NATIONAL CENTER ON
Performance Incentives

Teacher Incentives in Developing Countries:

Recent Experimental Evidence from Kenya

Paul Glewwe
Nauman Ilias
Michael Kremer

Prepared for Performance Incentives:
Their Growing Impact on American K-12 Education
in Nashville,Tennessee on February 28, 2008

Working Paper 2008-09
February 2008

LED BY
VANDERBILT
PEABODY COLLEGE

IN COOPERATION WITH:
RAND CORPORATION
Mizzou
University of Missouri - Columbia

The National Center on Performance Incentives (NCPI) is charged by the federal government with exercising leadership on performance incentives in education. Established in 2006 through a major research and development grant from the United States Department of Education's Institute of Education Sciences (IES), NCPI conducts scientific, comprehensive, and independent studies on the individual and institutional effects of performance incentives in education. A signature activity of the center is the conduct of two randomized field trials offering student achievement-related bonuses to teachers. The Center is committed to air and rigorous research in an effort to provide the field of education with reliable knowledge to guide policy and practice.

The Center is housed in the Learning Sciences Institute on the campus of Vanderbilt University's Peabody College. The Center's management under the Learning Sciences Institute, along with the National Center on School Choice, makes Vanderbilt the only higher education institution to house two federal research and development centers supported by the Institute of Education Services.

This working paper was supported by the National Center on Performance Incentives, which is funded by the United States Department of Education's Institute of Education Sciences (R30SA06034). This is a draft version of a paper that will be presented at a national conference, Performance Incentives; Their Growing Impact on American K-12 Education, in Nashville, Tennessee on February 28-29, 2008. The authors would like to thank two anonymous reviewers for detailed comments on this chapter. This paper reports selected results from Glewwe, Kremer and Ilias (2003).The reader should refer to that paper for a more detailed analysis of the teacher incentive program in Kenya. The views expressed in this paper do not necessarily reflect those of sponsoring agencies or individuals acknowledged. Any errors remain the sole responsibility of the author.

Please visit **www.performanceincentives.org** to learn more about our program of research and recent publications.

Teacher Incentives in Developing Countries: Recent Experimental Evidence from Kenya

PAUL GLEWWE
University of Minnesota

NAUMAN ILIAS
The Brattle Group

MICHAEL KREMER
Harvard University and Brookings Institution

Abstract

This paper reviews recent evidence on the impact of a teacher incentives program in Kenya. The results are based on a randomized trial, which removes many sources of bias that can arise in analyses of non-experimental data. One hundred schools in a rural area were randomly divided into 50 that participated in a teacher incentives program and 50 that served as controls. Students in the schools that were selected for the teacher incentives programs had higher test scores on exams linked to incentives during the time the program was in place, but test scores did not increase significantly on exams that were not linked to incentives, and test score gains on exams linked to incentives did not persist after the program ended. Teachers in the 50 program schools did conduct more test preparation sessions, and students in those schools were more likely to take exams. Overall, there is little evidence that the teacher incentives program increased student learning.

1. Introduction

Teacher incentive programs are policies that explicitly link teacher's compensation to their students' performance.[1] Until the 1990s, such programs were quite rare, but in the last 15 years they have become much more common. In the United States, examples include programs in Rhode Island (beginning in 1999), Denver (beginning in 1999-2000), and Douglas County, Colorado (1994) (Olsen, 1999; Education Commission of the States, 2000). They typically offer annual merit pay that amounts to from 10% to 40% of an average teacher's monthly salary (American Federation of Teachers, 2000). Teacher incentive programs have also been initiated in other countries. Two examples are Israel, which provides incentives to teachers based on students' scores (Lavy 2002b), and a World Bank-funded program in Mexico that provides performance incentives to primary school teachers.

Advocates of teacher incentive programs point out 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 et al., 1998; Hanushek, 1996; Lockheed and Verspoor, 1991). They claim that linking teachers' pay to students' performance will increase teacher effort. Opponents respond that, since teachers' tasks are multi-dimensional and tests can measure only some dimensions of learning, linking compensation to test scores gives teachers an incentive to sacrifice promoting curiosity and creative thinking in pursuit of skills that are tested on standardized exams (Holmstrom and Milgrom, 1991; Hannaway, 1992).

In many developing countries, incentives for teachers are usually very weak. One piece of evidence for this is high rates of teacher absenteeism. Chaudhury et al. (2006) show that

[1] Some U.S. school districts use the term "teacher incentive programs" more broadly, to include increasing teacher salaries in return for teaching certain subjects or teaching in schools where student performance is low. In this paper we define this term more narrowly, including only programs that link teacher pay to student performance.

primary school teachers were absent from school 27% of the time in Uganda, 25% in India, 14% in Ecuador and 11% in Peru. Our data from Kenya reveal a teacher absenteeism rate of 20% in rural primary schools. Even when Kenyan teachers are at the school they are often not in the classroom; classroom observation data show a teacher absenteeism rate of 27%.

Given this situation, advocates for teacher incentives seem to have a stronger case in developing countries than in developed countries. That is, in environments with very weak incentives, one could argue that the key problem is to get teachers to show up at all. Since most teaching in many developing countries is by rote, the risk of reducing efforts to stimulate creativity seems to be very low. On the other hand, teachers and school administrators in developing countries may respond to test score-based incentives in ways that are more pernicious than teaching to the test. For example, to raise average scores on exams they could deliberately force weak students to repeat grades or even drop out.

This paper reviews recent work by the authors that evaluates a teacher incentives program in Kenyan primary schools. For a more detailed analysis, see Glewwe, Ilias and Kremer (2003). Unlike virtually all analyses of teacher incentives in developed countries, our results are based on a randomized evaluation; that is, we compare 50 rural primary schools that were randomly assigned to implement the program to 50 schools that were randomly assigned to a group that did not implement the program. To discourage the potential for pernicious outcomes along the lines mentioned above, students who dropped out or did not take the test were assigned very low scores.

