

Teacher Incentives

Paul Glewwe*, Nauman Ilias**, and Michael Kremer***

Abstract

Advocates of teacher incentive programs argue that they can strengthen weak incentives, while opponents argue they lead to “teaching to the test.” We find evidence that existing teacher incentives in Kenya are indeed weak, with teachers absent 20% of the time. We then report on a randomized evaluation of a program which provided primary school teachers in rural Kenya with incentives based on students' test scores and dropout rates. Students in program schools had higher test scores, significantly so on at least some exams, during the time the program was in place. An examination of the channels through which this effect took place, however, provides little evidence of more teacher effort aimed at increasing long-run learning. Teacher attendance did not improve, homework assignment did not increase, dropout rates did not fall, and pedagogy did not change. There is, however, evidence that teachers increased effort at manipulating short-run test scores. Conditional on being enrolled, students in treatment schools were more likely to take tests. Teachers in treatment schools were more likely to conduct test preparation sessions. While students in treatment schools scored higher than their counterparts in comparison schools during the life of the program, they did not retain these gains after the end of the program, consistent with a model in which teachers focused on manipulating short-run scores.

* Department of Applied Economics, University of Minnesota. E-mail: pglewwe@appec.umn.edu

**Competition Economics, Inc., Washington, DC. E-mail: niliias@competitioneconomics.com

*** Department of Economics, Harvard University; The Brookings Institution; Center for Global Development NBER. E-mail: mkremer@fas.harvard.edu

We would like to thank Rachel Glennerster, Ed Kaplan, Janina Matuszeski, and Courtney Umberger for very helpful comments and assistance. We are especially grateful to Sylvie Moulin and Robert Namunyu for work in the field and to Emily Oster for outstanding research assistance in the U.S. We thank the World Bank and the MacArthur Foundation for financial support.

1. Introduction

Teacher incentive programs have enjoyed growing popularity. In the United States, a number of teacher incentive programs have been introduced in the past decade, generally offering annual merit pay on the order of 10% to 40% of an average teacher's monthly salary (American Federation of Teachers, 2000).¹ Under the No Child Left Behind (NCLB) act, passed in 2001, poorly performing schools face sanctions across the United States. Israel has provided incentives to teachers based on students' scores (Lavy 2002a, b) and a World Bank program in Mexico will provide performance incentives to primary school teachers.

Advocates of incentive pay for teachers note that teachers currently face weak incentives, with pay determined almost entirely by educational attainment, training, and experience, rather than performance (Harbison and Hanushek, 1992; Hanushek, Kain, and Rivkin, 1998; Hanushek, 1996; Lockheed and Verspoor, 1991), and argue that linking teachers' pay to students' performance would increase teacher effort.

Opponents of test score-based incentives argue that since teachers' tasks are multi-dimensional and only some aspects are measured by test scores, linking compensation to test scores could cause teachers to sacrifice promoting curiosity and creative thinking in order to teach skills tested on standardized exams (Holmstrom and Milgrom, 1991; Hannaway, 1992).

In many developing countries, incentives for teachers are even weaker than in developed countries. Teachers have very strong civil service protection and are not fired even for behavior that would be unacceptable in developed countries. Thus, for example, the PROBE report in India found that head teachers were absent from their school one-third of the time (PROBE,

¹ Examples include programs in Rhode Island in 1999, Denver in 1999-2000, Douglas County, Colorado beginning in 1994 and Iowa beginning in 2001 (Olsen, 1999; Education Commission of the States, 2000). A 1999 program in California offered a one-time award of \$25,000 to teachers in under-performing schools whose students showed substantial gains (Olsen, 1999).

1999). Our data suggests absenteeism is also a problem in our sample schools, nor is it limited to a few teachers with very high absentee rates. Instead, the majority of teachers are absent from school 20% of the time and absent from their classrooms even more frequently.

In environments with very weak incentives, it could be argued that the key problem is to get teachers to show up at all, and that the risk of diverting them from stimulating creativity is secondary. Given that most teaching in many developing countries is by rote, the risk of reducing efforts to stimulate creativity may seem remote. On the other hand, if incentive systems are very weak, schools could potentially respond to test score-based incentives in more pernicious ways than teaching to the test. For example, they could deliberately force students to repeat grades or even drop out in order to raise average scores on the exam.

Some local school committees in Kenya have responded to the existence of weak incentives by providing bonuses to teachers whose students perform well on exams. This paper reports on a randomized evaluation of a program along these lines that provided incentives to teachers in 50 rural schools (chosen quasi-randomly from a pool of 100) based on the average test score of students enrolled at the start of the program, with penalties for students who did not take the test, so as to discourage dropouts. The program provided prizes valued at up to 43% of typical monthly salary to teachers in grades 4 to 8 based on the performance of the school as a whole on the Kenyan government's districtwide exams that year. This ratio of prize to salary is similar to that used in typical U.S. incentive programs.

During the life of the program, students in treatment schools were more likely to take exams, and scored higher, at least on some measures. An examination of the channels through which this effect took place, however, provides little evidence of more teacher effort aimed at increasing long-run learning. Teacher attendance did not improve, homework assignment did not

increase, and pedagogy did not change. There is, however, evidence that teachers increased effort at manipulating short-run test scores. Conditional on being enrolled, students in treatment schools were more likely to take tests. Teachers in treatment schools were more likely to conduct test preparation sessions. While students in treatment schools scored higher than their counterparts in comparison schools during the life of the program, they did not retain these gains after the end of the program, consistent with a model in which teachers focused on manipulating short-run scores.

There is evidence that teachers learned how to adjust to the system over time. Test preparation sessions increased from the first to the second year of the program, as did the gap between treatment and comparison schools in exam participation rates and overall test scores.

1.1 Related Literature

A number of earlier papers examine the impact of linking teacher pay to students' test scores. Lavy (2002a) finds that an Israeli program providing teachers individual cash prizes for increases in student test scores on a high-school matriculation exam increased high school matriculation exam rates from 42% to 45.3%. At 60% to 300% of the average monthly salary, the prizes given in this case were much larger than those in most teacher incentive programs in the U.S. Lavy (2002b) finds that rewarding Israeli teachers based on school average performance (rather than individual performance) increased test scores and participation in matriculation exams, but not the percentage of students receiving matriculation certificates.

Jacob (2002) explores a Chicago program in which students with low test scores were not promoted to the next grade and schools and teachers were put on probation. He finds that the program increased student achievement, although the improvement was larger in skill sets used

on the high-stakes exam. Some schools manipulated scores by putting more students in special education classes. Figlio and Winciki (2002) show that school districts in Virginia increase the number of calories in school lunches on days when high-stakes tests are administered, thus artificially inflating test scores. Richards and Sheu (1992) argue that a South Carolina program in which schools were rewarded for increases in student test scores and both student and teacher attendance increased student achievement modestly, particularly in schools that began with lower achievement levels, but that it did not affect student or teacher attendance.

This paper differs from earlier work in three ways. First, since both advocates and opponents of teacher incentive programs agree that incentives can increase effort test scores, but disagree about whether these higher test scores would be due to increased overall teacher effort or more teaching to the test, we measure not only how teacher incentives affect test scores, but also how they affect different types of teacher effort. Second, since teacher incentive programs are likely to be introduced in areas where teacher performance is worse than expected, and since the introduction of teacher incentives may be correlated with other factors affecting teacher performance, it is difficult to econometrically identify the effect of such teacher incentive programs. We address this problem by examining a context with quasi-random assignment of schools to treatment and comparison groups. Third, we examine teacher incentives in a developing country context (where incentives are often particularly weak).² Finally, by collecting panel data on teacher absence, we are able to show that the problem is widespread, suggesting existing incentives are weak.

The remainder of the paper is organized as follows. Section 2 discusses primary education in Kenya and argues that the high rate of teacher absence suggests current incentives

are inadequate. Section 3 describes the teacher incentives program that we examined, and the process by which schools were selected for the program. Section 4 provides a simple Holmstrom-Milgrom model to interpret the results. Section 5 reports the impact of the program on teacher outcomes, while Section 6 reports the impact on student outcomes. Section 7 discusses how teacher behavior changed in response to incentives over time, and Section 8 concludes.

2. Background

This section provides some background on primary school teaching in Kenya and provides evidence that teacher absence is widespread in the area of the study.

2.1 Primary School Teaching in Kenya

Teacher hiring, firing, and transfer decisions in Kenya are made centrally by the Ministry of Education. Hiring is based primarily on academic qualifications.³ Salaries are set through collective bargaining between the government and the politically powerful Kenyan National Union of Teachers (KNUT). In 1997, the starting salary for teachers was Ksh 5,175 (\$88) per month, and a typical teacher in our sample earned approximately Ksh 7,000 (\$119) per month.⁴ Taking into account generous benefits, total teacher compensation was approximately \$2,000 a year, or more than five times annual GDP per capita.⁵

² Kingdon and Teal (2002) show that in private schools in India teacher pay and student achievement are linked, but they do not demonstrate specifically that linking pay to performance has any effect.

³ Primary school teachers in Kenya typically have completed two years of teacher training beyond secondary school. A small number of teachers were hired under an older system in which primary teachers had only a 7th grade education and two years of teacher training.

⁴ This is assuming an exchange rate of 58.7 shillings per dollar, the 1997 dollar-shilling exchange rate.

⁵ Authors' calculations based on value of housing allowance and other benefits. This is calculated from salary scales and represents a salary for a teacher with average education and experience.

Teachers' salaries depend primarily on education and experience. There is little opportunity for adjustments to salary or for promotion based on performance. Teachers have strong civil service and union protection and are difficult to fire. In some cases teachers who have performed very badly are transferred to less desirable locations, while the government may look more favorably on requests for transfers to more desirable postings or to home areas from teachers who perform well.

Although incentives provided to teachers by their employer are weak, every school is supposed to have a parent committee, and sometimes these committees provide gifts for teachers when schools perform well on the national exams. Similarly, communities sometimes refuse to allow exceptionally bad teachers to enter the school, thus putting pressure on the Ministry of Education to arrange a transfer for the teacher. However, only a minority of school committees provide supplemental bonuses, and school committees typically only attempt to influence the national authorities in extreme situations.

To the extent that incentives do exist, they are typically based on the system of national testing. Results on the national primary school leaving exams (the KCPE) are front-page news in Kenya, and newspapers publish the relative positions of different districts in the country on the national exam and lists of the highest-scoring schools. Results from this exam and from district exams administered in the upper grades of primary school are often posted in headmasters' offices. Since the primary school leaving exam determines what secondary schools, if any, will accept graduating primary school students, teachers devote considerable effort to preparing for these exams. An important method of preparation is to review books of questions from old exams. Some schools hold extra preparation sessions for the exams outside of normal class hours.

