



School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools[☆]



Esther Duflo^{a,b,c,d}, Pascaline Dupas^{b,c,d,e,*}, Michael Kremer^{b,c,f,g,h}

^a Massachusetts Institute of Technology, United States

^b BREAD, United States

^c NBER, United States

^d CEPR, United Kingdom

^e Stanford University, United States

^f Harvard University, United States

^g Brookings, United States

^h CGD, United States

ARTICLE INFO

Article history:

Received 31 March 2014

Received in revised form 18 September 2014

Accepted 24 November 2014

Available online 9 December 2014

JEL classification:

I21

O15

M51

D71

Keywords:

Contract teachers

PTA

School-based management

Teacher effort

Test score gain

Nepotism

Corruption

Decentralization

Institutions

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 school committees at randomly selected Kenyan 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 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 school committees 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.

© 2014 Published by Elsevier B.V.

[☆] We thank Josh Angrist, Paul Glewwe, Alaka Holla, Victor Lavy, Craig McIntosh, Naercio Menezes-Filho, Karthik Muralidharan, and seminar participants at the Harris School, Harvard, UC Santa Barbara, the World Bank, Dartmouth College, NEUDC, EGAP, and Itau International Seminar for their helpful comments and discussions. We thank the Kenya Ministry of Education, International Child Support Africa, and Matthew Jukes for their collaboration. We thank Willa Friedman, Jessica Morgan, Nicolas Studer, Ian Tomb, Victor Pouliquen and Paul Wang for their excellent research assistance. We are grateful to Grace Makana and her field team for collecting the data. We thank the World Bank and the Government of the Netherlands for the funding (Bank Netherlands Partnership Program (BNPP) grants 7135425 and 7135426) that made this study possible. The research protocol was approved by the following IRBs: MIT, Harvard, ICS and IPA Kenya.

* Corresponding author at: 579 Serra Mall, Stanford CA 94305.

E-mail addresses: eduflo@mit.edu (E. Duflo), pdupas@stanford.edu (P. Dupas), mkremer@fas.harvard.edu (M. Kremer).

1. Introduction

Despite rapid expansion in school participation over the last decade or so, learning achievements remain poor in developing countries, and students score very low on internationally comparable tests (Hanushek and Woessmann, 2010). Some policymakers seeking to address this problem focus on providing more resources (such as teachers or materials), while others focus on systemic reforms designed to improve governance and provider incentives.

In this paper, we examine how these two approaches can complement each other. In particular, we demonstrate that the impacts of centrally provided resources (in our setting, extra teachers) are muted by

the endogenous response of local officials who misuse them.¹ However, strengthening the power of parents vis-à-vis those officials reduces misappropriation and thereby improves the effectiveness of extra resources.

Specifically, we consider two policies that are very popular in the education sector in developing countries: the hiring of teachers on short-term contracts (subject to renewal by the community), and School Based Management training programs, designed to empower local communities (in particular, parents) in how school resources are used and monitored. Both of these programs are widespread in developing countries, but have seldom been evaluated. Contract teachers, hired sometimes centrally and sometimes locally, have been used in a dozen countries in sub-Saharan Africa, South Asia and Latin America over the past decade. Muralidharan and Sundararaman (2013) report that the share of contract teachers among public school teachers in India grew from 6% in 2003 to 30% in 2010; they also reference a study by Bourdon et al. (2010) estimating this share to be right around a third across a dozen of countries in sub-Saharan Africa. School-Based Management (SBM) programs have likewise been implemented in numerous countries since the mid-2000s.² SBM programs primarily consist in transferring decision-making authority over some school operations to local school community members. Some of the programs have focused on empowering headmasters and teachers, in particular with respect to spending (e.g. SBM reform in The Gambia came hand in hand with a school grant program, see Blimpo et al., 2011), but most SBM programs have a strong component aimed at strengthening parental involvement in the monitoring of the school, typically through the establishment or reinforcement of school committees. The SBM program we study falls in that category, but is on the “light” side in terms of its time and financial costs.

We set up an experiment to study how these two reforms can complement each other in the context of a randomized evaluation of the Extra Teacher Program (ETP) implemented in Kenya between 2005 and 2007. The ETP program provided 70 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. The ETP program followed a well-established practice, as prior to Kenya’s abolition of school fees in 2003, many of these committees used funds levied from parents to hire and pay extra teachers on short-term and low-pay contracts to supplement the civil-service teachers centrally assigned to schools. The contracts provided under the ETP program 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.

In a (randomly selected) subset of schools participating in the ETP program, members of the school committee and interested parents

received a short School-Based Management (SBM) empowerment training (90 min) on how to interview and select job applicants, monitor and assess teachers’ effort and performance, monitor the school’s management of the program, and perform a formal review of the contract teacher’s performance to decide whether to renew her contract.

Several features of the experimental design allow us to shed light on the interaction between extra teachers and governance. First, within schools sampled for the program, first-grade 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. This enables us to estimate the impact of the ETP program, alone or combined with SBM, separately for students assigned to civil-service teachers and those assigned to contract teachers. We find large differences across these groups.

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 (i.e. assigned to the civil-service teachers). This could be because a class size of 44 is still too big for much learning to take place. Indeed, class size experiments ran in the US in the early 20s had found little effect of class size reductions at that scale (Rockoff, 2009). However, our results suggest that an endogenous effort response by the teachers could also be part of 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. In contemporaneous work in India, Muralidharan and Sundararaman (2013) also find that teachers reduce effort in response to hiring additional teachers.

In contrast, the locally hired and monitored teachers exerted much greater effort, presumably because they faced stronger 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. In turn, students assigned to them also performed better at endline than those assigned to the civil service teachers, although this does not uniquely identify an incentive effect, since this could also be due to their different characteristics (in particular, they had much less experience) and the fact that they taught a single group of children.

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.

However, capture by teachers and their network was reduced by promoting parental involvement and control at the local level. Point estimates suggest that the SBM empowerment program cut by half both the reduction in the regular teacher effort in response to the program 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 et al., 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

¹ Das et al. (2013) find similarly that the impact of centrally provided resources (in their case school inputs) is mediated by parents’ response (who buy less inputs in response to the program). More generally, Alatas et al. (2013) document elite capture of a government program by officials (but not by informal elite) though they show that the welfare effects are modest.

² The list of countries that have experienced with contract teachers include 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). SBM programs have been conducted in Guatemala, Honduras, India, Lebanon, Lesotho, Macedonia, Madagascar, Mexico, Nicaragua, the Philippines, Senegal, Serbia, Sri Lanka, and The Gambia (see Bruns et al. (2011) for a review).

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 (1895 to 1964), churches and local communities started many independent schools. While the Harambee movement³ in the late 60s and 70s encouraged local communities to fundraise locally to start schools and retain representation of parents, the local community, and institutional sponsors (in particular the local 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 the long queue of qualified teachers waiting to be hired reflect the fact that they receive a packet of wages and benefits considerably above most private sector options for college graduates. 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, as 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

³ As per Wilson (1992), "The term Harambee has come to mean the provision of goods, usually social infrastructure, through the voluntary cooperation of members of the community."

than to press for increased teacher accountability at the national level.

PTA teachers are hired by school committees at each school, using funds raised from all parents. While all parents are formally members of the PTA, school committees are much smaller, and 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 school committee. 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, but school committee members may have other incentives (such as hiring relatives). Kremer et al. (2003) argue that the system by which school committee members are chosen may have 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, the few PTA teachers we found remaining in schools, who are typically just as qualified as the regular teachers, received compensation in the range of 2000 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 in 2003. Enrollment rose nearly 30% after the abolition of school fees. 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% had no locally-hired teachers. As a result of increased enrollment and decreased number of teachers, average class size in first grade in 2005 was 83, and median class size was 74; 28% of grade 1 classes had more than 100 pupils.⁴

3. The extra teacher program and study design

3.1. Program description

3.1.1. Extra Teacher Program

The Extra Teacher Program (ETP) was implemented in 2005 and 2006 in Kenya's Western Province, and provided funds to school committees to hire one extra teacher, a "contract teacher", to supplement TSC teachers teaching in lower grades. The program was

⁴ 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).

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. 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.⁵

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.

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.⁶

The monthly allowance for contract teachers was 2500 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.

