Scaling-Up Proven Education Interventions

Evidence from an RCT in Kenya

Tessa Bold
Mwangi Kimenyi
Germano Mwabu
Alice Ng'ang'a
Justin Sandefur

Scaling-up Proven Education Interventions: Evidence from an RCT in Kenya[*]

Tessa Bold, Mwangi Kimenyi, Germano Mwabu,
Alice Ng'ang'a and Justin Sandefur[†]

PRELIMINARY DRAFT - COMMENTS WELCOME

March 31, 2012

Abstract

Contract teachers have been shown to significantly raise test scores for primary students in previous randomized trials in both Western Kenya (Duflo, Dupas and Kremer 2009) and various locations in India (Banerjee, Cole, Duflo and Linden 2007, Muralidharan and Sundararaman 2010). We report on the initial, randomized phase of a Kenyan government program to scale-up the contract teacher model by (a) expanding geographically across all eight Kenyan provinces, and (b) shifting implementation from NGOs to the Ministry of Education. Overall, we find a significant, positive effect of 0.12 standard deviations on combined math and English scores across all treatment arms. Geographic heterogeneity in treatment effects does not appear to pose a significant threat to the external validity of earlier studies; effects are somewhat larger for schools with lower initial scores. However, overall effects were entirely due to a 0.19 standard deviation increase in scores in the randomly assigned sub-group of treatment schools where the contract teacher program was administered by an international NGO. Effects were significantly smaller and indistinguishable from zero in schools where the program was administered by the Ministry of Education. Evidence suggests this performance gap may reflect challenges faced by the Ministry in centrally administering a program that involved local recruitment and monitoring of contract teachers.

[†]Bold: Institute for International Economic Studies, Stockholm University, tessa.bold@iies.su.se. Kimenyi: Brookings Institution, Washington D.C., kimenyi@brookings.edu. Mwabu: Department of Economics, University of Nairobi, gmwabu@gmail.com. Ng'ang'a: Strathmore University, Nairobi, alicemnganga@yahoo.com. Sandefur: Center for Global Development, Washington D.C., jsandefur@cgdev.org.

1 Introduction

What options do policymakers in developing countries have to improve student learning in primary school? Despite decades of research on this core question, methodological shortcomings have undermined the credibility of policy guidance emerging from observational studies attempting to weigh the relative importance of factors such as class size, teacher training, etc. In recent years, new methods, in particular evaluations based on randomized assignments, have been successfully applied to deal with one of these concerns; i.e., the problem of identifying causal effects. While much remains to be done, the new literature is providing a menu of school-level interventions that have proven effective at raising test scores.

One of the most extensively tested, successful school-level interventions to raise student learning in primary schools has been the provision of contract teachers. Banerjee et al. (2007) present results from a randomized evaluation showing that an NGO program in urban India hiring young women to tutor lagging students in grades 3 and 4 led to a 0.28 standard deviation increase in tests scores. Muralidharan and Sundararaman (2010) evaluate a state-wide program Andhra Pradesh, finding that hiring an extra contract teacher leads to an increase in treatment schools of 0.15 and 0.13 standard deviations on math and language tests, respectively. In both cases, the additional teachers lead to significant learning gains despite salary costs that are a small fraction of civil service wages. Finally, of particular relevance for the present study given its geographic focus, Duflo et al. (2009) show that exposure to a contract teacher in government schools in Western Kenya raises test scores by 0.21 standard deviations relative to being taught by civil service teachers. Furthermore, their experimental design allows them to attribute this effect to contract teachers per se, rather than the accompanying reduction in class size from hiring an extra teacher.

This paper examines the ability of the Kenyan government to build on earlier research findings on the effect of contract teachers and scale-up a proven education intervention.

Generalizing the results of any randomized controlled trial to assess the effect of scaling up a given intervention raises several well-known questions (Shadish, Campbell and Cook 2002, Duflo 2004, Deaton 2010). The first is external validity. We illustrate how replication in new contexts and careful attention to heterogenous treatment effects can at least partially address this concern. We report on the randomized evaluation of the pilot phase of nationwide program which now employs over 18,000 contract teachers. The pilot was designed to test the Ministry of Education's ability to implement a fairly close variant of the NGO project described by Duflo et al. (2009) and to replicate the results across diverse conditions, spanning urban slums in Nairobi and nomadic communities in the remote Northeastern province.

A second question is whether successful NGO pilot projects can be replicated by governments – and in particular, within the institutional constraints of the Kenyan public sector bureaucracy. In Shadish et al.'s (2002) terminology, this is a question of 'construct validity' rather than external validity, i.e., of identifying the higher order construct represented by the experimental treatment. In most of the RCT literature, the treatment construct is defined to include only the school- or class-room level intervention, abstracting from the institutional context of these interventions which is often quite artificial relative to normal conditions in developing-country schools.

This study was designed to highlight the role of these larger institutional capacity constraints to scaling up. As part of the government's contract teacher pilot, 192 schools were chosen from across all eight Kenyan provinces: 64 were randomly assigned to the control group, 64 to receive a contract teacher as part of the government program, and 64 to receive a contract teacher under the coordination of the local affiliate of an international NGO, World Vision Kenya. The timing, salary levels, recruitment procedures and all other experimental protocols were held constant across the government and NGO arms of the evaluation.

Results confirm the hypothesis that school-level interventions cannot be adequately de-

3

scribed without attention to the broader institutional setting. While we find positive and significant effects of the program overall, these are concentrated entirely in schools where the contract teacher program was administered by an international NGO. We present suggestive evidence that this performance gap stems from the Ministry's centralized bureaucratic structure and challenges of communication and accountability in the middle-tiers of the bureaucracy that separate decision-makers in Nairobi from implementation in schools.

Examining the emerging literature on randomized trials in education in developing countries, and specifically at the sub-set of those studies which measure impacts on test scores or other learning outcomes, working with governments appears to place some constraints on the scope of interventions that can be tried.[1] NGO pilot programs have tested a wide range of interventions, particularly in India and Kenya, including a strong focus on changes to teacher incentives.[2] Programs with government involvement have, by and large, tended to focus on increasing school inputs of various kinds, and are more likely to occur in Latin

---

[1]Of the 31 studies we examined – all based on randomized trials in developing countries measuring impacts on test scores or other learning outcomes – 16 were conducted in Asia (13 of which in India), 11 in Africa (7 of which in Kenya) and 4 in Latin America. (See citations in subsequent footnotes.) Roughly half of the studies cite significant government involvement in project implementation (14 of 29), including all the Latin American studies, roughly half the Asian studies, and less than one third of the African studies.

[2]In India, RCTs have examined NGO programs to encourage parental involvement in schools (Pandey, Goyal and Sundararaman 2008, Banerjee, Banerji, Duflo, Glennerster, and Khemani 2010), changes to the English and reading curriculum (He, Linden and MacLeod 2008, He, Linden and MacLeod 2009), use of information technology in the classroom (Linden, Banerjee and Duflo 2003, Inamdar 2004, Linden 2008), student and parent incentives (Berry 2011), cameras in schools to discourage teacher absenteeism (Duflo, Hanna and Ryan 2010), and as already discussed, a form of contract teachers or tutors (Banerjee et al. 2007). Similarly in Kenya, NGO pilot programs have examined the impact of contract teachers and tracking students (Duflo, Dupas and Kremer 2011), teacher incentives (Glewwe, Ilias and Kremer 2010), student incentives (Kremer, Miguel and Thornton 2009), physical school inputs (Kremer, Moulin and Namunyu 2003, Glewwe, Kremer, Moulin and Zitzewitz 2004, Glewwe, Kremer and Moulin 2009), and school meals (Vermeersch and Kremer 2005), while in Uganda Barr, Mugisha, Serneels and Zeitlin (2011) report on an RCT of an NGO program to facilitate community monitoring of schools.

