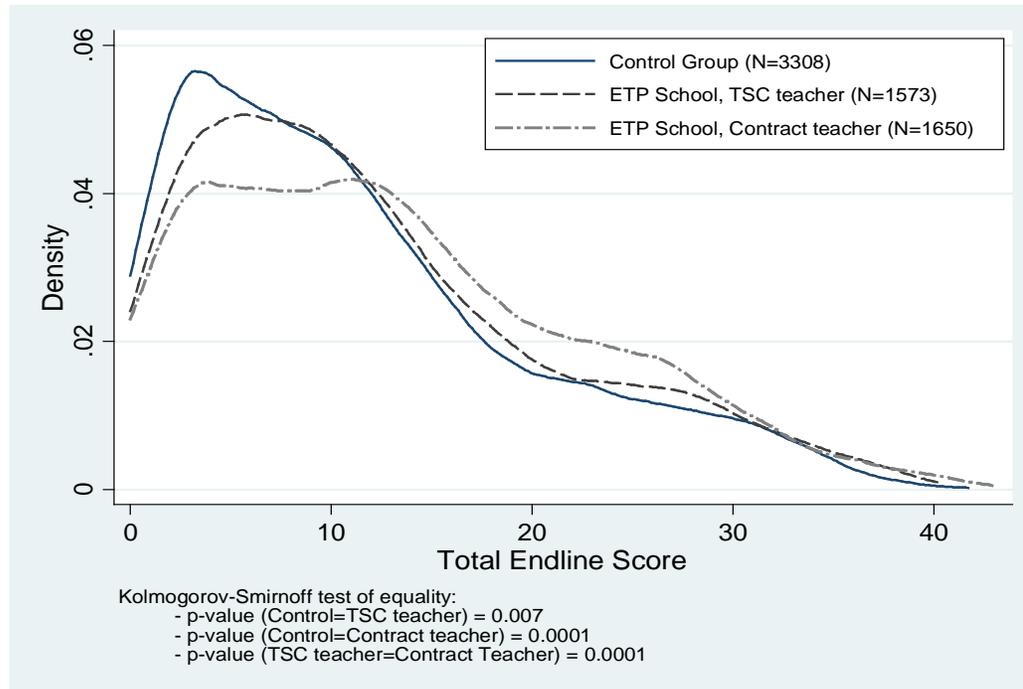
Table 1. Schools, Teachers and Students Characteristics, by Treatment Group, Pre- and Post-Program

	(1) Comparison Schools		(2) ETP Only Schools		(3) ETP + SBM Schools		p-value (1) = (2)	p-value (1) = (3)	p-value (2) = (3)
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.			
Panel A. School Characteristics Prior to Program Inception									
Primary School Exit exam score (out of 400)	260	*29*	258	*26*	260	*34*	0.719	0.930	0.702
Total student enrollment, 2004	598	*241*	646	*266*	574	*221*	0.344	0.635	0.221
Number of TSC teachers, 2004	12	*4*	12	*4*	12	*4*	0.649	0.582	0.928
School-level pupil/teacher ratio, 2004	43	*12*	46	*16*	42	*12*	0.168	0.821	0.167
Average enrollment in grade 1, March 2005	95	*41*	93	*34*	96	*39*	0.807	0.962	0.802
Proportion of female grade 1 students, March 2005	0.51	*0.06*	0.49	*0.06*	0.49	*0.06*	0.203	0.324	0.810
Average enrollment in grade 2, March 2005	97	*43*	98	*37*	99	*42*	0.907	0.762	0.870
Class-based (rather than subject-based) assignment of TSC teachers[a]	0.03	*0.18*	0.03	*0.17*	0.06	*0.25*	0.963	0.481	0.510
Proportion of female teachers among TSC teachers, March 2005	0.71	*0.46*	0.69	*0.46*	0.65	*0.48*	0.803	0.320	0.513
Years of experience among TSC teachers, March 2005	17.04	*9.26*	15.03	*8.97*	16.77	*10.22*	0.076*	0.817	0.190
Average age among TSC teachers, March 2005	43.17	*8.49*	41.84	*8.81*	42.42	*9.45*	0.214	0.489	0.647
Panel B. After Program Inception									
Average class size in grade 1, October 2005	81.5	*27.1*	42.1	*15.3*	44.6	*20.8*	0***	0***	0.664
Class-based (rather than subject-based) assignment of TSC teachers	0.05	*0.22*	0.18	*0.39*	0.19	*0.40*	0.055*	0.047**	0.943
If subject-based assignment: number of TSC teachers for grade 1	3.53	*1.07*	3.63	*1.21*	3.50	*1.03*	0.684	0.922	0.668
Average class size in grade 2, March 2006	67.8	*25.5*	41.4	*17.2*	41.6	*16.3*	0***	0***	0.965
Number of classes in grade 2, March 2006	1.4	*0.6*	2.3	*0.6*	2.3	*0.6*	0***	0***	0.731
Number of classes in grade 1, March 2006	1.0	*0.2*	1.0	*0.0*	1.0	*0.0*	0.250	0.255	1.000