3.1.2. 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 the school's implementation of the ETP program. The training lasted about 90 min, and immediately

followed the meeting that ICS held with parents to inform them of the Extra Teacher Program. 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.⁷ Two parents on the school committee 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. They were also put in charge of checking that the extra teachers focused on the duties as per their contract (i.e. teaching one section of first grade in the first year) and were not put to other uses in the school (i.e. replacing other teachers). Finally, in SBM schools the school committees were told that they should plan 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. All SBM schools were visited by an ICS officer around that time and reminded to schedule a meeting, which was to the extent possible attended by an ICS officer. Finally, all parents were informed that the school committee was receiving SBM training.⁸

3.1.3. 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 different 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 (these schools were already involved with ICS for another randomized controlled trial (RCT) concerning HIV prevention among upper grades). The 210 schools were randomly divided into a comparison group (70 schools), and an Extra Teacher Program (ETP) group (140 schools). Of these 140 schools, 70 were randomly assigned to the tracking program and are not the focus of this paper. Finally, thirty four of the 70 non-tracking ETP schools (and 36 of the 70 tracking schools) were randomly selected to participate in SBM training.

The non-evenness in the fraction sampled for SBM is by random chance due to the stratification by district. At the time we randomized the 210 schools into ETP vs. control, we stratified by administrative division (7 in total), by whether the school was above or below the sample median in terms of its performance at the 2001 primary school exit exam (the KCPE), and by the treatment status (4 in total) of the school in the prior RCT that ICS had conducted with

⁷ This training took place before students had been randomized into the class taught by the contract teacher or a regular class.

⁸ Specifically, grade 1 parents in SBM schools were told during the general ETP introduction meeting: "The program will be on for Terms 2 and 3 of this year, and again next year. This is a great opportunity for the children to benefit from smaller class size. For the pupils to fully benefit, it is very crucial, however, that the ETP teacher you hire is committed, hard-working and comes to school every day. To help your school choose the best candidate, we will have a small training session with the School Committee members at the end of this meeting. We will also train the School Committee members on how to monitor the teacher. But you, as parents, also have a role to play. If you feel like the ETP teacher is not dedicated enough, for instance if he is absent too often, you should talk to the School Committee members and ask for the ETP Teacher to be replaced. At the end of the year, we will organize a review meeting where the school community will assess the performance of the ETP teacher and decide whether to renew the contract or look for a new teacher."

⁵ 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.

⁶ 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.

the sample of schools. Thus we had $7 \times 2 \times 4 = 48$ strata with 4–5 schools each (not all administrative divisions have the same size). To randomize the 140 ETP schools into tracking vs. non-tracking, we stratified the schools by administrative district (2 in total), by whether the school was above or below the sample median 2001 KCPE, and by whether the school was above or below the sample median in terms of grade 1 enrollment in 2004. Finally, to randomize the 140 ETP schools into SBM or no SBM, we stratified the schools by administrative district (2 in total), by whether the school was above or below the sample median 2001 KCPE, and by whether the school had been sampled for tracking or non-tracking.

ICS held school meetings to explain the program to parents and teachers in March 2005. These meetings were held in all schools, including those sampled for the control group, as parental consent to test their children had to be sought from all schools. All schools offered the opportunity to participate in the ETP and SBM programs chose to do so. After they had hired a teacher, schools were informed that students would be randomly assigned to join the contract teacher section. All schools agreed. All but two of the 70 schools selected for the non-tracking Extra Teacher Program 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 42–44 in ETP schools.⁹ 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 41 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).¹⁰

Interestingly, a number of ETP schools decided that the students not assigned to the contract teacher would also 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. Although this is not random, we note that we see no statistically significant difference in test scores of students taught by a single TSC teacher and students taught by rotating TSC teachers.

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. School-maintained data on students' initial grades was collected from the school records at the onset of the study. This baseline performance data cannot be used to compare absolute performance between two students in different schools, since each school may have its own level of difficulty or grading scale, but it can be used to compare students within school. We can thus check that the random assignment of grade 1 students to sections within a given school indeed generated balanced sections — and we find that this is the case (Panel C

of Table 1). We will also use this data later to estimate the treatment effects separately for those performing really well or really poorly within their school, as well as to compare the characteristics of attritors and non-attritors (and statistically control for any imbalance among them). Note that there is a clear correlation between baseline position in one's class according to these grades and our endline test (0.46), which suggests that the baseline grade data has informational content.

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

3.4.1. Student participation, teacher effort and pedagogy

Over the course of the program, five unannounced visits were made to each school by the research team. Student presence in class was measured at all of them, and teacher effort was measured at four of the five visits. Teacher effort was measured by the teacher's presence in school and by whether she was in any classroom teaching (not just grade 1) when the observers entered the school compound.

During these unannounced school visits, we also verified whether the students were in their assigned section. Students typically attended their assigned section when both teachers were present, but when teachers were absent, sections were sometimes combined. On average across all visits, 94% of students were found in their assigned section, conditional on the two sections being taught separately. Sections were found pooled in 10.7% of visits.

Data on pedagogy was collected through classroom observations and structured interviews with teachers. The pedagogy data was collected in the first term of 2006 (year 2 of the program), and focused on grade 2 teaching. In each school, one TSC teacher was chosen to be surveyed and observed (while teaching grade 2 students). In ETP schools, the ETP contract teacher was also surveyed and observed while teaching grade 2 students.¹¹

3.4.2. 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 8000 students across all schools.

⁹ 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.

¹⁰ 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.

¹¹ The TSC teacher was selected as follows. Among grade 2 classes not taught by the ETP teacher, one class was selected at random. We then attempted to observe the TSC teacher teaching English to that class. If that teacher was not available or not teaching at the time of the visit, we tried to observe the Math, Swahili and Science teachers, in decreasing order of priority. Sometimes a teacher teaches multiple topics. The distribution of subjects observed in class was English 30%, Math 41%, Swahili 19%, and this is comparable between ETP and TSC teachers.

Table 1
Schools, teachers and students characteristics, by treatment group, pre- and post-programs.

	(1)		(2)		(3)		p-Value (1) = (2)	p-Value (1) = (3)	p-Value (2) = (3)
	Comparison schools		ETP only schools		ETP + SBM schools				
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.			
<i>Panel A. School characteristics prior to program inception</i>									
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
<i>Panel B. After program inception</i>									
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
Within ETP only and ETP + SBM schools:									
	Comparison schools		Section assigned to ETP contract teacher		Section assigned to TSC teachers		p-Value (2) = (3)		
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.			
<i>Panel C. Student and teacher characteristics by grade 1 sections</i>									
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) ^b	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 are presented in italics.

*** 1%.

** 5%.

* 10%.

^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.

^b Baseline scores come from tests administered by schools prior to the program and are not comparable across schools; they are normalized such that the mean score in each school is zero and the standard deviation is one.

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% for the endline test, and 23% for the long-run follow-up test.¹² While this level of attrition is higher than what we would have wished, it is in the common range for education studies in developing countries (attrition was 20.4% in year 2 in Muralidharan and Sundararaman (2013) and 22% after only a year in Duflo et al. (2012)). We discuss the patterns of attrition and their implications for the estimation of the treatment effects in the Results section.

3.4.3. 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 current teacher or born locally), and the hiring process. Unfortunately this

data is missing for two ETP teachers, thus tables using this data have only 68 observations instead of 70. 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 ETP program after the ICS subsidy ended.

4. Results

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

¹² School choice is not a concern in our context. We observe quasi in-existent 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.

form equation to estimate the effects of each school-level program on student outcomes:

$$Y_{ij} = a_1 \text{BasicETP}_j + a_2 \text{ETPwSBM}_j + X_{ij}'a_5 + \varepsilon_{ij1} \quad (1)$$

where Y_{ij} is the outcome of student i in school j (when the outcome is a test score, it is expressed in standard deviations of the distribution of scores in the comparison schools),¹³ BasicETP_j is a dummy equal to 1 if school j participated in the Basic Extra Teacher Program only, and 0 otherwise, 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, including age, gender, and baseline score.