America and Asia, and less so in Africa.[34]

The rest of the paper is organized as follows. Section 2 describes the public primary schooling system in Kenya. Section 3 outlines the experimental design, a clustered (school level) randomized trial with a partial factorial design in which schools were assigned to receive a contract teacher either from the Ministry of Education or an NGO and a subset of these treatment schools received additional training for their school management committees. Section 4 describes the randomization procedures based on a multivariate matching algorithm and reports tests for balance using baseline data. Section 5 examines the relative success of the government and NGO programs in recruiting teachers locally at relatively low wages (relative to civil service salaries) and in setting up a timely, centrally administered payment system.

Section 6 presents intention-to-treat (ITT) effects and average treatment effects for the treated (ATT), where actual treatment is defined as successfully recruiting a contract teacher and random assignment is used as an instrumental variable. Section 7 repeats this analysis, comparing the relative effectiveness of NGO and Ministry implementation. Section 8 tests for heterogeneous treatment effects, finding no geographic differences or differences by initial class size, but some evidence that schools with lower initial test scores benefitted more from a contract teacher. Section 9 concludes.

---

[3]Governments have been directly involved in evaluations of the learning impacts of conditional cash transfer programs in Ecuador (Paxson and Schady 2007), Malawi (Baird, McIntosh and Özler 2010), and Nicaragua (Macours, Schady and Vakis 2011). Other studies have evaluated government programs involving school meals (Kazianga, de Walque and Alderman 2009), use of ICT in the classroom in Chile (Rosas, Nussbaum, Cumsille, Marianov, Correa, Flores, Grau, Lagos, Lopez, Lopez, Rodriguez and Salinas 2003) and Colombia (Barrera-Osorio and Linden 2009), provision of eye-glasses in China Glewwe, Park and Zhao (2011), and school construction in Afghanistan (Burde and Linden 2010) and reforms to local school management in Madagascar (Glewwe and Maïga 2011).

[4]Notable exceptions to this pattern that we are aware of, where researchers have been able to involve government in rigorous testing of reforms to teacher incentives, are the study of contract teachers in Andhra Pradesh mentioned above (Muralidharan and Sundararaman 2010) and parallel work on teacher performance pay (Muralidharan and Sundararaman 2011, Muralidharan 2011), and the evaluation of World Bank-financed school management reform program in Madagascar, cited above (Glewwe and Maïga 2011).

2 The Kenyan primary schooling system

According to the most recent available national household survey from 2006, net primary enrollment was 81%, with government primary schools accounting for 72% (Bold, Kimenyi, Mwabu and Sandefur 2011). The public education system is highly centralized. Officially all resources for the operation and maintenance of public schools flow through the Ministry of Education via two channels: non-salary expenditures deposited in school bank accounts, and teacher salaries paid directly to civil servants.

2.1 School finance

In January 2003, the Kenyan government abolished all school fees in government primary schools. This "Free Primary Education" (FPE) policy established the current system of school finance in which government primary schools are prohibited from collecting revenue and instead receive a central government grant – commonly known as "FPE funds" – of approximately $13.50 per pupil per annum to cover non-salary costs.[5] At the school level, FPE funds are held in a school bank account administered by a governing body known as a School Management Committee (SMC). The SMC is chaired by the head teacher and comprised of representatives from the Ministry, parents from each grade, teachers, and in some cases local community or religious organizations.

Misappropriation of FPE funds was at the center of a major corruption scandal which emerged in 2009. An external audit commissioned by the Ministry of Education showed that actual funds disbursed to school bank accounts fell short of the allocated amount by $4.71 per pupil in 2005 and by smaller but significant amounts in other years. Press reports estimated that anywhere between $68 million and $590 million of the FPE budget had been misdirected between 2004 and 2008 (Teyie and Wanyama 2010), leading the President to

---

[5]Except where otherwise noted, we convert Kenyan shillings to U.S. dollars using the prevailing exchange rate at the time of the baseline survey in July 2009, 74.32 shillings per dollar.

suspend several top Ministry officials, and foreign donors including DfID and USAID to freeze aid disbursements in December 2009.

2.2 Civil service teachers

Formally, all teachers in Kenyan public primary schools are civil servants employed by the Teacher Service Commission (TSC), a centralized bureaucracy under the direction of the Ministry of Education. In practice, schools also informally contract local teachers known as parent-teacher association (PTA) teachers. In the sample of schools surveyed for this study, 83% of teachers were employed by TSC and the remaining 17% by PTAs. TSC teachers earned an average of $261 per month in 2009, compared to just $56 per month for PTA teachers.

The relatively high salaries of TSC teachers creates an extreme form of labor market disequilibrium. On the demand side, the high salaries and the Ministry's limited budget lead to unfilled teacher vacancies. At the beginning of 2011 the Ministry of Education reported a shortage of 61,000 teachers (across roughly 20,000 primary schools) relative to its target of a 40:1 pupil-teacher ratio. On the supply side, high salaries attract a long queue of job applicants. TSC hires on the basis of an algorithm that primarily rewards seniority: the first applicants to graduate from teacher training college are the first to be hired. In 2010, Ministry records show that most successful applicants to TSC positions had been in the job queue for 8 to 11 years. PTA teachers are often drawn from this queue of graduates.

These features contribute to limited accountability for TSC teachers vis-a-vis parents or School Management Committees. Salaries are paid directly from Nairobi to individual teachers' bank accounts. And because of the chronic teacher shortages, parents and schools have little incentive to pursue disciplinary action against teachers; if a teacher is reassigned or terminated, a school may wait months or years for a replacement.

2.3 Contract teachers

Motivated by a desire to fill teacher vacancies and regularize PTA teachers, in 2009 the Directorate of Basic Education within the Ministry of Education proposed an initiative to provide funds to schools to employ teachers on contract outside of the TSC system. A steering committee – including Ministry officials and the current authors – was formed to design a pilot program, evaluate its impacts, and report back to the Permanent Secretary.

Under pressure from the Ministry of Finance to spend funds as part of an economic stimulus package, the Ministry opted to scale-up the contract teacher program before the pilot was completed. Thus the randomized pilot program analyzed here was launched in June 2010, and in October 2010 the Ministry hired 18,000 contract teachers nationwide, nearly equivalent to one per school. These 18,000 teachers were initially hired on 2-year, non-renewable contracts, at salary levels somewhat higher than described below for the evaluation component (roughly \$135 per month). In 2011 the Ministry succumbed to political pressure and agreed to allow the contract teachers to unionize and to subsequently hire all 18,000 contract teachers into the civil service at the end of their contracts.[6]

3 Experimental design

The basic intervention to be evaluated is the hiring of additional teachers, under various contract arrangements. In order to disentangle the effect of the various contractual and programmatic arrangements, the research project will "cross cut" four project features: (i) government versus NGO implementation of the overall program, (ii) an SMC training com-

---

[6]From an evaluation perspective, an obvious concern is that the allocation of these 18,000 contract teachers contaminated the randomly allocated teachers from the pilot program. It is important to note that allocation of contract teachers to schools for the full-scale program – while not itself randomized – was done on the basis of pupil-teacher ratios measured during the first quarter of 2010, i.e., prior to the random assignment. This ensured that the influx of 18,000 new teachers did not offset or in anyway respond to the randomly allocated pilot component.

ponent which seeks to promote a link between teacher performance and job tenure, (iii) local versus centralized recruitment and payment of contract teachers, and (iv) two alternative salary offers, equivalent to approximately $121 and $67 per month, respectively. Each of these four dimensions involves a binary choice, yielding four potential treatment cells and a pure control cell.

Contract teachers were assigned to teach either grade 2 or 3.[7] As noted above, the contract teacher intervention combines both a class-size effect and the effect of changing teacher incentives. Head teachers were instructed to split the class to which the new contract teacher was assigned, maximizing the reduction in class sizes in the assigned grade rather than re-allocating teachers across grades. For example, a school which, prior to the experiment, had a single civil service teacher instructing 70 grade 3 pupils would have been asked to split grade 3 into two classes, one taught by the pre-existing civil service teacher and the other taught by the contract teacher. As discussed below, compliance with these instructions was high but imperfect. Field monitors were able to ensure experimental teachers were assigned to the correct class, but had difficulty ensuring that other teaching staff were not reallocated to spread the teaching load more evenly.