	Comparison Schools		*Within ETP Only and ETP+SBM schools:*				
			Section Assigned to ETP Contract Teacher		Section Assigned to TSC Teachers		p-value (2) = (3)
Panel C. Student and Teacher Characteristics by Grade 1 Sections							
Students' age	7.65	*0.38*	7.68	*0.50*	7.64	*0.48*	0.639
Proportion of female grade 1 students	0.50	*0.07*	0.49	*0.07*	0.49	*0.08*	0.841
Standardized baseline score (Mean 0, SD 1 at school level)	0.01	*0.05*	0.01	*0.11*	0.03	*0.10*	0.355
Proportion of female teachers (Average over program period)	0.70	*0.46*	0.49	*0.50*	0.66	*0.47*	0***
Teachers' Age (Average over program period)	41.88	*9.15*	27.29	*4.19*	42.02	*9.05*	0***
Teachers' Years of experience (Average over program period)	15.72	*9.70*	2.03	*3.98*	15.54	*9.47*	0***

Notes: Data from 70 comparison schools, 36 ETP Only schools and 34 ETP + SBM schools. Standard deviations in italics.

[a] Under "class-based assignment" of teachers, each teacher is assigned a specific class for which she teaches all subjects (math, reading, etc.). Under "subject-based assignment" of teachers, each teacher is assigned a specific subset of subjects and teaches those subjects in multiple classes and/or grades.

Table 2. Endline Test Scores

		(1)	(2)	(3)	(4)	(5)	(6)
		Total Score	Math Score	Literacy Score	Total Score	Math Score	Literacy Score
a_1	Basic ETP	0.189**	0.175**	0.165			
		(0.096)	(0.080)	(0.106)			
a_2	ETP + SBM	0.236**	0.250***	0.177*			
		(0.097)	(0.076)	(0.107)			
b_1	Basic ETP, Assigned to TSC Teacher				0.087	0.051	0.102
					(0.098)	(0.078)	(0.111)
b_2	Basic ETP, Assigned to Contract Teacher				0.294***	0.301***	0.231**
					(0.105)	(0.093)	(0.114)
b_3	ETP + SBM, Assigned to TSC Teacher				0.212*	0.233**	0.152
					(0.110)	(0.089)	(0.119)
b_4	ETP + SBM, Assigned to Contract Teacher				0.263***	0.270***	0.205**
					(0.095)	(0.081)	(0.103)
Region dummies		Yes	Yes	Yes	Yes	Yes	Yes
Linear trend in date of test		Yes	Yes	Yes	Yes	Yes	Yes
Observations		6,533	6,533	6,536	6,533	6,533	6,536
R-squared		0.032	0.027	0.024	0.034	0.031	0.025
p-val $(a_1 = a_2)$		0.656	0.367	0.923			
p-val $(b_1 = b_2)$					0.002***	0.000***	0.067*
p-val $(b_3 = b_4)$					0.46	0.627	0.413
p-val $(b_1 = b_3)$					0.293	0.053*	0.708
p-val $(b_2 = b_4)$					0.779	0.749	0.832

Notes: OLS regressions. The omitted category is the comparison group. The endline test was administered after the program had been in place for five school terms. Scores are normalized such that the mean and standard deviation of the comparison group are zero and one, respectively.