While the considerable attention given to results on national exams clearly spurs effort by some teachers to raise their scores, not all of this effort is necessarily desirable. For example, seventh graders who do not perform well are often required to repeat a grade rather than being allowed to go on to 8th grade and take the KCPE exam.

2.2 Data on Teacher Absence

One indication that teacher incentives are weak lies in high absence rates. Random visits conducted to check pupil attendance and observe pedagogy suggest that teachers in comparison schools were absent about 20% of the time. This absence rate is comparable to, or less than that in several other low-income countries (Chaudhury, Hammer, Kremer, Muraldhiran, and Rogers, 2003). While the 20% absence rate may reflect a variety of factors, including the high incidence of infectious diseases such as malaria and AIDS, these unavoidable absences are unlikely to account for all absence. For comparison, absence rates among staff at a non-profit organization working in the same area are around 6%.

Absences seem fairly broadly distributed among the population of teachers rather than primarily accounted for by a subset of teachers with very high absence rates. We calibrate two simple models to our panel data on teacher absence to assess the concentration of teacher absence.⁶ We first consider a non-parametric model in which the population of teachers consists of groups with distinct probabilities of attending. Second, we calibrate a model in which individual probabilities of attendance are drawn from a beta distribution. A robust result across these models is that although there is a small group of teachers with very high absence, these

⁶ We are grateful to Emily Oster for outstanding work calibrating these two models.

account for the minority of total absences. More than half of absences are due to teachers who attend more than 80% of the time.

To calibrate the non-parametric model we assume the population of teachers consists of 5 groups, with proportions α_1 through α_5 and probabilities of attendance p_1 through p_5 . There are nine independent unknowns in this model: $\alpha_1 \dots \alpha_4$ and $p_1 \dots p_5$. We use information from 8 visits to schools to calibrate the model. With 8 visits we observe a proportion x_0 of teachers who were never there, x_1 who were there once, etc, through x_8 who were there every time. We solve a system of equations in which the expected proportion of teachers present for $J < 8$ visits (based on $\alpha_1 \dots \alpha_4$ and $p_1 \dots p_5$) is set equal to the actual proportion x_J .

In an alternative, more structural, exploration of this issue, we assume that each teacher has an individual probability of attendance drawn from a beta distribution. For each teacher, the distribution of attendance at school is then binomial. The system is then modeled as a beta-binomial distribution, which is quite tractable because the beta distribution is a conjugate of the binomial. Given a total of n visits each person is observed to attend $k \leq n$ times and we can represent the likelihood of observing an individual who is present k times using a set of gamma distributions dependent on n , k and the parameters of the beta distribution (α and β). We then use information on observed attendance to produce maximum likelihood estimates of the beta distribution parameters. Table 1 shows estimates of the parameters of each model and the percentage of absences due to various groups in the population under each model. Figure 1 shows the estimated beta distribution.

Both models tell a similar story: more than half of the teachers are present less than 80% of the time. Although a few teachers are rarely present, the majority of absences appear to be due to those who attend between 50% and 80% of the time. In addition, in both models (although

more so in the beta distribution model) a large minority of absence is actually due to those who attend *more* than 80% of the time. Note that if most absences were due largely to the high AIDS rate in these areas we would expect to see the absences concentrated within a smaller group.

The widespread nature of absences suggests that teachers who are absent frequently may not be violating a social norm. This is not, however, an implicit contract in which the government pays teachers only for part-time work. As noted above, including benefits, Kenyan teachers are paid up to 5 times average per capita income and there is substantial queuing for teaching jobs and unemployment among people who would be qualified to become teachers. The government imposed a hiring freeze in 1998 so several cohorts of graduates of teacher training are unemployed. Some are working for free in schools in the hope of eventually obtaining a paying job.

Further evidence of weak incentives is that while teachers are absent from school about 20% of the time, they are absent from their classroom much more often. In 1998, the time of teacher arrival was not recorded during classroom visits to observe pedagogy, but in 1999, when it was, 45% of the time teachers never arrived in the classroom in only 55% of the classrooms we observed.

3. Program Description

As noted above, some school committees in Kenya provide bonuses to teachers whose students perform well in exams. This paper evaluates a program conducted by International Christelijk Steunfonds (ICS), a Dutch NGO, in Busia and Teso Districts of Western Kenya. The program offered schools the opportunity to provide gifts to teachers if students performed well. It

provided prizes to teachers in grades 4 to 8 based on the performance of the school as a whole on the district exams in each year. All teachers who taught these grades were eligible for the prize.⁷

ICS awarded prizes in two categories: "Top-scoring schools" and "Most-improved schools." Schools could not win in more than one category. Improvements were calculated relative to performance in the baseline year. Since the results of the district exams were not available for 1997, the scores for 1996 were used as the base to measure improvements. (Henceforth, we will refer to the last pre-program years for which we have data as year 0, the first (1998) and second (1999) years of the program as years 1 and 2 respectively, and the post-program year (2000) as year 3.) In each category, three first prizes were awarded, three second prizes were awarded, three third prizes were awarded, and three fourth prizes were awarded. Thus, overall, 24 out of the 50 schools participating in the program received prizes of some type, and teachers in most schools should have felt that they had a chance of winning a prize. Since Busia and Teso Districts had separate district exams, prizes were offered separately in each district in proportion to the number of schools in the district.

Education experts generally are more sympathetic to school-based incentives than to individual-based incentives since they feel these are more conducive to cooperation among teachers (Richards and Sheu, 1992; Hanushek, 1996). In order to encourage cooperation among teachers within schools and to avoid creating incentives for teachers to sabotage each other's work, ICS prizes were based on the performance of all of the grade 4 to 8 pupils in the school, with each subject weighted equally, rather than on a teacher-by-teacher basis. Thus, every teacher in grades 4 to 8 in the winning schools received the same prize. Setting prizes at the school level could potentially allow free-riding within the teaching staff. However, teachers are

⁷ Teachers of lower grades were not a part of the competition, because there were no district exams for those classes. They received a lantern as a token prize, whether or not they belonged to a winning school

in a relatively good position to monitor each other' s performance, particularly on easily observed aspects such as attendance. Moreover, since teachers can observe each other' s work at high frequency, they were in a repeated game with each other. Since the typical school in the sample had only 200 students and 12 teachers, about half of whom taught in the upper grades, teachers should have been able to enforce cooperation within the school.

In order to discourage schools from forcing weaker students to repeat, drop out, or not take the exam, students who did not take the exam were assigned low scores. Multiple choice exams were used in all subjects other than English. Students who did not take these exams were assigned a score of 15, whereas students who simply guessed would have obtained a score of 25 on average. On the English essay exam component, students who did not take the test were assigned a score of zero. In order to discourage schools from recruiting strong students to take the exams, only students enrolled in school as of February 1998 were included in the computation of the school mean score.⁸

Prizes ranged in value from 21 to 43% of typical teacher monthly salaries. This is comparable to merit pay programs in the United States. For example, the 1993-94 Dallas merit pay program, which was also based on school-wide performance, awarded \$1000 bonuses, which were 39% of an average monthly salary of Texas teachers that year, and presumably a somewhat lower percentage of salaries for teachers in Dallas (Clotfelter and Ladd, 1996; American Federation of Teachers, 2000). Similarly, a 1999 Rhode Island program awarded \$1,000 bonuses, worth about 25% of monthly salary (Olsen, 1999; American Federation of Teachers, 2000). Other programs in Colorado award between 10% and 50% of monthly salary in merit-based bonuses (Education Commission of the States, 2001).

⁸ Each winning school also received a briefcase for the headmaster, a wall clock, a time keeping clock, and a bell.

Schools were offered the opportunity to participate in this program in February 1998 and all accepted. (The academic year in Kenya runs from January to November.) The prizes were awarded during a ceremony held in November of each year, and all the schools in the program were invited to attend. In order to discourage teachers from arranging transfers into treatment schools in order to be eligible for the program, eligibility was restricted to teachers who were employed in the school as of March 1998. As discussed below, teacher entry and exit rates did not differ significantly between treatment and comparison schools.

When the program was originally announced in February 1998, it was scheduled to run for a single year and teachers were informed of this. Later the program was extended for an additional year. Because the NGO had conducted other programs in the area, we think most teachers found the commitment to provide the prizes credible. However, it is possible that teachers did not react fully to the program until after they had actually seen the prizes awarded during the first year. The awards ceremony presumably increased the salience of the program. It is not clear whether teachers expected the program to continue longer than the NGO had promised, since the NGO tries to be conservative in announcing benefits it will provide to avoid creating overly high expectations.

The short duration of the program and its surprise introduction allowed several elements to be included that might not otherwise have been possible in a permanent program. Half the prizes were based on improvements in performance. Teachers in many schools may have felt these were the only prizes for which they could realistically compete. The incentives created by these prizes for improvement were presumably larger than they would have been in a permanent program in which teachers would presumably understand that increasing their scores in one year would make it harder for them to win the prize in subsequent years. Moreover, the short duration

of the program made it possible to base prizes on the test scores of all students originally enrolled in the school, which allowed the program to discourage repetition and dropouts without risking the possibility that headmasters or teachers would seek to restrict admission to the school or selectively recruit strong students. Similarly, the program was restricted to the original teachers at this school. It thus reduced the opportunity for teachers to seek assignment to schools with high-scoring pupils due to non-teacher factors such as parental characteristics. All these factors imply that the incentives provided by the program were stronger than could have been provided with the same expenditure under a permanent program. On the other hand, teachers might have had stronger incentives to promote long-run learning if they expected the program to continue indefinitely.

The context seems particularly favorable for teacher incentives for several reasons: the level of teacher absence suggests that teacher effort was an issue in the schools; there was relatively little scope for diverting teacher attention away from creativity and towards teaching to the test, and the relatively short duration of the program (and the low scores assigned to dropouts) made it possible for ICS to design an incentive program that did not encourage schools to invest in manipulating the composition of the student body. In addition, the fact that the program was restricted to the original teachers avoided the temptation for teachers to sort into areas with favorable socioeconomic situations.

3.1 School Selection

The 50 schools in the program were selected from a group of 100 schools had originally been selected by the Ministry of Education because they were considered to be particularly in need of assistance, but had not participated in an earlier World Bank program which provided

textbooks to the schools judged to be in greatest need. These schools scored somewhat worse than average for the area before ICS began working with them. ICS had provided textbooks or modest grants to these schools before the inception of the teacher incentive program. Schools were numbered according to the year they received textbooks or grants from ICS, their geographic division, and their alphabetical order. Then the odd numbered schools were chosen to participate in the teacher incentive program. By construction, the odd and even numbered schools were split in comparable proportions across Busia and Teso districts, geographic divisions within these districts, and across schools which received textbooks and grants in different years.