The experimental sample focuses on schools with high pupil-teacher ratios. Within each of the eight provinces, districts were chosen non-randomly by the implementing partners, based in part on the location of the offices of the partnering NGO.[8] Within each province, schools with pupil-teacher ratio lower than the median were excluded from the sampling frame. Using this sampling frame of high PTR schools, schools were chosen through simple

---

[7]Half of the teachers in the experiment were randomly assigned to grade 2 in 2010, and half to grade 3 in 2010. In 2011, all the contract teachers were placed in grade 3. Thus there is some random variation in the length of direct exposure to the program within the treatment group.

[8]The sample draws from 14 districts in total, using multiple districts from the same province where necessary to reach sufficient sample size. These 14 districts were: Nairobi province (North, West, East); Central province (Muranga South); Coast province (Malindi); Eastern province (Moyale and Laisamis); North Eastern (Lagdera, Wajir South, Wajir West); Nyanza province (Kuria East and Kuria West); Rift Valley province (Trans Mara); Western province (Teso).

random sampling within the selected districts.

The effects of the randomized interventions will be measured by comparing baseline and follow-up academic assessments (exams) in math and English in 24 primary schools in each of Kenya's 8 provinces (192 total schools). The survey instruments were designed with the collaboration of Kenya National Examination Council (KNEC) to conform to the national curriculum.

The primary outcome variable of interest is pupil exam performance. The exams for this study were designed specifically for the evaluation by a team from KNEC and the Kenyan Institute for Education. The baseline survey - including pupil exams and questionnaires regarding pupil characteristics and school facilities - was conducted in July and August of 2009 by KNEC and the research team, with a sample of approximately 23,000 pupils. Teachers were placed, on a randomized basis, in either grade 2 and 3 in treatment schools in June 2010; their contracts ended in October 2011. Follow-up data collection was conducted in the same sample of schools in October 2011.

4   Randomization

To guarantee that the sample is balanced between treatment and control schools, an optimal multivariate matching algorithm was used (see Greevy, Lu, Silber and Rosenbaum (2004) and Bruhn and McKenzie (2009)). Treatment and control schools were matched along the following dimensions: results in nationwide end-of-primary leaving exams, results in Grade 1 baseline test, enrolment, no. of classrooms, no. of civil service teachers, no. contract teachers and average pay of PTA teachers at baseline. Table 1 shows that the randomization algorithm successfully balanced treatment and control schools along these dimensions. In practise, the algorithm created groups of 3 schools, which were matched along the above dimensions and then randomly assigned them to control, additional teacher with government

implementation and additional teacher with NGO implementation. The successful outcome of the randomization is reported in Table 1

In order to test whether the randomization was successful for our purposes, we check for balance using test score information collected at baseline that was not used in implementing the matching algorithm. Denote by Y_{ijt} the outcome of interest for pupil i in school j in period t. Let Z_j denote being randomly assigned treatment status, i.e. eligibility to receive an additional contract teacher. Let SMC_j denote the subset of treatment schools that are randomly assigned to receive SMC training. Finally, let $MOE_j = 1$ denote a treatment school where the intervention is implemented by the government and $(\text{NGO}_j = 0)$ a treatment school where the intervention is implemented by the NGO.

To examine whether the treatment and control schools are comparable prior to the intervention, we estimate

$$Y_{ij,t=0} = \alpha_0 + \beta_0 Z_j + \beta_0' Z_j \times \text{SMC}_j + \beta_0'' Z_j \times \text{MOE}_j + \beta_0''' Z_j \times \text{SMC}_j \times \text{MOE}_j + \varepsilon_{0ij,t=0}$$

using the baseline data. As seen from Table 2, none of the treatment dummies are significant, implying that test scores in treatment and control schools were indistinguishable prior to the intervention.

5 Compliance

Compliance with the intervention is described in Table 3. 56 (55) of the 64 schools assigned to the government (NGO) treatment arm were successful in hiring a contract teacher at some point during the programme. However, teachers did not necessarily stay with the school for the entire duration of the programme and when a vacancy opened up, it was not always filled. As a consequence, out of the 18 months of the programme, schools in the government

(NGO) arm actually employed a teacher for 11.59 (13) months on average. If we exclude schools that never employed a teacher from this calculation, the numbers rise to 13.25 and 15.13 months respectively. These differences are not significantly different across the NGO and government implementation arm.

Table 4 examines the vacancy rate more closely, modeling success in filling a vacancy as a function of various demand-side policies that were manipulated by the experiment, as well as other exogenous and/or predetermined school characteristics. The dependent variable is a binary indicator of whether a teacher was present and teaching in a given school in a given month, with monthly observations spanning the duration of the experiment from June 2010 to October 2011. We estimate both a linear probability model and a logit model, with and with-out controls for school characteristics.

We examine three experimental determinants of teacher labor supply. First, Table 4 shows that offering a "high" increases the probability of filling a teaching vacancy by just under 12%. This effect is significant and consistent between the LPM and logit models, but not robust to the inclusion of school-level controls. As noted above, a high and low salaries were equivalent to roughly $121 and $67 per month, respectively. Second, local control over teacher hiring and payment had an effect of similar magnitude to the salary differential, raising the probability of a filled vacancy by a robustly significant 14 to 16% across specifications. Third, NGO implementation led to between 12 and 17% more months with a filled vacancy, relative to the government treatment arm, and this effect is significant across all specifications. In addition, the correlation between the probability of filling the teacher vacancy in our intervention and the general thickness of the labor market – measured as the ratio of applicants to vacancies for the 18,000 teachers hired in 2010 – is very high at 0.1 [9]. This provides further evidence that failure to recruit a teacher was sensibly related to

---

[9]This is the coefficient in a regression of presence of a teacher on labor market thickness and a constant. It is significant at the 1% level with standard errors clustered at the school level.

local labor market conditions.

In addition to the difficulty in recruiting teachers, compliance with the experimental protocol was also imperfect due to salary delays. The final two rows of Table 3 summarize these delays in paying salaries. The average salary delay was 1 month in schools in the NGO implementation arm and more than twice as high – 2.33 months on average – in schools in the government implementation arm. In addition, there was large variation in the disbursement of salaries for schools where the intervention was administered by the Ministry of Education. The average maximum delay for that treatment arm was 5.56 months and 10% of teachers had to wait for their salaries for 10 months at some point.

The difference in salary delays highlights the different infrastructure for disbursing payments available to the government and the NGO. In the case of World Vision, salaries were disbursed by local staff in the Area Development Offices located in each of the 14 districts. These visited the schools on a monthly basis to collect names and payment details of teachers and then disbursed payments – overall in a timely fashion. In the case of government implementation, the Ministry of Education departed from its usual modality for paying teachers, which is done centrally by the Teacher Service Commission with essentially non-existent checks whether teachers on the central payroll are actually teaching in a school.[10] Instead, the Ministry requested its District Education and Staffing Officers to verify names and payment details of teachers and disbursed payments only after all the details were carefully cross-checked. This created long bottlenecks, which made the timely disbursement of salaries difficult.

---

[10]Accordingly, a recent report by the Kenya Anti Corruption Commission estimates that as many as 32,000 (or almost 14%) of the country's teachers could be 'ghost' workers (Siringi 2007).

6 Main results

We explore three different econometric specifications to measure the impact of an additional teacher on standardized test scores, Y_{ijt}: the intention-to-treat (ITT), a regression of test scores on a school's (endogenous) success in hiring and retaining a contract teacher, and an instrumental variables estimate of the local average treatment effect (LATE) for treated schools (i.e., the average treatment effect for the treated, or ATT).