At first glance, our teacher incentives program appears to have been successful. During the two years that the program operated, students in the 50 program schools were more likely to take exams, which suggests that the incentives to avoid pernicious effects worked. More importantly, the students in those schools scored higher on the exams that were linked to teacher incentives.

Yet further examination of the evidence raises serious doubts about the effectiveness of the program. First, little evidence is found showing increased teacher effort. Teacher attendance did not improve, homework assignment did not increase, and pedagogy did not change. Second, the test scores gains (on the exams linked to the incentives) did not persist after the program ended. Finally, test scores on exams that were not linked to teacher incentives did not increase significantly.

A particularly interesting finding is that teachers in the program schools were more likely to conduct test preparation sessions outside of normal class time. There is also some evidence that the program reduced the rate at which students left blank answers on multiple choice exams and increased students' propensity to complete the exams. This is consistent with an interpretation that teachers focused their efforts on actions that raise scores in the short run, but do not on efforts that result in long-run increases in students' cognitive skills. While it is impossible to rule out the possibility that the program increased students' learning, we find little evidence of this. The type of effort that we observe seems skewed towards activities that increase test scores per se, as distinct from learning, and this interpretation is supported by the lack of any program impact on tests given after the program ended and on a different set of tests (with a somewhat different format).

This paper differs from earlier work in several ways. First, since teacher incentive programs are likely to be introduced in areas where teacher performance is relatively low, the introduction of teacher incentives may be correlated with other, unobserved factors that affect teacher performance, which implies that it is difficult to identify econometrically the effect of such teacher incentive programs. We address this problem by randomly assigning schools to treatment (program) and comparison groups. Second, since both advocates and opponents of teacher incentive programs agree that incentives can increase test scores, but disagree about whether these higher test scores reflect increased overall teacher effort or just 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. In particular, we examine teacher absences, teacher behavior in the classroom, scores on the exams linked to incentives, scores on exams not linked to exams, and scores on exams given after the program ended. Third, while most of the research on teacher incentives takes place in the U.S., we examine a poor developing country, since 90% of the children born each year in the world are born in developing countries.

The remainder of the paper is organized as follows. Section 2 provides a brief discussion of primary education in Kenya. The next section describes the teacher incentives program that we implemented and the process by which schools were selected for it. Sections 4 and 5 report the impact of the program on teacher and student outcomes, respectively, and Section 6 concludes.

2. Education in Kenya

Almost all children in Kenya start primary school. Students in grades one, two, and three are taught in English, Kiswahili, and the local language (Kiluhya in two thirds of our sample and Ateso in the rest). From grade four on, all instruction is in English. At the end of grade eight, Kenyan students take a national exam that determines which secondary schools they can attend. In many primary schools, only the best students are promoted to grade eight, while the rest repeat grade seven or drop out. Many students drop out before reaching higher grades.

In Kenyan schools, the Ministry of Education sets the curriculum, administers national and district exams, and hires and pays teachers. Local school committees, composed primarily of parents, must raise funds for all other costs of running the school. Fundraising for major capital expenses, such as construction, takes place through *harambees*, which are large-scale fundraising drives. Recurrent costs, such as minor repairs, chalk, and books for teachers have historically been covered through various school fees. In practice, headmasters and parents often

bargain over how much of the official fee a particular family must pay. Physical facilities at schools in rural areas are minimal; classrooms are often dilapidated, and in some cases non-existent. Most students sit two or three to a bench, and some sit on the floor. Rural schools usually have textbooks for teachers, but few provide them to students.

In contrast to the U.S., the system for selecting teachers in Kenya is highly centralized. Teacher hiring, firing, and transfer decisions are made centrally by the Ministry of Education in Nairobi. Hiring is based primarily on academic qualifications.[2] Salaries are set through collective bargaining between the government and the politically powerful Kenyan National Union of Teachers (KNUT). At the time of the project,primary school teachers earned an average of Ksh 126,921 ($2162) in allowances and pay (World Bank, 2005).[3] This is more than six times the annual GDP per capita in that year.

Teacher salaries depend primarily on education and experience. There is little opportunity for performance-based promotion or increases to salary. 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, and the government may look more favorably on requests for transfers to desirable postings or to home areas from teachers who perform well.

Although the school system offers few incentives to teachers, every school is supposed to have a parent committee, and these committees sometimes provide gifts for teachers when schools perform well on the national exams. Similarly, there are sometimes public demonstrations against teachers who have behaved badly in some way, which creates pressure on the Ministry of Education to transfer the teacher to a different school. However, most school

[2] 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.
[3] This is assuming an exchange rate of 58.7 shillings per dollar, the 1997 dollar-shilling exchange rate.

committees do not provide supplemental bonuses, and school committees typically attempt to influence the national authorities only in extreme situations.

To the extent that incentives do exist, they derive from the system of national testing. Results on the national primary school leaving exam (the KCPE) are front-page news in Kenya; newspapers publish lists of the highest-scoring schools. Results from the KCPE and from district exams administered in the upper grades of primary school are often posted in headmasters' offices.