The 50 even-numbered schools that serve as the comparison group for this evaluation participated in a program designed to improve pre-schools by providing training, materials, and salary supplements conditional on pre-school teacher attendance. Pre-school teachers, unlike the primary school teachers, are semi-volunteers who are not hired or paid by the central Ministry of Education, but instead are hired locally by parents' committees and receive only contributions from parents, which are often irregular. Unlike primary school teachers, pre-school teachers often have no formal training. Since pre-school and primary school teachers are quite distinct, and since in any case the pre-school program had little effect on performance in the pre-school classes, it seems unlikely that this program affected outcomes in grades 4 to 8 during the time period we examine.

4. A Model of Productive and Signaling Effort

Holmstrom and Milgrom (1991) consider a model in which linking pay to observable signals of performance can potentially lead employees to focus on tasks for which output is

easily measured and divert effort away from tasks for which output is difficult to measure. They motivate their analysis using two main examples. In the first, linking teacher pay to test scores may cause teachers to teach to the test rather than encourage creativity. In the second, employees who are responsible both for producing output and for maintaining the value of an underlying asset, such as a piece of equipment or a firm's reputation, may neglect the long run value of the asset if they are provided with strong incentives to focus on current output.

We consider a model that combines elements of both Holmstrom and Milgrom's motivating examples, and can be considered a special case of their general model. Teachers can exert two types of effort: effort to promote long-run learning and effort to signal ability by measures that improve short-run scores. Employers observe only test scores. In particular, we assume test scores depend both on underlying learning (produced by teaching effort over time) and signaling effort. Suppose that $T_t = L(e_t, e_{t-1}, e_{t-2} \dots) + \eta_t + \varepsilon_t$, where T_t denotes test scores during period t , L denotes student learning, e_t denotes teaching effort during period t , s denotes signaling effort, and ε is a random shock. Thus, teaching effort produces long-run improvements in students' understanding, while signaling effort produces only short-run effects on test scores. (Teaching effort can thus be seen as unobservable effort to maintain asset value in Holmstrom and Milgrom's framework.)

Assume that teachers' utility is given by $U = M - C(e, s)$ where M is teacher pay and C is a utility cost that depends on both teaching and signaling effort. In this specification, e and s can be either substitutes or complements. For example, they may be substitutes because there is a fixed amount of time in the day that must be allocated between them. On the other hand, they may be complements if there is a fixed cost to teachers of attending school at all.

Suppose teacher pay is $M = \alpha + BT$. If $B = 0$, so pay is independent of performance, teachers will choose effort in teaching and signaling such that the marginal product of each is equal to zero. Following Holmstrom and Milgrom (1991), $C_1(0,0)$ and $C_2(0,0)$ may be negative, so some effort may be exerted even if $B = 0$. Teachers may care about their students, or enjoy exerting some effort even in the absence of performance incentives. In addition, there may be other incentive systems in place. For example, teachers may be motivated by a fear of punitive transfers.

If the government or an NGO makes a surprise announcement that pay will be linked to test scores for a single year, teachers will change both teaching and signaling effort to satisfy the first order conditions implied by the above equations. Specifically, teachers will exert teaching and signaling effort such that: $\frac{\partial L}{\partial e} B = \frac{\partial C}{\partial e}(e, s)$ and $\gamma B = \frac{\partial C}{\partial s}(e, s)$. If e and s are complements in the utility function, or if utility is additively separable, then both types of effort will increase. If they are substitutes in the utility function then incentives may increase one type of effort at the expense of the other. Thus in this model, incentives could potentially either increase or decrease teaching effort.⁹

Note that under the model, parents and local communities may not object to teachers' investing in short-run test preparation, since students' prospects for further education and labor market success depend on test scores as well as underlying learning. Test preparation, however,

⁹ Clearly, there is a continuum between exerting effort on promoting long-term learning and trying to manipulate short-run test scores. The extreme of manipulating short-run scores would be actually cheating at the time of the test; less extreme versions would include going over questions from previous years' exams, and teaching test-taking strategies such as guessing on multiple choice questions. Within the category of promoting learning, teachers could focus narrowly on the curriculum to be tested or could promote learning more broadly. One could imagine generalizing this model to allow teachers to choose from a menu of activities, with varying components of true and signaling effort, but results would presumably be similar.

is assumed to be socially wasteful, in that it requires teacher effort but does not improve the underlying learning that affects total output in society.

Before discussing evidence on the impact of the incentive program, it is worth noting that a simple OLS specification suggests that the marginal product of test preparation sessions, which can plausibly be interpreted as signaling effort, is much higher than that of teacher attendance, a plausible measure of teacher effort. The evidence, based on 1998 visits, suggests that teachers who attend 20% more of the time (are present for one additional visit of the five made) have students who score 0.0115 standard deviations higher (standard error 0.0123). However, those who report coaching in one additional time period (of the six inquired about) have pupils who score 0.044 standard deviations higher (standard error 0.0087). It is somewhat difficult to draw conclusions given that it is difficult to know how many days coaching entails. It is likely, however, that a 16% increase in coaching requires substantially fewer hours of work than a 20% increase in attendance, implying the returns from coaching are much larger.

Of course, it is not clear that these estimates should be interpreted causally. Nonetheless, the hypothesis that test preparation activities can raise test scores is consistent with evidence from the U.S. on the effects of commercial test preparation. These studies often show gains of 0.15 to 0.4 standard deviations on admissions tests, even though most U.S. admission tests are supposed to measure aptitude, rather than achievement, and thus to be difficult to study for. Extra coaching raised scores on an achievement test by 0.25 standard deviations (Bangert-Drowns *et al.*, 1983). In an environment in which teacher pay is not linked to test scores, it seems plausible that teachers might make only limited efforts at test preparation, leaving the marginal test score product of test preparation substantially greater than that of teacher attendance. In the model, if it were possible to easily and accurately observe individual performance on both tasks, then a wage

contract could induce teaching effort without inducing signaling effort. However, while distinguishing teaching and signaling effort would be expensive and inaccurate at the individual level, particularly if tied to an incentive program, there are potential ways to distinguish them empirically at the aggregate level. First, outside observers may be able to directly observe teachers' activities and categorize them as either promoting long-run learning or simply test preparation. For example, as noted above, in Kenya, some schools engage in "preps"—extra test preparation or coaching outside of normal class time, often during school vacations. One could potentially interpret these sessions as including a higher rate of signaling to teaching effort than ordinary classroom attendance. Second, improved learning should have a long-run effect on test scores, whereas test preparation may have either a short- or long-run effect. For example, helping students cheat will only increase scores in the short run, but teaching students to guess on multiple choice exams or better allocate their time could raise scores on other tests. Thus increases in test performance past the program date could be interpreted as being due either to increased effort in promoting learning or extra coaching on test taking techniques. However, a finding that test score gains do not persist is consistent with the hypothesis that the program led only to extra coaching specific to the test at hand. It is more difficult to reconcile this result with the hypothesis of increased long-run learning. A third potential way to distinguish efforts to increase long-run learning from test preparation activities is to check if test scores improved primarily in subjects prone to memorization.

5. Impact of Incentives on Teachers

This section examines the impact of the program on teacher behavior by comparing outcomes between treatment and comparison schools. It is worth noting, however, that

interviews of teachers in treatment schools about their satisfaction with the program suggest that teachers in participating schools like the program. In particular, in the middle of the second year of the incentives program, the headmaster and three other teachers were interviewed in each program school. All teachers interviewed supported the idea of motivating teachers by providing them with incentives. Most reported a change in school activities and teacher attitudes because of the program. 83% reported that prizes were justly awarded in 1998. 75% of teachers in program schools reported an increase in homework assignment due to the program, 67% reported an increase in cooperation among teachers, and 88% reported an increase in preps.

5.1 Teacher Assignment

The incentives program was designed so as not to affect incentives for teachers to join program schools. Only teachers who were already assigned to an incentive school as of March of the first year were eligible for a prize. However, the program could potentially have reduced the exit rate of teachers from the incentive schools.

In fact, the exit rate was not significantly different between program and comparison schools (Table 2, Columns 1 and 2).¹⁰ The entry rate was higher in the incentive schools for the first year of the program, and lower for the second, although in neither case was the difference statistically significant (Table 2, Columns 3 and 4).¹¹

We also considered the possibility that teachers in treatment schools in lower grades would attempt to transfer into teaching higher grades, even though this would not actually make

¹⁰ All regressions in this paper allow for school-level random effects to take account of the possibility that there may be correlation between error terms for students or teachers in the same school. Note that in the random effect regression framework the coefficient on the constant term is not exactly equal to the mean of the omitted category.

¹¹ The transfers include voluntary transfers due to family reasons (such as marriage) or involuntary transfers such as disciplinary actions against teachers or staff balancing needs (to replace teachers who retire, die, or move.)

them eligible for the program. There is no evidence of differential transfers across grades. In treatment schools 7.36% of teachers transfer from a non-incentive grade to an incentive grade during the program; in comparison schools 7.30% do.

5.2 Teacher Attendance

Teacher attendance was not affected by the incentive program. In the year prior to the establishment of the program, each of the 100 schools was visited two times and the present/absent status of each teacher in grades 4 to 8 was recorded. Similar visits were made five times in year 1 and three times in the year 2.¹² For each teacher in each year, an attendance rate was computed as the proportion of visits during which the teacher was present. Note that teachers would be recorded as present if they were at the school, even if they were not teaching when the visit took place. Following standard intention to treat (ITT) methodology, the sample included only those teachers who were assigned to program or comparison schools in year 0. Any teachers who changed schools between year 0 and year 1 or between year 1 and year 2 were classified with their initial schools.¹³

Prior to the program, schools that would later be selected to be program schools have slightly higher teacher attendance, although the difference was insignificant (Table 3, Column 1). In year 1 of the program, teacher attendance was actually slightly lower in the incentive schools,

¹² We focus on teacher absence data based on visits to schools, rather than on official school logs, because school logs are often not filled out. However, school log data also suggest no effect of the program on absence.

¹³ This could only be done for those teachers who switched schools and remained in the sample of 100 schools. Since there is no data on the teachers who switched to other schools, they were dropped from the analysis.

and in year 2 the attendance was slightly higher in incentive schools (Table 3, columns 2 and 3), although both coefficients are insignificant and quite small.¹⁴

5.3 Homework Assignment

For a random subset of students in grades 4 to 8 for each school, information was collected from the students on whether they were assigned any homework on the previous day. Treatment schools assigned slightly more homework than comparison schools prior to the program, although the difference is far from significant (Table 4, Column 1). After the launch of the program, treatment schools assigned slightly less homework, although the gap was insignificant both in levels and in differences (Columns 2 and 3).