The intention-to-treat (ITT) effect is measured by the coefficient on the random assignment variable Z_{jt} in equation (1).

$$Y_{ijt} = \alpha_1 + \beta_1 Z_{jt} + \gamma_1 \mathbf{X}_{ijt} + \varepsilon_{1ijt} \tag{1}$$

The coefficient β_1 measures the causal effect of being assigned to treatment status, averaging over schools with varying degrees of success in recruiting contract teachers.

For comparison we also present a 'naive' OLS regression of test scores on treatment status, where T_{jt} measures the number of months (out of a possible 18 months total duration of the program) that a contract teacher was in place in a given school.

$$Y_{ijt} = \alpha_2 + \beta_2 T_{jt} + \gamma_2 \mathbf{X}_{ijt} + \varepsilon_{2ijt} \tag{2}$$

Lastly, we use the random assignment to instrument actual treatment.

$$Y_{ijt} = \alpha_3 + \beta_3 \hat{T}_{jt} + \gamma_3 \mathbf{X}_{ijt} + \varepsilon_{3ijt} \tag{3}$$

where $\hat{T}_{jt}$ are the predicted values from the first-stage regression

$$T_{jt} = \alpha_4 + \delta_4 Z_{jt} + \gamma_4 \mathbf{X}_{ijt} + \varepsilon_{4ijt} \tag{4}$$

We estimate three variations of each of these equations – (1), (2), and (3) – with varying sets of controls ($\mathbf{X}_{ijt}$): first, a simple cross-sectional OLS regression with no controls; second, controlling for initial tests scores, $\bar{Y}_{j,t-1}$; and third, a school-level fixed effects regression.

The top panel of table 5 presents the results for each of these three estimates of impact. Columns 1 to 3 report an ITT effect of approximately 0.1 standard deviation. The point estimate is fairly consistent across all three specifications, though marginally significant only in the lagged dependent variable model.

Columns 4 to 6 of in the top panel of table 5 show the 'naive' OLS estimate of the effect of exposure to treatment on test scores. As seen, the effect is slightly larger than the ITT effect at between 0.1 to 0.14 standard deviations, but is insignificant across all three specifications. The treatment variable ranges from zero to one, where one implies a school employed a teacher for all 18 months of the program. Thus the point estimates can be interpreted as the comparison of a school with no teacher to one with a full 18-months' exposure to treatment.

Finally, columns 7 to 9 in the top panel of table 5 instrument the treatment variable using the random assignment. Under the assumption of no spillovers between schools, this measures the causal effect of the months of treatment on the treated – i.e., on pupils in schools which successfully recruited a teacher. The effect varies across specifications from roughly 0.12 to 0.17 standard deviations, and is statistically significant in the lagged dependent variable model.

To test whether the impact of an additional teacher varied according to any of the other cross-cutting interventions – SMC training, local versus centralized recruitment and payment of teachers, and high versus low salaries – we estimate treatment effects from each of the possible two-way interactions between Z and each individual cross-cut. (Our design lacks statistical power to explore three-way interactions, and the cross-cuts were not motivated by

hypotheses about substitutability or complementarity between the cross-cuts.)

$$Y_{ijt} = \alpha_5 + \beta_5 Z_{jt} + \beta_5' Z_{jt} \times \text{Cross-cut}_{jt} + \gamma_5 \mathbf{X}_{ijt} + \varepsilon_{5ijt} \tag{5}$$

The intention-to-treat (ITT) effect of the contract teacher in isolation is measured by the co-efficient on the random assignment variable Z_{jt}, and the coefficient $Z_{jt} \times \text{Cross-cut}_{jt}$ measures the marginal effect of each cross-cut training. Note that SMC training was not implemented in schools not receiving a contract teacher (and the recruitment and salary variations are meaningless in the absence of a contract teacher), thus the cross-cutting interventions do not enter the regressions except as interaction terms.

As above, we estimate three variations of each of these equations – (1), (2), and (3) – with varying sets of controls ($\mathbf{X}_{ijt}$): first, a simple cross-sectional OLS regression with no controls; second, controlling for initial tests scores, $\bar{Y}_{j,t-1}$; and third, a school-level fixed effects regression.

The top panel of table 6 presents the results for each of these three estimates of impact. In columns 1-3 the "cross-cut" variable is defined as SMC training, in columns 4-6 as local hiring, and in columns 7-9 as high salary. We do not find a significant marginal effect of any of the cross-cutting interventions in any of the specifications. Interestingly, while paying higher salaries and giving SMCs direct control over teacher employment led to greater success in filling vacancies (as shown in Table 4), this did not have a discernible effect on test scores.

7   Comparative effectiveness of government and NGO programs

In this section we examine the differential effect of the contract teacher program when administered by an international NGO and the Ministry of Education (MOE). The bottom

panel of table 5 repeats estimation from the top panel, allowing for the effect to differ by implementing agency. In each case, we regress scores on a treatment variable and the treatment variable interacted with a dummy for MOE implementation. Thus for the ITT we estimate

$$Y_{ijt} = \alpha_6 + \beta_6 Z_{jt} + \beta_6'' Z_{jt} \times \text{MOE}_{jt} + \gamma_6 \mathbf{X}_{ijt} + \varepsilon_{6ijt} \tag{6}$$

and analogously for equation (2). In the IV model, we instrument the endogenous treatment variable T_{jt}, and its interaction with the MOE dummy, $T_{jt} \times \text{MOE}_{jt}$, with the random assignment Z_{jt} and $Z_{jt} \times \text{MOE}_{jt}$.

Across all specifications, the results consistently suggest that the overall effect of the CSTP is driven by the NGO program, with a weaker effect of MOE implementation. Columns 1 to 3 in the bottom panel of table 5 compare the causal of effect of assignment to NGO versus MOE implementation of the CSTP. The coefficient on Z_{jt} shows that NGO implementation raises scores by 0.13 to 0.16 standard deviations. (This coefficient is statistically significant at the 5% level in the lag dependent model and at the 10% level in the fixed effects model.) The coefficient on $Z_{jt} \times \text{MOE}_{jt}$ shows the relative effect of moving from NGO to MOE implementation. This effect consistently negative, but imprecisely estimated.

Columns 4 to 6 in the bottom panel of table 5 report the results from the 'naive' OLS estimates comparing the effect of NGO and MOE treatment on test scores. The point estimates range from 0.16 to 0.2 standard deviations, and are significant at the 5% level in both the lag dependent and fixed effects models. Finally, columns 7 to 9 reported the ATT based on the instrumented regression. Here we see a large, causal effect of NGO implementation – equivalent to 0.22 to 0.23 standard deviations – which is statistically significant at the 5% level in the lag dependent variable model and at the 10% level in the fixed effects model. It is noteworthy that in all of the specifications where the NGO treatment effect is statistically significant, the marginal effect of MOE treatment is of approximately

equal magnitude and opposite sign – implying roughly zero impact of an additional teacher under MOE implementation.

To assess whether the impact of SMC training and the other cross-cuts differs by implementing agency, we repeat the analysis in equation (5), including additional interaction terms as follows:

$$
\begin{aligned}
Y_{ijt} \;=\; & \alpha_7 + \beta_7 Z_{jt} + \beta_7' Z_{jt} \times \text{Cross-cut}_{jt} + \beta_7'' Z_{jt} \times \text{MOE}_{jt} + \qquad (7) \\
& \beta_7''' Z_{jt} \times \text{MOE}_{jt} \times \text{Cross-cut}_{jt} + \gamma_7 \mathbf{X}_{ijt} + \varepsilon_{7ijt}
\end{aligned}
$$

and analogously for the OLS and instrumented regression. The coefficient on the three-way interaction term, β_7''', measures the differential impact of, e.g., SMC training in a school where the contract teacher program was implemented by the government compared with a school where it was implemented by the NGO.