Since the KCPE determines what secondary schools, if any, will accept graduating primary school students, teachers devote considerable effort to preparing for it. Many schools hold extra preparation sessions outside of normal class hours to prepare students for the exams. These "prep" sessions take place during evenings, weekends or vacation periods. They consist of many different activities, ranging from class-work similar to normal classroom sessions to direct test preparation activities like going over old exams. Yet these sessions are generally more heavily weighted to specific test-preparation activities than normal classroom sessions. Teachers in the rural schools in Kenya sometimes receive some additional compensation from parents for these sessions, but the compensation is minimal; a typical student will pay only 10-20 Ksh (about 16 to 33 US cents) per school term to attend "preps" held in the school, and students whose parents cannot pay are not excluded. The amount of time students spend in preps varies widely, but a typical amount of time during a term (when school is in session) is about five hours per week, while a typical amount during vacation periods is about 25 hours per week (five hours per day for five days).

While the considerable attention given to results on the national KCPE exam clearly spurs effort by some teachers to raise average 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 8^{th} grade and take the KCPE exam.

One indication that teacher incentives are weak is high teacher absenteeism. Random visits conducted to check pupil attendance and observe pedagogy suggest that teachers in comparison schools were absent from school about 20% of the time. For comparison, absence rates among staff at a non-profit organization working in the same area are around 6%. In our opinion, these high absence rates reflect the fact that teachers have political power and are able to prevent much enforcement by school administrators. Teachers in Kenya have a very strong union. As noted above, they are paid roughly five times per capita GDP, a high ratio even for poor countries. Given these benefits, it is not surprising that there is considerable queuing for teaching positions, with many qualified applicants waiting for jobs.

High rates of teacher absenteeism is not the only manifestation of low incentives for teachers to devote effort to teaching. We find that teachers are often in school, but absent from their classrooms. Although teachers were absent from school only 20% of the time, enumerators who visited classes to observe pedagogy found teachers absent from class about 27% of the time. Even when teachers are present in class, they usually arrive late; only a small percentage of teachers were in the classroom at the time that the class officially started. Casual observation indicates that teachers are often drinking tea in the staff room with other teachers during class time. Similarly in India, public school teachers are absent from school only one-quarter of the time but absent from class about half the time (Chaudhury, et al., 2006). Finally, as explained in Glewwe, Ilias and Kremer (2003), absences seem fairly broadly distributed among the population of teachers rather than concentrated among a subset of teachers with very high absence rates.

3. Description of the Teacher Incentives Experiment

The teacher incentives program evaluated by the authors was implemented in Busia and Teso districts in Western Kenya. International Child Support (ICS), a Dutch Non-Governmental

Organization, offered 50 schools in these districts the opportunity to participate in a program that provided gifts to teachers and headmasters whose students performed well. The program provided prizes to teachers in grades 4 to 8 based on the performance of their school as a whole on the district exams in each year. All teachers who taught these grades were eligible for the prize. Teachers of lower grades were not a part of the competition, because no district-wide exams are offered for those grades.[4]

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 average performance of all of the grade 4 to 8 pupils in each school, with each subject weighted equally, rather than on a teacher-by-teacher basis. Education experts generally are more sympathetic to this type of incentive (Richards and Sheu, 1992; Hanushek, 1996). In order to create incentives for headmasters, and to further encourage cooperation among teachers at each school, each winning school also received a briefcase for the headmaster, a wall clock, a time keeping clock, and a bell. A potential disadvantage of setting prizes at the school level is that this weakens the incentives of each individual teacher, since each teacher's actions have a relatively small effect on school average test scores. However, since the typical school in the sample had only 200 students and 12 teachers, about half of whom taught in grades 4-8, coordination within the school seems potentially feasible. Moreover, headmasters have various means to encourage teachers to cooperate with each other, since they control teaching assignments within the school and have other powers, such as the power to choose a deputy from among the teachers.

ICS awarded prizes in two categories: "top-scoring schools" and "most-improved schools." Schools could not win in more than one category, and prizes were the same across the two categories. Improvements (for judging which schools were "most improved") were

[4] They received a flashlight as a token prize, whether or not they belonged to a winning school.

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 refer to the last pre-program year for which we have data as year 0,[5] 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 (a suit, worth about $51), three second prizes were awarded (plates, glasses and cutlery worth about $43), three third prizes were awarded (a tea set, abour $34), and three fourth prizes were awarded (bed linens and a blanket, about $26). Thus, , 24 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.[6]

The value of these prizes ranged from 21 to 43% of a typical teacher's monthly salary. This is comparable to other merit pay programs. For example, the 1993-94 Dallas merit pay program, which was also based on school-wide performance, awarded $1000 annual bonuses that 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 annual bonuses, worth about 25% of monthly salary (Olsen, 1999; American Federation of Teachers, 2000), and programs in Colorado awarded from 10% to 50% of monthly salary in merit-based annual bonuses (Education Commission of the States, 2001). In Israel, the annual bonuses examined by Lavy (2002) ranged from $250 to $1000, which corresponds to 10% to 40% of the average teacher's monthly salary.

[5] This is either 1996 or 1997, depending on the type of data.
[6] 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.

The program in Kenya created incentives not only to raise test scores but also to reduce dropout rates. All students who were enrolled at the beginning of the program were included in the computation of scores. Students who did not take the test were assigned very low scores for the purpose of assigning prizes. More specifically, students who did not take the English essay exam test were assigned a score of zero. Students who did not take the multiple choice exams used in all subjects other than English composition were assigned a score of 15, whereas students who simply guessed randomly would have obtained, on average, a score of 25. 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.

Fifty schools were offered the opportunity to participate in this program in February 1998, and all accepted. When the NGO initially announced the program, schools were told it could commit to funding only for a single year, yet based on the favorable reports from the field the NGO later extended the program for a second year. The prizes were awarded during ceremonies held in November of years 1 and 2, and all the schools in the program were invited to attend. (The academic year in Kenya runs from January to November.)

Overall, the context seems particularly favorable for a teacher incentives program. First, the level of teacher absence suggests that teacher effort was an issue in these schools. Second, standardized curriculum and the prevalence of rote teaching meant that there was little scope for diverting teacher attention away from creativity and towards teaching to the test. Finally, the short duration of the program made possible a design that was less susceptible to manipulation of the student body or of the set of teachers in the school.