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:

$$Y_{ij} = b_1 \text{BasicETP_TSC}_{i,j} + b_2 \text{BasicETP_CT}_{i,j} + b_3 \text{ETPwSBM_TSC}_{i,j} + b_4 \text{ETPwSBM_CT}_{i,j} + X_{ij}'a_5 + \varepsilon_{ij2} \quad (2)$$

where $\text{BasicETP_TSC}_{i,j}$ is a dummy equal to 1 if student i was assigned to a class taught by TSC teachers in a school that received basic ETP, $\text{BasicETP_CT}_{i,j}$ is a dummy equal to 1 if student i was assigned to the class taught by the contract teacher in a school that received basic ETP, $\text{ETPwSBM_TSC}_{i,j}$ is a dummy equal to 1 if student i was assigned to a class taught by TSC teachers in a school that received ETP with SBM, and finally $\text{ETPwSBM_CT}_{i,j}$ is a dummy equal to 1 if student i was assigned to the class taught by the contract teacher in a school that received ETP with SBM.

We present results on student attendance, attrition and endline test scores in Table 2. Panel A presents estimations of Eq. (1) and Panel B presents estimations of Eq. (2).

In column 1 of Table 2 Panel A, a_1 is the average effect on student attendance of being in a school sampled for the basic ETP program. Compared to the mean attendance rate in the comparison schools (85.6%, shown at the bottom of the panel), the estimated effect of the basic ETP program is very small and insignificant (0.002). The ETP + SBM program (coefficient a_2) has a somewhat larger, positive effect on attendance (+2.1 percentage points, p-value = 0.098). When we breakdown the effect by teacher type in panel B, we find that only students of contract teachers in SBM schools were significantly more likely to attend. As will be discussed in Table 3 below, this mirrors the higher effort rate we observed among contract teachers in SBM schools.

The next column in Table 2 presents estimates of attrition at the endline test. Attrition was lower in SBM schools, consistent with the higher attendance rate observed in those schools. Attrition is also lower among students of contract teachers, which is again consistent with the attendance pattern of students, and with the higher effort observed among contract teachers, which we will discuss later. To the extent that attrition is driven at least in part by low school attachment (students who have either completely dropped out or are absent often), the reduction in attrition is itself an important outcome, and indication that those two programs were successful. However, they potentially confound the test score results.

To investigate this pattern more deeply, Table A1 shows the attrition results separately for students initially in the bottom and top half of the test scores distribution. Overall, the attrition pattern in ETP and ETP with SBM schools is similar for high and low achievement student. However, low achievement students assigned to TSC teachers in SBM schools were less likely to attrit than those in the control group, and

high achievement students assigned to the contract teacher in SBM schools were less likely to attrit than those in the control group. This may bias any estimated effect of being assigned to the contract teacher in SBM schools upwards and any effect of being assigned to the TSC teachers in SBM schools downwards. To address this problem, we do two things: First, we implement DiNardo, Fortin and Lemieux's (hereafter DFL) reweighting scheme, using the baseline data to group students by quintile in their school, and reweighting (within each region) so that the distribution of students baseline scores those who took the endline test resembles that of all students (DiNardo et al., 1996). These results are presented in the last three columns of Table 2. Second, we present non-parametric Manski-Lee bounds (Lee, 2008), in Table A4.

In Eq. (1) estimated in columns 3–5 of Table 2 Panel A, a_1 is the average effect on test scores of being in a school sampled for the basic ETP program. The effect on test scores of the basic ETP program is 0.14, not quite significant, with a standard error of 0.098 (p-value – 0.112). The coefficient estimate a_2 is the average effect of being in a school sampled for the ETP with SBM program. The effect is larger, an increase of 0.196 standard deviation, significant at the 5% level. We cannot reject the null hypothesis 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 Panel A, at 0.63). The gains were stronger in math than literacy (columns 4 and 5 of Table 2). The results are robust to including a linear trend in test scores date, shown in Table A2 (there is a small imbalance in test score dates by group) and also robust to including school-level and student-level controls, shown in Table A3.

Fig. 1 and panel B of Table 2 break down the program effects by teacher type. Focusing first on the TSC vs. contract teachers split, Fig. 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 the ETP contract teacher 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 panel B 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 relatively small (0.043 of a standard deviation) and insignificantly different from zero (standard error: 0.102). In contrast, the effect of the basic ETP program on students assigned to the contract teacher is three times as large, at 0.244

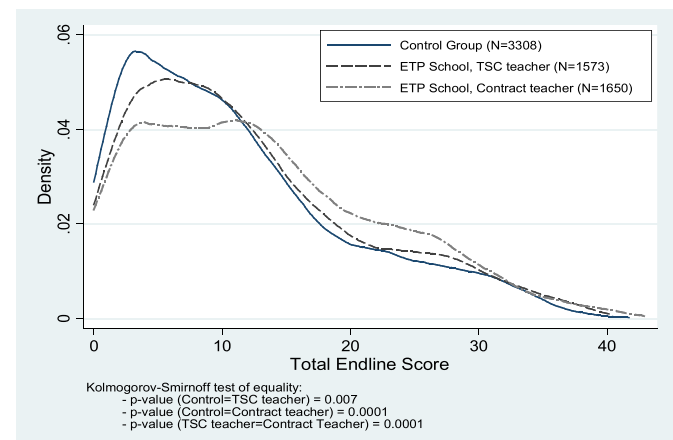


Fig. 1. Distribution of endline test scores, by school and teacher types.

¹³ 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.

Table 2
Student attendance, endline attrition and endline test scores.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Found in class during unannounced spot checks	Attrited at endline test	Total score	Math score	Literacy score	Using weights to correct for differential attrition by quintile of initial distribution		
							Total score	Math score	Literacy score
<i>Panel A</i>									
a_1	Basic ETP	0.002 (0.013)	-0.015 (0.016)	0.142 (0.098)	0.133 (0.083)	0.123 (0.106)	0.146 (0.097)	0.136 (0.082)	0.127 (0.105)
a_2	ETP + SBM	0.021* (0.012)	-0.031* (0.017)	0.196** (0.098)	0.214*** (0.078)	0.141 (0.107)	0.195** (0.098)	0.213*** (0.078)	0.140 (0.106)
	Observations	52,622	8,012	6,533	6,533	6,536	6,533	6,533	6,536
	Mean in comparison schools	0.856	0.196	0.00	0.00	0.00	0.00	0.00	0.00
	p-Val ($a_1 = a_2$)	0.172	0.34	0.63	0.346	0.887	0.66	0.366	0.916
<i>Panel B</i>									
b_1	Basic ETP, assigned to TSC teacher	-0.006 (0.013)	-0.001 (0.019)	0.043 (0.102)	0.012 (0.081)	0.063 (0.112)	0.047 (0.101)	0.014 (0.081)	0.067 (0.112)
b_2	Basic ETP, assigned to contract teacher	0.011 (0.013)	-0.031* (0.018)	0.244** (0.107)	0.256*** (0.096)	0.185 (0.113)	0.249** (0.105)	0.260*** (0.095)	0.190* (0.112)
b_3	ETP + SBM, assigned to TSC teacher	0.014 (0.013)	-0.023 (0.020)	0.175 (0.111)	0.200** (0.090)	0.119 (0.119)	0.186* (0.111)	0.211** (0.090)	0.128 (0.120)
b_4	ETP + SBM, assigned to contract teacher	0.027** (0.013)	-0.039** (0.018)	0.220** (0.097)	0.231*** (0.083)	0.166 (0.102)	0.204** (0.097)	0.215** (0.083)	0.154 (0.102)
	Observations	52,622	8012	6533	6533	6536	6531	6531	6534
	p-Val ($b_1 = b_2$)	0.021**	0.079*	0.003***	0.000***	0.094*	0.003***	0.000***	0.093*
	p-Val ($b_3 = b_4$)	0.066*	0.397	0.526	0.685	0.479	0.802	0.962	0.703
	p-Val ($b_1 = b_3$)	0.16	0.309	0.294	0.051*	0.691	0.266	0.039**	0.665
	p-Val ($b_2 = b_4$)	0.274	0.679	0.838	0.805	0.879	0.7	0.657	0.772

Notes: column 1: linear probability model regressions. Data comes from five unannounced spot checks performed by the research team (two in 2005 and three in 2006). Students identified as dropouts are coded as absent. Those reported to have moved to another school are excluded from the analysis. Columns 2–8: OLS regressions. The endline test was administered after the program had been in place for five school terms. In each school, 60 students were randomly drawn from the baseline sample to participate in the tests. Scores are normalized such that the mean and standard deviation of the comparison group are zero and one, respectively. Columns 6–8 reweigh the observations by the inverse of the non-attrition rate in their quintile of the initial performance distribution.