The results are reported in Table 6. As above, there is no significant effect of SMC training overall and no significant difference between the two implementing agencies – though it should be noted that the sign of β_7''' is positive in the fixed effects and lagged dependent variable specification indicating that the government was perhaps slightly more successful in implementing the training. Overall, however, the conclusions remain unchanged: hiring an additional teacher has a positive and significant impact only for the NGO treatment arm.

8 Heterogeneous effects

In addition to the institutional considerations raised above, a more traditional concern about the generalizability of RCT results is external validity. The estimates in Table 5 provide an unbiased estimate of the intention-to-treat effect for schools within the sampling frame – i.e., schools with high pupil-teacher ratios in the 14 study districts. If the treatment

effect varies with school or pupil characteristics, and the sampling frame differs from the population of interest for policymaking, the results will not be broadly applicable. Estimation of heterogeneous treatment effects, combined with knowledge of the distribution of exogenous characteristics in the sample and population, may provide a bridge from internal to external validity .

Two issues to be addressed in estimating heterogeneous effects are (i) selecting the dimensions of heterogeneity, and (ii) hypothesis testing with multiple comparisons (Green and Kern 2010). On the former question, the literature on medical trials commonly takes a data-driven approach based on boosting algorithms (Friedman, Hastie and Tibshirani 2000). Boosting is particularly well-suited to the design of optimal treatment regimens for a particular sub-group. An alternative approach to studying heterogeneity, more common in the social sciences and which we use here, is hypothesis driven. Specific interaction terms, $\mathbf{X}_{jt}$, are proposed based on *ex ante* hypotheses and tested in an extension of equation (1) including school fixed effects.

$$ Y_{ijt} = \alpha_8 + \beta_8 Z_{jt} + \beta_8^x \left(Z_{jt} \times \frac{\mathbf{X}_{jt} - \mu_x}{\sigma_x} \right) + \gamma_8 \mathbf{X}_{ijt} + \varepsilon_{8ijt} \tag{8} $$

We explore three hypotheses. The first is that the intervention's effect will be stronger where the supply of teachers is higher, reducing the risk of unfilled vacancies and potentially increasing contract teachers' motivation to maintain employment. As a rough proxy for the supply of teachers in a given area, we use the count of other primary schools within a 5-mile radius of the school.

Our second hypothesis about heterogeneity is that the addition of a contract teacher will have a larger effect in schools with a higher initial pupil-teacher ratio, as these schools will experience a larger reduction in class size due to treatment. Finally, our third hypothesis the treatment will be more effective in schools with lower initial test scores. This hypothesis is

more speculative, but is motivated by the attention paid to tracking and remedial education in the contract teacher literature (Banerjee et al. 2007, Duflo et al. 2009).

Table 7 shows the results from estimating the heterogeneous ITT effects in equation (8). Because the variables measuring exogenous heterogeneity have been standardized, the all coefficients can be interpreted as the change in the treatment effect implied by a one standard-deviation change in the independent variable. For instance, column 1 shows that the ITT is roughly 4 percentage points smaller in locations with a higher density of schools, contradicting our hypothesis – though this effect is entirely insignificant. Column 2 shows no consistent relationship between initial pupil-teacher ratios and the treatment effect. Turning to our third hypothesis, we explore two measures of schools' initial level of academic achievement: scores on an independent national standardized test administered to grade 8 pupils in 2005, and scores on the baseline test used in the primary analysis here. Column 3 shows no relationship between scores on the national test and treatment effects. Column 4, however, shows a significantly negative relationship between initial test scores in the baseline and subsequent treatment effects. While the average ITT for schools with NGO-implementation was roughly $1/5^{th}$ of a standard deviation, column 4 implies this effect was only half as large in schools one standard deviation above the mean.

So far we have ignored the issues raised by conducting multiple comparisons. Testing m null hypotheses at a significance level of α, Boole's inequality implies that at least one null will be rejected with probability less than or equal to $m\alpha$. The Bonferroni correction limits this probability, known as the family-wise error rate (FWER), by testing each individual hypothesis against the corrected critical value $\alpha' = \alpha/m$. As Fink, McConnell and Vollmer (2011) show, the Bonferroni correction is quite conservative, in the sense of controlling Type I errors at the expense of more Type II errors (less power) vis-a-vis available alternatives. Benjamini and Hochberg (1995) and subsequent authors have proposed alternatives which minimize the false-discovery rate (FDR) rather than the FWER, defined as the proportion

of the rejected null hypothesis which are erroneously rejected, leading to greater power.[11]

We apply Benjamini and Hochberg (1995) method to the estimates in Table 7. The correction does not affect the coefficients or standard errors, but rather the critical value (p-value) used to establish statistical significance. As shown in Figure 3, this amounts to literally 'raising the bar' for statistical significance. In our particular example, the results are fairly unremarkable: only one interaction term in Table 7 was statistically significant at the 5% level when considered in isolation – the interaction of the Ministry of Education treatment arm with baseline test scores. This effect remains significant at the equivalent of the 5% level (now 0.625%) using corrected p-values.

9   Conclusion

Contract teachers have been shown to significantly raise test scores for primary students in previous randomized trials in both Western Kenya (Duflo et al. 2009) and various locations in India (Banerjee et al. 2007, Muralidharan and Sundararaman 2010). This paper sought to identify the relevant obstacles to successfully scaling up this proven intervention to the national level in Kenya.

We reported on the randomized evaluation of the pilot phase of a Kenyan government program to scale-up the contract teacher model by (a) expanding geographically across all eight Kenyan provinces, and (b) shifting implementation from NGOs to the Ministry of Education.

On the first point, our results are overwhelmingly positive. We find a significant, positive ITT effect of 0.12 standard deviations on combined math and English scores across all treatment arms in our national sample. Geographic heterogeneity in treatment effects does not appear to pose a significant threat to the external validity of earlier studies. The effect

---

[11]For an intuitive exposition of the advantages of FDR versus FWER corrections see Fink et al. (2011). For a comprehensive classification of the corrections proposed to date, see Newson (2003).

of the program was not notably diminished in remote areas or under-performing schools. Indeed, effects are somewhat larger for schools with lower initial scores.

Rather than focusing on external validity as commonly conceived – i.e., out of sample predictions – our results suggest that the more relevant obstacle to scaling up successful education interventions in a context such as Kenya may be the broader institutional structure through which school-level interventions are implemented. As noted in Section 7, contract teachers were exclusively effective in the sub-group of treatment schools administered by an international NGO. Effects were significantly smaller and indistinguishable from zero in schools where the program was administered by the Ministry of Education.

References

Baird, Sarah, Craig McIntosh, and Berk Özler, "Cash or Condition? Evidence from a Cash Transfer Experiment," *World Bank Policy Research Working Paper*, 2010, *5259.*

Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster, , and Stuti Khemani, "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," *American Economic Journal: Economic Policy*, 2010, *2* (1), 1–30.

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

Barr, Abigail, Frederick Mugisha, Pieter Serneels, and Andrew Zeitlin, "Information and collective action in the community monitoring of schools: Field and lab experimental evidence from Uganda," mimeo, Centre for the Study of African Economies, Oxford 2011.

Barrera-Osorio, Felipe and Leigh Linden, "The Use and Misuse of Computers in Education: Evidence from a Randomized Controlled Trial of a Language Arts Program," 2009.

Benjamini, Y. and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, 1995, *Series B*, 289–300.

Berry, James, "Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India," 2011.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, and Justin Sandefur, "Why Did Abolishing Fees Not Increase Public School Enrollment in Kenya?," *Center for Global Development Working Paper Series*, 2011, *271.*

Bruhn, Miriam and David McKenzie, "In Pursuit of Balance: Randomization in Practice in Development Field Experiments," *American Economic Journal: Applied Economics*, October 2009, *1* (4), 200–232.

Burde, Dana and Leigh Linden, "The Effect of Village-Based Schools: Evidence from a Randomized Controlled Trial in Afghanistan," 2010.