5.4 Pedagogy

Teacher behavior was not significantly different between the incentive and comparison schools. Trained observers observed each teacher annually, spending one period in a class and recording teacher behavior on a number of measures, including both objective information about what the teacher was doing and subjective information about their energy level and caring for the students.¹⁵ We examined a wide variety of pedagogy measures, and the results for four are presented here. In particular, we present results for two objective measures (blackboard use and teaching aid use) and two subjective ones (teacher caring and energy).

¹⁴ Results are similar when lower primary school teachers are used as a control in a regression in which attendance of all the teachers in the schools is regressed on a dummy for the program, on whether they are a lower primary teacher or an upper primary teacher and on the interaction between the two.

¹⁵ Teacher' s caring and energy level are categorized as follows: Caring (5=Very caring and attentive, 4=fairly caring and attentive, 3=neutral, 2=somewhat uncaring, 1=mean or rude); Energy level (5=very enthusiastic/alert/active, 4=fairly enthusiastic, 3=neutral, 2=bit dull/tired, 1=lazy/very dull).

There was no significant difference in pedagogy between the incentive and comparison schools for any of the classroom observations prior to the program (Table 5, Column 1). We also find no significant difference during the intervention period between the two school groups in any of the pedagogical practices (Columns 2 and 3). The point estimates are close to zero for each observation type. The difference-in-difference estimates shown in the last two columns were computed at the school-grade level since it was not possible to match individual teachers across observations years. These estimates are also close to zero for every observation type.

5.5 Extra Coaching

Extra coaching rates in the incentive schools were higher than in comparison schools. Headmasters in each school provided information on whether there were any preps for grades 4 through 8 in six time periods during the year – the three school vacations (April, August and December) and out of school hours during each of the three terms.

Prior to the program, incentive schools were slightly less likely to offer preps (Table 6, Column 1), but after the introduction of the program, treatment schools started to conduct more preps (Columns 2 and 3). They were 4.2 percentage points more likely to conduct preps in the first year and 7.4 percentage points more likely in the second, with the latter estimate being significant at the 5% level. The coaching results shown above are driven primarily by the coaching over vacations, as can be seen by the stronger results in the lower panel of Table 6.

6. The Impact of Incentives on Students

Consistent with the hypothesis that teachers responded to the program primarily by seeking to manipulate the variables determining prize allocation rather than by increasing efforts

to promote long-run learning, the program had little impact on dropout and repetition rates, but increased student participation in exams. During the period the program was in place, student scores increased, significantly so on some test measures. There is some suggestive evidence that the effect was larger in the subjects more vulnerable to coaching. After the end of the program the effect on test scores did not persist. Students who had been in program schools during the program scored no higher than their counterparts who had been in comparison schools. Below we discuss the program impact on dropout and repetition (subsection 6.1), the Kenyan exams (subsection 6.2), exam participation (subsection 6.3) and test scores (subsection 6.4).

6.1 Dropout and Repetition

As mentioned above, schools were penalized for those students who did not take the test. However, dropout and repetition rates in incentive schools were not significantly different from those in comparison schools. Dropout dummy variables were set equal to one if a student enrolled in the previous year did not continue schooling in the current year. Repetition dummy variables were equal to one if the student repeated the same class in the following year. Dropout rates were insignificantly higher in treatment schools (Table 7, Columns 1 and 2), while the repetition rate was insignificantly lower in incentive schools (Table 7, Columns 3 and 4).

6.2 The Kenyan Exams

Incentives for teachers were based on their students' performance on the district exams, which are administered in seven subjects: English, Math, Science, Swahili, Geography-History-Christian Religion (G.H.CR.), Arts-Crafts-Music (A.C.M.), and Home Science-Business Education (HS.BE.) Students in grades 4 through 8 take these exams in October of each year.

Participation is incomplete since students have to pay a fee of 120 Shillings (US \$2) to participate in the exams. Since the rules of the incentive program stipulated that any student who did not take the district exams would be assigned lower scores than students could obtain by guessing, teachers in program schools had incentives to encourage their students to take the district exams.

In addition to the district exams, students also took KCPE and ICS exams, which were not tied to the teacher incentive program. Thus, these tests provide us with an independent way to assess the impact of the program.

The Kenyan primary school leaving exam (KCPE) is administered by the Ministry of Education each year to pupils completing grade 8. It determines what secondary school, if any, students attend. ICS also administered exams to students in grades 3 through 8.¹⁶ In year 1, the ICS tests were administered in English, Math, and Science. In years 2 and 3, they were administered only in English and Math.¹⁷ We have data on the district exam scores and the KCPE test scores from both intervention years (1 and 2) as well as the post-program year (3). Finally, we have information on the ICS tests for all participating schools for both of the intervention years.¹⁸ We were unable to obtain the data for comparison schools in Teso District for this year. Consequently, analysis of the district exam scores for year 1 is restricted to schools in Busia District.

¹⁶ For a complete description of the ICS tests and their administration, see Glewwe, Kremer, and Moulin (2001). These tests were also administered in 1996 and 1997.

¹⁷ In year 0 and year 1, all three exams were fairly similar in format and content. Separate exams were given for each grade and the exams had a multiple-choice format. However, the ICS exams in year 2 and 3 were "multilevel," with the same test given to all students in grades 3 through 8. Easy questions in the beginning of the test could be answered by all students, including those in grade 3, while questions became progressively harder. The final questions were based on material seen only in the eighth grade. These exams also had a "fill in the blank," as opposed to a multiple choice format.

Security is generally tight in Kenyan exams to prevent cheating. District exams were supervised by three to four teachers from a neighboring school. Because the KCPE exam determines the future scholastic paths of the eighth grade students who take it, these exams are even more strictly monitored and supervised.

6.3 Exam Participation

Exam participation is important both as an outcome in its own right and because differential exam participation could complicate the interpretation of test score differences between program and comparison schools. Exam participation rates were higher in program schools than in comparison schools for the district exams (on which the incentives were based), but were similar between treatment and comparison schools on the ICS and KCPE exams.

Following standard ITT methodology, we restrict attention to only those students who were enrolled as of February 1988 (year 1) and assign students who switched schools during the program to their original schools.

Baseline participation on the district exams was around 70%; on the ICS and KCPE exams it was around 85%. In year 1, participation in the district exams was higher by 6.0 percentage points in the incentive schools, a difference which is statistically significant at the 10% level (Table 8, Column 1). By year 2, participation was higher by 10.8 percentage points in the treatment schools, a significant difference at the 5% level. In the post-program year, when there was no longer an incentive to encourage students to take the test, the participation rate was actually 2.2% lower in the incentive schools than in the comparison schools, though the difference was not significant. In contrast, the participation rates in the ICS and KCPE exams

¹⁸ In the post-program year, 27 of the 100 schools were involved in a de-worming project. This enabled us to collect

were not different between the two school groups in either year (Columns 2 and 3). The main differences between incentive and treatment schools in exam participation were in grades 4 through 7; participation in grade 8 on the district exam was close to 90% prior to the program.

This pattern of exam participation matches the incentive faced by the program schools to have more of their students take the district exams.

Theoretically, efforts by teachers in treatment schools to increase exam participation could potentially bias scores either upwards or downwards, but available evidence suggests the bias is likely downward. If teachers in the treatment schools put equal effort into encouraging all students who would not otherwise have taken the exam to do so, then the addition of marginal students would likely have dragged down average test scores, since poorer students are less likely to pay the fee to take the district exam. But if teachers selectively chose to concentrate on convincing potentially high-scoring students and their parents of the exam's importance, then average scores in the treatment schools could be potentially biased upward. To get a sense of the direction of bias, we compared pre-scores of test takers treatment and comparison schools. The additional students who took the tests in the incentive schools had lower pre-test scores, but not significantly so.

In 1998, one case was identified in which the headmaster of a program school colluded with the teachers assigned to supervise the schools to allow cheating on the exam. That school was disqualified from the competition in 1998 but was allowed to participate in 1999. The scores from that school were not included in the analysis in 1998, but their scores were included in 1999.

ICS test score data from that subset of schools for that program year.

6.4 Test Scores

Test scores on the district exam are higher in the incentive schools during the years of the program (significantly higher in the second year). Scores on other exams are also higher in program schools during the duration of the program, but not significantly higher. In the post-program year, however, we see no persistence of the test score gains. This provides some support for the hypothesis that teachers are focusing primarily on extra coaching specifically for the test in question. On the other hand, there is little evidence of outright cheating, and we did not see any evidence that teaching effort actually decreased during the program, such as a decline in long-run test scores for students exposed to the program.

We examine differences in test scores between the incentive and comparison schools using a random effect regression framework that allows for the possibility that scores of students in the same grade and same school might be correlated due to unobserved characteristics of teachers and headmasters. In particular, we use an error components econometric model with school, grade, and subject random effects: random effects at the school level and at the level of individual subjects and grades within the school.

$$(1) t_{ijks} = \alpha_{4k}D_{4i} + \alpha_{5k}D_{5i} + \dots + \alpha_{8k}D_{8i} + \beta_k p_s + u_{ks} + v_{jks} + e_{ijks}$$

where k = English, Math, Science, Swahili, G.H.CR., A.C.M., and HS.BE.

Equation (1) combines data from several grades to measure the impact of the incentive program for a given subject. The test score of student i in grade j in subject k in school s is t_{ijks} . The dummy variables D_{ji} indicate whether child i is in grade j . The variable p_s is a dummy variable that equals 1 if school s is an incentive school (i.e. a school which was eligible for teacher incentives) and 0 if not. Thus if the impact of the incentive program varies across grades,

β_k will measure the (weighted) average impact of the program across all grades. The error term contains three components, the school-specific error term (for subject k), u_{ks} , a grade-specific term conditional on being in that school, v_{jks} , and a child specific term, e_{ijks} .

We estimate these equations using Generalized Least Squares (GLS) without imposing a specific distribution (e.g. the normal distribution) on the error terms. The regressions also include controls for sex and geographic division within Busia. Given the prospective design of the program, regressions without such controls are consistent, but adding controls to the regression may increase the precision of the estimates. As a check, we ran regressions without the controls for region and sex; they yield similar results.