In both panels the omitted category is the comparison group. Panel A: “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 equal to 1 if the school was sampled for the ETP program with the SBM training. Panel B: “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.

All regressions control for region dummies. Robust standard errors are clustered at the school level in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. Columns 2–8: there are only 139 schools/clusters because tests could not be administered in one of the ETP schools.

standard deviation (coefficient b_2 , standard error: 0.107). 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 ETP contract 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 benefited students assigned to the contract teachers. Thus the very cheap SBM intervention appears to increase the effectiveness of the Extra Teacher Program.¹⁴

The last three columns in Table 2 show the test score results after reweighting to account for differential attrition. The results are virtually identical, suggesting the effects are unlikely to be driven by potential biases caused by attrition (as mentioned above there is a strong

correlation between baseline test scores and endline scores, so the reweighting would make a difference if differential unobserved ability was a concern).

Finally, Table A4 shows the estimated lower bounds for these specifications accounting for attrition, using the approach first proposed in Lee (2008).¹⁵ Here again, 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. This is quite notable, because in many cases where such analyses are presented, Lee bounds are so wide that most results become insignificant at the lower bound.

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

¹⁴ For completeness, it should 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 itself mitigated the endogenous TSC teacher effort response (at least in the top track), thus leaving little scope to further increase effort and test scores.

¹⁵ 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 Table A4 columns 4, 8 and 12.

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*** (0.038)	−0.017 (0.024)	−0.174*** (0.040)	
b_2 Basic ETP, contract teacher	0.117** (0.047)	0.011 (0.037)	0.143*** (0.049)	
b_3 ETP + SBM, TSC teacher	−0.083** (0.041)	−0.024 (0.026)	−0.079 (0.048)	
b_4 ETP + SBM, contract teacher	0.206*** (0.046)	0.093*** (0.026)	0.154*** (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)				
	Number of students who went to blackboard	Number of students who answered a teacher's question	Number of students who asked a question	
b_1 Basic ETP, TSC teacher	1.036** (0.510)	2.272 (3.726)	−1.822* (1.047)	
b_2 Basic ETP, contract teacher	0.623 (0.669)	−0.398 (2.255)	−1.585 (1.106)	
b_3 ETP + SBM, TSC teacher	0.490 (0.599)	1.160 (2.285)	−1.895 (1.157)	
b_4 ETP + SBM, contract teacher	0.982** (0.448)	−0.816 (2.172)	−2.095* (1.179)	
Observations	172	172	172	
Mean in comparison schools	0.714	9.127	2.19	
p-Val ($b_1 = b_2$)	0.59	0.485	0.568	
p-Val ($b_3 = b_4$)	0.51	0.276	0.311	
p-Val ($b_1 = b_3$)	0.475	0.753	0.855	
p-Val ($b_2 = b_4$)	0.641	0.825	0.273	
Panel C. Pedagogy (source: teacher surveys)				
	Teacher has a lesson plan prepared for the class	Teacher has time to mark individual work for children	Teacher has time to help children individually	Teacher has time to pay attention to slow learners
b_1 Basic ETP, TSC teacher	−0.001 (0.119)	0.243* (0.127)	0.168 (0.132)	0.230* (0.123)
b_2 Basic ETP, contract teacher	0.219** (0.102)	0.335*** (0.121)	0.298** (0.122)	0.636*** (0.089)
b_3 ETP + SBM, TSC teacher	−0.030 (0.121)	0.118 (0.112)	0.212* (0.115)	0.280** (0.114)
b_4 ETP + SBM, contract teacher	0.084 (0.113)	0.260** (0.115)	0.192* (0.110)	0.507*** (0.099)
Observations	166	167	166	166
Mean in comparison schools	0.597	0.258	0.323	0.274
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 are 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.

teachers in large classes, the effects were no longer statistically significant, and some have completely disappeared (Table A5). 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, 2012, 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). However, it 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), though it very well may.¹⁶

¹⁶ At the post-secondary level, evidence from Carrell and West (2010) suggests short-term positive test score effects can be driven by changes in teaching style that hurt students as they advance to a more challenging material. To the extent that we focus on second-graders and whether they acquired very basic skills without which further learning would be difficult, the potential for negative long-term effects seems low but we can only speculate.

Interestingly, the fact that the achievement results faded out, even though attrition patterns are similar across the endline and long-run follow-up surveys (see Table A1), suggests that the endline achievement impacts shown in Table 2 are unlikely to be entirely due to differential selection. If they were, then as the differential attrition persisted, the achievement effects would likely also have persisted.¹⁷

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.

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 teaching in a classroom (*any* classroom – not specifically grade 1) 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 (coefficient b_1). This effect corresponds to a 27% 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 of the difference = 0.081). Thus, it seems that the SBM program was partly effective in ensuring that the extra resources were put to their intended use, rather than captured by the incumbents.

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, most often our monitors reported finding teachers 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 teaching effort.¹⁸ 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.

The gap in effort between TSC and contract teachers remained large in SBM schools however, with TSC teachers almost 30 percentage points less likely to be found teaching ($0.083 + 0.206$). Given the results in Table 2, showing no difference in the scores of their students, this implies that TSC teachers, when they teach, can be very efficient teachers.

Besides teacher effort, the ETP and SBM programs could have led to changes in pedagogy. This was not the aim of the SBM program, but it is possible it indirectly affected teachers' commitment to the students. For the ETP program, pedagogy may have changed because it made the class 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 (and thus 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 only modest evidence that the ETP program changed pedagogy, and no pedagogy effect of SBM whatsoever. The absolute number of students called to the blackboard and called on to answer questions somewhat increased in both types of ETP schools (the effects are large in magnitude compared to the base, but most are not statistically significant; see Panel B). Since class size was reduced by close to half, the rate at which any given student was called increased, without the teachers having to change their pedagogy. The number of students who asked questions decreased more than proportionally than the class size reduction, possibly because teachers were able to address issues before they arose. Indeed in ETP schools teachers (whether civil servants or contract teachers) report having more time to provide individual attention to children, including slow learners, and provide feedback on individual work (Panel C).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 we will show in Section 5 below, 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

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 And we provide evidence that the SBM program, just as it mitigated the

¹⁷ We thank an anonymous referee for making that point.

¹⁸ Recall that we observe whether the teacher was teaching in any class, not just grade 1. So our measure would capture re-allocation to other grades. Conditional on not being found in class teaching, TSC teachers were found in the staff room 79% of the time and under a tree 21% of the time.

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.

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.

Table 4
Contract teacher selection.

	(1)	(2)	(3)	(4)	(5)
	Characteristics of ETP contract teacher hired by the school				
	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 (0.10)	−0.12 (0.12)	−0.12 (0.13)	0.14 (0.11)	−0.02 (0.10)
Prop. TSC teachers female	−0.48*** (0.15)	−0.25 (0.19)	−0.16 (0.20)	−0.04 (0.17)	−0.01 (0.16)
Average experience among TSC teachers	0.02* (0.01)	−0.02* (0.01)	0.00 (0.01)	−0.01 (0.01)	−0.01 (0.01)
School performance on 2004 national exam	−0.02 (0.06)	−0.01 (0.08)	−0.01 (0.09)	0.04 (0.07)	−0.02 (0.07)
School size (/100)	0.08** (0.03)	0.00 (0.04)	−0.04 (0.04)	0.03 (0.04)	0.03 (0.03)
Number of teachers assigned to lower grades	0.03 (0.03)	0.03 (0.04)	0.03 (0.04)	−0.04 (0.03)	−0.01 (0.03)
School had at least one PTA teacher in 2004	−0.17 (0.11)	−0.06 (0.14)	0.02 (0.15)	−0.28** (0.12)	−0.04 (0.12)
Observations	68	68	68	68	68
R-squared	0.37	0.1	0.09	0.19	0.06
Mean in schools without SBM	0.34	0.74	0.51	0.17	0.2

Notes: sample restricted to the 70 schools sampled for the extra-teacher program. unit of observation: contract teacher hired by school committee at onset of ETP program. Teacher survey data are missing for two ETP teachers. Linear probability model regressions with standard errors presented in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. Region dummies are included but not shown.

effort response, mitigated such nepotism. This is another channel through which empowering parents increased the effectiveness of providing additional resources to the school.