Deaton, Angus, "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature*, 2010, *48* (2), 424–455.

Duflo, Esther, "Scaling Up and Evaluation," *Annual World Bank Conference on Development Economics*, 2004, pp. 341–369.

—— , **Pascaline Dupas, and Michael Kremer**, "Additional Resources versus Organizational Changes in Education: Experimental Evidence from Kenya," 2009.

—— , —— , **and** —— , "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 2011, *101* (5).

—— , **Rema Hanna, and Stephen Ryan**, "Incentives Work: Getting Teachers to Come to School," 2010.

Fink, Gunther, Margaret McConnell, and Sebastian Vollmer, "Testing for Heterogeneous Treatment Effects in Experimental Data: False Discovery Risks and Correction Procedures," 2011.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, 2000, *28* (2), 337–407.

Glewwe, Paul, Albert Park, and Meng Zhao, "A Better Vision for Development: Eyeglasses and Academic Performance in Rural Primary Schools in China," 2011.

____ **and Eugenie Maïga**, "The Impacts of School Management Reforms in Madagascar: Do the Impacts Vary by Teacher Type?," 2011.

____ **, Michael Kremer, and Sylvie Moulin**, "Many Children Left Behind? Textbooks and Test Scores in Kenya," *American Economic Journal: Applied Economics*, 2009, *1* (1), 112–135.

____ **, ____ , ____ , and Eric Zitzewitz**, "Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya," *Journal of Development Economics*, 2004, *74*, 251–268.

____ **, Nauman Ilias, and Michael Kremer**, "Teacher Incentives," *American Economic Journal: Applied Economics*, 2010, *2*, 205–227.

Green, Donald P. and Holger L. Kern, "Modeling Heterogenous Treatment Effects in Large-Scale Experiments using Bayesian Additive Regression Trees," 2010.

Greevy, Robert, Bo Lu, Jeffrey Silber, and Paul Rosenbaum, "Optimal multivariate matching before randomization," *Biometrika*, 2004, *5* (2), 263–275.

He, Fang, Leigh Linden, and Margaret MacLeod, "How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program," 2008.

____ **, ____ , and ____** , "A Better Way to Teach Children to Read? Evidence from a Randomized Controlled Trial," 2009.

Inamdar, Parimala, "Computer skills development by children using 'hole in the wall' facilities in rural India," *Australasian Journal of Educational Technology*, 2004, *20* (3), 337–350.

Kazianga, Harounan, Damien de Walque, and Harold Alderman, "Educational and Health Impacts of Two School Feeding Schemes: Evidence from a Randomized Trial in Rural Burkina Faso," *World Bank Policy Research Working Paper*, 2009, *4976.*

Kremer, Michael, Edwar Miguel, and Rebecca Thornton, "Incentives to Learn," *The Review of Economics and Statistics*, 2009, *92* (3), 437–456.

⸻ , **Sylvie Moulin, and Robert Namunyu**, "Decentralization: A Cautionary Tale," *Poverty Action Lab Paper No. 10*, 2003.

Linden, Leigh, "Complement or Substitute? The Effect of Technology on Student Achievement in India," 2008.

⸻ , **Abhijit V Banerjee, and Esther Duflo**, "Computer-Assisted Learning: Evidence from a Randomized Experiment," *Poverty Action Lab Paper No. 5*, 2003.

Macours, Karen, Norbert Schady, and Renos Vakis, "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment," *Human Capital and Economic Opportunity: A Global Working Group*, 2011, *2011-007.*

Muralidharan, Karthik, "Long-Term Effects of Teacher Performance Pay: Experimental Evidence from India," 2011.

⸻ **and Venkatesh Sundararaman**, "Contract Teachers: Experimental Evidence from India," 2010.

____ **and** ____ , "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39–77.

Newson, Roger, "Multiple-test procedures and smile plot," *The Stata Journal*, 2003, *3* (2), 109–132.

Pandey, Priyanka, Sangeeta Goyal, and Venkatesh Sundararaman, "Community Participation in Public Schools: The Impact of Information Campaigns in Three Indian States," *World Bank Policy Research Working Paper*, 2008, *4776.*

Paxson, Christina and Norbert Schady, "Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador," *World Bank Policy Research Working Paper*, 2007, *4226.*

Rosas, Ricardo, Miguel Nussbaum, Patricio Cumsille, Vladimir Marianov, Monica Correa, Patricia Flores, Valeska Grau, Francisca Lagos, Ximena Lopez, Veronica Lopez, Patricio Rodriguez, and Marcela Salinas, "Beyond Nintendo: design and assessment of educational video games for rst and second grade students," *Computers and Education*, 2003, *40*, 71–94.

Shadish, William R., Thomas D. Campbell, and Donald T. Cook, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin Company, 2002.

Siringi, Samuel, "Kenya: Exposed – Country's 32,000 Ghost Teachers," August 2007. Published online, Aug. 11, 2007, at http://allafrica.com/stories/200708110007.html.

Teyie, Andrew and Henry Wanyama, "Losses in FPE rise to Sh5.5 billion," *Nairobi Star*, 2010, *8 January.* Available online at http://allafrica.com.

Vermeersch, Christel and Michael Kremer, "School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation," *World Bank Policy Research Working Paper*, 2005, *3523*.

10   Appendix: Figure and Tables

Table 1: Results of optimal multivariate matching algorithm

	Control	Treatment	Difference
Enrolment	43.33	53.26	9.935
			(7.418)
No. of classrooms	11.76	12.48	.715
			(1.046)
No. of civil service teachers	10.02	10.21	.195
			(1.002)
No. of contract teachers	1.90	2.27	.369
			(.347)
Average pay for contract teacher	2843	3393	550.103
			(531.535)
KCPE	239.48	235.083	-4.396
			(6.783)
Grade 1 English	.028	.074	.046
			(.166)
Grade 1 Maths	.060	.063	.003
			(.156)

Regressions based on 192 schools, collapsed at school level

Table 2: Differences in test scores in treatment and control schools prior to the intervention

	(1)	(2)	(3)	(4)
Z	.087	.039	.070	-.018
	(.083)	(.095)	(.094)	(.108)
Z × MOE		.099		.180
		(.098)		(.133)
Z × SMC			.034	.113
			(.099)	(.137)
Obs.	6,264	6,264	6,264	6,264

Regressions based on 174 schools. Standard errors are clustered at the school level.

Table 3: Compliance with the intervention protocol

	Control	Government	NGO
Schools that (ever) employed a teacher	0	56	55
Months of teacher	0	11.59	13
Months of teacher (conditional on employing a teacher)	0	13.25	15.13
Avg. months salary delay	0	2.33	NA
Avg. maximum months of salary delay	0	5.56	NA
No of. obs	64	64	64

Figures on salary delays are not currently available for the NGO treatment arm.

Table 4: Labor supply of contract teachers

	Linear Probability Model			Logit Model	
	(1)	(2)		(3)	(4)
High salary	.116	.087		.115	.089
	(.064)*	(.068)		(.064)*	(.068)
NGO implementation	.123	.166		.124	.170
	(.065)*	(.064)***		(.066)*	(.067)**
Local recruitment	.143	.162		.144	.157
	(.065)**	(.063)**		(.066)**	(.067)**
Geographic density		-.004			-.003
		(.002)**			(.002)*
Lagged KCPE score		.001			.002
		(.001)			(.001)
Pupil-teacher ratio		.003			.004
		(.002)			(.003)
Obs.	2,044	1,977		2,044	1,977

The unit of observation is the school, with monthly observations from June 2010 to October 2011. The dependent variable is a binary indicator of whether a teacher was present and teaching in a given school in a given month. Columns 1 and 3 restrict the determinants of teacher presence to factors controlled by the experiment, while columns 2 and 4 include other exogenous and/or predetermined school characteristics. For the logit model, the table reports marginal effects and their standard errors. All standard errors are clustered at the school level.