The 50 schools given the option to participate in the program were randomly selected from a group of 100 schools in Busia and Teso districts that had been designated in 1995 by the Ministry of Education as being particularly in need of assistance but had not participated in a previous World Bank textbooks program. These 100 schools scored somewhat worse than

average for the area before ICS began working with them. ICS had also provided textbooks or modest grants to these 100 schools before or during the teacher incentive program as determined by their random assignment into four groups of 25 schools. In each of the four groups, half of the schools were randomly chosen to participate in the teacher incentive program. By construction, the schools that did and did not participate in the incentive program were divided in the same proportions across Busia and Teso districts, by geographic divisions within these districts, and by their status regarding receipt of textbooks or grants in earlier years. We do not believe that the fact that these schools had received some other NGO assistance seriously compromises external validity, since the previous ICS assistance was very small relative to overall school budgets, and these previous programs had only a modest impact on the schools (see, *inter alia*, Glewwe, Kremer and Moulin, 2007). Moreover, many NGOs assist schools in Kenya, and while the schools in the sample received more support than average they were not in the upper tail of the "assistance" distribution. For example, some schools in the district received about 600,000 Ksh (about $10,000) in the mid 1990s to support classroom construction.

All 50 schools that serve as the comparison group for this evaluation participated in another program that provided pre-schools training, materials, and salary supplements (the last was conditional on pre-school teacher attendance). Unlike primary school teachers, pre-school teachers are semi-volunteers who are neither hired nor paid by the central Ministry of Education. Rather, they are hired locally by parents' committees. Their salaries come from parent contributions, which are often irregular. Pre-school teachers typically have no formal training. In fact, the pre-school program had little effect on performance in the pre-school classes **(citation)**, so it seems unlikely that that program affected outcomes in grades 4 to 8 during the time period we examine. Resources provided by the pre-school program went to the pre-school teacher or pre-school specific supplies that pre-schools would not have purchased in the absence of the program; thus, those resources could not have leaked in any substantial way to the upper

grades. Moreover, the locally-run pre-schools are administratively separate from the Ministry of Education operated schools, and maintain separate accounts.

Interviews with the headmaster and three teachers in each treatment school in the middle of the second year suggest that they were pleased with the program. All teachers interviewed supported the idea of motivating teachers by providing them with incentives; 83% of the teachers reported that prizes were justly awarded; 75% reported an increase in homework assignment due to the program (although, as noted below, this is not borne out by our data); 67% reported increased cooperation among teachers; and 88% reported an increase in prep sessions.

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 any grade) in the school as of March 1998 (the beginning of the first year of the program). Exit and entry of teachers were not significantly different between program and comparison schools (see Glewwe, Ilias and Kremer, 2003, for details). There is also no evidence of differences in the way teachers were reassigned to different grades within the schools. More specifically, in treatment schools 7.4% of teachers transferred from a non-incentive grade to an incentive grade during the program; in comparison schools 7.3% did.

4. Impact of the Teacher Incentives Program on Teacher Behavior

This section examines the impact of the program on teacher attendance, pedagogy and homework assignment, and on teachers' propensity to offer prep sessions.

Teacher attendance was not affected by the incentive program. In the year before the teacher incentives program was implemented, each of the 100 schools received two random, unannounced visits during which the attendance of each teacher in grades 4 to 8 was recorded. Similar visits were made five times in year 1 and three times in year 2. An attendance rate was

computed, for each teacher in each year, as the proportion of visits during which the teacher was present. Teachers were considered to be present if they were at the school, even if they were not in the classroom 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 by their initial schools.[7]

The first column of Panel A of Table 1 shows that, prior to the program, schools that were later selected to be program schools had slightly higher teacher attendance than the control schools, although the difference is statistically insignificant. In year 1 of the program, teacher attendance was slightly lower in the incentive schools, and in year 2 the attendance was slightly higher in incentive schools, as seen in columns 2 and 3 of Table 1, although both coefficients are small and completely insignificant. Difference-in-difference estimates, which compare *changes* over time in the teacher incentive schools with *changes* over time in the comparison schools, are presented in columns 4 and 5 in Panel A of Table 1; they also yield insignificant results.

To analyse the impact of the teacher incentives program on the behaviour of teachers in classroom, trained observers watched each teacher once per year, spending one period in a class recording several measures of teacher behavior. The first column of Panels B, C and D of Table 1 show that there were no significant differences in pedagogy between the teacher incentive schools and the comparison schools prior to the program in terms of teacher presence in the classroom, teacher energy levels and homework assignment. Moving to columns 2 and 3 of those panels in Table 1, there are no significant differences during the two years that the program operated between the two school groups in any of these pedagogical practices. The difference-in-difference estimates shown in the last two columns of Table 1, which were computed at the

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

school-grade level for panels B, C and D since it was not possible to match individual teachers across observations years, are also close to zero and far from statistically significant for these three pedagogical techniques.

In contrast to the lack of program impact on the teacher attendance, pedagogy and homework assignment, the program did lead teachers at the teacher incentive schools to offer more preps than were offered by their counterparts in they comparison schools. Headmasters in each school provided information on whether there were any preps for grades 4 through 8 in six time periods for each year: during each of the three terms (outside of normal school hours) and during the three school vacation periods (April, August and December).

Before the program, teacher incentive schools were slightly less likely to offer preps, as seen in column 1 of Table 2, but the second and third columns of that table show that treatment schools started to conduct more preps after the introduction of the program. 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, and the latter estimate is significant at the 5% level. Difference-in-difference estimates, shown in columns 4 and 5, are similar but less precisely estimated.