Because the units in which test scores are measured are arbitrary, for each subject and grade combination we normalize all test scores (including district, ICS, and KCPE tests) by subtracting the mean test score in the comparison schools and then dividing by the corresponding standard deviation for those schools. Thus, a student with a normalized score of 0.1 was 0.1 standard deviations above the mean score in the comparison schools. Note that for a normal distribution, an increase of 0.1 standard deviation would move a student from the 50th percentile to the 54th. Since the district test exams were different for Busia and Teso Districts, the normalization of these tests was done separately for each district.

There is no significant difference in pre-program scores on the district exam between incentive and comparison schools for any subject or grade (details in Appendix Table C). Overall, prior to the program, treatment schools scored almost the same as comparison schools (Appendix Table C, Columns 1 and 2). For students who were in grade 4 in 1998 and who were therefore in grade 2 in 1996, we did not have mock test scores in 1996. For such students, we used the 1997 ICS tests as pre-tests whenever available.

The difference in test scores between treatment and comparison schools, as well as the difference in difference estimator of the effect of the program, are shown in Tables 9 (for the district exam) and 10 (for the ICS and KCPE exams).¹⁹ (Appendix Table B shows results by grade.) The difference results were calculated using the full sample, i.e. all the students in grades 4 through 8 who took the exams in either intervention year, while the difference-in-difference results use a restricted sample, i.e. those students who took exams in at least one subject, in both the pre-program year and at least one of the intervention years. As discussed, we restrict attention to only those students who were enrolled prior to the announcement of the program, as of February 1998 (year 1). We also restrict the sample to those students who did not repeat or dropout in any year, since students who repeated would be taking a different exam.

Before going into detail on individual tests, we report some general results. The difference estimate for the district exam is insignificantly negative in the first year of the program (point estimate -0.04), but this could potentially be due to the differential exam participation between treatment and comparison schools on the district exams. The difference-in-difference estimate is positive, although not significant. Both the difference and difference-in-difference estimate of the treatment effect in year 2 are significantly positive, and quite large, at around 0.136 . For the ICS exams, differences are not significant in either year, and the point estimates across years are similar, at around 0.085 . For the KCPE, the overall difference estimate in year 1 is 0.138 and the difference-in-difference is 0.104 , with the former being significant at the 5% level. In year 2, the effect is stronger, with the difference estimate equal to 0.165 , significant at the 5% level. By year 3, students in the incentive schools had been exposed to the program for all of year 1 and year 2. If the increases in test scores in year 1 and year 2 of the

¹⁹ Normalized district test scores from year 0 (1996) were used as the KCPE pre-program scores, since the KCPE exam is taken by grade 8 students only.

program were due to increases in students' underlying long-term learning, then students in incentive schools should also have scored higher in year 3, after the program ended. However, estimates of the program effect in year 3 on the district and KCPE exams were close to zero.²⁰

Since the teacher incentives were based on performance on the district tests only, data on ICS and KCPE tests can be used to check whether the difference in student outcomes between the incentive and comparison schools was due to factors specific to the district exams, such as outright cheating or altering the pool of students taking the test. Program schools scored insignificantly higher on the ICS test (Table 10, Panel A), and significantly higher on the KCPE (Table 10, Panel B) in both program years.

Breaking down the results by subject, the average effect for the two program years was strongest for the subject test on Geography, History, and Christian Religion, which is arguably the subject that involves the most memorization (Tables 9 and 10). In year 1, the difference-in-difference estimate on G.H.CR was 0.205 for the district exam and 0.149 for the KCPE exam, with the former being significant at the 10% level. In year 2, the program impact on G.H.CR scores is even stronger, with the difference-in-difference estimates being 0.341 for the District Exams and 0.336 for the KCPE exams. Both estimates are significant at the 5% level. The next biggest effects were for science and math, with no significant effect for other subjects. Arguably, GHCR is the subject with most memorization and is particularly susceptible to extra-coaching and short-run teaching strategies. Primary school science also involves a fair amount of memorization, but math presumably requires less memorization.²¹

²⁰ Data for the ICS exams in year 3 was only available for 27 of the schools – those that participated in a de-worming project that year. Point estimates are positive, but t-statistics are less than or equal to one.

²¹ ICS staff members familiar with the curriculum suggested that G.H.CR and HS.BE. require the most memorization, science requires a medium amount of memorization and English, Math and Swahili require the most creative thinking.

For the 1999 ICS test, we have item-level data on whether students had correct answers to individual questions in their English and math tests. Treatment students scored higher than comparison students on the later part of the test but not on the early part of the test. This may be because teachers in treatment schools taught students to better allocate time across sections of the test. (However, given that both tests were designed so as to increase in difficulty as the test progressed, another possibility is that the program induced increased teaching effort, but that this was most effective at raising scores on more difficult questions.)

Analysis of the item responses to detect cheating using techniques developed by Jacob and Levitt (2002) provide little evidence of suspicious strings of questions for which all students in the class got the question right, suggesting cheating was not widespread, although it is worth noting that there was one instance in which cheating was discovered. Were cheating widespread we likely would have seen much larger test score differentials on the district exam (on which incentives were based) than on other exams. Although the estimated program impact on the district exam is somewhat larger than on the ICS exams, it is lower than on the heavily-monitored KCPE exams, suggesting that cheating was not the main source of the program effect.

7. Time-path of Changes in Teacher Behavior

There is some evidence that teachers learned over time how better to take advantage of the program. Estimated differences between treatment and comparison schools grew between the first and second year, as did differences in the exam participation.. Anecdotal evidence from the first year's prize award ceremonies suggests that prior to these ceremonies some teachers did not fully understand that having students drop out or not take the test would reduce their chances of receiving a prize. After this experience, exam participation rates rose in the second year,

presumably because teachers worked harder to get students to take the exam. The test score gap between treatment and comparison schools was greater in the second year than in the first year.

8. Conclusion

Students in schools with a teacher incentive program in Kenya were more likely to take exams and had higher test scores in the short run. There is little evidence, however, that teachers responded to the program by taking steps to reduce dropouts or increasing effort on stimulating long-run learning. Teachers in program schools had no higher attendance rates or homework assignment rates. Pedagogy and student dropout rates were similar across schools. Instead, teachers in program schools increased test preparation activities and encouraged students enrolled in school to take the test. Following the end of the program, the test score difference between students who had attended treatment and comparison schools disappeared, consistent with a model in which teachers increased signaling effort but did not significantly increase effort to promote long-run learning.

It is worth noting several caveats (as well as caveats to these caveats). First, we cannot rule out the possibility that a larger incentive program or teacher-specific incentives would have induced not only increased test preparation, but also increased effort to improve underlying learning. However, the program provided an annual bonus of up to 40% of monthly income and was thus comparable in magnitude to most U.S. programs. Although the bonuses were a small percentage of yearly salary so the implied increase in daily wages was modest, if teachers chose attendance optimally prior to the program given their intrinsic motivation to teach, other incentives implicit in the system, and their value of time in other activities, they should have been indifferent to small changes in attendance, and hence modest incentives could potentially

have had a substantial effect. Moreover, while larger incentives might induce more effort by teachers, they could also induce more effort at counter-productive signaling, for example through cheating on tests or forcing weak students to drop out. They would also force teachers to bear more risk. Individual-level incentives for teachers could potentially undermine cooperation within the school.

A second caveat is that incentives may work as much by encouraging people who will be good teachers to enter the profession as by eliciting higher effort from those who would join the profession in any case. However, given the queuing for teaching positions, it is unlikely that people who either already have teaching jobs or who have the academic qualifications to enter teacher training colleges (but not universities) will select out of the profession. Any effect on this margin in Kenya and other developing countries with queues for teaching jobs is therefore likely to be small.

Third, the program was explicitly temporary. If teachers expected the program to continue indefinitely, and if they expected to remain at the schools for many years, they would have had more incentive to make long-run investments in learning.²² On the other hand, because the program was temporary it was possible to base incentives on improvements over baseline performance, to incorporate incentives to prevent students from dropping out, and to restrict the program to teachers already in school and thus to avoid strengthening incentives for teachers to seek transfers to schools with pupils from more advantaged backgrounds. A program without these features would be much less attractive since it would be difficult to provide incentives to teachers in weak schools, to prevent teachers from trying to influence the pool of pupils entering

²² In practice, many teachers transfer between.

their school, or to avoid increasing incentives for good teachers to try to transfer to the best schools.

Fourth, teachers in program schools may have exerted little effort because they believed that the test design is such that learning has only a small impact on test scores. Alternative tests that better measure long-run learning might potentially create better incentives. However, since the incentives provided by ICS were based on the official government of Kenya exams, which in turn are based on the official curriculum, any incentive program based around these exams is likely to run into similar difficulties.

Whatever the problems with teacher incentives, the status quo - with its 20% absence rate - is inadequate.

One strategy for improving incentives would be to attach incentives to measurable inputs, such as teacher attendance. In many countries, teachers' pay is not linked to pupils' test scores, but a teacher who is absent 20% of the time would typically face some sort of disciplinary sanction. An obvious problem with this approach is that attaching incentives to inputs rather than outputs can lead to undersupply of badly measured inputs. For example, since random audits can verify whether a teacher is in school, but not whether a teacher is in class, teachers might come to school, but not come to class, since outside inspectors cannot easily monitor presence in class.

While monitoring teacher absence is not sufficient to ensure good teaching it does seem like this should be part of an incentive scheme. It would be prohibitively expensive for outside inspectors to visit schools regularly enough to keep track of teachers' presence for incentive purposes, but headmasters could keep these records. However, headmasters currently do not have incentives to risk getting into a fight with teachers over absences. One indication of this comes from an evaluation of the program the NGO implemented at pre-schools. Preschool

teachers are not formal Ministry of Education employees, but instead are locally hired and paid by parents. In this program, school committees were given funds with which to provide bonuses to pre-school teachers, conditional on their not missing more than a specified number of days of class. If the funds were not spent on the teachers, school committees could use them for other pre-school purposes, so the school committees and school headmasters arguably had a strong incentive to monitor pre-school teachers. Nonetheless, the program yielded little if any improvement in absence rates, and it is clear that headmasters did not strictly enforce the rules requiring teacher attendance as a condition of the bonus being provided. This suggests that headmasters may find it costly to crack down on teacher absences under the current regime.

If there were in fact political will to address the problem of teacher absences, much could presumably be done. Schools are already required to keep log books of teacher attendance, but these are often not even filled out. An inspection system that penalized headmasters for failing to keep accurate records could allow monitoring of absences. The fact that such a system has not been implemented suggests that an analysis of the problem must ultimately turn to the political economy of education.