"Basic ETP" is a dummy equal to 1 if the school was sampled for the basic ETP program. The coefficient on this dummy provides the average effect of the basic ETP program on students. "ETP+SBM" is a dummy qual to 1 if the school was sampled for the ETP program with the SBM training. "Basic ETP, Assigned to TSC Teacher" is a dummy equal to 1 if the school was sampled for the Basic ETP program and the student was assigned to the TSC teacher. The coefficient on this dummy provides the average effect of the Basic ETP program on students assigned to TSC teachers.

*Robust standard errors clustered at the school level in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. There are only 139 schools/clusters because tests could not be administered in one of the ETP schools.*

Table 3. Teacher Effort and Pedagogy

		(1)	(2)	(3)	(4)

Panel A. Teacher effort (Source: unannounced spot checks)

		Teacher Found in Class Teaching	Teacher Present at School	If Present at School: Found in Class Teaching	
b_1	Basic ETP, TSC Teacher	-0.157***	-0.017	-0.174***	
		(0.038)	(0.024)	(0.040)	
b_2	Basic ETP, Contract Teacher	0.117**	0.011	0.143***	
		(0.047)	(0.037)	(0.049)	
b_3	ETP + SBM, TSC Teacher	-0.083**	-0.024	-0.079	
		(0.041)	(0.026)	(0.048)	
b_4	ETP + SBM, Contract Teacher	0.206***	0.093***	0.154***	
		(0.046)	(0.026)	(0.042)	
Observations		2,240	2,240	1,880	
Mean in comparison schools		0.579	0.84	0.688	
p-val ($b_1 = b_2$)		0.000***	0.485	0.000***	
p-val ($b_3 = b_4$)		0.000***	0.000***	0.000***	
p-val ($b_1 = b_3$)		0.081*	0.794	0.067*	
p-val ($b_2 = b_4$)		0.106	0.041**	0.842	

Panel B. Pedagogy (Source: Classroom observations)

		Share of Students Who Went to Blackboard to Write an Answer	Share of Students Who Answered a Question	Share of Students Who Asked a Question	
b_1	Basic ETP, TSC Teacher	0.041***	0.152*	-0.036	
		(0.015)	(0.083)	(0.025)	
b_2	Basic ETP, Contract Teacher	0.025	0.127**	-0.027	
		(0.017)	(0.062)	(0.027)	
b_3	ETP + SBM, TSC Teacher	0.037	0.185***	-0.039	
		(0.034)	(0.059)	(0.027)	
b_4	ETP + SBM, Contract Teacher	0.031**	0.084*	-0.043	
		(0.012)	(0.049)	(0.027)	
Observations		169	169	169	
Mean in comparison schools		0.012	0.154	0.044	
p-val ($b_1 = b_2$)		0.474	0.777	0.451	
p-val ($b_3 = b_4$)		0.879	0.123	0.319	
p-val ($b_1 = b_3$)		0.921	0.704	0.79	
p-val ($b_2 = b_4$)		0.764	0.503	0.194	

Panel C. Pedagogy (Source: Teacher Surveys)

		Teacher had a lesson plan prepared for the class	Teacher did not have time to mark individual work for children	Teacher reports not having time to help children individually	Teacher reports not having time to pay attention to slow learners
b_1	Basic ETP, TSC Teacher	-0.001	-0.243*	-0.168	-0.230*
		(0.119)	(0.127)	(0.132)	(0.123)
b_2	Basic ETP, Contract Teacher	0.219**	-0.335***	-0.298**	-0.636***
		(0.102)	(0.121)	(0.122)	(0.089)
b_3	ETP + SBM, TSC Teacher	-0.030	-0.118	-0.212*	-0.280**
		(0.121)	(0.112)	(0.115)	(0.114)
b_4	ETP + SBM, Contract Teacher	0.084	-0.260**	-0.192*	-0.507***
		(0.113)	(0.115)	(0.110)	(0.099)
Observations		166	167	166	166
Mean in comparison schools		0.597	0.742	0.677	0.726
p-val ($b_1 = b_2$)		0.033**	0.45	0.357	0.000***
p-val ($b_3 = b_4$)		0.391	0.277	0.885	0.073*
p-val ($b_1 = b_3$)		0.838	0.395	0.772	0.744
p-val ($b_2 = b_4$)		0.247	0.6	0.446	0.253