To summarize this section, the teacher incentives program did not reduce teacher absenteeism and led to little or no change in teachers pedagogical practices. Yet they did increase the number of "prep" courses, which are designed to increase students test scores on the district tests and especially the KCPE, that schools offered outside of normal school hours.

5. Impact of the Teacher Incentives Program on Student Outcomes

This section reports estimates of the impact of the teacher incentives program on student learning, as measured by test scores. The results distinguish between the tests on which incentives were based (district tests) and the tests that were not tied to incentives (ICS and KCPE

tests). See Glewwe, Kremer and Ilias (2003) for estimates of the impact of the program on dropout and repetition rates; that paper found no significant impact on either of these education outcomes.

Students in the 100 schools that participated in this evaluation took a variety of 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 (recall that the school year ends in November). Participation is incomplete since students must pay a fee of 120 Shillings (US $2) to take 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 a strong incentive to encourage their students to take the district exams.

Pupils in grade 8 also took the Kenyan Certificate of Primary Education (KCPE) exam, which determines what secondary school, if any, students can attend. The district exams, on which the incentives were based, are explicitly designed to prepare students for the KCPE exams, and are often referred to as "mocks" because they are designed to be as similar to the KCPE exam as possible. We have test score data on the district exams and the KCPE from both intervention years (1 and 2) as well as from the post-program year (3). There are no district exam scores for the 32 schools in the Teso district for year 1 (that district chose not to offer exams that year, due to the cost to parents). Consequently, analysis of the district exam scores for year 1 is restricted to schools in Busia district, which includes 68 out of the 100 schools.

The district exams are not an ideal way to evaluate the impact of the program, since a considerable proportion of students do not pay the exam fee, and hence do not take the exams. It is also useful to measure performance using a source of data that was not directly tied to

incentives. Hence students were also given exams prepared by ICS. These exams were not tied to the teacher incentive program, and thus provide an independent assessment of the impact of the program. The ICS exams were administered to students in grades 3 through 8. In years 0 and 1, ICS administered tests in English, Math, and Science. In year 2, ICS administered only English and Math tests. In years 0 and 1, separate exams were given for each grade, and the exams had a multiple-choice format. However, the ICS exams in years 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 tests could be answered by all students, while later questions became progressively harder. The last questions were based on material covered in grade 8. We have information on the ICS tests for all 100 schools for years 0, 1 and 2. Another change in the ICS exams for year 2 is that most of the questions were "fill in the blank", as opposed to multiple choice.

To prevent cheating, security is generally tight in Kenyan exams. All exams are supervised by teachers from other schools to prevent teachers from assisting students on the exam. In year 1 (1998), one case was discovered in which the headmaster of a program school colluded with the teachers assigned to supervise that school to allow cheating on the district exam. That school was disqualified from the competition in year 1 but was allowed to participate in year 2. The scores from that school were not included in the analysis in year 1, but its scores were included in year 2.

Table 3 examines the impact of the teacher incentives program on students' participation in the district, KCPE and ICS exams. 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. Following standard intention to treat (ITT) methodology, we restrict attention to only those students who were enrolled as of

February 1998 (the beginning of year 1) and assign the few students who switched schools during the program to their original schools.

Baseline participation (year zero) was around 80% on the district exams and around 85% on the ICS and KCPE exams. There is no statistically significant difference in the participation rates between the two school groups for any of the exams in year 0. In year 1, the point estimate in column 2 of Table 3 indicates that participation in the district exams was higher by 5.7 percentage points in the incentive schools, although this is not statistically significant. By year 2 (column 3), participation was higher by 11.2 percentage points in the incentive schools, a difference significant 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 1.9 percentage points lower in the incentive schools than in the comparison schools, though the difference was not significant. In contrast, panels B and C of Table 3 show that the participation rates in the KCPE and ICS exams, which were not linked to teacher incentives, were similar between the two school groups in all years. The results for years 2 and 3 in Table 3 include students who repeated in order to minimize the decline in the sample size (dropping repeaters reduces the sample to only 48% of the 14,982 students in the year 1 sample); the results are almost identical when repeaters are dropped.

Examination of the pre-test scores of district test takers in the treatment and comparison groups does not suggest substantial bias. Theoretically, efforts by teachers in the teacher incentive schools to increase exam participation could bias scores either upwards or downwards. 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 academically weaker 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 biased upward. To get a sense of the potential bias, we compared pre-test scores of students in treatment and comparison schools. Of the 11,122 students eligible to participate in the district exams in year 1, the mean of the normalized year 0 test score of the treatment group students was 0.022 standard deviations higher than that of the comparison group students, but this difference is not statistically significant (t-statistic of 0.81). For the 7,259 students eligible for the year 2 district exams, the mean of the year 0 score of the treatment group was 0.032 standard deviations higher than that of the comparison group, but again this was not significant (t-statistic of 0.93).

We examine differences in test scores between the teacher incentive schools and the comparison schools using a random effects regression framework that allows the scores of students in the same grade and same school to be correlated due to unobserved characteristics of teachers and schools. Thus we use the following error components econometric model with school random effects and grade random effects within each school:

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

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 a teacher incentives school. Thus if the impact of the incentive program varies across grades, β_k will measure the (weighted) average impact of the program for subject k 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 and Teso. Given the randomized assignment of schools to 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.

The units in which test scores are measured are arbitrary, so for each year and each subject and grade combination we normalize all test scores (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. For reference, note that for a normal distribution an increase of 0.1 standard deviations would move a student from the 50th percentile to the 54th. Since the district test exams differed in Busia and Teso districts, these normalizations were done separately for each district.