One possible reform would be greater decentralization and control by local school committees. Local school committees could more easily monitor teacher attendance and would arguably have incentives to do so. Since students' placement in secondary school depends on performance on the primary-school leaving exam, local communities and parents would share some of the same incentives to focus on test preparation as teachers (see Acemoglu, Kremer, and Mian 2002). Nonetheless, since teachers are transferred fairly frequently, local communities might have incentives to focus on longer-term outcomes than teachers' receiving bonuses based on current scores.

References

- Acemoglu, Daron, Kremer, Michael R., and Atif Mian (2002), "Markets, Firms and Governments." Unpublished.
- American Federation of Teachers, Teacher Salary Survey Archives at <http://www.aft.org/research/survey00/salarysurvey00.pdf>
- Chapman, David W, Snyder, Conrad W. and Shirley A. Burchfield (1991), "Teacher Incentives in the Third World," Agency for International Development Report no. 143 (Washington, DC).
- Chaudhury, Kremer, Muraldhiran, and Rogers (2003), "Absence Among Service Providers: A Preliminary Note." unpublished, World Bank.
- Clotfelter, Charles T. and Helen F. Ladd (1996), "Recognizing and Rewarding Success in Public Schools." in Helen F. Ladd ed. *Holding Schools Accountable: Performance-Based Reform in Education*. Brookings Institution. Washington, D.C.
- Education Commission of the States (2000), "Pay-for-Performance: Key Questions and Lessons from Five Current Models." ECS Issue Paper. Education Commission of the States, at www.ecs.org/clearinghouse/28/30/2830.htm.
- Glewwe, Paul W, Kremer, Michael R. and Sylvie Moulin (2001), "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya," mimeo.
- Gramlich, Edward and Patricia Koshel (1975), "Educational Performance Contracting," Brookings Institution. Washington, D.C.
- Hannaway, Jane (1992), "Higher Order Thinking, Job Design, and Incentives: An Analysis and Proposal," American Education Research Journal, vol. 29, 1, pp. 3-21.
- Harbison, Ralph W. and Eric A. Hanushek (1992), *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*, NY: Oxford University Press.
- Hanushek, Eric A., Kain, John F. and Steven R. Rivkin (1999), "Do Higher Salaries Buy Better Teachers?" unpublished.
- Hanushek, Eric A., Kain, John F. and Steven R. Rivkin (1998), "Teachers, Schools, and Academic Achievement," National Bureau of Economic Research Working Paper no. 6691 (Cambridge, MA).
- Hanushek, Eric A. (1996), "Outcomes, Cost, and Incentives in Schools," in Hanushek, Eric A. and Dale W. Jorgenson, eds. *Improving America's schools: The Role of Incentives*. National Academy. Washington, D.C.

Hanushek, Eric A. with Charles S. Benson et al (1994), *Making Schools Work: Improving performance and controlling costs*. Brookings Institution. Washington, D.C.

Holmstrom, Bengt and Paul Milgrom (1991), "Multi-Task Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization*, vol. 7, 0, p 24-52.

Jacob, Brian (2002), "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools," National Bureau of Economics Working Paper 8968.

Kingdon, Geeta and Francis Teal (2002), "Does performance related pay for teachers improve student performance? Some evidence from India," mimeo.

Lavy, Victor (2002a), "Paying for Performance: The Effect of Teachers' Financial Outcomes on Students' Scholastic Outcomes." mimeo.

Lavy, Victor (2002b), "Evaluating the Effect of Teacher Group Performance Incentives on Students Achievements." Forthcoming, *Journal of Political Economy*, December 2002.

Lockheed, Marlaine E. and Adriaan M. Verspoor (1991), *Improving Primary Education in Developing Countries*, NY: Oxford University Press.

Olsen, Lynn (1999), "Pay-Performance Link in Salaries Gains Momentum." *Education Week* October 13, 1999.

PROBE (Public Report on Basic Education for India) (1999), Oxford University Press.

Richards, Craig and Tian M. Sheu (1992), "The South Carolina School Incentive Reward Program: A Policy Analysis," *Economics of Education Review*, vol. 11, 1, pp. 71-86.

Table 1: Concentration of Teacher Absences

(1)	(2)	(3)
Five Group, Non-Parametric Model		
<i>Share of teachers</i>	<i>Attendance Percentage</i>	<i>% of Total Absence From This Group</i>
		<i>Total Absence ≈ 20%</i>
3.8%	1.1%	17.0%
2.8%	7.9%	11.7%
12.2%	58.1%	23.2%
42.5%	81.5%	35.6%
38.7%	92.9%	12.4%

MLE of Beta Distribution: $\alpha = 8.62$; $\beta = 1.57$

0.73%	0% - 50%	2.7%
2.4%	51% - 60%	7.1%
7.4%	61% - 70%	16.7%
17.9%	71% - 80%	28.6%
34.0%	81% - 90%	32.1%
37.5%	91% - 100%	12.7%

Notes: Column 1 shows the share of teachers in each group. Column 2 shows the percent of the time each group attends school.

Table 2: Program Effect on Teacher Entry and Exit

	(1)	(2)	(3)	(4)
<i>Dependent Variable:</i>	Exit Current School (0/1)		Enter New School (0/1)	
	Exit in 1997	Exit in 1998	Enter in 1998	Enter in 1999
Incentive	0.041	0.007	0.026	-0.002
School	(0.033)	(0.026)	(0.030)	(0.034)
Male	0.047	0.020	0.043	-0.092
	(0.032)	(0.031)	(0.032)	(0.035)***
Constant	0.137	0.209	0.190	0.234
	(0.024)***	(0.019)***	(0.022)***	(0.025)***
Observations	1157	1227	1227	1228

Notes:

Standard errors in parentheses; regressions include school-level random effects.

* significant at 10%; ** significant at 5%;

For exiting regressions, the incentive/non-incentive refers to the originating school; for entry regressions the incentive/non-incentive refers to the destination school. The unit of observation in all regressions is the teacher.

Table 3: Program Effect on Teacher Attendance

<i>Dependent Variable:</i>	(1)	(2)	(3)	(4)	(5)
	Teacher Attendance Percentage			Attendance Differences (Attendance _{program year} – Attendance _{pre-program year})	
	Year 0	Year 1	Year 2	Year 1 - Year 0	Year 2 – Year 0
Incentive	0.012	-0.008	-0.011	-0.007	-0.063
School	(0.043)	(0.019)	(0.022)	(0.048)	(0.049)
Grade	-0.005	-0.010	0.000	-0.009	0.002
	(0.012)	(0.007)	(0.009)	(0.015)	(0.016)
Male (0/1)	0.015	0.007	-0.108	-0.028	-0.095
	(0.045)	(0.022)	(0.025)***	(0.053)	(0.055)*
Constant	0.828	0.882	0.904	0.049	0.064
	(0.073)***	(0.044)***	(0.049)***	(0.090)	(0.094)
Observations	466	397	320	396	319

Notes:

Standard errors in parentheses; regressions include school-level random effects.

* significant at 10%; ** significant at 5%

The dependent variable is the percentage of the visits for which the teacher was present, based on two visits in 1997, five visits in 1998 and three visits in 1999. The unit of observation is the teacher.

Table 4: Program Effect on Homework Assignment

<i>Dependent Variable:</i>	(1)	(2)	(3)	(4)	(5)
	Homework Assignment (0/1)			Homework Differences ($Homework_{program\ year} - Homework_{pre-program\ year}$)	
	Year 0	Year 1	Year 2	Year 1 - Year 0	Year 2 - Year 0
Incentive	0.012	-0.052	-0.009	-0.092	-0.042
School	(0.042)	(0.045)	(0.047)	(0.055)*	(0.059)
Grade	0.079	0.062	0.149	-0.017	0.036
	(0.007)**	(0.007)**	(0.007)**	(0.017)	(0.017)**
Constant	-0.176	-0.060	-0.586	0.137	-0.155
	(0.049)**	(0.053)	(0.055)**	(0.111)	(0.111)
Observations	1914	1676	2371	431	427

Notes:

Standard errors in parentheses; regressions include school-level random effects.

* significant at 10%; ** significant at 5%;

In columns 1 through 3 each observation represents a student asked about homework assignment in the previous day; in columns 4 and 5 differences across years are calculated at the school-grade level.

Table 5: Program Effects on Pedagogy

	(1)	(2)	(3)	(4)	(5)
	Year 0	Year 1	Year 2	Year 1 - Year 0	Year 2 – Year 0
Panel A					
<i>Dependent Variable: Use of Blackboard (0/1)</i>					
Incentive	0.018	-0.032	0.038	-0.051	0.078
School	(0.031)	(0.026)	(0.051)	(0.036)	(0.069)
Grade	0.010	-0.003	-0.018	-0.001	-0.029
	(0.009)	(0.007)	(0.013)	(0.014)	(0.022)
Constant	0.875	0.973	0.989	0.021	0.098
	(0.054)**	(0.044)**	(0.084)**	(0.085)	(0.133)
Observations	404	598	237	246	149
Panel B					
<i>Dependent Variable: Use Teaching Aid (0/1)</i>					
Incentive	-0.026	-0.006	0.012	0.025	0.052
School	(0.032)	(0.031)	(0.035)	(0.052)	(0.067)
Grade	-0.021	-0.006	-0.004	-0.016	0.002
	(0.012)*	(0.009)	(0.013)	(0.021)	(0.025)
Constant	0.235	0.143	0.094	0.093	-0.040
	(0.073)**	(0.059)**	(0.080)	(0.124)	(0.151)
Observations	399	567	235	241	147
Panel C					
<i>Dependent Variable: Teacher Caring (1 to 5: 1=very caring, 5=mean or rude)</i>					
Incentive	-0.080	-0.065	-0.051	0.052	-0.058
School	(0.104)	(0.062)	(0.125)	(0.133)	(0.178)
Grade	0.018	-0.010	0.125	-0.025	0.093
	(0.034)	(0.022)	(0.031)**	(0.048)	(0.062)
Constant	1.586	1.701	1.184	0.122	-0.280
	(0.204)**	(0.135)**	(0.205)**	(0.292)	(0.375)
Observations	382	571	234	238	146
Panel D					
<i>Dependent Variable: Teacher Energy (1 to 5: 1=energetic, 5=lazy/dull)</i>					
Incentive	-0.030	-0.041	0.164	0.050	0.070
School	(0.096)	(0.080)	(0.120)	(0.167)	(0.195)
Grade	-0.023	-0.019	0.070	-0.017	0.092
	(0.035)	(0.023)	(0.027)**	(0.052)	(0.062)
Constant	1.926	1.870	1.126	0.073	-0.798
	(0.211)**	(0.146)**	(0.180)**	(0.324)	(0.377)**
Observations	383	570	233	239	146

Notes:

Standard errors in parentheses; regressions include school-level random effects.