Notes: Standard errors clustered at school level. Controls included in all regressions but not shown are: Region dummies and dummies for year, month in the year and day in the week on which the survey was completed.
***, **, * indicate significance at the 1%, 5%, and 10% levels respectively.
Panel A: Linear probability model regressions. Multiple observations per teacher.
Panels B and C: In each school, two or three grade 1 teachers (including the ETP teacher in ETP schools) were selected for classroom observation and for an interview. OLS regressions in Panel B and LPM regressions in Panel C.

Table 4. Contract Teacher Selection

	(1)	(2)	(3)	(4)	(5)
	Characteristics of ETP Contract Teacher hired by the PTA				
	Relative of a teacher in the school	Born or married locally	Female	<1 year of experience	Previously PTA teacher at the school
SBM	-0.15	-0.12	-0.16	0.15	-0.02
	(0.10)	(0.12)	(0.13)	(0.10)	(0.10)
Prop. TSC teachers female	-0.48	-0.25	-0.14	-0.06	-0.01
	(0.15)***	(0.19)	(0.20)	(0.17)	(0.16)
Average experience among TSC teachers	0.02	-0.02	0.00	-0.01	-0.01
	(0.01)*	(0.01)*	(0.01)	(0.01)	(0.01)
School performance on 2004 national exam	-0.02	-0.01	0.02	0.03	-0.02
	(0.06)	(0.08)	(0.08)	(0.07)	(0.07)
School Size (/100)	0.08	0	-0.04	0.02	0.03
	(0.03)**	(0.04)	(0.04)	(0.04)	(0.03)
Number of teachers assigned to lower grades	0.03	0.03	0.02	-0.04	-0.01
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
School had at least one PTA teacher in 2004	-0.17	-0.06	-0.02	-0.26	-0.04
	(0.11)	(0.14)	(0.15)	(0.12)**	(0.12)
Observations	68	68	70	70	68
R-Squared	0.37	0.1	0.09	0.17	0.06
Mean in schools without SBM	0.34	0.74	0.53	0.17	0.2

*Notes: Sample restricted to the 70 schools sampled for the Extra-Teacher Program. Unit of bservation: contract teacher hired by school committee at onset of ETP program. Linear probability model regressions with standard errors presented in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. Region dummies included but not shown.*

Table 5. "Relative" effect: Performance of ETP Contract Teachers related to existing TSC teachers

	(1)	(2)	(3)	(4)
	Student Performance		Teacher Effort	
	Dep. Var: Standardized Endline Score for Students assigned to…		Dep. Var: Teacher Found in Class Teaching	
	ETP Contract Teacher	TSC Teachers	ETP Contract Teacher	TSC Teachers
School based management (SBM)	-0.065	0.094	0.020	-0.034
	(0.131)	(0.136)	(0.067)	(0.041)
(a) ETP contract teacher is the relative of a TSC teacher	-0.339	-0.102	-0.277	-0.136
	(0.175)*	(0.191)	(0.143)*	(0.109)
(b) SBM x ETP contract teacher is the relative of a TSC teacher	0.385	0.220	0.242	0.242
	(0.165)**	(0.177)	(0.131)*	(0.086)***
Proportion of females among TSC teachers	0.141	0.551	-0.019	-0.011
	(0.173)	(0.178)***	(0.098)	(0.069)
Proportion of females among TSC teachers x ETP contract teacher is the relative of a TSC teacher	0.507	0.119	0.237	0.067
	(0.261)*	(0.247)	(0.178)	(0.123)
ETP contract teacher has < 1 year of experience	-0.159		-0.017	
	(0.092)*		(0.076)	
Observations	1548	1631	217	1718
p-val a+b=0	0.770	0.464	0.794	0.148
Mean in schools without SBM	0.236	0.006	0.690	0.526
Unit of Observation	Student	Student	Teacher-Day	Teacher-Day

*Notes: Standard errors in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively.*

Columns 1-2: OLS. Sample restricted to students assigned to ETP contract teachers in ETP schools. Include controls for student age, gender, region and date of tests. Data source: endline test administered by research team.