Table 4 reports estimates of the impact of the teacher incentives program on the three tests (district, KCPE and ICS).[8] Analysis of pre-program data indicate no differences between the teacher incentive schools and the control schools. More specifically, in year 0 (before the program was implemented) there were no significant differences in scores on any of these tests between the teacher incentive schools and the comparison schools for any subject or grade (column 1).

[8] For brevity, Table 4 reports only results averaged across all subjects. In addition, Table 4 reports only regressions that exclude students who repeated. See Glewwe, Ilias and Kremer (2003) for separate results by subject and for results that include repeaters (which are very similar to the results reported here).

Now consider the impact of the teacher incentives program during the two years it was operating. The difference estimates in columns 2 and 3 in Table 4 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 estimates in columns 5 and 6 use a restricted sample, namely those students who took exams in at least one subject in the pre-program year and in at least one of the intervention years.[9] As discussed above, we restrict attention to those students who were enrolled prior to the announcement of the program in February 1998 (year 1).

Averaging across all subjects and grades, the difference estimate for the district exam is insignificantly negative (-0.04) in the first year of the program, but this could be due to the differential exam participation between treatment and comparison schools on the district exams. The difference-in-difference estimate, which should be less subject to bias from attrition, is positive (0.05), but not significant. Both the difference and difference-in-difference estimate of the treatment effect in year 2 are significantly positive (about 0.14 for both). For the KCPE, the difference estimate in year 1 is 0.14 and the difference-in-difference estimate is 0.10, with the former being significant at the 10% level. In year 2, the effects are stronger, with the difference estimate of 0.17 that is also significant at the 10% level.

As discussed above, it is instructive to examine the ICS tests because the incentives were designed around the district tests, because there is differential exam participation on the district tests but not on the ICS tests, and because the later ICS test was designed to be used for students who were not all in the same grade. For the ICS exams, differences are not significant in either year, and the point estimates across years are similar, at 0.09 and 0.08, respectively. The ICS exam difference in differences results are even smaller (about 0.02) and completely insignificant.

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

Summarizing the overall pattern of difference-in-difference results in Table 4, test scores improved the most on the similar district and KCPE exams, while they did not increase as much, if at all, on the ICS test. This suggests that some of the test preparation activities may have been aimed specifically at the district exams, on which the teacher incentives were based, and that they were also effective in preparing students for the similar KCPE exams. In contrast, there were no significant gains on the non-incentive ICS exam, with its different format that included fill-in-the-blank questions.

By year 3, students in the teacher incentive schools had been exposed to the program for all of year 1 and year 2. If the increases in test scores in years 1 and 2 of the 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. This offers even stronger support for the hypothesis that teachers were exerting effort to increase test scores on the incentive exam only during the program years and not after the incentives were removed.[10]

There is also evidence that the teacher incentives program improved students' test taking techniques. As noted above, one thing teachers emphasized in prep sessions is that students should not leave blanks on multiple choice questions and should make sure that they reach the end of exams. 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. Although the English test has mostly a fill-in-the-blank format, there are some multiple choice questions (all math test questions were fill-in-the- blank questions). The test began with 20 relatively easy multiple choice questions, followed by 74 questions that become progressively more difficult. Of the 74

[10] Data for the ICS exams in year 3 was only available for 27 of the schools—those that participated in the evaluation of a de-worming project that year. These schools were evenly distributed across the teacher incentive schools and the comparison schools. Point estimates are positive, but none of the t-statistics exceeds one.

more difficult questions, 70 are fill-in-the-blank and 4 are multiple choice. The item by item data on the 1999 ICS test do not indicate whether a student answered a question, but of course a student who does not answer a question will receive a score of zero. We constructed two variables, one indicating the percentage of the four relatively hard multiple choice questions that were answered correctly, and the other indicating the percentage of the 70 relatively hard fill-in-the-blank questions answered correctly. Regressing the ratio of the former over the latter on the program dummy and the same set of control variables, we found a positive program impact of 0.042 points that was statistically significant at the 5% level (t-stat. of 2.16). Students in incentive schools were also less likely to get none of the last 10 answers correct, and none of the last 20 questions correct (these results are significant at the 10% level), which suggests that they learned to answer all questions, guessing on those for which they did not know the answer. Finally, analysis of school level data on leaving questions blank (based on revisiting schools and examining randomly drawn exams) suggests a reduction in leaving answers blank, although this is not quite significant at the 10% level (t-statistic of 1.50). These results suggest that the extra prep classes in the program schools increased students test taking skills, such as how best to answer multiple choice questions, and that these skills had a small impact on the ICS exams (which was not the test that prep classes were designed to teach to), and only on those questions whose format was commonly used in the district tests and the KCPE.

In summary, the teacher incentives program increased the proportion of students who participated in the district exams (on which the incentives were based) but not on the KCPE and ICS exams (which were not connected to the incentives). There were no differences between the teacher incentive schools and the comparison schools on any of the three tests in the year before the program began. During the two years the program was implemented, scores increased in the teacher incentives schools on the district exams but not on the KCPE and ICS exams. After the

program ended, there were no difference on any exams across the two types of schools. Overall, there is little evidence of an impact on student learning from the teacher incentives program.

6. Conclusion

Students in schools with the teacher incentives program were more likely to take exams and had higher test scores in the short run on exams linked to incentives. 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 neither higher attendance rates nor higher homework assignment rates. Pedagogy and student dropout rates were similar in program and comparison schools. Instead, teachers in program schools increased test preparation sessions and encouraged students enrolled in school to take the test. Test scores did not differ significantly on exams that were not linked to incentives (the KCPE and ICS exams). Following the end of the program, the test score difference, even on tests linked to incentives, between students in treatment and comparison schools disappeared.