As a 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 patrilineal 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, while women will have fewer obligation towards their birth family and their husbands' relatives.²²

To understand the factors influencing hiring and retention of contract teachers it is useful to consider the following framework. PTA hiring committees are composed of parents and teachers. In the hiring decisions, assume that 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, it will reduce the weight PTA committees place on delivering rents to relatives of teachers. The idea that local beneficiaries will put more weight on the collective good and less on rent diversion is a key motivation behind the effort to strengthen beneficiary committees (such as school committees) (World Bank, 2004), although some express the concern that local committees would be more likely to try and share the rents locally than teachers who are not from the area (Bardhan and Mokherjee, 2000, 2005). Consistent with our assumption, however, Alatas et al. (2013) find that while formal elites to capture rents for themselves and their relative when they are put in charge of allocating a program, informal elites (those without an official position) are less likely to do so.

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 better selection of relatives. However,

they also have an incentive to hire relatives even if they are bad: this is the same key trade off that employers face when using workers to refer individuals from their social network (see, e.g., Beaman and Magruder, 2012). 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.

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. Moreover, we might see lower effort among teachers' relatives (due to lower incentives), but SBM should also improve this. *Ceteris paribus*, this will increase the test scores among teacher relatives in SBM schools. The overall impact of SBM on test scores of relatives is however potentially ambiguous because of the information effect: if teachers have valuable information about their relatives but SBM-trained committees do not use this information, the relatives they hire may be worse (see Fisman et al., 2014 for a discussion of a similar tradeoff for CEOs and their boards).

We have limited power to test these predictions due to the small sample size, but we present some evidence that is consistent with the predictions.

4.3.1. Hiring of relatives

In ETP schools without SBM, 34% of contract teachers hired at the onset of the program were relatives of existing TSC teachers. In SBM schools the share of relatives among contract teachers was 15 percentage points lower, at 19% (Table 4, column 1). This difference is not statistically significant at conventional levels, however, given the limited sample size (the p-value for the SBM effect is 0.133).

As noted, 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,

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, especially male ones, to be in posts in their home area.

Table 5

"Relative" effect: performance of ETP contract teachers related to existing TSC teachers.

	(1)	(2)	(3)	(4)	(5)	(6)
	Student performance			Teacher effort		
	Dep. var: standardized endline score for student assigned to...		<i>P-val coeff on ETP contract teacher = Coeff on TSC teacher</i>	Dep. var: teacher found in class teaching		<i>P-val coeff on ETP contract teacher = Coeff on TSC teacher</i>
	ETP contract teacher	TSC teachers		ETP contract teacher	TSC teachers	
School based management (SBM)	−0.098 (0.149)	0.089 (0.141)	0.023	0.018 (0.068)	0.099** (0.047)	0.251
(a) ETP contract teacher is the relative of a TSC teacher	−0.355* (0.184)	−0.092 (0.201)	0.126	−0.304** (0.143)	−0.057 (0.102)	0.107
(b) SBM × ETP contract teacher is the relative of a TSC teacher	0.441** (0.209)	0.260 (0.181)	0.265	0.249* (0.129)	0.096 (0.090)	0.238
Proportion of females among TSC teachers	0.231 (0.200)	0.571*** (0.191)	0.015	−0.018 (0.099)	−0.125 (0.082)	0.341
Proportion of females among TSC teachers × ETP contract teacher is the relative of a TSC teacher	0.642** (0.272)	0.192 (0.255)	0.119	0.360** (0.175)	0.097 (0.101)	0.192
ETP contract teacher has <1 year of experience	−0.175* (0.094)			−0.018 (0.084)		
Average experience level among TSC teachers	−0.007 (0.011)	−0.009 (0.012)		0.000 (0.006)	0.000 (0.005)	
School standardized KCPE score at baseline	0.114 (0.081)	0.017 (0.074)		0.001 (0.040)	−0.027 (0.023)	
School size (/100)	−0.036 (0.038)	−0.016 (0.034)		−0.014 (0.020)	0.017 (0.015)	
Female teacher	0.005 (0.108)			−0.086 (0.059)	0.106* (0.056)	
Female student	0.094* (0.051)	0.045 (0.056)				
Student's baseline score	0.551*** (0.035)	0.486*** (0.040)				
Observations	1548	1631		217	852	
(c) Total effect of contract teacher being a relative under SBM (<i>a + b</i>)	0.086	0.168	0.604	−0.055	−0.055	0.465
<i>p</i> -Val <i>a + b = 0</i>	0.702	0.390		0.673	0.616	
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 regressions. controls for student age, gender, region and date of tests are included. Data source: endline test administered by a research team.

Columns 3–4: LPM regressions. There are up to 5 observations per teacher. Data source: Five unannounced spot checks. Controls for region and spot check date as well as the log number of TSC teachers in lower grades are included.

point estimates suggest SBM schools were 12 percentage points less likely to hire contract teachers from the local area, although here again the difference is not significant (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 Table A6).

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).

Table 4 also shows that, as predicted, the share of women among TSC teachers is indeed negatively correlated with the hiring of a relative. Note that this could also be driven by omitted characteristics of schools, themselves responsible for a lesser rate of nepotism. For example, schools in poorer conditions may have a harder time retaining male teachers, and in those schools the benefit associated with being a contract teacher would be lower, reducing the incentives to secure the position for a relative.

4.3.2. 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 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. The only reason why this would not be the case is if SBM prevent information to flow effectively from teachers to the committee and this effect is sufficiently strong to outweigh the two others.

Looking at students' test 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 −0.355 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 +0.441 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

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
<i>Panel A. Effect of SBM</i>				
SBM	−0.15 (0.12)	−0.02 (0.09)	0.16 (0.11)	0.09 (0.16)
Observations	70	70	70	42
<i>Panel B. Effect of teacher characteristics</i>				
SBM	−0.17 (0.11)	0.05 (0.10)	0.22* (0.13)	−0.02 (0.20)
Female	0.03 (0.11)	0.17* (0.09)	−0.06 (0.12)	−0.01 (0.18)
Years of experience	0.12* (0.06)	0.05 (0.05)	−0.19*** (0.07)	−0.15* (0.09)
Relative of a teacher in the school	−0.07 (0.15)	−0.12 (0.13)	0.20 (0.17)	0.10 (0.25)
Average presence	0.25 (0.23)	−0.48** (0.19)	−0.1 (0.25)	−0.24 (0.36)
Mean score of contract teacher students at baseline	0.16 (0.55)	−0.11 (0.47)	0.28 (0.60)	0.39 (1.09)
Mean score of contract teacher students at endline	0.34* (0.17)	−0.11 (0.14)	−0.07 (0.19)	0.10 (0.35)
Mean score of TSC teacher students at endline	0.01 (0.17)	−0.2 (0.15)	0.15 (0.19)	0.45 (0.35)
Prop. TSC teachers female	−0.12 (0.19)	−0.14 (0.16)	0.09 (0.21)	0.03 (0.35)
Average experience among TSC teachers	0.03*** (0.01)	−0.01 (0.01)	−0.02* (0.01)	−0.02 (0.02)
School performance on 2004 national exam	−0.07 (0.07)	0.1 (0.06)	−0.04 (0.08)	−0.05 (0.12)
School Size (/100)	0.01 (0.04)	0.02 (0.03)	−0.04 (0.04)	−0.04 (0.06)
Number of teachers assigned to lower grades	0 (0.04)	−0.01 (0.03)	0.02 (0.04)	0.02 (0.06)
School had at least one PTA teacher in 2004	0.13 (0.13)	0.02 (0.11)	0.19 (0.15)	0.33 (0.21)
Observations	68	68	68	42
R-squared	0.44	0.38	0.29	0.36
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 observation: contract teacher hired by school committee at onset of ETP program. Teacher characteristics survey data are missing for two ETP teachers. Linear probability model regressions with standard errors are presented in parentheses. Panels A and B correspond to two separate regressions. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. Region dummy are included but not shown.

specific effect of nepotism or the lack thereof. To check this, in column 2 of Table 5 we test whether the “relative” and “relative × 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, although the “relative” effect is not significantly different for student assigned to TSC and ETP teacher in non-SBM schools (p-value = 0.126, column 3).