Columns 3-4: LPM. Sample restricted to ETP contract teachers in ETP schools. Up to 5 observations per teacher. Data source: Five unannounced spot checks. Include controls for region and spot check date as well as the log number of TSC teachers in lower grades.

Table 6. Contract Teacher Retention and Promotion

	(1)	(2)	(3)	(4)
	What happened to the ETP Contract Teacher?			
	Hired by TSC during or shortly after ETP program	Left (not for TSC) and had to be replaced	Employed by school by beginning of year 3	If not hired by TSC: Employed by school by beginning of year 3
SBM	-0.17	0.05	0.21	-0.02
	(0.11)	(0.10)	(0.12)*	(0.20)
Female	0.04	0.16	-0.06	-0.01
	(0.11)	(0.09)*	(0.12)	(0.18)
Years of experience	0.11	0.05	-0.19	-0.16
	(0.06)*	(0.05)	(0.06)***	(0.09)*
Relative of a teacher in the school	-0.07	-0.11	0.20	0.10
	(0.15)	(0.13)	(0.16)	(0.25)
Average presence	0.24	-0.47	-0.12	-0.26
	(0.22)	(0.19)**	(0.24)	(0.35)
Mean score of Contract teacher students at endline	0.35	-0.11	-0.05	0.14
	(0.16)**	(0.14)	(0.18)	(0.33)
Mean score of TSC teacher students at endline	-0.01	-0.19	0.11	0.40
	(0.15)	(0.13)	(0.17)	(0.30)
Prop. TSC teachers female	-0.11	-0.15	0.1	0.07
	(0.19)	(0.16)	(0.21)	(0.32)
Average experience among TSC teachers	0.03	-0.01	-0.02	-0.02
	(0.01)***	(0.01)	(0.01)*	(0.02)
School performance on 2004 national exam	-0.07	0.1	-0.04	-0.04
	(0.07)	(0.06)	(0.08)	(0.12)
School Size (/100)	0.01	0.02	-0.04	-0.04
	(0.04)	(0.03)	(0.04)	(0.06)
Number of teachers assigned to lower grades	0.01	-0.01	0.02	0.03
	(0.03)	(0.03)	(0.04)	(0.05)
School had at least one PTA teacher in 2004	0.13	0.02	0.18	0.32
	(0.13)	(0.11)	(0.14)	(0.21)
Observations	68	68	68	42
R-Squared	0.44	0.38	0.29	0.35
Mean in schools without SBM	0.47	0.19	0.25	0.47

Notes: Sample restricted to the 70 schools sampled for the Extra-Teacher Program. Unit of bservation: contract teacher hired by school committee at onset of ETP program. Linear probability model regressions with standard errors presented in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. Region dummy included but not shown.