Table 5: Treatment effects

	ITT			OLS			ATT		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pooling treatment arms:									
Z	.141	.115	.083						
	(.077)*	(.068)*	(.076)						
T				.139	.129	.122	.205	.166	.119
				(.089)	(.075)*	(.081)	(.113)*	(.098)*	(.108)
NGO vs MOE implementation:									
Z	.163	.190	.180						
	(.094)*	(.078)**	(.084)**						
$Z\times$ MOE	-.043	-.151	-.197						
	(.102)	(.083)*	(.085)**						
T				.167	.219	.238	.224	.259	.245
				(.111)	(.086)**	(.088)***	(.129)*	(.106)**	(.114)**
$T\times$ MOE				-.063	-.199	-.258	-.039	-.201	-.270
				(.133)	(.110)*	(.111)**	(.151)	(.121)*	(.122)**
Lag dependent variable		X			X			X	
School fixed effects			X			X			X
Obs.	8,711	8,154	14,975	8,711	8,154	14,975	8,711	8,154	14,975

The dependent variable in all columns is a standardized score on a math and English test administered to pupils in grades 1, 2 and 3 in 2009 and grades 3 and 4 in 2011. Columns 1, 4 and 7 use only the 2011 (follow-up) test data. Z represents an indicator variable for random assignment to any treatment arm; T is a continuous, and potentially endogenous, treatment variable measuring months of exposure to a contract teacher; MOE is an indicator variable for the Ministry of Education treatment arm. Standard errors are clustered at the school level.

32

Table 6: Intention to treat effects of cross-cutting interventions

	SMC Training			Local Hiring			High Salary		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pooling treatment arms:									
Z	.113	.099	.085	.150	.158	.135	.125	.117	.091
	(.086)	(.076)	(.087)	(.088)*	(.079)**	(.089)	(.081)	(.072)	(.081)
$Z\times$ Cross-cut	.057	.033	-.003	-.018	-.084	-.101	.065	-.009	-.036
	(.103)	(.084)	(.086)	(.102)	(.083)	(.086)	(.128)	(.092)	(.083)
NGO vs MOE implementation:									
Z	.129	.193	.219	.090	.185	.211	.162	.192	.188
	(.100)	(.088)**	(.101)**	(.098)	(.089)**	(.108)*	(.095)*	(.084)**	(.093)**
$Z\times$ MOE	-.033	-.185	-.272	.122	-.055	-.160	-.074	-.148	-.194
	(.128)	(.107)*	(.116)**	(.132)	(.117)	(.126)	(.113)	(.096)	(.102)*
$Z\times$ Cross-cut	.068	-.007	-.079	.147	.010	-.062	.003	-.010	-.034
	(.149)	(.114)	(.113)	(.147)	(.113)	(.113)	(.202)	(.131)	(.106)
$Z\times$ Cross-cut $\times$ MOE	-.022	.073	.153	-.329	-.183	-.064	.124	-.017	-.017
	(.206)	(.167)	(.169)	(.203)	(.163)	(.168)	(.255)	(.182)	(.161)
Lag dependent variable		X			X			X	
School fixed effects			X			X			X
Obs.	8,711	8,154	14,975	8,711	8,154	14,975	8,711	8,154	14,975

See notes for table 5. Columns 1, 4 and 7 use only the 2011 (follow-up) test data. Z represents an indicator variable for random assignment to any treatment arm; MOE is an indicator variable for the Ministry of Education treatment arm; SMC is an indicator variable for the SMC training treatment arm. In each column, the 'cross-cut' variable – denoting a cross-cutting experimental treatment or variation of the contract-teacher treatment – is defined according to the column heading. Standard errors are clustered at the school level.

33

Table 7: Heterogeneous treatment effects

	(1)	(2)	(3)	(4)
$Z\times$ MOE	.045	-.017	-.017	.011
	(.089)	(.089)	(.089)	(.085)
$Z\times$ NGO	.223	.193	.181	.173
	(.086)***	(.084)**	(.084)**	(.084)**
$Z\times$ MOE $\times$ Density	-.039			
	(.068)			
$Z\times$ NGO $\times$ Density	-.043			
	(.057)			
$Z\times$ MOE $\times$ PTR		-.046		
		(.054)		
$Z\times$ NGO $\times$ PTR		.088		
		(.056)		
$Z\times$ MOE $\times$ KCPE			-.033	
			(.057)	
$Z\times$ NGO $\times$ KCPE			.046	
			(.054)	
$Z\times$ MOE $\times Y_1$				-.185
				(.060)***
$Z\times$ NGO $\times Y_1$				-.101
				(.055)*
Obs.	14,475	14,975	14,975	14,418