There is evidence that teachers learned over time how better to take advantage of the program. Estimated differences in preparation sessions between treatment and comparison schools grew between the first and second year. 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, differences in exam participation rates between program and comparison schools rose, presumably because teachers worked harder to persuade students to take the exam. Moreover, the test score gap between treatment and comparison schools was greater in the second year than in the first year.

While these results cast doubt on the effectiveness of teacher incentive programs, one should be careful about drawing general conclusions. 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, at up to 40% of monthly income, the incentives were comparable in magnitude to those in most U.S. programs and in the Israeli program analyzed by Lavy (2002). On the other hand, some have argued that the incentives offered in U.S. programs are too small to have a sizeable impact on teachers' behavior.

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 become teachers in any case. However, given the queuing for teaching positions in Kenya, it is unlikely that people who either have teaching jobs or who have the academic qualifications to enter teacher training colleges (but not universities) are opting out of the profession in the current system. 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.[11] 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 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 was 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.

[11] In practice, many teachers transfer between schools.

References

American Federation of Teachers, Teacher Salary Survey Archives at http://www.aft.org/research/survey00/salarysurvey00.pdf

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

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, Kremer, Michael, and Sylvie Moulin (2007), "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* (forthcoming).

Glewwe, Paul, Ilias, Nauman, and Michael Kremer (2003), "Teacher Incentives," National Bureau of Economics Working Paper 9671.

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 (1998), "Teachers, Schools, and Academic Achievement," National Bureau of Economic Research Working Paper 6691 .

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.

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, pp. 24-52.

Lavy, Victor (2002), "Evaluating the Effect of Teacher Group Performance Incentives on Students Achievements," *Journal of Political Economy,* vol. 110, 6, pp. 1286-1318.

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.

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.

World Bank, (2005), "Kenya - Public Expenditure Review 2004," Washington, D.C.

Table 1: Program Effects on Teacher Behavior

	(1)	(2)	(3)	(4)	(5)
	Year 0	**Year 1**	**Year 2**	**Year 1 - Year 0**	**Year 2 – Year 0**
Panel A - *Dependent Variable:* Teacher Attendance (percentage)					
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
Panel B - *Dependent Variable:* Teacher Present in the Classroom					
Incentive	-0.066	-0.034	0.132	0.060	0.078
School	(0.157)	(0.139)	(0.167)	(0.153)	(0.130)
Grade	-0.061	0.038	-0.096	-0.092	-0.065
	(0.042)	(0.036)	(0.044)**	(0.052)*	(0.047)
Constant	0.797	0.313	0.211	0.589*	0.850
	(0.315)**	(0.276)	(0.338)	(0.356)	(0.330)**
Observations	631	826	481	380	400
Panel C – *Dependent Variable:* Teacher Energy (1 to 5: 1=energetic)					
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
Panel D – *Dependent Varuable :* Homework Assignment					
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, based on regressions that 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. Observations vary by year because the number of visits made to schools varied across years.
For homework regressions, 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 2: Program Effect on Test Preparation Sessions

Dependent Variable:	(1)	(2) Preparations	(3)	(4) Preparation Differences *(Preparation $_{program\ year}$ − Preparation $_{pre\text{-}program\ year}$)*	(5)
	Year 0	**Year 1**	**Year 2**	**Year 1 - Year 0**	**Year 2 – Year 0**
Preparations (Vacation and During School)					
Incentive School	-0.007	0.042	0.074	0.049	0.081
	(0.044)	(0.037)	(0.034)**	(0.042)	(0.047)*
Grade	0.155	0.135	0.103	-0.021	-0.052
	(0.009)***	(0.007)***	(0.007)***	(0.009)**	(0.009)***
August Holiday (0/1)	-0.020	0.058	-0.122	0.078	-0.102
	(0.016)	(0.030)*	(0.034)***	(0.037)**	(0.036)***
December Holiday (0/1)	-0.370	-0.452	-0.534	-0.082	-0.164
	(0.025)***	(0.029)***	(0.030)***	(0.031)**	(0.036)***
Term Visit 1 (0/1)	0.094	0.126	0.130	0.032	0.036
	(0.035)***	(0.040)***	(0.038)***	(0.035)	(0.052)
Term Visit 2 (0/1)	0.094	0.168	0.282	0.074	0.188
	(0.035)***	(0.041)***	(0.040)***	(0.052)	(0.052)***
Term Visit 3 (0/1)	0.096	0.158	0.242	0.062	0.146
	(0.035)***	(0.040)***	(0.043)***	(0.051)	(0.055)***
Constant	-0.502	-0.372	-0.121	0.130	0.381
	(0.064)***	(0.053)***	(0.052)**	(0.064)**	(0.064)***
Observations	3000	3000	3000	3000	3000

Notes:
Standard errors in parentheses, based on regressions that include school-level random effects.
* significant at 10%; ** significant at 5%;
Preparations are 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.
The mean number of times that preps were held in year zero, by grade, were: 1.0 for grade 4, 1.1 for grade 5, 2.4 for grade 6, 3.5 for grade 7, and 4.5 for grade 8.