Appendix Table A1: Attrition

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Dep. Var:	Attrited at Endline Test (after 19 Months in Program)						Attrited at Long-Run Follow-up Test (a Year after Program Ended)					
	Subsample:	All		Bottom half of baseline distribution		Top half of baseline distribution		All		Bottom half of baseline distribution		Top half of baseline distribution	
a_1	Basic ETP	-0.016		-0.014		0.012		-0.008		0.001		-0.003	
		(0.016)		(0.024)		(0.015)		(0.018)		(0.026)		(0.021)	
a_2	ETP + SBM	-0.030*		-0.030		-0.027		-0.030**		-0.045**		-0.026	
		(0.017)		(0.023)		(0.016)		(0.014)		(0.022)		(0.016)	
b_1	Basic ETP, Assigned to TSC Teacher		-0.001		0.000		0.021		0.007		-0.013		0.028
			(0.019)		(0.030)		(0.021)		(0.021)		(0.031)		(0.027)
b_2	Basic ETP, Assigned to Contract Teacher		-0.031*		-0.029		0.003		-0.023		0.015		-0.034
			(0.018)		(0.027)		(0.018)		(0.020)		(0.027)		(0.024)
b_3	ETP + SBM, Assigned to TSC Teacher		-0.023		-0.046*		-0.000		-0.027		-0.061**		-0.008
			(0.020)		(0.028)		(0.021)		(0.017)		(0.026)		(0.019)
b_4	ETP + SBM, Assigned to Contract Teacher		-0.039**		-0.012		-0.058***		-0.033**		-0.029		-0.046*
			(0.018)		(0.026)		(0.021)		(0.015)		(0.023)		(0.024)
Region dummies		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations		8,009	8,009	3,467	3,467	3,656	3,656	8,009	8,009	3,467	3,467	3,656	3,656
R-squared		0.001	0.002	0.001	0.002	0.002	0.004	0.001	0.001	0.002	0.003	0.001	0.002
p-val ($a_1 = a_2$)		0.352		0.525		0.027**		0.236		0.09*		0.312	
p-val ($b_1 = b_2$)			0.079*		0.367		0.44		0.115		0.261		0.029**
p-val ($b_3 = b_4$)			0.397		0.244		0.025**		0.68		0.145		0.17
p-val ($b2 = b_4$)			0.679		0.587		0.011**		0.629		0.118		0.7

Notes: OLS regressions. The omitted category is the comparison group. The endline test was administered after the program had been in place for five school terms. Scores are normalized such that the mean and standard deviation of the comparison group are zero and one, respectively.

"Basic ETP" is a dummy equal to 1 if the school was sampled for the basic ETP program. The coefficient on this dummy provides the average effect of the basic ETP program on students. "ETP+SBM" is a dummy qual to 1 if the school was sampled for the ETP program with the SBM training. "Basic ETP, Assigned to TSC Teacher" is a dummy equal to 1 if the school was sampled for the Basic ETP program and the student was assigned to the TSC teacher. The coefficient on this dummy provides the average effect of the Basic ETP program on students assigned to TSC teachers.

Robust standard errors clustered at the school level in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. There are only 139 schools/clusters because tests could not be administered in one of the ETP schools.

Appendix Table A2. Robustness check: Impacts on Endline Test Scores, controlling for school-level and individual-level controls

		(1)	(2)	(3)	(4)	(5)	(6)
		Total Score	Math Score	Literacy Score	Total Score	Math Score	Literacy Score
a_1	Basic ETP	0.184**	0.135*	0.191**			
		(0.088)	(0.075)	(0.095)			
a_2	ETP + SBM	0.226**	0.207***	0.198**			
		(0.092)	(0.076)	(0.100)			
b_1	Basic ETP, Assigned to TSC Teacher				0.073	0.014	0.111
					(0.088)	(0.073)	(0.096)
b_2	Basic ETP, Assigned to Contract Teacher				0.315***	0.274***	0.288***
					(0.103)	(0.090)	(0.109)
b_3	ETP + SBM, Assigned to TSC Teacher				0.204**	0.201**	0.166
					(0.096)	(0.082)	(0.103)
b_4	ETP + SBM, Assigned to Contract Teacher				0.276***	0.237***	0.256**
					(0.100)	(0.087)	(0.108)
	School size (/100)	-0.059	-0.208***	0.083	-0.051	-0.201***	0.089
		(0.088)	(0.074)	(0.096)	(0.088)	(0.075)	(0.096)
	Share female among TSC teachers	0.499***	0.308***	0.569***	0.497***	0.306***	0.567***
		(0.104)	(0.093)	(0.107)	(0.104)	(0.093)	(0.107)
	Average years of experience (/10) among TSC teach	0.018	-0.021	0.047	0.020	-0.018	0.049
		(0.067)	(0.057)	(0.070)	(0.068)	(0.058)	(0.070)
	Girl	0.057**	-0.012	0.106***	0.056**	-0.013	0.106***
		(0.023)	(0.021)	(0.026)	(0.023)	(0.021)	(0.026)
	Age	-0.042***	-0.007	-0.065***	-0.043***	-0.007	-0.065***
		(0.012)	(0.010)	(0.013)	(0.012)	(0.011)	(0.013)
	Baseline score	0.498***	0.496***	0.401***	0.498***	0.496***	0.401***
		(0.018)	(0.015)	(0.022)	(0.018)	(0.015)	(0.022)
	Teacher has <1 year of experience	0.022	0.067	-0.024	-0.065	-0.012	-0.100
		(0.081)	(0.078)	(0.096)	(0.091)	(0.085)	(0.107)
Region dummies		Yes	Yes	Yes	Yes	Yes	Yes
Linear trend in date of test		Yes	Yes	Yes	Yes	Yes	Yes
Observations		6,533	6,533	6,536	6,533	6,533	6,536
R-squared		0.248	0.242	0.181	0.251	0.246	0.183
p-val ($a_1 = a_2$)		0.667	0.362	0.945			
p-val ($b_1 = b_2$)					0.000***	0.000***	0.011**
p-val ($b_3 = b_4$)					0.259	0.624	0.15
p-val ($b_1 = b_3$)					0.22	0.033**	0.637
p-val ($b_2 = b_4$)					0.719	0.692	0.779