Looking at teacher effort, the point estimates in column 4 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.

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 are 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 5% level). Relatives hired in these schools were also present and teaching in class significantly more often (Table 5, column 3). Of course, as mentioned above, schools with a higher share of female TSC teachers may be systematically different

on other dimensions than their female share, therefore here again it is difficult to interpret these effects as causal.

5. Potential dynamic effects of local probationary teacher hiring

The results presented above suggest important potential gains from 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 by trained PTAs before becoming eligible for TSC positions. We estimate it could not only save roughly 75% on teacher salaries during teachers' initial period as locally-hired contract employees but could also 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 promote better-performing contract teachers (among those who had been hired). 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.

5.1. Performance-based retention

Table 6 presents evidence on how contract teachers' performance in the ETP program affected their subsequent 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 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 34 percentage points, after controlling for performance among students of TSC teachers, and student baseline characteristics (Table 6, column 1).²⁵ Endline test score and prior experience are the only contract teacher characteristics that predict future hiring.

Another 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 6, 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.²⁶

More telling is what happened at the end of the program. The funds for the program stopped, but school committees could try to 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 $0.25 + 0.22 = 47\%$ among SBM schools (see Table 6, 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 6, column 4). Relatives are not less likely to be retained, despite worse performance, so conditional on performance, relatives are more likely to be retained.

5.2. 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 Appendix A, we do a calibration exercise and estimate, for a reasonable set of assumptions, the potential dynamic impact of a local contract teacher hiring program 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)²⁷ 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 quite speculative: our estimates of the impact of the program on teacher selection are not the most precise, the calculations rely on a number of assumptions (e.g. that teachers serve for 27 years), the extent of capture by teachers or local elites may differ if a probation system is formalized and scaled, and much would clearly depend on the details of how such a system was implemented; but it provides a useful way to think about 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 such 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 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 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 contemporary to ours and carried out across the state of Andhra Pradesh in India (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.

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 ours. But in

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.

²⁵ We find a positive but not significant correlation between TSC hiring and the gap in scores between students assigned to TSC and those assigned to ETP teachers, thus it's not clear that benchmark competition between contract teachers and existing civil service teachers is responsible for the tendency for teachers with better scores to obtain TSC positions.

²⁶ School committees never explicitly voted against renewing a contract in year 2.

²⁷ Novice teachers typically have lower value added (Kane et al., 2008; Rockoff and Staiger, 2010).

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.

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 a lower pay, and then are eventually brought in as civil-service employees (or fired, in an up or out system akin to the higher education system in the US (Pritchett and Murgai, 2006).

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 likely undermined both the incentive and selection effects of local hiring of contract teachers by parent committees.

Appendix A. Calibrating the potential dynamic impact of contract teachers on teacher workforce

We follow an approach similar to that of Gordon et al. (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 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.²⁸

To get a sense of how big the total effect might be, suppose teachers work for 27 years.²⁹ If new entrants to the profession worked three years as a contract teacher and half were subsequently hired as civil

²⁸ 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.

²⁹ 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.

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 $0.275 \times 0.2 - 0.0022 = 0.052$ 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,³⁰ 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 $0.16 \times 0.8 = 0.13$ 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 a positive selection among those initially hired as contract teachers compared to civil servants, which would mean that 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.³¹

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.³²

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.

³⁰ 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). 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.

³¹ 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.

³² 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.

Table A1
Attrition: breakdown by initial position and attrition at long-run follow-up.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	10
Dep. var:		Attrited at endline test (after 19 months in program)				Attrited at long-run follow-up test (a year after program ended)					
Subsample		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	-0.013 (0.024)		0.012 (0.015)		-0.008 (0.018)		0.001 (0.026)		-0.002 (0.022)	
a_2	ETP + SBM	-0.030 (0.023)		-0.027 (0.016)		-0.030** (0.014)		-0.045** (0.022)		-0.026 (0.016)	
b_1	Basic ETP, assigned to TSC teacher		-0.000 (0.030)		0.021 (0.021)		0.007 (0.021)		-0.013 (0.031)		0.028 (0.027)
b_2	Basic ETP, assigned to contract teacher		-0.029 (0.027)		0.003 (0.018)		-0.023 (0.020)		0.016 (0.027)		-0.035 (0.024)
b_3	ETP + SBM, assigned to TSC teacher		-0.047* (0.028)		0.000 (0.021)		-0.027 (0.017)		-0.061** (0.026)		-0.008 (0.019)
b_4	ETP + SBM, assigned to contract teacher		-0.013 (0.026)		-0.058*** (0.021)		-0.033** (0.015)		-0.029 (0.023)		-0.047* (0.024)
Region dummies		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations		3469	3469	3657	3657	8012	8012	3469	3469	3657	3657
R-squared		0.001	0.002	0.002	0.004	0.001	0.001	0.002	0.003	0.001	0.002
p-Val ($b_1 = b_2$)		0.196		0.196		0.238		0.238		0.238	
p-Val ($b_1 = b_2$)											
p-Val ($b_3 = b_4$)		0.367		0.44		0.115		0.261		0.029**	
p-Val ($b_1 = b_3$)		0.244		0.025**		0.679		0.145		0.17	
p-Val ($b_2 = b_4$)		0.182		0.411		0.155		0.168		0.219	

Notes: see Table 2 notes.

Table A2
Student attendance, endline attrition and endline test scores, controlling for linear trend in visit/test date.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Found in class during unannounced spot checks	Attrited at endline test	Total score	Math score	Literacy score	Using weights to correct for differential attrition by quintile of initial distribution		
							Total score	Math score	Literacy score
Panel A									
a_1	Basic ETP	0.002 (0.012)	-0.015 (0.016)	0.189** (0.096)	0.175** (0.080)	0.165 (0.106)	0.192** (0.095)	0.177** (0.079)	0.167 (0.105)
a_2	ETP + SBM	0.020 (0.012)	-0.030* (0.016)	0.236** (0.097)	0.250*** (0.076)	0.177* (0.107)	0.233** (0.097)	0.247*** (0.077)	0.174 (0.106)
Observations		8012	6533	6533	6536	6533	6533	6536	6534
Mean in comparison schools		0.196	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p-Val ($a_1 = a_2$)		0.181	0.656	0.367	0.923	0.69	0.391	0.955	
Panel B									
b_1	Basic ETP, assigned to TSC teacher	-0.006 (0.013)	-0.001 (0.018)	0.087 (0.098)	0.051 (0.078)	0.102 (0.111)	0.089 (0.098)	0.052 (0.078)	0.104 (0.110)
b_2	Basic ETP, assigned to contract teacher	0.011 (0.013)	-0.031* (0.018)	0.294*** (0.105)	0.301*** (0.093)	0.231** (0.114)	0.298*** (0.104)	0.304*** (0.092)	0.235** (0.113)
b_3	ETP + SBM, assigned to TSC teacher	0.014 (0.013)	-0.023 (0.020)	0.212* (0.110)	0.233** (0.089)	0.152 (0.119)	0.221** (0.110)	0.242*** (0.089)	0.159 (0.120)
b_4	ETP + SBM, assigned to contract teacher	0.026** (0.013)	-0.039** (0.017)	0.263*** (0.095)	0.270*** (0.081)	0.205** (0.103)	0.247** (0.095)	0.253*** (0.082)	0.192* (0.102)
Observations		52,622	8,012	6,533	6,533	6,536	6,531	6,531	6,534
p-Val ($b_1 = b_2$)		0.023**	0.08*	0.002***	0.000***	0.067*	0.002***	0.000***	0.067*
p-Val ($b_3 = b_4$)		0.075*	0.398	0.46	0.627	0.413	0.716	0.89	0.62
p-Val ($b_1 = b_3$)		0.167	0.307	0.293	0.053*	0.708	0.265	0.042**	0.682
p-Val ($b_2 = b_4$)		0.288	0.678	0.779	0.749	0.832	0.639	0.603	0.722

Notes: Column 1: Linear probability model regressions. Data comes from five unannounced spot checks performed by the research team (two in 2005 and three in 2006). Students identified as dropouts are coded as absent. Those reported to have moved to another school are excluded from the analysis. Columns 2–8: OLS regressions. The endline test was administered after the program had been in place for five school terms. In each school, 60 students were randomly drawn from the baseline sample to participate in the tests. Scores are normalized such that the mean and standard deviation of the comparison group are zero and one, respectively. Columns 6–8 reweigh the observations by the inverse of the non-attrition rate in their quintile of the initial performance distribution.