* significant at 10%; ** significant at 5%.

Each observation in columns 1 through 3 represents a classroom; differences in columns 4 and 5 are calculated at the school-grade level. There are fewer observations in 1999 because only one class per school/grade was observed that year.

Table 6: Program Effect on Extra Coaching

<i>Dependent Variable:</i>	(1)	(2)	(3)	(4)	(5)
	Year 0	Year 1	Year 2	Year 1 - Year 0	Year 2 - Year 0
	Coaching (0/1)			Coaching Differences (<i>Coaching</i> _{program year} - <i>Coaching</i> _{pre-program year})	
<i>All Coaching (Vacation and During School)</i>					
Incentive School	-0.007 (0.044)	0.042 (0.037)	0.074 (0.034)**	0.049 (0.042)	0.081 (0.047)*
Grade	0.155 (0.009)***	0.135 (0.007)***	0.103 (0.007)***	-0.021 (0.009)**	-0.052 (0.009)***
August Holiday (0/1)	-0.020 (0.016)	0.058 (0.030)*	-0.122 (0.034)***	0.078 (0.037)**	-0.102 (0.036)***
December Holiday (0/1)	-0.370 (0.025)***	-0.452 (0.029)***	-0.534 (0.030)***	-0.082 (0.031)**	-0.164 (0.036)***
Term Visit 1 (0/1)	0.094 (0.035)***	0.126 (0.040)***	0.130 (0.038)***	0.032 (0.035)	0.036 (0.052)
Term Visit 2 (0/1)	0.094 (0.035)***	0.168 (0.041)***	0.282 (0.040)***	0.074 (0.052)	0.188 (0.052)***
Term Visit 3 (0/1)	0.096 (0.035)***	0.158 (0.040)***	0.242 (0.043)***	0.062 (0.051)	0.146 (0.055)***
Constant	-0.502 (0.064)***	-0.372 (0.053)***	-0.121 (0.052)**	0.130 (0.064)**	0.381 (0.064)***
Observations	3000	3000	3000	3000	3000
<i>Vacation Coaching</i>					
Incentive School	0.035 (0.034)	0.089 (0.031)***	0.091 (0.035)**	0.055 (0.038)	0.056 (0.046)
Grade	0.156 (0.008)***	0.139 (0.006)***	0.118 (0.007)***	-0.017 (0.008)*	-0.038 (0.009)***
August Holiday (0/1)	-0.020 (0.016)	0.058 (0.030)*	-0.122 (0.034)***	0.078 (0.037)**	-0.102 (0.036)***
December Holiday (0/1)	-0.370 (0.025)***	-0.452 (0.029)***	-0.534 (0.030)***	-0.082 (0.031)**	-0.164 (0.036)***
Constant	-0.527 (0.049)***	-0.425 (0.049)***	-0.219 (0.054)***	0.103 (0.057)*	0.308 (0.059)***
Observations	1500	1500	1500	1500	1500

Notes:

Standard errors in parentheses; regressions include school-level random effects. * significant at 10%; ** significant at 5%;

Coaching is reported at 6 times during the year for each grade: 3 vacation terms and three periods during the year; each observation represents a school grade at a given time during the year. Rates for given time periods are reported compared to the omitted time period, the April holiday.

Table 7: Program Effect on Dropout and Repetition Rates

	(1)	(2)	(3)	(4)
	Year 1		Year 2	
	Dropout	Repetition	Dropout	Repetition
Incentive School	0.004 (0.010)	-0.012 (0.019)	0.002 (0.015)	-0.010 (0.026)
Male (0/1)	0.021 (0.005)**	0.010 (0.008)	0.018 (0.005)**	-0.009 (0.008)
Constant	0.073 (0.007)**	0.295 (0.014)**	0.116 (0.011)**	0.256 (0.018)**
Observations	14153	12686	14545	12671

Notes:

Standard errors in parentheses; regressions include school-level random effects.

* significant at 10%; ** significant at 5%;

Each observation represents an upper primary school student.

Table 8: Program Effects on Participation in Exams

	(1) Year 1	(2) Year 2	(3) Year 3 (Post-Program)
Panel A			
<i>Dependent Variable: Take District Exam (0/1)</i>			
Incentive School	0.060 (0.033)*	0.108 (0.028)**	-0.022 (0.034)
Male (0/1)	0.015 (0.008)*	-0.011 (0.009)	0.007 (0.016)
Grade	0.058 (0.003)**	0.036 (0.004)**	0.010 (0.010)
Constant	0.389 (0.029)**	0.525 (0.034)**	0.560 (0.069)**
Observations	10690	7158	3642
Panel B			
<i>Dependent Variable: Take ICS Exam (0/1)</i>			
Incentive School	0.005 (0.007)	0.031 (0.024)	
Male (0/1)	0.004 (0.004)	0.000 (0.007)	
Grade	0.010 (0.001)**	0.009 (0.003)**	
Constant	0.880 (0.010)**	0.822 (0.027)**	
Observations	14397	7158	
Panel C			
<i>Dependent Variable: Take KCPE Exam (0/1)</i>			
Incentive School	-0.014 (0.047)	0.026 (0.017)	0.021 (0.039)
Male (0/1)	-0.000 (0.019)	0.001 (0.015)	0.013 (0.029)
Constant	0.779 (0.034)**	0.912 (0.014)**	0.813 (0.029)**
Observations	1624	1265	681

Note: Standard errors in parentheses; school-level random effects included.

* significant at 10%; ** significant at 5%;

District test data was not available for Teso District in 1998.

ITT methodology employed.

Each observation represents an upper primary school pupil in year 0; columns 2 and 3 are limited to pupils who did not repeat or drop out in any year.

Table 9: Program Effect on Test Scores by Subject (District Exam)

<i>Dependent Variable</i>	(1)	(2)	(3)	(4)	(5)	(6)
	Test Scores (Standardized relative to comparison schools)			Test Score Differences $Test Score_{program\ year} -$ $Test Score_{pre-program\ year}$		
	Year 1	Year 2	Year 3	Year 1 – Year 0	Year 2 – Year 0	Year 3 – Year 0
English	-0.059 (0.107)	0.094 (0.094)	0.017 (0.112)	-0.024 (0.071)	-0.003 (0.086)	-0.091 (0.122)
Math	0.058 (0.089)	0.099 (0.084)	-0.077 (0.089)	0.076 (0.054)	0.150** (0.064)	-0.106 (0.089)
Science	0.015 (0.091)	0.155 (0.102)	0.121 (0.115)	0.050 (0.076)	0.206* (0.094)	0.194 (0.128)
Swahili	-0.052 (0.093)	0.105 (0.072)	0.091 (0.084)	0.023 (0.083)	0.019 (0.094)	-0.134 (0.221)
G.H.CR.	-0.039 (0.089)	0.202** (0.097)	0.055 (0.105)	0.205* (0.107)	0.341** (0.129)	-0.021 (0.262)
A.C.M.	-0.007 (0.096)	0.010 (0.092)	-0.049 (0.102)	0.116 (0.121)	0.108 (0.154)	-0.218 (0.249)
HS. BE.	0.049 (0.092)	0.073 (0.107)	-0.079 (0.113)	0.073 (0.161)	0.167 (0.196)	-1.232** (0.525)
All Subjects & Grades	-0.040 (0.079)	0.136* (0.077)	-0.087 (0.083)	0.054 (0.054)	0.139** (0.065)	-0.008 (0.084)
Observations	50,842	37,620	15,893	24,677	15,641	5,330

Note: Standard errors in parentheses; regressions include school-level random effects.

* significant at 10%; ** significant at 5%;

Year 1 district test results are available only for Busia

Each row represents a random effects regression of test scores on a dummy variable for teacher incentive schools and on region and sex dummy variables, based on data on the 100 schools in Teso and Busia Districts. For each grade/subject combination, test scores were standardized by subtracting the mean score and dividing by the standard deviation of the test score from the comparison schools.

Each observation represents a test score in a particular subject for an upper primary school pupil; columns 2 and 3 are limited to pupils who were enrolled in year 1 and did not repeat or drop out. Columns 4, 5, 6 impose the additional restriction that a pre-test score is available.

The district exam was only given to grade 8 students in Teso District in 1998. Moreover, the test score information was only collected for the program schools in Teso in 1998. 7,848 students (grades 4 to 8) took at least one district exam in year 1. Of these, 5,751 had pre-test scores from a pre-program year, in this case 1996. In year 2, when exam results are also available for Teso, 10,927 students (grade 4 to 8) took at least one exam and 6,365 of these students also had pre-test scores from the same pre-program year. In the post-program year, 9,613 students (grade 4 to 8) took at least one exam and 4,016 of these had pre-test scores. In later years more students have no pre-test scores because students who enter the sample (by reaching 4th grade) after the first year will not have pre-test scores. So, for example, in the post-program year students in 4th and 5th grade will, by definition, not have pre-test scores.

Table 10: Program Effect on Test Scores by Subject (Non-Incentive Tests)

<i>Dependent Variable</i>	(1)	(2)	(3)	(4)	(5)	(6)
	Test Scores (Standardized relative to comparison schools)			Test Score Differences $Test\ Score_{program\ year} - Test\ Score_{pre-program\ year}$		
	Year 1	Year 2	Year 3	Year 1 – Year 0	Year 2 – Year 0	Year 3 – Year 0
Panel A						
<i>Dependent Variable: ICS Subject Test Scores</i>						
English	0.077 (0.090)	0.079 (0.121)		0.001 (0.040)	0.011 (0.085)	
Math	0.053 (0.074)	0.077 (0.070)		-0.042 (0.041)	0.010 (0.056)	
Science	0.129 (0.082)			0.091** (0.043)		
All Subjects & Grades	0.089 (0.079)	0.085 (0.090)		0.017 (0.033)	0.014 (0.057)	
Observations	39,510	18,558		32,993	15,225	
Panel B						
<i>Dependent Variable: KCPE Test Scores</i>						
English	0.116 (0.094)	0.103 (0.126)	0.002 (0.125)	-0.045 (0.105)	-0.120 (0.137)	-0.130 (0.192)
Math	0.166 (0.102)	0.120 (0.099)	0.044 (0.124)	0.123 (0.103)	0.145 (0.124)	0.071 (0.185)
Science	0.132 (0.098)	0.113 (0.114)	0.040 (0.142)	0.200* (0.114)	0.189 (0.142)	0.156 (0.194)
Swahili	0.212* (0.121)	0.226** (0.112)	-0.126 (0.131)	0.107 (0.126)	0.081 (0.123)	-0.473* (0.270)
G.H. CR	0.167* (0.088)	0.257** (0.115)	0.004 (0.120)	0.149 (0.108)	0.336** (0.136)	-0.053 (0.253)
A.C.M.	0.054 (0.098)	0.169 (0.117)	-0.027 (0.133)	0.139 (0.121)	0.169 (0.168)	-0.058 (0.267)
HS. BE.	0.125 (0.091)	0.154 (0.111)	0.052 (0.128)	-0.008 (0.146)	0.268 (0.175)	-1.285* (0.658)
All Subjects & Grades	0.138* (0.074)	0.165* (0.090)	-0.009 (0.101)	0.104 (0.080)	0.152 (0.097)	-0.006 (0.138)
Observations	10,430	8,427	4,053	7,152	5,247	1,505

Note: Standard errors in parenthesis; regressions include school-level random effects.