Table 3: Program Effects on Participation in Exams

	(1) Year 0 (Pre-Program)	(2) Year 1	(3) Year 2	(4) Year 3 (Post-Program)
	Panel A			
	***Dependent Variable:* Take District Exam (0/1)**			
Incentive School	-0.002	0.057	0.118	-0.007
	(0.025)	(0.037)	(0.027)**	(0.030)
Male (0/1)	-0.0007	0.015	-0.007	-0.004
	(0.006)	(0.008)	(0.008)	(0.011)
Grade	0.018	0.059	0.041	0.007
	(0.002)**	(0.003)**	(0.003)**	(0.005)
Constant	0.709	0.356	0.477	0.563
	(0.022)**	(0.031)**	(0.028)**	(0.040)**
Observations	15,224	11,122	10,654	8,055
	Panel B			
	***Dependent Variable:* Take KCPE Exam (0/1)**			
Incentive School	0.038	-0.002	0.022	0.034
	(0.023)	(0.022)	(0.020)	(0.027)
Male (0/1)	-0.016	-0.017	0.007	0.006
	(0.016)	(0.015)	(0.015)	(0.019)
Constant	0.861	0.896	0.899	0.823
	(0.018)**	(0.016)**	(0.015)**	(0.020)**
Observations	1,736	1,656	1,276	1,439
	Panel C			
	***Dependent Variable:* Take ICS Exam (0/1)**			
Incentive School	0.005	0.023	0.045*	0.007
	(0.013)	(0.023)	(0.024)	(0.042)
Male (0/1)	-0.004	0.003	0.002	0.021
	(0.006)	(0.005)	(0.006)	(0.019)**
Grade	0.012	0.018	0.010	0.028
	(0.002)**	(0.002)**	(0.003)**	(0.009)**
Constant	0.767	0.757	0.793	0.540
	(0.016)**	(0.020)**	(0.023)**	(0.068)**
Observations	15,718	14,982	10,654	2,226

Note: Standard errors in parentheses, based on regressions that include school-level random effects.
* significant at 10%; ** significant at 5%;
District test data were not available for Teso District in Year 1 (1998), and the ICS data for Year 3 are limited to 27 schools.
Each observation represents an upper primary school pupil in year 0; columns 2 and 3 are limited to pupils who did not drop out or transfer to another school in those years. Panel B includes pupils who were in 8th grade.

Table 4: Program Effect on Average Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Test Scores				Test Score Differences		
	Year 0	Year 1	Year 2	Year 3	Year 1 – Year 0	Year 2 – Year 0	Year 3 – Year 0
Panel A – District Tests							
Average over all Subjects & Grades	0.011	-0.040	0.136	-0.087	0.054	0.139	-0.008
	(0.091)	(0.079)	(0.077)*	(0.083)	(0.054)	(0.065)**	(0.084)
Observations	24,716	50,842	37,620	15,893	24,677	15,641	5,330
Panel B – KCPE Tests							
Average over all Subjects & Grades	0.011	0.138	0.165	-0.009	0.104	0.152	-0.006
	(0.091)	(0.074)*	(0.090)*	(0.101)	(0.080)	(0.097)	(0.138)
Observations	24,716	10,430	8,427	4,053	7,152	5,247	1,505
Panel C – ICS Tests							
Average over all Subjects & Grades	0.071	0.089	0.083	- -	0.017	0.016	- -
	(0.086)	(0.079)	(0.090)		(0.033)	(0.063)	
Observations	33,162	39,510	12,996		32,993	10,512	

Note: Standard errors in parentheses, regressions include school-level random effects.
* significant at 10%; ** significant at 5%;
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.
Normalized district test scores from year 0 (1996) were used as the KCPE pre-program scores.
Columns 2, 3, and 4 are limited to pupils who were enrolled in year 1 and did not repeat or drop out. Columns 5, 6 and 7 impose the additional restriction that a pre-test score is available.
Year 3 ICS tests were given only in 27 schools, so scores are not reported.

Faculty and Research Affiliates

Matthew G. Springer
Director
National Center on Performance Incentives

Assistant Professor of Public Policy and Education
Vanderbilt University's Peabody College

Dale Ballou
Associate Professor of Public Policy and Education
Vanderbilt University's Peabody College

Leonard Bradley
Lecturer in Education
Vanderbilt University's Peabody College

Timothy C. Caboni
Associate Dean for Professional Education and External Relations
Associate Professor of the Practice in Public Policy and Higher Education
Vanderbilt University's Peabody College

Mark Ehlert
Research Assistant Professor
University of Missouri – Columbia

Bonnie Ghosh-Dastidar
Statistician
The RAND Corporation

Timothy J. Gronberg
Professor of Economics
Texas A&M University

James W. Guthrie
Senior Fellow
George W. Bush Institute

Professor
Southern Methodist University

Laura Hamilton
Senior Behavioral Scientist
RAND Corporation

Janet S. Hansen
Vice President and Director of Education Studies
Committee for Economic Development

Chris Hulleman
Assistant Professor
James Madison University

Brian A. Jacob
Walter H. Annenberg Professor of Education Policy
Gerald R. Ford School of Public Policy
University of Michigan

Dennis W. Jansen
Professor of Economics
Texas A&M University

Cory Koedel
Assistant Professor of Economics
University of Missouri-Columbia

Vi-Nhuan Le
Behavioral Scientist
RAND Corporation

Jessica L. Lewis
Research Associate
National Center on Performance Incentives

J.R. Lockwood
Senior Statistician
RAND Corporation

Daniel F. McCaffrey
Senior Statistician
PNC Chair in Policy Analysis
RAND Corporation

Patrick J. McEwan
Associate Professor of Economics
Whitehead Associate Professor of Critical Thought
Wellesley College

Shawn Ni
Professor of Economics and Adjunct Professor of Statistics
University of Missouri-Columbia

Michael J. Podgursky
Professor of Economics
University of Missouri-Columbia

Brian M. Stecher
Senior Social Scientist
RAND Corporation

Lori L. Taylor
Associate Professor
Texas A&M University

NATIONAL CENTER ON

Performance Incentives

EXAMINING PERFORMANCE INCENTIVES IN EDUCATION

National Center on Performance Incentives
Vanderbilt University Peabody College

Peabody #43
230 Appleton Place
Nashville, TN 37203

(615) 322-5538
www.performanceincentives.org

VANDERBILT
PEABODY COLLEGE