In both panels the omitted category is the comparison group. Panel A: “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 equal to 1 if the school was sampled for the ETP program with the SBM training. Panel B: “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.

All regressions control for region dummies. Robust standard errors are clustered at the school level and shown in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively. Columns 2–8: There are only 139 schools/clusters because tests could not be administered in one of the ETP schools.

Table A3
Impacts on attrition and endline test scores, controlling for school-level and individual-level controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Attrited at endline	Total score	Math score	Literacy score	Attrited at endline	Total score	Math score	Literacy score
a_1 Basic ETP	−0.036* (0.021)	0.145 (0.093)	0.104 (0.079)	0.152 (0.099)				
a_2 ETP + SBM	−0.048** (0.020)	0.188** (0.092)	0.177** (0.075)	0.161 (0.100)				
b_1 Basic ETP, assigned to TSC teacher					−0.029 (0.023)	0.042 (0.094)	−0.011 (0.078)	0.080 (0.101)
b_2 Basic ETP, assigned to contract teacher					−0.047** (0.021)	0.262** (0.107)	0.232** (0.095)	0.237** (0.111)
b_3 ETP + SBM, assigned to TSC teacher					−0.045** (0.021)	0.177* (0.096)	0.179** (0.080)	0.140 (0.104)
b_4 ETP + SBM, assigned to contract teacher					−0.054*** (0.020)	0.220** (0.098)	0.192** (0.087)	0.201* (0.105)
School size (/100)	0.000 (0.018)	−0.170** (0.085)	−0.297*** (0.073)	−0.026 (0.092)	−0.000 (0.018)	−0.166* (0.085)	−0.293*** (0.073)	−0.022 (0.092)
Share female among TSC teachers	0.030 (0.024)	0.505*** (0.111)	0.312*** (0.098)	0.575*** (0.113)	0.030 (0.024)	0.503*** (0.111)	0.311*** (0.098)	0.573*** (0.113)
Average years of experience (/10) among TSC teachers	−0.002 (0.011)	0.042 (0.069)	−0.001 (0.060)	0.071 (0.070)	−0.003 (0.012)	0.045 (0.069)	0.002 (0.061)	0.073 (0.071)
Girl	0.010 (0.008)	0.057** (0.024)	−0.012 (0.022)	0.107*** (0.027)	0.011 (0.008)	0.057** (0.024)	−0.013 (0.022)	0.106*** (0.027)
Age	−0.022*** (0.004)	−0.043*** (0.012)	−0.008 (0.010)	−0.066*** (0.013)	−0.022*** (0.004)	−0.044*** (0.012)	−0.008 (0.010)	−0.066*** (0.013)
Baseline score	0.001 (0.005)	0.497*** (0.018)	0.495*** (0.015)	0.400*** (0.021)	0.001 (0.005)	0.497*** (0.018)	0.496*** (0.015)	0.400*** (0.021)
Teacher has <1 year of experience	0.005 (0.020)	0.065 (0.092)	0.101 (0.087)	0.018 (0.103)	0.013 (0.023)	−0.006 (0.101)	0.036 (0.094)	−0.042 (0.113)
Observations	8012	6533	6533	6536	8012	6533	6533	6536
R-squared		0.230	0.230	0.164		0.232	0.234	0.166
p-Val ($a_1 = a_2$)	0.509	0.675	0.371	0.938				
p-Val ($b_1 = b_2$)					0.169	0.002***	0.000***	0.032**
p-Val ($b_3 = b_4$)					0.525	0.473	0.857	0.297
p-Val ($b_1 = b_3$)					0.433	0.229	0.032**	0.633
p-Val ($b_2 = b_4$)					0.706	0.711	0.684	0.768

Notes: see Table 2.

Table A4
Lee (2008) 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 [†]	Lower bound	Upper bound	N before trimming	Trimming proportion [†]	Lower bound	Upper bound	N before trimming	Trimming proportion [†]
<i>Difference between treatment group [...] and control group</i>												
a_1 Basic ETP	0.056 (0.101)	0.173 (0.107)	4965	0.020	0.087 (0.101)	0.28*** (0.097)	4965	0.020	0.062 (0.115)	0.135 (0.113)	4967	0.020
a_2 ETP + SBM	0.097 (0.106)	0.238** (0.099)	4876	0.037	0.141 (0.092)	0.368*** (0.076)	4876	0.037	0.029 (0.114)	0.174 (0.108)	4879	0.037
<i>Difference between treatment group [...] and control group</i>												
b_1 Basic ETP, assigned to TSC teacher	−0.022 (0.112)	0.054 (0.111)	4143	0.004	−0.043 (0.100)	0.143 (0.101)	4143	0.004	0.045 (0.121)	0.058 (0.118)	4145	0.003
b_2 Basic ETP, assigned to contract teacher	0.124 (0.117)	0.304** (0.120)	4129	0.038	0.188* (0.099)	0.418*** (0.096)	4129	0.038	0.07 (0.129)	0.218* (0.126)	4131	0.038
b_3 ETP + SBM, assigned to TSC teacher	0.087 (0.120)	0.211* (0.114)	4123	0.029	0.14 (0.101)	0.347*** (0.079)	4123	0.029	0.012 (0.127)	0.135 (0.120)	4125	0.029
b_4 ETP + SBM, assigned to contract teacher	0.112 (0.104)	0.29*** (0.101)	4060	0.045	0.173* (0.092)	0.39*** (0.081)	4060	0.045	0.049 (0.111)	0.218** (0.108)	4063	0.046
<i>Difference between...</i>												
e Contract teacher (b_2) and TSC teacher (b_1) in basic ETP	0.107 (0.082)	0.256*** (0.073)	1656	0.035	0.172** (0.074)	0.41*** (0.065)	1656	0.035	0.019 (0.090)	0.162** (0.079)	1656	0.035
f Contract teacher (b_3) and TSC teacher (b_4) in ETP + SBM	−0.002 (0.085)	0.097 (0.071)	1567	0.016	−0.003 (0.090)	0.198** (0.086)	1567	0.016	−0.019 (0.083)	0.075 (0.068)	1568	0.018

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). These estimated upper and lower bound estimates do not control for the date of the test. Bootstrapped standard errors clustered at the school level in parentheses (1000 replications). 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.

†: the trimming proportion corresponds to the proportion of trimmed observations in the treatment group (or the group with less attrition for rows e and f).