See notes for table 5. All equations include school fixed effects and standard errors are clustered at the school level.
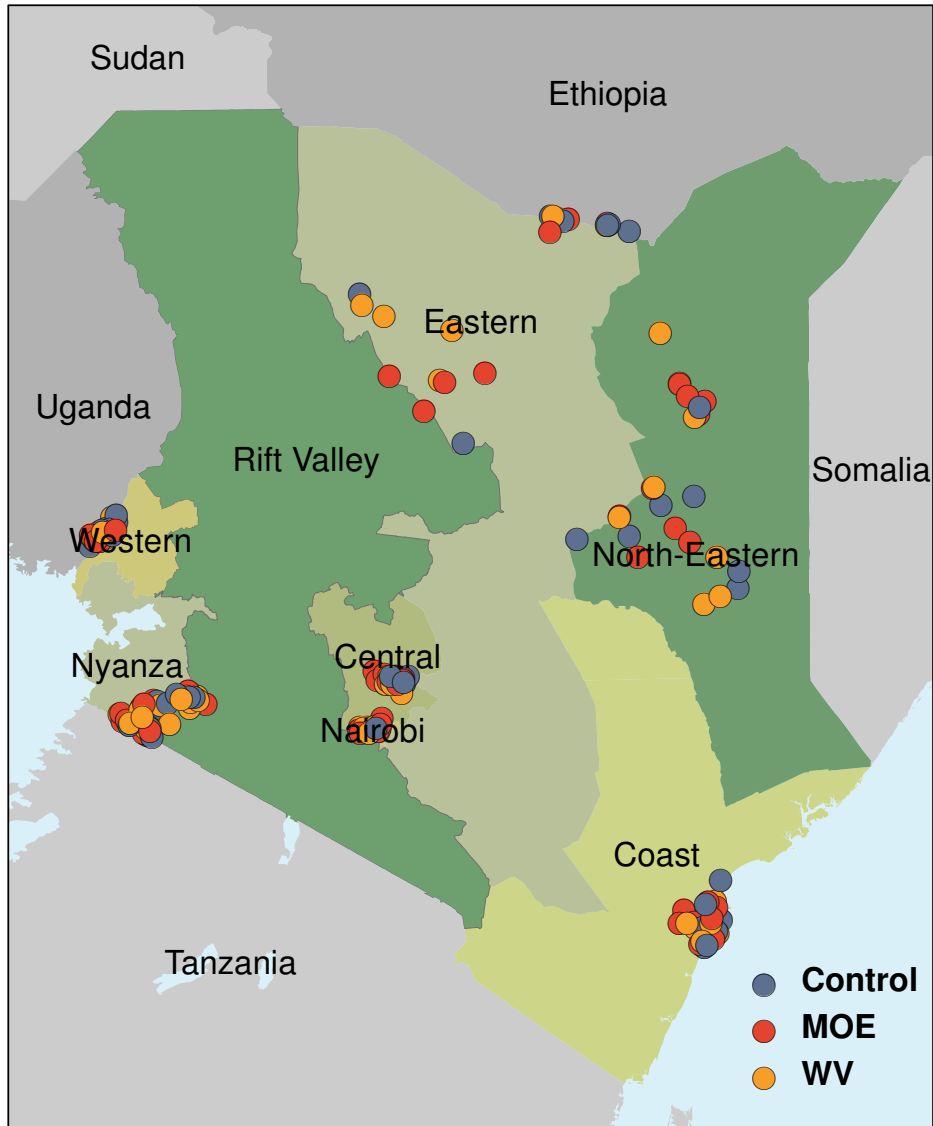
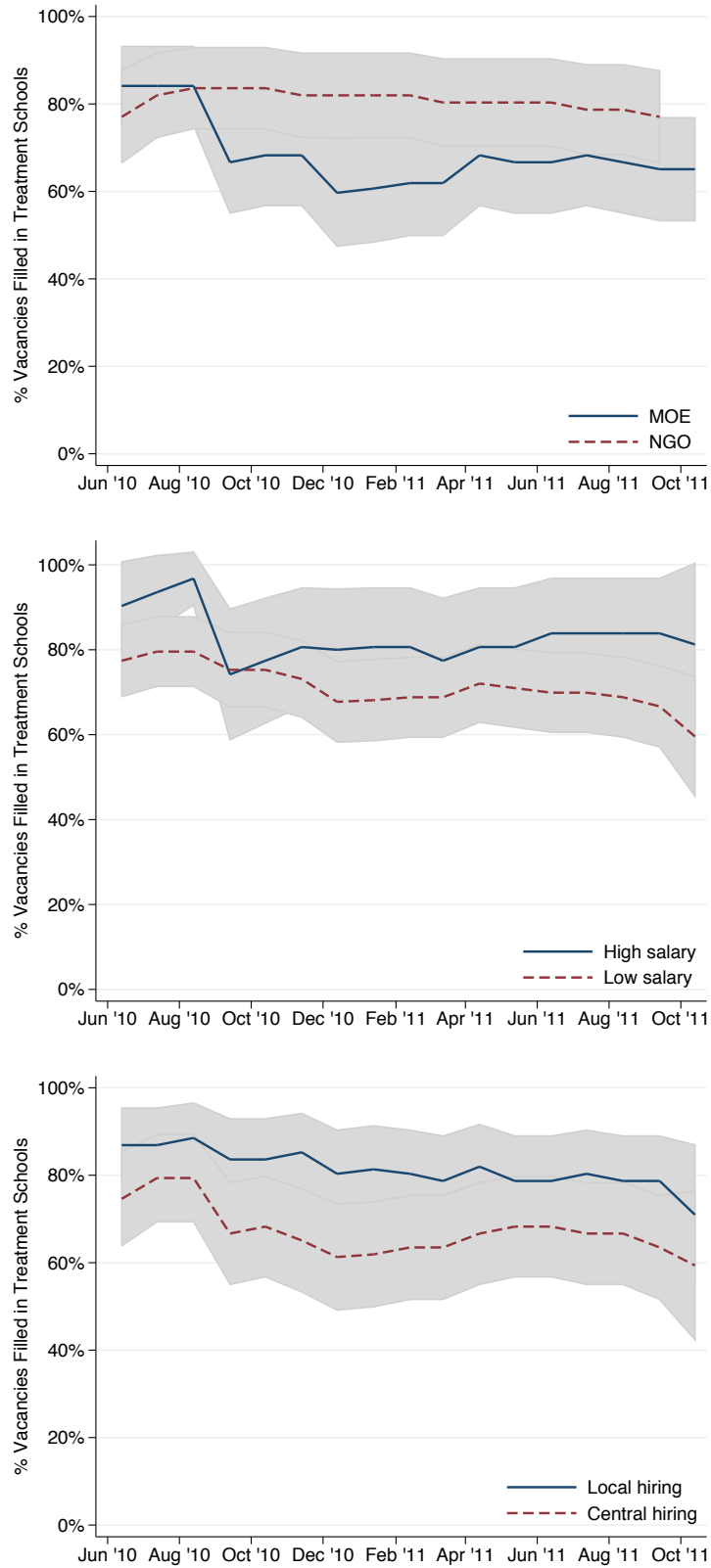Figure 1: Treatment & control sites across Kenya's 8 provinces

Figure 2: Proportion of contract teacher vacancies filled during evaluation
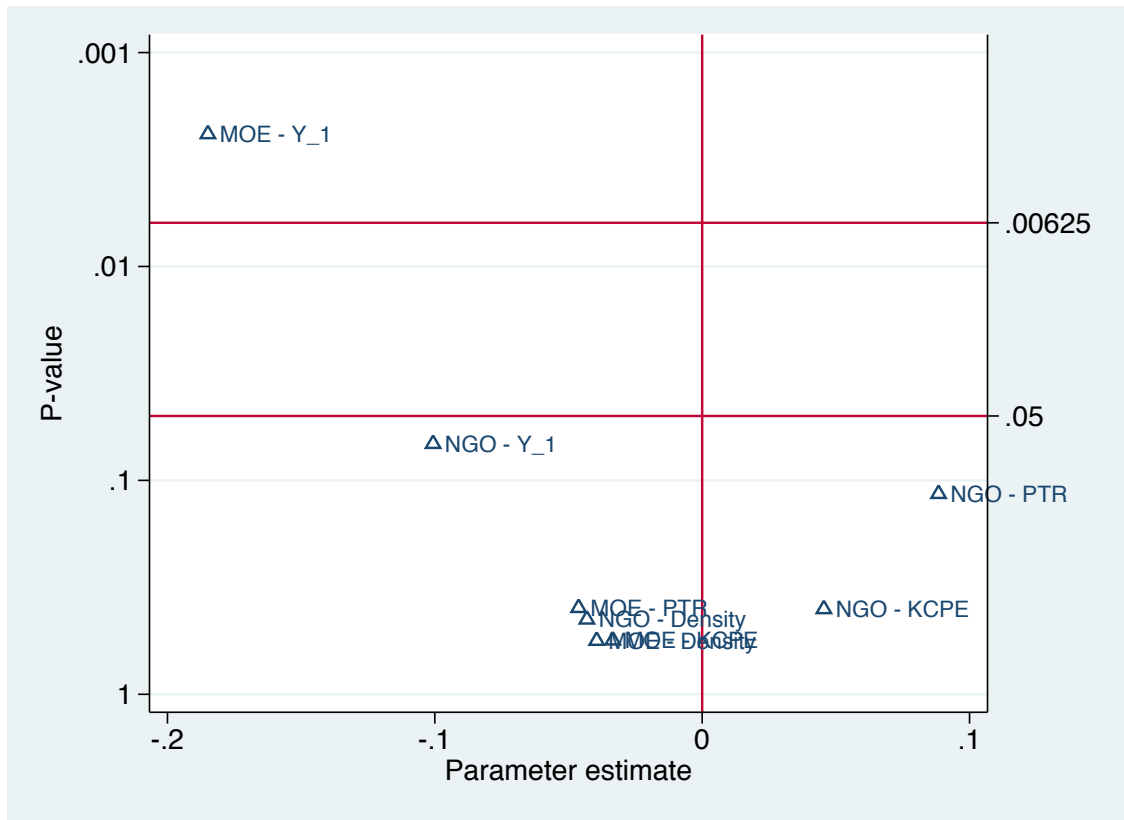
Figure 3: Heterogeneous treatment effects with p-values corrected for multiple comparisons. Each point represents a coefficient reported in Table 7. Points above the lower horizontal line are statistically significant when considered in isolation; points above the upper horizontal line remain significant with correct p-values.

The International Growth Centre
(IGC) aims to promote sustainable
growth in developing countries
by providing demand-led policy
advice based on frontier research.

Find out more about
our work on our website
www.theigc.org

For media or communications
enquiries, please contact
mail@theigc.org

Subscribe to our newsletter
and topic updates
www.theigc.org/newsletter

Follow us on Twitter
@the_igc

Contact us
International Growth Centre,
London School of Economic
and Political Science,
Houghton Street,
London WC2A 2AE

IGC
**International
Growth Centre**