Notes: see Table 2.

41

Appendix Table A3. Lee (2009) Bounds for Treatment Effects at Endline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Total score				Math score				Literacy Score			
	Lower bound	Upper bound	N before trimming	Trimming proportion[τ]	Lower bound	Upper bound	N before trimming	Trimming proportion[τ]	Lower bound	Upper bound	N before trimming	Trimming proportion[τ]
Difference betweeen treatment group [...] and control group												
a Basic ETP, Assigned to TSC Teacher	-0.022	0.054	4143	0.004	-0.043	0.143	4143	0.004	0.045	0.058	4145	0.003
	(0.108)	(0.106)			(0.099)	(0.098)			(0.12)	(0.118)		
b Basic ETP, Assigned to Contract Teacher	0.124	0.304**	4129	0.038	0.188*	0.418***	4129	0.038	0.07	0.218*	4131	0.038
	(0.114)	(0.121)			(0.103)	(0.099)			(0.126)	(0.119)		
c ETP + SBM, Assigned to TSC Teacher	0.087	0.211*	4123	0.029	0.14	0.347***	4123	0.029	0.012	0.135	4125	0.029
	(0.121)	(0.111)			(0.104)	(0.087)			(0.13)	(0.123)		
d ETP + SBM, Assigned to Contract Teacher	0.112	0.29***	4060	0.045	0.173*	0.39***	4060	0.045	0.049	0.218**	4063	0.046
	(0.103)	(0.102)			(0.101)	(0.085)			(0.114)	(0.107)		
Difference betweeen...												
e Contract teacher (b) and TSC teacher (a) in Basic ETP	0.107	0.256***	1656	0.035	0.172**	0.41***	1656	0.035	0.019	0.162**	1656	0.035
	(0.085)	(0.073)			(0.077)	(0.066)			(0.091)	(0.074)		
f Contract teacher (d) and TSC teacher (c) in ETP+SBM	-0.002	0.097	1567	0.016	-0.003	0.198**	1567	0.016	-0.019	0.075	1568	0.018
	(0.086)	(0.072)			(0.09)	(0.086)			(0.087)	(0.07)		

Notes: Bounds à la Lee (2009) obtained by trimming upper or lower tail in treatment group (see text for details). Baseline covariates (gender and age below median) are used to narrow the bounds, as explained in Lee (2009). Standard errors clustered at the school level in parentheses. Standard errors are calculated following Lee (2009), that is to take into account the trimmed estimate. Scores are normalized such that the mean and standard deviation of the comparison group are zero and one, respectively. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. There are only 139 schools/clusters because tests could not be administered in one of the ETP schools.

τ : The trimming proportion correspond to the the proportion of trimmed observation in the Treatment group (or the group with less attrition for rows e and f).