Table A5
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 (0.080)	−0.016 (0.069)	−0.010 (0.089)			
a_2 ETP + SBM	0.099 (0.097)	0.104 (0.074)	0.079 (0.105)			
b_1 Basic ETP, assigned to TSC teacher				−0.062 (0.089)	−0.059 (0.079)	−0.054 (0.095)
b_2 Basic ETP, assigned to contract teacher				0.039 (0.084)	0.029 (0.071)	0.038 (0.093)
b_3 ETP + SBM, assigned to TSC teacher				0.113 (0.109)	0.128 (0.084)	0.084 (0.118)
b_4 ETP + SBM, assigned to contract teacher				0.085 (0.092)	0.080 (0.074)	0.075 (0.100)
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes
Linear trend in date of test	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6171	6171	6175	6171	6171	6175
R-squared	0.004	0.003	0.004	0.005	0.003	0.005
p-Val ($a_1 = a_2$)	0.239	0.109	0.41			
p-Val ($b_1 = b_2$)				0.105	0.139	0.13
p-Val ($b_3 = b_4$)				0.646	0.417	0.886
p-Val ($b_1 = b_3$)				0.127	0.046**	0.271
p-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 equal 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 A6
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	Overall hiring score ^a
SBM	0.12 (0.08)	−0.13 (0.19)	0.33 (0.34)	0.55 (0.45)	0.57 (0.36)	0.71* (0.41)
Bungoma district	−0.02 (0.11)	0.12 (0.27)	−0.62 (0.48)	2.06*** (0.68)	1.31** (0.50)	0.36 (0.59)
School performance on 2004 national exam	0.00 (0.05)	0.05 (0.13)	−0.10 (0.23)	−0.54* (0.31)	−0.48* (0.25)	−0.21 (0.29)
Prop. TSC teachers female	0.20* (0.12)	0.23 (0.30)	−0.33 (0.52)	−0.32 (0.68)	−0.58 (0.55)	−0.16 (0.64)
Average experience among TSC teachers	0.00 (0.00)	0.00 (0.02)	0.01 (0.03)	−0.02 (0.04)	−0.01 (0.03)	−0.01 (0.04)
School size (/100)	−0.02 (0.03)	−0.01 (0.07)	0.17 (0.12)	0.01 (0.16)	0.13 (0.13)	0.22 (0.15)
Number of teachers assigned to lower grades	−0.02 (0.02)	0.02 (0.06)	−0.09 (0.10)	0.04 (0.14)	0.08 (0.11)	−0.10 (0.12)
School had at least one PTA teacher in 2004	0.09 (0.09)	0.34 (0.21)	−0.71* (0.37)	0.56 (0.52)	0.49 (0.40)	−0.06 (0.46)
Observations	63	63	63	62	63	63
R-squared	0.19	0.07	0.13	0.27	0.27	0.10
Mean in schools without SBM	0.03	0.97	0.69	2.75	2.00	2.28

Notes: the post-hire survey was not administered in 7 of the 70 ETP schools. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

^a The overall hiring score is the sum over 5 dummies: the three dummies in columns 1–3 as well as two dummies for whether the number of applications (resp, candidates) is above the median number of applications (candidates).

References

- Alatas, Vivi, Banerjee, Abhijit, Hanna, Rema, Purnamasari, Ririn, Wai-Poi, Matthew, 2013. Does elite capture matter? Local elites and targeted welfare programs in Indonesia. NBER Working Paper #18798.
- Andrabi, Tahir, Das, Jishnu, Khwaja, Asim Ijaz, Zajonc, Tristan, 2011. Do value-added estimates add value? Accounting for learning dynamics. *Am. Econ. J. Appl. Econ.* 3 (3), 29–54.
- Banerjee, Abhijit V., Cole, Shawn, Duflo, Esther, Linden, Leigh, 2007. Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, MIT Press 122 (3), 1235–1264.
- Bardhan, Pranab, Mokherjee, Dilip, 2000. Capture and governance at local and national levels. *Am. Econ. Rev.* 90 (2), 135–139.
- Bardhan, Pranab, Mokherjee, Dilip, 2005. Decentralizing antipoverty program delivery in developing countries. *J. Public Econ.* 89, 675–704.
- Beaman, Lori, Magruder, Jeremy, 2012. Who gets the job referral? Evidence from a social networks experiment. *Am. Econ. Rev.* 102 (7), 3574–3593.
- Blimpo, M.P., Evans, D.K., 2011. School-based management and educational outcomes: lessons from a randomized field experiment. Unpublished manuscript, <http://siteresources.worldbank.org/EDUCATION/Resources/Blimpo-EvansWSD-2012-01-12.pdf> Accessed 12/2/2014.
- Bold, Tessa, Kimenyi, Mwangi, Mwabu, Germano, Ng'ang'a, Alice, Sandefur, Justin, 2013. Pitfalls of scaling-up proven education interventions: evidence from an RCT in Kenya. CGD Working paper 321.
- Bourdon, J., Frölich, M., Michaelowa, K., 2010. Teacher Shortages, Teacher Contracts and Their Impact on Education in Africa, *Journal of the Royal Statistical Society. Series A* 173, 93–116.
- Bruns, Barbara, Filmer, Deon, Patrinos, Harry A., 2011. Making schools work. new evidence on accountability reforms. The World Bank, Washington D.C.
- Carrell, Scott, West, James, 2010. Does professor quality matter? Evidence from random assignment of students to professors. *J. Polit. Econ.* 118 (3), 409–432.
- Cascio, Elizabeth, and Douglas Staiger (2012). "Knowledge, Tests, and Fadeout in Educational Interventions". Mimeo.
- Chetty, Raj, Friedman, John N., Hilger, Nathaniel, Saez, Emmanuel, Whitmore Schanzenbach, Diane, Yagan, Danny, 2010. How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Q. J. Econ.* 126 (4), 1593–1660 (2011).
- Das, Jishnu, Dercon, Stephane, Habyarimana, James, Krishnan, Pramila, Muralidharan, Karthik, Sundararaman, Venkatesh, 2013. School inputs, household substitution, and test scores. *Am. Econ. J. Appl. Econ.* 5 (2), 29–57.
- DiNardo, John, Fortin, Nicole, Lemieux, Thomas, 1996. Labor market institutions and the distribution of wages, 1973–1993: a semi-parametric approach. *Econometrica* 64 (5), 1001–1045.
- Duflo, Esther, Dupas, Pascaline, Kremer, Michael, 2011. Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. *Am. Econ. Rev.* 101 (5), 1739–1774.
- Duflo, Esther, Hanna, Rema, Ryan, Stephne P., 2012. Incentives work: getting teachers to come to school. *Am. Econ. Rev.* 102 (4), 1241–1278.
- Durante, Ruben, Labartino, Giovanna, Perotti, Roberto, 2011. Academic dynasties: decentralization and familism in the Italian academia. CEPR Discussion Paper 8645.
- Duthilleul, Yael, 2005. Lessons learnt in the use of "contract" teachers. synthesis report. International Institute for Educational Planning (IIEP) UNESCO, Paris.
- Fisman, Raymond, Khurana, Rakesh, Rhodes-Korpf, Matthew, Yim, Sookin, 2014. Governance and CEO turnover: do something or do the right thing? *Manag. Sci.* 60 (2), 319–337.
- Glewwe, Paul, Ilias, Naumam, Kremer, Michael, 2010. Teachers incentives. *Am. Econ. J. Appl. Econ.* 2 (3), 205–227.
- Goldhaber, D., Hansen, M., 2013. Is it just a bad class? Assessing the stability of measured teacher performance. *Economica* 80 (319), 589–612.
- Gordon, Robert, Kane, Thomas J., Staiger, Douglas O., 2006. "Identifying effective teachers using performance on the job" The Hamilton Project. Brookings Institute, Washington, DC.
- Hanushek, Eric A., Woessmann, Ludger, 2010. The economics of international differences in educational achievement. In: Hanushek, Eric A., Machin, Stephen, Woessmann, Ludger (Eds.), *Handbook of the Economics of Education* vol. 3. North Holland, Amsterdam.
- Kane, Thomas J., Rockoff, Jonah E., Staiger, Douglas O., 2008. What does certification tell us about teacher effectiveness? Evidence from New York City. *Econ. Educ. Rev.* 27, 619–631.
- Kremer, Michael, Moulin, Sylvie, Namunyu, Robert, 2003. Decentralization: a cautionary tale. Working Paper Harvard University.
- Lee, David, 2008. Training, wages, and sample selection: estimating sharp bounds on treatment effects. *Rev. Econ. Stud.* 76 (3), 1071–1102.
- Lucas, Adrienne, Mbiti, Isaac, 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. *Educ. Finan. Policy* 4 (4), 572–606.
- Muralidharan, K., Sundararaman, V., 2013. Contract Teachers: Experimental Evidence from India. Mimeo, UC San Diego.
- Pritchett, Lant, Murgai, Rinku, 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) pp. 123–177
- Rockoff, Jonah E., 2009. Field experiments in class size from the early twentieth century. *J. Econ. Perspect.* 23 (4), 211–230.
- Rockoff, Jonah E., Staiger, Douglas O., 2010. Searching for the effective teachers with imperfect information. *J. Econ. Perspect.* 24 (3), 97–118.
- Wilson, L.S., 1992. The Harambee movement and efficient public good provision in Kenya. *J. Public Econ.* 48 (1), 1–19.
- World Bank, 2004. World development report 2004: making services work for poor people. World Bank, Washington, D.C.