* significant at 10%; ** significant at 5%;

Year 3 ICS tests were given only in 27 schools so scores are not reported; KCPE tests are taken by grade 8 students only. Each row represents a random effects regression of test scores on a dummy variable for teacher incentive schools and on region and sex dummy variables, based on data on the 100 schools in Teso and Busia Districts. For each grade/subject combination, test scores were standardized by subtracting the mean score and dividing by the standard deviation of the test score from the comparison schools.

Each observation represents an upper primary school pupil; columns 2 and 3 are limited to pupils who were enrolled in year 1 and did not repeat or drop out.

13,339 students (grades 4 to 8) took at least one subject of the ICS exams in year 1. Of these, 11,298 had pre-test scores from year 0, in the form of normalized district exam scores from year 0. 15,647 students took at least one ICS exam in year 2, of which 8,638 had pre-test scores from year 0. 1,490 eight graders took at least one KCPE exam in year 1, of which 1,026 had pre-test scores from year 0. 1,584 students took at least once KCPE exam in year 2, of which 944 had pre-test scores. 1,537 students took at least one KCPE exam in year 3, of which 839 had pre-test scores.

Appendix Table A: Pre-Program Means

	Incentive Schools		Comparison Schools	
	Mean (Std. Dev)	Num. of observations	Mean (Std. Dev)	Num. of observations
Teacher Outcomes				
Attendance	0.833 (0.334)	180	0.862 (0.314)	167
Home Work Assignment	0.310 (0.463)	900	0.302 (0.459)	947
Pedagogy -Teacher present	0.673 (0.469)	377	0.708 (0.455)	394
Pedagogy -Teacher's caring	4.400 (0.827)	227	4.332 (0.812)	262
Pedagogy -Teacher's energy level	4.237 (0.834)	227	4.224 (0.847)	263
Pedagogy –Teacher's Organization	2.572 (0.608)	226	2.580 (0.618)	262
Pedagogy -Pupil's Participation	4.151 (0.837)	225	4.110 (0.833)	263
Extra-Coaching	0.484 (0.500)	1,224	0.489 (0.500)	1,198
<u>Student Outcomes</u>				
Dropout Rates	0.133 (0.339)	6,711	0.129 (0.336)	7,382
Repetition Rates	0.267 (0.442)	5,920	0.262 (0.440)	6,352
Participation in District Exams	0.800 (0.399)	9,507	0.800 (0.400)	10,472

Note: Teacher attendance = 1 if teacher present on the day of visit, = 0 otherwise.

Homework assignment = 1 if homework assigned on previous day, = 0 otherwise.

Pedagogy; Teacher present (1=yes, 2=no). Teacher' s behavior, energy level, and organization level are categorized as follows: Behavior (5=Very caring and attentive, 4=fairly caring and attentive, 3=neutral, 2=somewhat uncaring, 1=mean or rude); Energy level (5=very enthusiastic/alert/active, 4=fairly enthusiastic, 3=neutral, 2=bit dull/tired, 1=lazy/very dull); Organizational level (3=yes, 2=somewhat, 1=no); Pupil's participation (5=very enthusiastic/alert/active, 4=fairly enthusiastic, 3=neutral, 2=a bit passive, 1=very passive/inactive/tired).

Extra-Coaching = 1 if coaching took place during school term or school vacation for a particular grade, = 0 otherwise.

Dropout = 1 if student enrolled in the last year drops out in the current year, = 0 if student continues.

Repetition = 1 if student continues in the same grade in the following year, =0 if student promoted to higher grade.

Participation = 1 if student took the exam, =0 otherwise.

Appendix Table B: Program Effect on Test Scores by Grade

<i>Dependent Variable</i>	Test Scores (Standardized relative to comparison schools)			Test Score Differences $Test\ Score_{program\ year} - Test\ Score_{pre-program\ year}$		
	Year 1	Year 2	Year 3	Year 1 – Year 0	Year 2 – Year 0	Year 3 – Year 0
Panel A						
<i>Dependent Variable: District Subject Test Scores</i>						
Grade 4	-0.021 (0.087)			0.009 (0.086)		
Grade 5	-0.096 (0.112)	0.131 (0.090)	-0.313 (0.427)	-0.081 (0.089)	0.120 (0.087)	0.101 (0.795)
Grade 6	-0.006 (0.113)	0.059 (0.105)	-0.025 (0.091)	0.183** (0.085)	0.100 (0.106)	-0.087 (0.102)
Grade 7	0.008 (0.095)	0.068 (0.090)	0.009 (0.111)	0.106 (0.080)	0.077 (0.096)	0.062 (0.115)
Grade 8	0.082 (0.082)	0.159* (0.095)	-0.096 (0.120)	0.052 (0.083)	0.152 (0.099)	0.029 (0.159)
Observations	50,842	37,620	15,893	24,677	15,641	5,330
Panel B						
<i>Dependent Variable: ICS Subject Test Scores</i>						
Grade 4	0.118 (0.078)			0.105 (0.068)		
Grade 5	0.002 (0.103)	0.092 (0.099)	-0.085 (0.696)	0.035 (0.083)	-0.004 (0.086)	-1.178 (0.682)
Grade 6	0.094 (0.087)	0.068 (0.129)	0.277* (0.146)	0.132* (0.073)	0.128 (0.138)	0.085 (0.191)
Grade 7	0.053 (0.083)	0.082 (0.089)	0.013 (0.143)	-0.029 (0.072)	0.057 (0.095)	0.229 (0.230)
Grade 8	0.124 (0.092)	-0.001 (0.154)	0.128 (0.263)	0.083 (0.092)	-0.073 (0.169)	0.303 (0.269)
Observations	39,510	12,996	1,593	32,993	15,225	1,142

Note: Standard errors in parenthesis; regressions include school-level random effects.

* significant at 10%; ** significant at 5%;

Year 1 district test results are available only for Busia

Each row represents a random effects regression of test scores on a dummy variable for teacher incentive schools interacted with dummy variable for grade and on region and sex dummy variables.

Each observation represents an upper primary school pupil; columns 2 and 3 are limited to pupils who were enrolled in year 1 and did not repeat or drop out.

Appendix Table C : Pre-Program Difference in Mean District Test Scores Among Students Who Had Post-Test Scores in District and ICS Exams in Year 1

Subject/Grade	District Exam	ICS Exam
	Difference	Difference
	(1)	(2)
English	0.041 (0.122)	0.085 (0.083)
Math	-0.032 (0.099)	0.032 (0.070)
Science	-0.046 (0.098)	-0.023 (0.072)
Swahili	-0.020 (0.102)	—
G.H.CR.	-0.139 (0.116)	—
A.C.M.	-0.128 (0.114)	—
HS.BE.	-0.008 (0.152)	—
2	0.209 (0.126)	0.064 (0.079)
3	-0.118 (0.143)	-0.042 (0.108)
4	-0.100 (0.156)	-0.065 (0.110)
5	-0.044 (0.114)	0.107 (0.089)
6	-0.111 (0.105)	0.216 (0.087)
7	-0.024 (0.096)	0.082 (0.100)
8	-0.656 (0.305)	0.046 (0.314)
All Subjects & Grades	0.011 (0.091)	0.035 (0.062)
	(obs=24,716)	(obs=30,774)

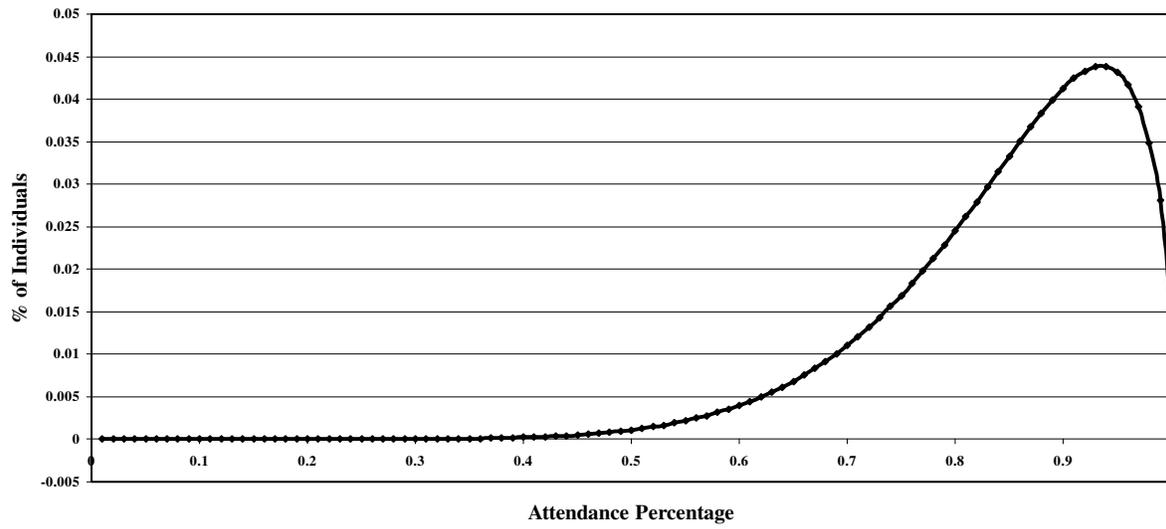
Note: * = 10% significance; ** = 5% significance.

Standard errors in parentheses.

Each row represents a random effects regression of test scores on a dummy variable for teacher incentive schools and on region and sex dummy variables, based on data on the 100 schools in Teso and Busia Districts. For each grade/subject combination, test scores were standardized by subtracting the mean score and dividing by the standard deviation of the test score from the comparison schools.

The sample includes all upper primary students (grade 4-8) in year 1 who had pre-program scores in year 0.

Figure 1
Beta Distribution of Attendance Probabilities



Notes: The graph represents the distribution of attendance percentages across the population with the ML estimates of the beta distribution ($\alpha = 8.62$, $\beta = 1.57$).