Appendix Table A4. Test Scores at long-run follow-up

		(1)	(2)	(3)	(4)	(5)	(6)
		Total Score	Math Score	Literacy Score	Total Score	Math Score	Literacy Score
a_1	Basic ETP	-0.013	-0.016	-0.010			
		(0.080)	(0.069)	(0.089)			
a_2	ETP + SBM	0.099	0.104	0.079			
		(0.097)	(0.074)	(0.105)			
b_1	Basic ETP, Assigned to TSC Teacher				-0.062	-0.059	-0.054
					(0.089)	(0.079)	(0.095)
b_2	Basic ETP, Assigned to Contract Teacher				0.039	0.029	0.038
					(0.084)	(0.071)	(0.093)
b_3	ETP + SBM, Assigned to TSC Teacher				0.113	0.128	0.084
					(0.109)	(0.084)	(0.118)
b_4	ETP + SBM, Assigned to Contract Teacher				0.085	0.080	0.075
					(0.092)	(0.074)	(0.100)
Region dummies		Yes	Yes	Yes	Yes	Yes	Yes
Linear trend in date of test		Yes	Yes	Yes	Yes	Yes	Yes
Observations		6,171	6,171	6,175	6,171	6,171	6,175
R-squared		0.004	0.003	0.004	0.005	0.003	0.005
$p\text{-val}\,(a_1 = a_2)$		0.239	0.109	0.41			
$p\text{-val}\,(b_1 = b_2)$					0.105	0.139	0.13
$p\text{-val}\,(b_3 = b_4)$					0.646	0.417	0.886
$p\text{-val}\,(b_1 = b_3)$					0.127	0.046**	0.271
$p\text{-val}\,(b_2 = b_4)$					0.618	0.509	0.724

Notes: OLS regressions. The omitted category is the comparison group. The long-run test was administered one year after the 18-month long program had ended. Scores are normalized such that the mean and standard deviation of the comparison group are zero and one, respectively.

"Basic ETP" is a dummy equal to 1 if the school was sampled for the basic ETP program. The coefficient on this dummy provides the average effect of the basic ETP program on students. "ETP+SBM" is a dummy qual to 1 if the school was sampled for the ETP program with the SBM training. "Basic ETP, Assigned to TSC Teacher" is a dummy equal to 1 if the school was sampled for the Basic ETP program and the student was assigned to the TSC teacher. The coefficient on this dummy provides the average effect of the Basic ETP program on students assigned to TSC teachers.

Robust standard errors clustered at the school level in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. There are only 139 schools/clusters because tests could not be administered in one of the ETP schools.

Appendix Table A5. ETP Hiring Procedures: Results from post-hire survey with Headmaster and PTA members

	(1)	(2)	(3)	(4)	(5)
	Hiring Procedures				
	Ad for ETP position was posted at Area Education Office (AEO)	Parents asked to spread information about ETP position	Local chief asked to spread information about ETP position	Number of applications received	Number of candidates interviewed
SBM	0.05	-0.09	0.15	0.34	0.50
	(0.04)	(0.10)	(0.21)	(0.33)	(0.25)**
Bungoma District	0.00	0.07	-0.03	0.65	0.60
	(0.05)	(0.14)	(0.28)	(0.49)	(0.35)*
School performance on 2004 national exam	0.00	-0.03	-0.08	-0.30	-0.33
	(0.02)	(0.06)	(0.12)	(0.21)	(0.15)**
Prop. TSC teachers female	0.09	0.16	-0.08	-0.29	-0.27
	(0.06)	(0.18)	(0.36)	(0.58)	(0.43)
Average experience among TSC teachers	0.00	-0.01	-0.01	-0.03	0.00
	(0.00)	(0.01)	(0.02)	(0.03)	(0.02)
School Size (/100)	-0.01	0.00	0.01	-0.07	0.12
	(0.01)	(0.04)	(0.07)	(0.13)	(0.09)
Number of teachers assigned to lower grades	-0.01	0.02	-0.05	0.09	0.04
	(0.01)	(0.03)	(0.07)	(0.11)	(0.08)
School had at least one PTA teacher in 2004	0.06	0.19	-0.32	0.00	0.08
	(0.04)	(0.12)	(0.24)	(0.40)	(0.29)
Observations	138	138	138	125	126
R-Squared	0.09	0.05	0.03	0.08	0.11
Mean in schools without SBM	0.01	0.84	0.53	2.75	1.92

Notes: Sample includes non-tracking and tracking schools. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively.