The Impact of Financial Incentives on Academic Achievement and Household Behavior: Evidence from a Randomized Trial in Nepal

Dhiraj Sharma*

Department of Agricultural, Environmental, and Development Economics

The Ohio State University

November 21, 2010

*I acknowledge support from the Ohio Agricultural and Research Development Center at the Ohio State University. This project would not have been possible without the cooperation of principals, teachers, parents, and students in the sample schools. I am especially indebted to my advisor Dr. Brian Roe for his guidance, support, and encouragement throughout the process.

Abstract

We present evidence on the impact of piece rate financial incentives on students' school outcomes from a randomized field experiment in Nepal. Despite several experimental and institutional factors making it less likely of finding a positive treatment effect, I find that incentives increase average aggregate scores by 0.09 standard deviations. The bulk of the aggregate gain is constituted of large gains in two subjects with relatively high baseline scores. There is no noticeable difference in gains between males and females and incentives have a relatively higher impact on students from higher socioeconomic strata and the lower quartiles. My study contributes to the growing literature on short term impact of academic incentives by recording household response to the incentives. I show that the proportion of students who received help with schoolwork at home, either from a hired tutor or a household member, increased among the subgroups that exhibit the most gains in scores. Finally, financial rewards do not have an adverse impact on students' intrinsic motivation to learn.

JEL Codes: O15, I21, C93

Keywords: Academic Achievement, Financial Incentives, Intrinsic Motivation, Field Experiment

1. Introduction

Education policies in developing countries have focused largely on increasing enrollment and attendance rates with large scale programs like Education for All (EFA), the UN's Millennium Development Goals (MDGs), and conditional cash transfer programs. As a result, enrollment rates have increased substantially in poor countries in recent years (UNESCO Institute of Statistics). However, as enrollment rates rise, developing countries face the larger challenge of improving students' learning in schools. Literacy and numeracy is low even among children attending school in the least developed countries, both in the absolute and relative sense (Nepal Department of Education, 2008; PIRLS International Report, 2006; TIMSS Advanced International Report, 2008).

Recently, a number of studies have evaluated the effectiveness of short term monetary incentives in stimulating students' learning. Economic theory suggests that incentives can improve academic outcomes by acting as a price subsidy to school effort. Furthermore, short term rewards can mitigate the likely underinvestment in schooling and cognitive abilities if students and parents have low perceived returns to education or high discount rates for future returns. Such incentives are usually structured as tournaments where students have to score in the top percentiles or cross a pre-specified threshold to be eligible for the rewards. These contract structures may not affect the effort and performance of the most able and least able students for opposite reasons: better prepared students may find the criteria too easy to meet and ill prepared students may have no incentive to put in more effort if they find the criteria too stringent. A simple alternative is the piece rate payment scheme where students are paid linearly in proportion to their grades or test scores.

This paper presents evidence on the impact of piece rate financial incentives on students' testing outcomes from a randomized field experiment in Nepal.[1] From a pool of 33 public schools, 11 schools were randomly assigned to the treatment group while the remaining schools constituted the control group. Grade 8 students in treatment schools were offered cash incentives based on their aggregate scores in two semester exams and the end-of-the-year district level exam during the academic year 2009/10. Our estimate of the impact of incentives on average aggregate scores is 0.09 standard deviations. The result is significant given that various experimental and institutional factors in our study make it less likely to find a significant treatment effect. The bulk of the gain in aggregate score is constituted of large gains in two subjects with relatively high baseline scores. This observation is consistent with the evidence from psychology literature that finds rewards have the most impact on tasks with high intrinsic interest. There is no noticeable difference in gains between males and females and incentives had a relatively higher impact among students from the highest socioeconomic strata and the lower quartiles. The results from our experiment shed new light into the behavioral responses of students and households. The proportion of students who received helped with schoolwork at home, either from a hired tutor or a household member, increased among the subgroups that exhibited the largest gains in scores. Finally, financial rewards did not have a negative impact on students' intrinsic motivation to learn.

Our paper adds to the growing literature on the impact of short term incentives on students' outcomes. Studies from Canada and the US show mixed results on the effects of merit based financial aid on academic outcomes of college students (Angrist et al., 2006; Cornwell et al., 2003; Cornwell et al., 2006). The Quantum Opportunity Program (QOP) and Ohio's Learning, Earning, and Parenting (LEAP) program are examples of large scale initiatives for

[1] However, as we later discuss, our reward structure also has kinks.

high risk youth in the US that have improved graduation rates, academic achievement, and other behavioral outcomes (Hahn et al., 1994; Long et al., 1996). Conditional cash transfer programs in developing countries have increased school enrollment rates, but have had little or no impact on academic outcomes (Behrman, Sengupta, and Todd, 2000; Behrman, Parker, and Todd, 2005, Duryea and Morrison, 2004; Ponce and Bedi, 2008; Rawlings and Rubio, 2005; Reimers et al., 2006; Schultz, 2004). Student learning has improved due to institutional reforms like school voucher programs in Colombia (Angrist et al., 2002; Angrist et al., 2006), girls' scholarship programs in Kenya (Kremer et al., 2009) and incentives for high school matriculation in Israel (Angrist and Lavy, 2009).

There are two studies that are most similar to ours. The first is an experiment conducted at the University of Amsterdam where first year economics and business students were awarded cash incentives for fulfilling all the first year requirements before starting the second year (Leuven, Oosterbeek, and Van der Klaauw, 2003). The authors find no significant differences in the passing rates and credit points collected across high reward, low reward, and control group. The second is a series of studies conducted in urban schools in the United States (Fryer, 2010). Fryer concludes that when incentives are conditioned on inputs like reading a book or completing homework on time, there are significant gains in students' achievements but when incentives are conditioned on outputs like grades and standardized test scores, incentives do not have an impact.

Our paper makes a number of original contributions to the literature. First, as far as we are aware, our study is the first to test for linear pay-for-grades incentives scheme in a developing country. Second, we document grade inflation in the end-of-the-year exams. Third, we also analyze student and household level response to cash incentives.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature to place the study in context. Section 3 outlines the experiment design and program structure in detail. Section 4 sketches a theoretical model that provides a framework to estimate the impact of financial incentives. Section 5 describes the data. Section 6 and 7 present the estimates of the impact of cash incentives on scores and student and household behavior respectively. Section 8 compares the cost effectiveness of piece rate cash incentives with other programs. Section 9 outlines the key findings and outstanding issues. Finally, section 10 summarizes and concludes.

2. Related Literature

This study relates to at least three strands of literature in economics. The first is the literature on education production in general, and the determinants of quantity and quality of schooling in developing countries in particular. Since the earliest studies on the determinants of education production, findings on the relative importance of educational inputs in determining educational outcomes have been controversial (Coleman, 1966). In the context of developing countries, studies have found no consistently positive relationship between academic outcomes and inputs conventionally thought to be important like class size, teacher training, school infrastructure, and expenditure per student (Fuller 1987; Hanushek, 1995; Hanushek, 1997; Hanushek, 2003; Harbison and Hanushek, 1992; Heyneman et al., 1978; Schiefelbein and Simmons, 1981; Simmons and Alexander, 1980; Velez, Schiefelbein, and Valenzuela, 1993). Some attribute the inconsistencies to methodological problems inherent in “retrospective” studies and advocate the use of well-executed randomized experiments to overcome the problems (Glewwe, 2002; Glewwe and Kremer, 2006; Krueger, 2003; Pritchett and Filmer, 1999). Experimental studies have again shown mixed results on the impact of inputs on students’ achievement (Angrist and Lavy, 1999; Duflo et al., 2007; Glewwe et al., 2004; Glewwe et al., 2007; Krueger, 1999).

Some argue that the focus on exclusively input-based policies is flawed because it ignores the role of incentives in education production (Hanushek, 1995; Kremer, 2003; Kremer et al., 2003; Pritchett and Filmer, 1999). A number of studies in recent years have investigated the effectiveness of various institutional reforms, including contract teachers (Muralidharan and Sundararaman, 2008), pay for performance schemes for teachers (Glewwe et al., 2003; Lavy, 2002; Muralidharan and Sundararaman, 2006), parental oversight in schools (Duflo et al., 2007), and school vouchers (Angrist et al., 2002). These studies suggest that the reform of incentives may yield substantial gains in academic outcomes. In the spirit of institutional reforms, we hypothesize that an important determinant of students' achievement is their own effort and provide cash incentives conditional on grades to stimulate effort.

The second strand of literature relevant to our study is the literature on agency or contract theory. Over the past few decades, a large body of literature has analyzed various compensation mechanisms that can be used to optimally modify agents' behavior in favor of principal's interest (Prendergast, 1999). Most merit scholarships are structured as rank-order tournaments where students have to score in the top percentiles or cross a pre-specified threshold to be eligible for rewards. In contrast, our study provides incentives on a piece-rate basis where all students receive incentives in proportion to grades. Theoretically, tournaments are superior to schemes based on individual outputs when agents face common risk because the vector of output is informative of the common uncertainty parameter (Lazear and Rosen, 1981; Green and Stokey 1982; Holmstrom 1982; Nalebuff and Stiglitz 1983). However, when the agents are of heterogeneous abilities, tournaments may not affect the performance of the most able agents if they find the criteria for rewards too easy to meet and of the least able agents if they find the criteria out of reach. Furthermore, there is evidence that agents' efforts are clustered closer around the mean under piece rate schemes than under tournaments (Bull et al., 1987), possibly because piece-rate schemes simply require solving a maximization problem whereas a tournament is a game that involves making choices strategically. Therefore, in contrast to the theoretical superiority of tournaments for merit scholarships, piece-rate incentives could elicit more effort on any given occasion.

Much of the literature on contract theory is based on the assumption that people are motivated by pecuniary gains alone. A sub-branch of the literature examines non-pecuniary motivations behind human actions and finds that along with reciprocity and desire for social approval, intrinsic joy of performing a task is a powerful drive for people's actions (Fehr and Falk, 2002; Frey and Jegen, 2001). A concern raised by this literature is that external rewards may undermine intrinsic motivation, and even lead to lower supply of effort when the negative effect dominates (Deci, 1971, Deci et al., 1999; Frey and Felix, 1997). To check for the impact of incentives on intrinsic motivation, our year-end survey consists of a section that elicits students' interest in learning.

The third strand of literature relevant to the current study is the vast literature on human capital and its impact on economic and noneconomic outcomes. Early work on human capital in developing countries found high rates of returns to an additional year of schooling (Psacharopoulos, 1972, 1981, 1985, and 1994), while later studies also found positive returns to the quality of schooling (Behrman and Birdsall, 1983) and cognitive skills (Hanushek and Woessmann and references therein, 2008). Besides private earnings, schooling and cognitive skills also affect nonmarket outcomes like personal health, children's schooling, children's health and nutrition, and women's fertility (Glewwe and references therein, 2002; Wolfe and Zuvekas and references therein, 1995). A number of influential studies also find that economies with a labor force that has more years of schooling and higher cognitive skills grow at a faster rate (Barro 1991; Hanushek and Kimko, 2000; Romer 1990; Romer, Mankiw, and Weil, 1992). Therefore, these studies provide the justification for an investment in improving educational outcomes in developing countries,

3. Experiment Design and Program Structure

The experiment was conducted among grade 8 students in 33 public schools in Lalitpur district in Nepal. Lalitpur is one of the three districts that overlap the Kathmandu valley. Schools were included in the sample if they were farther than a pre-specified distance from the city center and within the valley of Kathmandu. Towards the end of the academic year 2008/09, we visited each of the 33 sample schools and explained the incentive program to the principals. They were also briefed on the randomization scheme and the possibility that their school may not be assigned to the treatment group. All principals provided their verbal commitment to support the initiative regardless of the actual assignment.

Randomization was conducted on a stratified sample where schools were stratified on the basis of the highest grade of instruction: lower secondary (grades 1 – 8), secondary (grades 1-10), or higher secondary (grades 1-12). Out of 6 lower secondary, 17 secondary, and 11 higher secondary schools in the sample, 2 lower secondary, 6 secondary, and 3 higher secondary schools were randomly assigned to the treatment group. [2] One lower secondary school originally

[2] Given a sample size, power of the experiment is maximized by assigning 50 percent of the sample to the treatment group. However, due to logistical and financial constraints, the ratio of treatment to control schools in our study is 1:2.

assigned to the control group had to be dropped from the sample because no student enrolled in grade eight for the academic year.[3]

There are nine officially mandated courses for students in grade eight. The core courses include Nepali, English, Math, Science, and Social Studies and the non-core courses are Health and Physical Education, Population and Environment, Moral Education, and Vocational Studies. Exams are worth 100 points each for the core courses and 50 points for the remaining courses while the passing scores for core and non-core courses are 32 and 16 respectively.

The incentive program provided cash rewards based on students' average aggregate scores in each of the two semester exams and the end-of-the-year district level exam. For example, a student who scored 280 aggregate points would have been said to secure an average of 40 percent (280 = 40% of 700). Under our scheme, if a student passed all courses, she received 5 rupees (~6 cents) per percent; if she failed any course, the price per percent was 2.5 rupees (~3 cents). For instance, if the student passed all courses while securing 40 percent, she received 200 rupees (Rs. 5.00*40 = Rs. 200), but if she failed one or more courses, she only received 100 rupees (Rs. 2.50*40 = Rs. 100). Those who failed this semester could earn the amount withheld from them by passing all subjects the next semester. To continue with the previous example, if the student who failed in the first semester scored 280 points again the next semester, she received 300 rupees: 200 rupees for the second semester *and* the 100 rupees withheld from her the previous semester. Similarly, a student who failed in both the first and second semesters but passed all subjects in the district level exam received her rewards for the final semester along with the rewards withheld in the first two semesters. This stacked incentive structure is similar to that in The Paper Project carried out in Chicago (Fryer, 2010).

[3] Zero enrollment for the school was not due to its control status. Our program was announced one month into the academic year 2009/10, and no students were enrolled in the school even prior to the date.

Although there are nine officially mandated subjects for grade 8, schools have some leeway in substituting the non-core subjects (Health and Physical Education, Population and Environment, Moral Education, and Vocational Studies). Schools would rather teach subjects like Computer Science that are more practical and “in demand” than subjects like Moral Education that are perceived to be out of touch with the changing needs and desires of students. Of the four non-core subjects, Health and Physical Education, and Population and Environment are taught in all treatment schools and all but one control school. Fifteen schools teach Moral Education (5 treatment, 10 control) and 25 teach Vocational Studies (9 treatment, 16 control). To keep the comparison valid, scores of only the nine officially mandated subjects are compared and average aggregate score is computed by averaging the scores over an appropriate total.

The average amount of incentives received by students over the course of the program is Rs. 590 (~$7.50). This sum is not significant relative to household income or expenditure in the semi urban areas of Kathmandu.[4] However, the amount is a significant portion of annual educational expenditure as school fees are about 600 rupees (~$6.50 – 8.00) per year. Furthermore, cash was distributed to students so the amount is not insignificant from students' perspective.[5]

To ensure that the sample was not contaminated by endogenous transfer of students into treatment schools, the treatment assignment was announced in May 2009, a full month into the academic year 2009/10. One month is approximately the time it takes for school enrollments to stabilize, and schools are loath to admit students after this time.[6] We explained the incentive

[4] Nepal Living Standard Survey 2003/04 (NLSS II) found real mean per capita expenditure in Kathmandu valley to be Rs. 26,832 (~ $ 360) (base year 1995/96), although the comparable figure for households in our sample is likely to be much smaller.

[5] Anecdotal evidence suggests that students received Rs. 5 – 10 per day as stipend for afternoon snacks. Thus the incentives represent 2 to 4 months' worth of stipend for an average student.

[6] Our data shows no transfers of students between schools.

scheme to students in classroom, and worked through a number of examples on the blackboard to cover all the possible payment scenarios. In addition, each student received an information sheet that described the incentive program in detail. The students were also informed that if they had any questions about the program, they could contact their teachers for clarification. Throughout this process, phrases like "scholarship," "prize," and "award" were used to describe the incentives instead of words like "reward" and "payment."

After each semester, when the exam results were announced, we visited the treatment schools and distributed the rewards in class. Cash was handed out to students in a closed envelope and the amount was not disclosed publicly.[7] To leave no room for confusion about the incentive scheme, the envelope also contained a sheet with information on scores, pass/fail status, amount eligible, and amount withheld.

Two surveys were conducted among all students, one at the beginning of the academic year (May – July, 2009) and one towards the end (February – March, 2010).[8] The surveys contained questions on three general themes: socioeconomic status, availability of educational material, and daily use of time. Socioeconomic variables include, among other things, parents' education and occupation, number of siblings, and siblings' education. Key questions on educational inputs include whether students received any help at home for schoolwork, either from a family member or a hired tutor. Finally, the survey asked students to report how they spent their time during off-school hours. To gauge if the incentives scheme had any adverse impact on students' intrinsic motivation to learn, the end-of-the-year survey asked students to report how interested they were in learning each subject. In addition, students in treatment

[7] However, students were eager to compare their rewards with each other therefore, to a large extent, students knew their performance relative to their peers.
[8] Refer to table 1 for the timeline of the experiment.

schools were also asked to answer questions intended to measure their understanding of the incentive scheme.[9]

4. Theoretical Model

Key components of the theoretical model consist of students' effort, costs and rewards of effort, and external monetary incentives. A more complex model than presented here is due to Benabou and Tirole (2003) where agents infer the nature of the task or their suitability to it from the principal's offer of external rewards. Therefore, a principal's offer affects agents' behavior not only through its effect on the agents' payoff function, but also via its impact on the agents' inference process. This information-based strategic analysis allows for a precise description of the conditions under which external rewards crowd out intrinsic motivation. I adopt the reduced form approach as sketched by Frey and Oberholzer-Gee (1997) for its simplicity.

Let the learning that takes place during a period (e.g., a school year) be described by the function $y = f(S, F, \mu, e)$ where

S= vector of school inputs

F= vector of family inputs

μ = innate ability of student, and

e = student's effort (e.g. time spent doing homework)

We assume that f is strictly concave in e i.e. $f_e > 0 \; and \; f_{ee} < 0$. The private monetary reward of learning is given by wy where w can be interpreted as the present discounted value of lifetime after-tax earnings. Students also derive satisfaction from learning itself, regardless of the monetary rewards. Nonmonetary rewards to learning are indexed by $j(e)$ where $j_e(e) > 0$,

[9] Data from the exit survey shows that 95 percent of students correctly understood the incentive scheme. The common confusion among the remaining 5 percent was regarding the stacked incentive structure.

$j_{ee}(e) < 0$. There are costs to exerting effort, which could be monetary, like the opportunity cost of forgone market earnings, or nonmonetary, like loss of free time and boredom. Cost of effort is described by the function $c(e)$, $c'(e) > 0$ and $c"(e) > 0$. Students choose effort to maximize net benefit $wy + j(e) - c(e)$. The first order condition for optimization is:

$$wf_e + j_e - c_e = 0$$

Here we have assumed that there are no explicit rewards to learning other than the higher lifetime earnings. Now suppose students are offered monetary incentives as a function of their level of learning y in the current period and there are no interactions between joy of learning and external rewards. Students' new net payoff function is $(w + p)y + j(e) - c(e)$. The new first order condition is:

$$(w + p)f_e + j_e - c_e = 0$$

Effort level under the new regime is higher than when there were no incentives. The interesting case arises when nonmonetary reward of learning is a function of monetary incentives. Let nonmonetary benefits depend both on level of effort and external rewards i.e. $j(e, p)$. The first order condition is:

$$(w + p)f_e + j_e(e, p) - c_e(e) = 0$$

Students put in effort e^* that solves the above equation. To see how the optimal level of effort changes with external rewards, we compute the total derivative of $e*$ with respect to p:

$$\frac{de^*}{dp} = -\frac{\frac{\partial}{\partial p}(w + p)f_e + j_e(e, p) - c_e(e)}{\frac{\partial}{\partial e}(w + p)f_e + j_e(e, p) - c_e(e)}$$

$$= \frac{f_e + j_{ep}(e, p)}{-[(w + p)f_{ee} + j_{ee}(e, p) - c_{ee}(e)]}$$

The first two terms in the denominator are negative and the third term is positive, so the denominator is unambiguously positive. The first term in the numerator is also positive therefore the sign of the expression depends on the sign of j_{ep}. If $j_{ep} > 0$ i.e., the joy of learning increases with incentives, the expression is positive. If the joy of learning and incentives move in opposite directions, i.e., when incentives crowd out intrinsic motivation, the equilibrium level of effort may decrease with an increase in external rewards.

5. Data Description

In initial experimental design, final semester scores from grade seven were intended to be the baseline scores and grade eight district level-exam scores were to be the final measure of achievement. Several patterns in the data led us to question the suitability of these scores for our analysis. First, grade seven scores are missing for about 20 percent of the students, either because they transferred from schools outside the experiment basin or were out of school the previous year. Second, about 10 percent of students are repeating grade eight so it is not obvious which scores should be used as baseline scores. Third, grade seven scores cannot be compared across schools because tests and curricula are not uniform in grade seven. Fourth, we observe grade inflation in the end-of-the-year exams. To illustrate this last problem, histogram of mathematics scores from grade seven final exams, grade eight first and second semester exams, and grade eight district level exams are presented in Figure 1. There is a sharp spike in density around 32, the passing grade, in the end-of-the-year scores while the spike is much muted in the semester scores.

There are a number of possible reasons for grade inflation in year-end exams. Schools have an incentive to nudge a marginal student over the threshold because the funding they receive from the district education office is partly based on the number of students that pass the exams. It is also accepted and expected of teachers to award a few bonus points if it puts a marginal student over the passing threshold. Principals and teachers who have close ties to the community often find it difficult to resist the guardians' pleas to pass their ward, especially if the student has failed by a small margin. These interactions are absent in the first and second semester exams; schools do not report the scores from the semester exams to district authorities and students' grade transition does not depend on these scores.

Therefore, we treat grade eight first semester scores as baseline measure and second semester scores as our final measure of achievement in our analyses. This stacks the deck *against* finding positive treatment effects for several reasons. First, the incentives program was announced before the first semester exams so to the extent that incentives had positive impact on first semester scores, we are less likely to find positive treatment effect.[10] Also, if it takes time for students to internalize the program, change their habits and attitudes, and take actions to translate their enthusiasm into measurable outcomes, the impact of the program is likely to be weaker seven months after the initiation of the program than after a full year.

While we avoid year-end exams due to grade inflation, we cannot rule out grade inflation even in the semester exams for other reasons, though we believe the incentives are substantially weaker. Teachers may award more points simply out of sympathy. However, many teachers mentioned that the semester exams are graded fairly and are used as a disciplining device. The often subjective discussion on whether to pass a student in the year-end exams is determined largely by the student's attitude towards her studies during the year. Students who show no concern despite failing multiple subjects and being repeatedly warned by teachers to work harder are made to repeat the grade. Therefore it is unlikely that teachers gave up the effective

[10] At the time of the first round of incentives distribution, some teachers confided to us that they did not find our promise of cash rewards credible. This anecdotal evidence suggests that the first semester scores may be an appropriate measure of baseline achievement.

disciplining device just so the students could have some extra cash. More nefariously, one might worry that teachers could strike a deal with students to receive a portion of the cash rewards in return for higher scores. This is also unlikely to occur in our experiment because the marginal increase in cash rewards due to grade inflation in just one subject would be insignificant because the payments were based on average aggregate scores. It would take concerted action and collusion among multiple teachers to inflate the grades in all subjects and share the proceeds. The professional norms of teaching are established enough in our experimental area for such an outright collusion to be unlikely. Furthermore, the spike in density in the first and second semester scores in the neighborhood of the passing score for students in the treatment group is much muted compared to year-end scores. Thus we are confident that the semester exams are appropriate measures of students' learning.

Table 2 reports the baseline averages of key school and socioeconomic variables for treatment and control groups. Schools are balanced in terms of pupil-teacher ratio and availability of school infrastructure, the two largest areas of school expenditure, but we find that average years of schooling of teachers qualified to teach at lower secondary level is slightly lower in treatment schools. There is also little difference in the proportion of students whose family owns a home, a key indicator of socioeconomic status. However, indicators for households' tastes for education, like parents with secondary education, households hiring a tutor for students to study with after school, or family members helping children with schoolwork, suggest students in treatment schools are likely to come from relatively disadvantaged homes.

Estimates of the impact of incentives on scores could be biased if the attrition pattern is different between treatment and control groups. Table 3 shows that there is little difference in the proportion of students whose scores were available in each of the three exams.[11] The estimates could still be biased if the composition of the dropouts is different across treatment groups. To check for the difference in the composition of the dropouts, we compare pre-program percentile ranks of the dropouts in treatment and comparison schools using students' roll numbers as an approximate measure of their class rank.[12] As shown in table 3, we find little difference in the composition of the dropouts. Therefore our estimates are not likely to be biased due to nonrandom attrition across treatment arms.

6. Reduced Form Estimates

Given two periods of data, causal effect of the program could be estimated using an individual fixed effects model, a difference-in-difference equation, and a lagged dependent variable model under different identifying assumptions. If the randomization was successful, we would expect to find broadly similar results from alternative specifications. Therefore, we check the robustness of our results from the three models before providing a justification for the most appropriate one.

An individual fixed effects specification accounts for all time-invariant confounders such as innate ability of students, unobserved teacher and school characteristics, family background, and peer attributes that do not change over time. In our study, it is also important to account for the variation in the difficulty level of the semester exams and the grading scheme within schools that are likely to remain unchanged between the two semesters. Formally, the fixed effects specification is:

---

[11] The dropout rate observed in our sample is also similar to the districtwide dropout rates in Lalitpur district (Department of Education, 2010).

[12] Roll number in the current year usually reflects students' class rank in the previous years' year-end exams. As a diagnostic test, we check the correlation between roll number and class rank in contemporaneous exams and find it to be high (0.73 in the first semester, 0.71 in the second semester, and 0.69 in the district exams).

$$y_{ijt} = \beta_1 * treatment * 2nd\ Semester + \mu_i + X'_{ijt}\delta + \varepsilon_{jt} + \varepsilon_{ijt}\ , t = 0{,}1$$

where y_{ijt} is the normalized score of student *i* in school *j* in time *t*, *treatment* and *2nd Semester* are dummy variables for treatment schools and the second semester respectively, X_{ijt} is a vector of variables that vary over time, and μ_i is the time-invariant individual fixed effect for student *i*. The scores are normalized such that first semester scores in comparison schools have a mean of 0 and standard deviation of 1. Thus the obtained estimates represent the standard deviation (s.d.) gain in scores. Errors are clustered at the school level to allow for intra-cluster correlation.

The identifying assumption for the difference-in-difference specification is that the average scores follow the same trend in treatment and control groups in absence of the treatment. Thus, any observed difference in trends can be attributed as the causal impact of the treatment. Difference-in-difference equation is estimated as following:

$$y_{ijt} = \beta_0 + \beta_1 * treatment + \beta_2 * 2nd\ Semester + \beta_3 * treatment * 2nd\ Semester$$
$$+ X'_{ijt}\delta + \varepsilon_{jt} + \varepsilon_{ijt}\ , t = 0{,}1$$

where β_3, the parameter of interest, represents the between-group difference in within-group differences in pre- and post-test scores.

Finally, the lagged dependent variable model controls for a lagged measure of achievement and the identifying assumption is of conditional independence. The specification measures the gain in scores for treated students relative to that of control students with the same baseline achievement level. Conceptually, the baseline achievement measure is assumed to be a sufficient statistic for the history of family and school inputs and the unobserved innate ability of students. The lagged dependent variable model is:

$$y_{ijt=1} = \beta_0 + \beta_1 * y_{ijt=0} + \beta_2 * treatment + X'_{ijt}\delta + \varepsilon_j + \varepsilon_{ij}$$

with the parameter of interest β_2.

Estimates of the impact of incentives on second semester scores from all three specifications are reported in table 4. The fixed effects estimate, reported in the first column (0.09; *p*-value 0.08), is marginally significant while the estimates from the other two specifications are relatively small and insignificant.[13] Point estimates for the difference-in-difference and lagged dependent variable specification do not change significantly with the inclusion of school type dummies (lower secondary, secondary, and higher secondary), home variables (highest level of education completed by parents and availability of help in schoolwork, either from a hired tutor or a household member), and school level variables (pupil-teacher ratio, infrastructure index, and average years of teachers' education). The specifications, nevertheless, do not correct for bias due to unobserved variables like difficulty level of the semester exams, grading scheme, and other factors that are correlated to observed school and home characteristics.[14] Therefore, we conduct all further analyses of test scores using individual fixed-effects model.

To put the estimate from the fixed effects model in perspective, we compare our estimates with standard deviation gains due to other interventions. The girls' scholarship program in Kenya increased academic exam scores among girls by 0.12 to 0.19 s.d. (Kremer et al., 2009) depending on the sample. At the end of the first year of implementation, teacher incentives in Kenya had an effect of 0.05 s.d. on test scores (Glewwe et al., 2003), and textbook provision in Kenyan schools had an impact 0.04 s.d. (Glewwe et al., 2007). Pay for performance schemes for teachers in India improved students' scores in math and language tests by 0.19 and 0.12 s.d. respectively at the end of the first year of the program (Muralidharan and

[13] We do not estimate fixed-effects model with time varying covariates because there are few such variables and with small variation over time.
[14] Unobserved variables have likely negative bias because of their positive correlation with observed variables that show students have an unfavorable learning environment both at home and in school.

Sundararaman, 2006). Remedial education in urban India increased average test scores of all children in treatment schools by 0.14 s.d. at the end of the first year (Duflo et al., 2005). Finally, in the work most similar to ours, Fryer (2010) finds that when incentives are conditioned on grades, the impact on grades is between 0.09 and 0.13 s.d depending on the parameter of interest ("intent to treat" or "treatment on treated").[15] Thus our result is on par with many of the school intervention experiements implemented in the last decade, especially given that the use of first semester scores as our baseline and second semester scores as final measures likely understates our total impact.

We also analyze the impact of incentives on average scores aggregated over only the core subjects. As shown in the second column of table 5, the point estimate is zero, suggesting that the gains in non-core subjects constitute the source of observed gains. We have no reason to believe that schools choose to keep non-core courses based on students' performance. It is also unlikely that the gains reflect an artificial increase in scores because if teachers were to inflate the grades in the second semester, the scope to do so would be far more obvious and meaningful in core subjects like English, Mathematics, and Science where the baseline pass rate is low.

We observe that gains in two subjects, Health and Physical Education and Population and Environment, constitute the bulk of the aggregate gain. One possible explanation of this finding is that in response to incentives, students concentrate their effort in subjects that do not require higher-order thinking skills. The two subjects are also among the subjects with relatively high baseline scores. A possible explanation for this observation comes from the psychology literature where there is evidence that when rewards are made contingent on manifesting competence, rewards have the most impact on tasks with highest intrinsic interest (Arkes, 1979). It is

[15] Fryer (2010) finds no impact on standardized test scores.

plausible that students' intrinsic motivation is high in subjects with relatively high baseline scores.

Next we investigate if there are heterogeneous treatment effects across subsamples of gender, socioeconomic status, and baseline achievement levels.[16] To analyze if treatment effects differed by initial academic performance, students were placed into one of four quartiles based on average aggregate first semester scores.[17] Socioeconomic strata were constructed on the basis of students' report of their guardians' education.[18] The sample was divided into four mutually exclusive groups: guardians are illiterate, at least one guardian has at least some primary education, at least one guardian has at least some secondary education, and information on guardians' education is missing.

Table 6 presents the differential impacts of incentives on males and females. There is no significant difference in gains across gender: the gains for boys and girls are 0.10 s.d. (*p*-value 0.11) and 0.08 s.d. (*p*-value 0.09) respectively. This finding contrasts with the pattern observed in the literature where girls respond relatively strongly to incentives (Angrist and Lavy, 2009).

Heterogeneous treatment effects by ability are reported in table 7. Students in the bottom half of the distribution gain more from incentives than those in the top half. Point estimates for students in the third and bottom quartiles are 0.10 s.d. (*p*-value 0.08) and 0.14 s.d. (*p*-value 0.01) respectively whereas students in the top quartile gain 0.03 s.d. (*p*-value 0.51) and those in the second quartile gain 0.07 s.d. (*p*-value 0.29). Students in the bottom half show significant gains even when the scores are aggregated over the core subjects, and the lowest quartile is the only

[16] Information on gender was filled out by researchers for students who did not complete the baseline survey but had unambiguous male or female names.

[17] To the extent that incentives affected the distribution of scores in the first semester, the quartiles may not correspond to the "true" baseline quartiles.

[18] Students were asked to report the highest level of education completed by their guardians. The responses were coded under the following six categories at the time of data entry: (i) no schooling/illiterate, (ii) some primary education/adult literacy classes, (iii) primary education completed, (iv) some secondary education, (v) secondary education completed, (vi) more than secondary education.

subgroup to post significantly positive gains in any core subject (0.19 s.d. in Math; *p*-value 0.02). Thus our study shows that linear incentives scheme can be effective in improving academic outcomes of the least able student.

Other reasons why students in the bottom quartiles could have gained more than their counterparts in the top quartiles include the relatively stronger incentives faced by low ability students in the second semester because of their failure to pass all subjects in the first semester. Alternatively, it is possible that students in the top half find it harder to exhibit gains because they are already performing at a maximum level. Furthermore, it is also plausible that students in the top half internalized the program early on and the gains to them accrued in the first semester while it took longer for the gains to materialize for students in the bottom half.

The interaction effects of treatment with socioeconomic status are reported in table 8. The gains for students who come from relatively disadvantaged background, i.e., whose guardians are either illiterate (0.07 s.d.; *p*-value 0.24) or have primary schooling (0.08 s.d.; *p*-value 0.25) are statistically insignificant and lower than that for students whose guardians have secondary education (0.10 s.d.; *p*-value 0.03). This finding is significant because it suggests the *channel* through which the incentives affect students' outcomes. It is possible that there exist multiple equilibria in learning environment at home. In the absence of incentives, students and families are locked in a lower equilibrium because of the mutual expectations of parents and children: students do not work hard because they believe their families will not invest in academic reinforcement at home and parents do not provide extra reinforcement because of their belief that their children are not motivated to work hard. The provision of incentives may serve to break this self-fulfilling cycle of beliefs and move the system to a higher equilibrium. If this mechanism were at work, we would expect students with better educated parents to show the most significant gains because the parents are likely to be more interested in their children's schooling. Our analysis of the impact of incentives on passing rate and academic reinforcement at home provides further support for the hypothesis.

According to our incentives structure, students who passed all subjects received twice the price per score compared to students who failed one or more subjects. To measure the impact of incentives on passing rate, we estimate a difference-in-difference equation for the proportion of students passing all subjects in the first and second semesters. We use difference-in-difference specification because the coefficients have a simple and straightforward interpretation. We also check the robustness of the results using school and household variables as controls. Formally, the difference-in-difference specification is:

$$Pass_{ijt} = \beta_0 + \beta_1 * \ 2nd\ Sem + \beta_2 * treatment + \beta_3 * treatment * 2nd\ Sem + \ X'_{ijt}\delta + \varepsilon_{jt} + \varepsilon_{ijt}$$

where $Pass_{ijt}$ is the indicator variable for whether a student passed in the first and second semesters, $2nd\ Sem$ and $treatment$ are the dummy variables for second semester and treatment schools respectively, and X is a vector of home and school level variables .

The difference-in-difference coefficients are reported in table 9. We find that in the overall sample, students in treatment schools were no more likely to pass all subjects in the second semester than students in the control group. The observed point estimates could have been small because of the significant positive impact incentives might have had in the first semester. It is also possible that the threshold of passing all subjects to reach the higher price regime was out of reach for most students so they made little effort to earn rewards at the extensive margin.[19] We see that incentives had a large and significant impact on passing rate of

[19] Only a quarter of the students passed all subjects in the first semester.

students from the highest socioeconomic strata (0.10; *p*-value 0.05), while the estimates for students from lower socioeconomic strata are practically and statistically insignificant. The coefficients are robust to controlling for home and school variables.

If the incentives did have an impact on passing rates, it would be concentrated among the marginal students who fell short of the higher price regime by failing one or two subjects. We conduct separate analyses on passing rates by partitioning the sample into subgroups based on the number of subjects failed in the first semester. The results are reported in table 10. As per our expectation, we find that the difference in difference estimate for students who failed just one subject in the first semester is large and significant (0.22; *p*-value 0.00) and it is insignificant for all other subgroups. When we control for home and school variables, the point estimate for students who failed 3 courses becomes negative and marginally significant but there is no net loss in scores for the students. Analysis of aggregate scores for each subgroup shows that even students who failed multiple subjects and had little chance of switching to the higher reward regime gained substantially at the intensive margin.

7. Behavioral Response

One of the original contributions of our study is the analysis of student and household response to cash incentives. Our exit survey allows us to estimate if there were any changes in students' daily time use at home. We find that students in the treatment schools spend about 2 and 6 minutes less in schoolwork in the morning and evening respectively (not shown).[20] Two explanations are consistent with the reduction in study time. First, cash incentives could have had an adverse impact on students' intrinsic motivation thus lowering their effort level. As we

[20] We drop observations if the reported time use is not credible (for example, if the reported time on schoolwork is more than the total morning or evening hours available at home).

discuss subsequently, a direct check of intrinsic motivation shows no such effect. Second, students in the treatment group are more likely to be helped by a tutor, so it is also possible that the quantity of study time is traded off in favor of quality but we cannot verify this claim. In any event, both the magnitudes are practically and statistically insignificant for it to be of concern.

Table 11 presents the difference-in-difference estimates of the proportion of students receiving help at home, either from a hired tutor or a household member (older siblings, uncles/aunts, parents, etc.). For the overall sample, the estimate is practically large but insignificant (0.13; *p*-value 0.13). The coefficient for students from the bottom quartile (0.26; *p*-value 0.02) and students whose parents are relatively more educated (0.13; *p*-value 0.03) are large and significant while the coefficients for all other subgroups are statistically insignificant. These are also the subgroups that post the largest gains in aggregate scores. This observation provides further support for the hypothesis of multiple equilibria. If households performed a cost-benefit analysis before investing in children, we would expect academic reinforcement to increase the most for students in the top quartiles who received considerably more cash relative to students in the bottom quartiles. Instead, we find no relation between rewards and the change in availability of help at home but see a large point effect on students from higher socioeconomic strata.

It has been recognized that the official attendance information available from school register is not always reliable in developing countries. Therefore, we use the number of students who filled out the baseline and exit surveys as measures of attendance.[21] Difference-in-difference

[21] When conducting surveys, we generally contacted the principals for an appointment either the previous evening or the same morning and our requests were mostly met so our visits are “unannounced” for all practical purposes. Attendance rates might be nonrandom in the few occasions when our request was not met and were asked to visit at a later time.

estimates show no significant increase in attendance rate in treated schools (0.03; p-value 0.44).[22] Financial rewards had no significant impact on the availability of other educational inputs like pens, pencils, geometrical instruments, school bag, school shoes, school dress, books, and notebooks (not shown). This is perhaps not so surprising because educational materials were sufficiently available at baseline and marginal returns to additional materials is likely to diminish sharply.[23]

To check if external rewards had an adverse impact on intrinsic motivation, students were asked to report their interest in learning each subject. The question on the exit survey read "How interested are you in learning the following subjects" and the respondents had four options to choose from: (1) highly interested; (2) moderately interested; (3) moderately uninterested; and (4) highly uninterested.[24] Table 12 reports the results from an ordered probit analysis.[25] We find that the treatment status does not significantly predict the probability of choosing any option. This analysis suggests that financial rewards had no noticeable adverse impact on students' motivation to learn.

---

[22] This finding is reasonable because of high preprogram attendance rates in sample schools. Average attendance rate at baseline was 78 percent.

[23] Mean number of educational materials available at baseline were as follows: pens – 2.45; pencils – 1.30; erasures – 0.91; pencil sharpeners – 0.87; geometrical instruments kit – 1.43; school bags – 1.02; school shoes – 1.11 pairs; school dress – 1.37 pairs; books – 10.17; notebooks – 10.17. The number of books is higher than the number of subjects because more than 1 book is required for some subjects.

[24] A more comprehensive measure of students' intrinsic motivation is the interest/enjoyment subsection from the Intrinsic Motivation Inventory (IMI) as developed by Ryan (1982) which elicits responses for seven statements: (i) I enjoyed doing this activity very much; (ii) this activity was fun to do; (iii) I thought this was a boring activity; (iv) this activity did not hold my attention at all; (v) I would describe this activity as very interesting; (vi) I thought this activity was quite enjoyable; (vii) while I was doing this activity, I was thinking about how much I enjoyed it. We chose to use a simple question instead to simplify the task for students who filled out the questionnaires themselves. Our question nevertheless captures the essence of the seven statements.

[25] The ordered probit analysis requires larger values to represent "higher" outcomes. Therefore, we recoded the survey responses as following: (1) not interested at all; (2) not so interested; (3) moderately interested; and (4) highly interested.

8. Cost Effectiveness of Cash Incentives

As researchers experiment with various school interventions, programs can be now compared in terms of their cost effectiveness. Insofar as policymakers face budget constraints, comparative cost analysis allows them to choose the most cost effective programs. In the following exercise, we adopt the cost effectiveness methodology developed by researchers at Poverty Action Lab (J-PAL, 2010).

361 students in treatment schools took the first semester exams, and were eligible for a total of 70,067 rupees (~ $891.44).[26] Similarly, 351 students took the second semester exams in treatment schools and were eligible for a total of 71,817 rupees (~ $913.70). Thus the average eligible reward per student for the two semesters is $5.07.[27] Our estimate of average gains in students' scores between the first and second semesters is 0.09 standard deviations. Therefore, cost per pupil per 0.1 standard deviation gains is $5.63. These figures do not include the administrative costs associated with the program. Accounting for administrative costs, cost per pupil per 0.1 standard deviation gains jump to $8.87.[28]

Cost analysis in our study is complicated by the fact that students were rewarded for the first semester scores, yet these scores serve as our measure of baseline performance. Assuming that the provision of incentives during the second semester alone was responsible for all the gains in

[26] The total amount distributed after the first and second semester exams were rupees 46,318 (~$617.57) and 55,050 (~734.00) respectively. Recall that the eligible amount is different from distributed amount because those who failed received only half of what they were entitled to. Disbursed amount in the second semester includes the amount released to students who failed in the first semester and but passed all subjects in the second semester.

[27] Average exchange rate between US dollar and Nepali rupees was 1 to 78.6 for the year 2009 (http://www.oanda.com/currency/historical-rates). The first and second rounds of payments were distributed in September 2009 and February 2010 respectively (see table 1).

[28] Administrative costs up to the second round of incentives distribution is Rs. 90617 (~ $1152). Costs include transportation and stationery cost, cost of small incentives like pencils and erasures distributed to students for filling out the survey questionnaire, and most importantly, imputed payments for enumerators from the time of initial contact with schools up to the second round of incentives distribution and exit survey. Payments are imputed because all survey work was performed by researchers themselves. Enumerators' daily wage is set to be Rs. 700 (~ $8.9) per day as per the existing rates in the experiment area. Administrative costs associated with conducting the semester exams are not included because schools would have incurred them even in absence of the program.

the second semester exams, cost per pupil per 0.1 standard deviation gain is $2.89 without administrative costs. This figure is lower than the figure for textbook provision and teacher incentives in Kenya ($4.01 and $3.41 respectively), comparable to the girls' scholarship program in ($3.53), and higher than remedial education in India ($0.67 - $1.77 depending on the year).[29] When administrative costs are taken into account, cost per pupil per 0.1 standard deviations gain is $5.78.[30]

There is also a discussion in the literature on whether to treat financial rewards as costs or transfers - they are costs from the program implementer's point of view, but they are mere transfers for the society. If all cash rewards are treated as transfers, cost per pupil per 0.1 standard deviation gains in our study is $3.24 including the administrative costs. Therefore, a design like ours that utilizes test score data from schools is cheaper to implement while it is also liable to be manipulated.

9. Discussion

Our study has highlighted a number of interesting facets of incentives and education. We find that although there are positive gains in aggregate scores, the gains are concentrated among subjects in which students have relatively high baseline scores and that do not require higher order thinking skills.[31] Future studies should be designed to test if making rewards conditional on measures of higher order thinking skill improves students' performance.[32]

[29] Costs reported in Kremer et al. (2004) and Duflo et al. (2005) are in nominal US dollars for the year in which the programs were carried out (between 1996 to 2002 in Kenya and between 2001 and 2003 in India). The costs are likely to be higher in 2009 prices.
[30] Administrative cost is 80817 (~$1028) when costs associated with the first round of incentives distribution are excluded.
[31] Although students in the bottom quartile do show improvement in math which requires higher-order thinking skills and has low baseline scores.
[32] For example, Muralidharan and Sundararaman (2006) design tests to measure for "mechanical" and "conceptual" skills.

Our study finds that when incentives are structured as piece rates, low ability students gain substantially more relative to high ability students. We are aware of only one other study that finds large and positive (but statistically insignificant) test score gains among students in the lower tail of ability distribution (Kremer et al., 2009). This suggests that linear schemes could be effective in improving the academic outcomes of the least able students. From a policy perspective, linear schemes are also feasible for implementation because of their simplicity.

The design of our experiment allows us to evaluate the impact of incentives both at the extensive (passing all subjects) and intensive (improving aggregate scores) margins. We find that at the extensive margin, incentives affected students at the threshold of the higher reward regime while at the intensive margin, it improved academic outcome of the low ability students who had little chance of crossing over to the higher regime.

We also document the incidence of grade inflation in year-end exams which necessitated our shift to use second semester scores as the basis for measuring treatment effects. Researchers have often wondered why substantial increases in schooling have had no impact on the growth rate of output per worker and consequently on economic growth (Easterly, 2002). One explanation put forward by Pritchett (2001) is that schooling quality in developing countries is too low for additional years of schooling to increase cognitive skills or productivity. To the extent that the practice of grade inflation can be generalized, our finding supports the assertion that years of schooling grossly overestimate the endowed human capital of the labor force in developing countries.

The low power and scale of our experiment did not allow for differential levels of incentives and therefore we are unable to provide an estimate of the price elasticity of achievement gains. Under an ideal experiment, price per score would be randomized among treatment schools before comparing the gains in outcomes. This issue is particularly important in light of the evidence that that paying too little may actually be counterproductive (Gneezy and Rustichini, 2000).

Our finding on the behavioral response of households is also revealing of the interaction between incentives and households' actions. Incentives increased the proportion of students that received help with schoolwork at home, and it did so significantly for students from the upper socioeconomic strata. This suggests the presence of multiple equilibria in learning dynamics at home which implies the program could have been more effective if it were designed to involve guardians explicitly. Increasing guardians' participation and involvement by explaining the program in their presence, interacting with them regularly, and actively encouraging them to provide a better learning environment at home might improve the effectiveness of incentives.[33]

The results presented in this paper are short-term impacts of incentives. Researchers believe that though rewards may modify people's behavior in the short-run, such change is temporary and incentives may lead to poorer outcomes in the long-term (Kohn, 1993).[34] To measure the long term impact of incentives, we plan to use scores from national level exams students take at the end of grade 10.

10. Conclusion

In recent years, researchers have experimented with innovative school interventions to increase school attendance and academic outcomes of students. This paper reports the results of a piece-rate cash incentives program where students were provided monetary rewards in proportion to their scores in academic exams. We find that monetary incentives raise academic achievement, with the weakest students and students from higher socioeconomic strata exhibiting

[33] At the end of the year, when asked of their general impression of the program and what changes could make it more effective, several teachers responded that the family's involvement should be increased.
[34] If we do observe such fade-out effect, it would not be unique to our study.

the most gains. Our finding of a positive treatment effect is remarkable especially because a number of experimental and institutional factors make it less likely to find significant treatment effects. Incentives also bring about favorable changes in the learning environment at home and our findings suggest the presence of synergy between students' effort and households' investment in after-school academic reinforcement. If this were indeed the case, incentives schemes should be designed to ensure parents' participation in the program. Indeed, the only study in the literature that finds a robust positive impact of incentives on academic outcomes makes parents an explicit part of the program and even provides them a portion of the rewards (Kremer et al., 2009).

We also recognize that there is a scope for implementation of cash incentives program like ours in developing countries because of its cost effectiveness. Incentives programs where rewards are contingent on educational inputs are likely to have higher administrative and monitoring costs. At the same time, incentive schemes based on outputs are more amenable to gaming and fraud. Relative superiority of input versus output based incentives is therefore not obvious *a priori* and should be the subject of further inquiry.

Our study adds to the new generation of "determinants of achievement" studies that has begun to shed more light into the "black box" of education production function and tease out the *mechanisms* of education production. Thus we are now better positioned to assert *how, why,* and *when* certain interventions work, as some commentators insist theories of economic development must do (Deaton, 2010a, 2010b). Future studies should be designed based on richer and more nuanced behavioral models of cognitive achievement.

References

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos, "Lead them to Water and Pay them to Drink: An Experiment with Services and Incentives for College Achievement" (NBER Working Paper No. 12790, 2006).

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *American Economic Review* 92 (5): 1535-58.

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. Long-term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia. *American Economic Review* 96 (3): 847-62.

Angrist, Joshua, and Victor Lavy. 2009. The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial. *American Economic Review* 99 (4): 1384-414.

———. 1999. Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* 114 (2): 533-75.

Arkes, Hal R. 1979. Competence and the Overjustification Effect. *Motivation and Emotion* 3 (2): 143-50.

Banerjee, Abhijeet, Shawn Cole, Esther Duflo, and Leigh Linden, "Remedying Education: Evidence from Two Randomized Experiments in India" (NBER Working Paper No. 11904, 2005).

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

Barro, Robert J. 1991. Economic Growth in a Cross Section of Countries. *Quarterly Journal of Economics* 106 (2): 407-43.

Behrman , Jere R., Susan W. Parker, and Petra E. Todd, "Long-Term Impacts of the *Oportunidades* Conditional Cash Transfer Program on Rural Youth in Mexico" (Discussion Paper No. 122, Ibero-America Institute for Economic Research, 2005).

Behrman, Jere, and Nancy Birdsall. 1983. The Quality of Schooling - Quantity Alone is Misleading. *American Economic Review* 73 (5): 928-46.

Bull, Clive, Andrew Schotter, and Keith Weigelt. 1987. Tournament and Piece Rates - An Experimental Study. *Journal of Political Economy* 95 (1): 1-33.

Cornwell, Christopher W., Kyung Hee Lee, and David B. Mustard, "The Effects of Merit-Based Financial Aid on Course Enrollment, Withdrawal and Completion in College" (IZA Discussion Paper 820, 2003).

Cornwell, Christopher W., David B. Mustard, and Deepa J. Sridhar. 2006. The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE program. *Journal of Labor Economics* 24 (4) (10/01): 761-86.

Deaton, Angus. 2010a. Instruments, Randomization, and Learning about Development. *Journal of Economic Literature* 48 (2): 424-55.

———. 2010b. Understanding the Mechanisms of Economic Development. *Journal of Economic Perspectives* 24 (3): 3-16.

Deci, Edward L. 1971. Effects of Externally Mediated Rewards on Intrinsic Motivation. *Journal of Personality and Social Psychology* 18 (1): 105.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin* 125 (6): 627-68.

Department of Education. 2008. *National Assessment of Grade Eight Students.*

Deparment of Education. 2010. Flash I Report 2009-10.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2007. "Peer Effects, Pupil-Teacher Ratios, and Teacher Incentives - Evidence from a Randomized Evaluation in Kenya" (Unpublished Manuscript).

Easterly, William. 2002. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. United States: Massachusetts Institute of Technology.

Fehr, Ernst, and Armin Falk. 2002. Psychological Foundations of Incentives. *European Economic Review* 46 (4-5): 687-724.

Frey, Bruno S., and Reto Jegen. 2001. Motivation Crowding Theory. *Journal of Economic Surveys* 15 (5): 589-611.

Frey, Bruno, and Felix Oberholzer-Gee. 1997. The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-out. *American Economic Review* 87 (4): 746-55.

Fryer, Roland J. Jr.,. 2010. "Financial Incentives and Student Achievement: Evidence from Randomized Trials."

Fuller, Bruce. 1987. What School Factors Raise Achievement in the Third-World. *Review of Educational Research* 57 (3): 255-92.

Glewwe, Paul. 2002. Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes. *Journal of Economic Literature* 40 (2): 436-82.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives" (NBER Working Paper No. 9671, 2003).

Glewwe, Paul, and Michael Kremer. 2006. Schools, Teachers, and Education Outcomes in Developing Countries. In *Handbook of the Economics of Education*, eds. Eric A. Hanushek, Finis Welch. Vol. 2, 945-1017. Amsterdam: North-Holland.

Gneezy, Uri, and Aldo Rustichini. 2000. Pay Enough or Don't Pay At All. *Quarterly Journal of Economics* 115 (3): 791-810.

Green, Jerry R., and Nancy L. Stokey. 1983. A Comparison of Tournaments and Contracts. *Journal of Political Economy* 91 (3): 349-64.

Hahn Andrew, Tom Leavitt, and Paul Aaron, "Evaluations of the Quantum Opportunities Program (QOP): Did the Program Work?" (Evaluation Report, Brandeis University, Center for Human Resources, 1994).

Hanushek, Eric A. 2003. The Failure of Input-based Schooling Policies. *The Economic Journal* 113 (485): 64-98.

———. 1997. Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis* 19 (2): 141-64.

———. 1995. Interpreting Recent Research on Schooling in Developing Countries. *World Bank Research Observer* 10 (2): 227-46.

Hanushek Eric A., and Ludger Woessmann, 2008. The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature* 46 (3): 607-68.

Hanushek, Eric A., and Dennis D. Kimko. 2000. Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review* 90 (5): 1184-208.

Harbison, Ralph W., and Eric A. Hanushek. 1992. *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. New York: Oxford University Press.

Heyneman, Stephen P., Joseph P. Farrell, and Manuel A. Sepulveda-Stuardo, "Textbooks and Achievement: What we Know" (Working Paper No. 298, The World Bank, 1978).

Holmstrom, Bengt. 1982. Moral Hazard in Teams. *Bell Journal of Economics* 13 (2): 324-40.

Jameel Poverty Action Lab. 2010. “J-PAL Cost Effectiveness Methodology” (Unpublished Manuscript).

Kohn, Alfie. 1993. *Punished by Rewards*. Boston: Houghton Mifflin.

Kremer, Michael. 2003. Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons. *The American Economic Review* 93 (2, Papers and Proceedings of the One Hundred Fifteenth Annual Meeting of the American Economic Association, Washington, DC, January 3-5, 2003) (May): 102-6.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. Incentives to learn. *Review of Economics and Statistics* 91 (3): 437-56.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. “Incentives to learn” (Working Paper No. 10971, NBER, 2004).

Kremer, Michael, Sylvie Moulin and Robert Namunyu, "Decentralization: A Cautionary Tale" (Working Paper No. 10, Poverty Action Lab, 2003).

Krueger, Alan B. 1999. Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics* 114 (2): 497-532.

Lavy, Victor. 2002. Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement. *Journal of Political Economy* 110 (6): 1286-317.

Lazear, Edward P., and Sherwin Rosen. 1981. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy* 89 (5): 841-64.

Leuven, Edwin, Hessel Oosterbeek and Bas Van der Klaauw, "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment" University of Amsterdam, 2003.

Long, David et al., "LEAP: Three-Year Impacts of Ohio's Welfare Initiative to Improve School Attendance among Teenage Parents. Ohio's Learning, Earning, and Parenting Program" (Evaluation Report, Manpower Demonstration Research Corp., New York, 1996).

Mankiw, Gregory, David Romer, and David N. Weil. 1992. A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics* 107 (2): 407-37.

Mullis, Ina V. S., Michael O. Martin, Ann M. Kennedy, and Pierre Foy. 2007. *PIRLS 2006 International Report: IEA's Progress in International Reading Literary Study in Primary School in 40 countries.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, .

Mullis, Ina V. S., Michael O. Martin, D. F. Robitaille, and Pierre Foy. 2009. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2006. "Teacher Incentives in Developing Countries: Experimental Evidence from India” (Unpublished Manuscript).

Nalebuff, Barry J., and Joseph E. Stiglitz. 1983. Prizes and Incentives: Towards a General Theory of Compensation and Competition. *Bell Journal of Economics* 14 (1): 21-43.

OANDA Historical Exchange Rates. 2010 [cited 08/08/2010]. Available from http://www.oanda.com/currency/historical-rates.

Ponce, Juan, and Arjun S. Bedi, "The Impact of a Cash Transfer Program on Cognitive Achievement: The Bono De Desarrollo Humano of Equador" (Discussion Paper No. 3658, IZA, 2008).

Prendergast, Canice. 1999. The Provision of Incentives in Firms. *Journal of Economic Literature* 37 (1): 7-63.

Pritchett, Lant, and Deon Filmer. 1999. What Education Production Functions Really Show: A Positive Theory of Education Expenditures. *Economics of Education Review* 18 (2): 223-39.

Psacharopoulos, George. 1994. Returns to Investment in Education: A Global Update. *World Development* 22 (9): 1325-43.

———. 1985. Returns to Education: A Further International Update and Implications. *Journal of Human Resources* 20 (4): 583-604.

———. 1981. Returns to Education: An Updated International Comparison. *Comparative Education* 17 (3): 321-41.

———. 1972. Rates of Return to Investment in Education Around the World. *Comparative Education Review* 16 (1) (Feb.): 54-67.

Reimers, Fernando, Carol DeShano da Silva and Ernesto Trevino, "Where is the "Education" in Conditional Cash Transfers in Education?" (Working Paper No. 4, UNESCO Institute for Statistics, 2006).

Romer, Paul M. 1990. Endogenous Technological Change. *The Journal of Political Economy* 98 (5, Part 2: The Problem of Development: A Conference of the Institute for the Study of Free Enterprise Systems) (Oct.): S71-S102.

Sappington, David E. M. 1991. Incentives in Principal-Agent Relationships. *Journal of Economics Perspectives* 5 (2): 45-66.

Schiefelbein, Ernesto, and John Simmons, "The Determinants of School Achievements: A Review of the Research for Developing Countries" (Manuscript Report, International Development Research Centre, 1981).

Schultz, T. Paul. 2004. School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program. *Journal of Development Economics* 74 (1): 199-250.

Simmons, John, and Leigh Alexander. 1978. Determinants of School Achievement in Developing Countries: A Review of the Research. *Economic Development and Cultural Chage* 26 (2): 341-56.

UNESCO Institute for Statistics. 2010 [cited 07/30 2010]. Available from http://www.uis.unesco.org/ev.php?ID=2867_201&ID2=DO_TOPIC.

Velez, Eduardo, Ernesto Schiefelbein and Jorge Valenzuela, "Factors Affecting Achievement in Primary Education" (Working Paper No. 2, Human Resources Development and Operations Policy, The World Bank, 1993).

Wolfe, Barbara and Samuel Zuvekas, "Nonmarket Outcomes of Schooling" (Discussion Paper No. 1065-95, Institute for Research on Poverty, University of Wisconsin-Madison, 1995).

Figure 1: Distribution of Math Scores (Histogram)

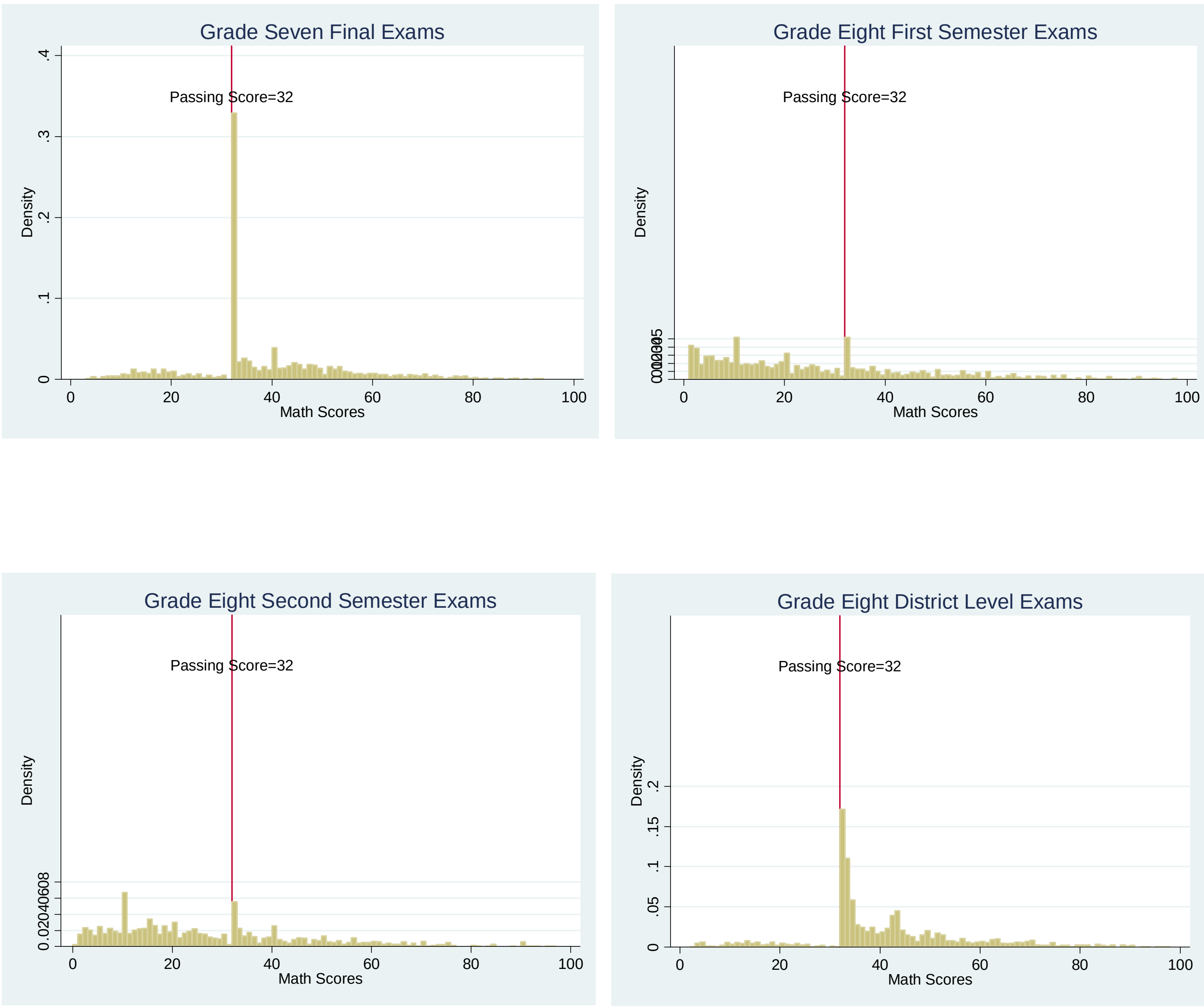

Table 1: Timeline of Achievement Incentive Program, 2009 – 2010

Date	**Activity**
2009	
February – March	Initial contact with schools
March	Grade 7 final semester exams
April 12	Start of the academic year 2009/10
May – June	Intake survey in program schools
	Incentive program announced
June – July	Intake survey in control schools
August	Revisit of program schools to remind students of the program
	First semester exams
September	Distribution of the first round of incentives
December	Second semester exams
2010	
January	Winter vacation
February	Distribution of the second round of incentives
	Exit survey in treatment schools
February – March	Exit survey in control schools
March	Third semester exams
April 15	Start of the academic year 2010/2011
April – May	Distribution of the third round of incentives

Table 2: Summary Statistics of School and Socioeconomic Variables

	Treatment	Control	Difference	Obs.
Pupil-teacher ratio	18.48	18.37	0.11	32
School infrastructure index (0-7)	5.54	6.00	-0.46	33
Teachers' schooling (years)	13.84	14.29	-0.45*	31
School attendance rate	0.79	0.78	-0.00	33
Grade eight repeaters (proportion)	0.10	0.09	0.01	1402
Boys (proportion)	0.45	0.43	0.02	1479
Age of students (years)	14.49	14.35	0.14	1227
Students living with parents (proportion)	0.92	0.90	0.02	1228
Family owns the place of residence (proportion)	0.77	0.72	0.05	1194
Number of siblings	2.18	2.26	-0.08	1228
Guardians illiterate (proportion)	0.18	0.16	0.02	1227
Guardians have primary education (proportion)	0.34	0.30	0.04	1227
Guardians have secondary education (proportion)	0.36	0.44	-0.08*	1227
Guardians' education missing (proportion)	0.11	0.10	0.01	1227
Tutor helps with schoolwork (proportion)	0.03	0.11	-0.08***	1227
Household member helps with schoolwork (proportion)	0.38	0.52	-0.14***	1208

Notes: Information on number of teachers is unavailable for 1 school. Information on teachers' education is unavailable for 2 schools. Infrastructure index is obtained by summing seven binary variables: availability of sufficient drinking water, toilet, urinal, playground, electricity, fence in school compound, and computer facilities.
Difference between group means is tested using a *t*-test. Standard errors used in testing difference between group means are clustered at the school level.
*Significant at 10% level **Significant at 5% level ***Significant at 1% level

Table 3: Attrition Pattern

	1st Semester			2nd Semester			3rd Semester		
	Treatment	Control	Difference	Treatment	Control	Difference	Treatment	Control	Difference
Attrition Rates	0.07	0.09	-0.02 (0.02)	0.10	0.12	-0.02 (0.03)	0.10	0.10	-0.00 (0.03)
Percentile Rank of Dropouts	37.04	39.11	-2.07 (6.00)	35.11	39.93	-4.82 (5.46)	36.13	36.20	-0.07 (5.15)

Notes: The figures in the first row refer to the proportion of students whose exam scores for all subjects are unavailable for the first, second, and third semesters. The figures in the second row refer to the pre-program percentile rank of the dropouts. Students' roll numbers are used an approximate measure of pre-program class rank.
Standard errors are clustered at the school level and reported in parentheses.

Table 4: Estimates of the Impact of Cash Incentives on Average Aggregate Scores

	Fixed Effects	Diff-in-Diff	Diff-in-Diff	Lagged Model	Lagged Model
Causal Effect	0.09* (0.05)	0.03 (0.07)	0.04 (0.08)	0.04 (0.07)	0.04 (0.05)
School Type Controls			X		X
Household Variables			X		X
School Variables			X		X
Obs.	2624	2624	1850	1312	925

Notes: Dependent variables are normalized scores (with respect to mean and standard deviation of first semester scores in comparison schools). School type controls include dummy variables for the type of school (lower secondary, secondary, higher secondary). Household variables include dummy variables for the highest level of education completed for parents, tutor help, and help from a household member. School variables include pupil-teacher ratio, infrastructure index, and years of teachers' education. Standard errors are clustered at the school level and reported in parentheses.

*** Significant at 1% level; **Significant at 5% level; *Significant at 10% level

Table 5: Estimates of the Impact of Cash Incentives

	Aggregate (All Subjects)	**Aggregate (Core Subjects)**	Nepali	English	Math	Science	Social Studies	Health and Physical Education	Population and Environment	Moral Science	Vocational Studies
Treatment* Second Semester	**0.09*** **(0.05)**	**-0.00** **(0.06)**	0.06 (0.06)	-0.03 (0.13)	0.02 (0.07)	0.09 (0.16)	-0.18 (0.12)	0.52*** (0.17)	0.42** (0.16)	0.22 (0.15)	-0.16 (0.20)
Obs.	**2624**	**2670**	2706	2738	2706	2726	2728	2454	2442	1464	1902
Raw Baseline Score	**37.7**	**35.8**	40.1	36.6	26.0	34.2	41.6	19.5	21.2	24.2	21.7

Notes: Dependent variables are normalized scores (with respect to mean and standard deviation of first semester scores in comparison schools). Standard errors are clustered at the school level and reported in parentheses.
*** Significant at 1% level; **Significant at 5% level; *Significant at 10% level

Table 6: Estimates of the Differential Impact of Cash Incentives by Gender

	Aggregate (All Subjects)	Aggregate (Core Subjects)	Nepali	English	Math	Science	Social Studies	Health and Physical Education	Population and Environment	Moral Science	Vocational Studies
Boys	0.10 (0.06)	0.00 (0.07)	0.09 (0.07)	-0.03 (0.12)	0.02 (0.07)	0.11 (0.18)	-0.17 (0.13)	0.54*** (0.17)	0.44** (0.20)	0.23 (0.21)	-0.24 (0.23)
Obs.	1074	1102	1114	1136	1122	1128	1130	1018	1016	642	794
Girls	0.08* (0.04)	-0.01 (0.07)	0.04 (0.05)	-0.02 (0.14)	0.03 (0.08)	0.07 (0.15)	-0.19 (0.12)	0.50*** (0.18)	0.40*** (0.15)	0.21** (0.09)	-0.10 (0.19)
Obs.	1532	1550	1572	1582	1566	1578	1578	1418	1408	814	1098

Notes: Dependent variables are normalized scores (with respect to mean and standard deviation of first semester scores in comparison schools). Standard errors are clustered at the school level and reported in parentheses.
*** Significant at 1% level; **Significant at 5% level; *Significant at 10% level

Table 7: Estimates of the Differential Impact of Cash Incentives by Achievement Level

	Aggregate (All Subjects)	Aggregate (Core Subjects)	Nepali	English	Math	Science	Social Studies	Health and Physical Education	Population and Environment	Moral Science	Vocational Studies
Top Quartile	0.03 (0.04)	-0.06 (0.06)	-0.03 (0.07)	0.06 (0.22)	-0.15 (0.13)	0.08 (0.22)	-0.21 (0.23)	0.45 (0.31)	0.37 (0.22)	0.18 (0.12)	-0.19 (0.26)
Obs.	652	652	656	656	658	656	658	592	590	350	454
Second Quartile	0.07 (0.07)	0.07 (0.07)	0.04 (0.08)	-0.01 (0.16)	0.03 (0.07)	-0.01 (0.13)	-0.15 (0.11)	0.51*** (0.16)	0.44** (0.19)	0.26* (0.13)	-0.14 (0.22)
Obs.	670	670	682	684	682	682	686	618	614	366	480
Third Quartile	0.10* (0.05)	0.09* (0.05)	0.12 (0.08)	-0.10 (0.14)	0.03 (0.08)	0.25 (0.20)	-0.21 (0.12)	0.50*** (0.16)	0.40** (0.17)	0.14 (0.19)	-0.20 (0.18)
Obs.	660	660	674	686	676	682	686	612	612	364	474
Bottom Quartile	0.14** (0.05)	0.13** (0.05)	0.10 (0.07)	-0.05 (0.10)	0.19** (0.07)	0.03 (0.13)	-0.16* (0.08)	0.62*** (0.11)	0.47*** (0.12)	0.28* (0.16)	-0.12 (0.23)
Obs.	642	642	664	674	666	670	664	596	592	358	460

Notes: Dependent variables are normalized scores (with respect to mean and standard deviation of first semester scores in comparison schools). Standard errors are clustered at the school level and reported in parentheses.
*** Significant at 1% level; **Significant at 5% level; *Significant at 10% level

Table 8: Estimates of the Differential Impact of Cash Incentives by Socioeconomic Status

	Aggregate (All Subjects)	Aggregate (Core Subjects)	Nepali	English	Math	Science	Social Studies	Health and Physical Education	Population and Environment	Moral Science	Vocational Studies
Guardians Illiterate	0.07 (0.05)	-0.02 (0.06)	0.06 (0.06)	-0.19 (0.14)	-0.05 (0.08)	0.33 (0.23)	-0.26 (0.21)	0.57** (0.23)	0.45 (0.27)	0.06 (0.20)	-0.28 (0.21)
Obs.	370	384	388	396	388	390	394	360	356	230	192
Guardians with Primary Education	0.08 (0.06)	-0.04 (0.07)	0.07 (0.06)	-0.03 (0.12)	0.02 (0.08)	0.07 (0.15)	-0.18** (0.09)	0.50** (0.17)	0.40** (0.18)	0.18 (0.18)	-0.11 (0.25)
Obs.	822	710	848	854	844	850	854	756	750	420	560
Guardians with Secondary Education	0.10** (0.05)	0.03 (0.05)	0.04 (0.07)	0.05 (0.12)	0.07 (0.07)	0.05 (0.12)	-0.13 (0.12)	0.49*** (0.17)	0.44** (0.16)	0.25** (0.11)	-0.21 (0.18)
Obs.	964	980	982	996	990	992	994	994	892	420	662
Guardians' Education Missing	0.07 (0.09)	0.03 (0.10)	0.12 (0.09)	0.01 (0.21)	-0.00 (0.11)	-0.15 (0.24)	-0.27** (0.10)	0.67*** (0.19)	0.39*** (0.09)	0.55*** (0.12)	-0.00 (0.21)
Obs.	468	596	488	492	484	494	994	444	444	306	388

Notes: Dependent variables are normalized scores (with respect to mean and standard deviation of first semester scores in comparison schools). Standard errors are clustered at the school level and reported in parentheses.
*** Significant at 1% level; **Significant at 5% level; *Significant at 10% level

Table 9: Difference in Difference Estimate of the Proportion of Students Passing All Subjects

	Passed All Subjects	Obs.	Passed All Subjects	Obs.
Full Sample	0.03 (0.03)	2624	0.04 (0.04)	1850
Top Quartile	-0.05 (0.07)	652	-0.05 (0.08)	488
Second Quartile	0.11 (0.07)	670	0.12 (0.09)	492
Third Quartile	0.08 (0.05)	660	0.08 (0.06)	462
Fourth Quartile	-0.00 (0.00)	642	-0.00 (0.00)	408
Guardians Illiterate	-0.04 (0.06)	370	-0.04 (0.06)	358
Guardians with Primary Education	-0.02 (0.04)	700	-0.03 (0.04)	624
Guardians with Secondary Education	0.10** (0.05)	964	0.11** (0.05)	850
School Type Controls			X	
Home Variables			X	
School Variables			X	

Notes: School type effects include dummy variables for the type of school (lower secondary, secondary, higher secondary). Home variables include dummy variables for the highest level of education completed by guardians, tutor help, and help from a household member. School variables include pupil-teacher ratio, infrastructure index, and years of teachers' education. Standard errors are clustered at the school level and reported in parentheses.

*** Significant at 1% level; **Significant at 5% level; *Significant at 10% level

Table 10: Difference in Difference Estimate of the Proportion of Students Passing All Subjects
(by Number of Courses Failed in 1st Semester)

	Passed All Subjects	Obs.	Passed All Subjects	Obs.	Gains in Aggregate Scores
Full Sample	**0.03** (0.03)	2624	**0.04** (0.04)	1850	**0.09*** (0.05)
Failed 0 Course	**0.05** (0.04)	102	**0.03** (0.04)	80	**-0.02** (0.07)
Failed 1 Course	**0.22***** (0.08)	232	**0.26**** (0.09)	176	**0.06** (0.06)
Failed 2 Courses	**0.09** (0.18)	258	**0.16** (0.18)	214	**0.17**** (0.07)
Failed 3 Courses	**-0.17** (0.14)	244	**-0.32*** (0.16)	178	**0.10** (0.10)
Failed 4 Courses	**0.07** (0.10)	398	**-0.05** (0.12)	296	**0.06** (0.07)
Failed 5 Courses	**0.06** (0.07)	392	**0.07** (0.06)	294	**0.08** (0.07)
Failed 6 Courses	**0.03** (0.04)	376	**0.02** (0.04)	252	**0.09*** (0.05)
Failed 7 Courses	**0.02** (0.02)	384	**0.02** (0.03)	234	**0.07**** (0.04)
Failed 8 Courses	-	186	-	100	**0.20***** (0.00)
Failed 9 Courses	-	52	-	26	**0.26**** (0.11)
School Type Controls			X		
Home Variables			X		
School Variables			X		

Notes: School type effects include dummy variables for the type of school (lower secondary, secondary, higher secondary). Home variables include dummy variables for the highest level of education completed by guardians, tutor help, and help from a household member. School variables include pupil-teacher ratio, infrastructure index, and years of teachers' education. Standard errors are clustered at the school level and reported in parentheses.
*** Significant at 1% level; **Significant at 5% level; *Significant at 10% level

Table 11: Difference in Difference Estimate of the Proportion of Students Receiving Help at Home

	Help at Home	Obs.	Help at Home	Obs.
Full Sample	0.13 (0.08)	2456	0.11 (0.08)	2002
Top Quartile	0.14 (0.10)	608	0.15 (0.11)	494
Second Quartile	0.07 (0.08)	614	0.02 (0.07)	510
Third Quartile	0.07 (0.11)	586	0.04 (0.11)	484
Bottom Quartile	0.26** (0.11)	568	0.24** (0.12)	450
Parents Illiterate	0.12 (0.12)	406	0.11 (0.12)	396
Parents with Primary Education	0.08 (0.12)	752	0.04 (0.11)	668
Parents with Secondary Education	0.13** (0.06)	1036	0.13** (0.06)	918
School Type Controls			X	
Home Variables			X	
School Variables			X	

Notes: School type effects include dummy variables for the type of school (lower secondary, secondary, higher secondary). Home variables include dummy variables for the highest level of education completed by guardians, tutor help, and help from a household member. School variables include pupil-teacher ratio, infrastructure index, and years of teachers' education. Standard errors are clustered at the school level and reported in parentheses.

*** Significant at 1% level; **Significant at 5% level; *Significant at 10% level

Table 12: Impact of Cash Incentives on Intrinsic Motivation to Learn

	Mean (Treatment Group)	Mean (Control Group)	Marginal Effects of Treatment				Obs.
			Not Interested at All	Not Very Interested	Moderately Interested	Highly Interested	
Nepali	3.47	3.47	-0.00 (0.00)	-0.00 (0.00)	-0.01 (0.06)	0.01 (0.07)	1123
English	3.35	3.39	0.00 (0.00)	0.00 (0.01)	0.01 (0.03)	-0.01 (0.05)	1123
Math	3.37	3.39	0.00 (0.00)	0.00 (0.01)	0.00 (0.03)	-0.02 (0.05)	1120
Science	3.57	3.54	0.00 (0.00	0.00 (0.00)	0.00 (0.04)	-0.00 (0.05)	1120
Social Studies	3.48	3.40	-0.00 (0.00)	-0.02 (0.01)	-0.05 (0.04)	0.06 (0.05)	1119
Health and Physical Education	3.31	3.39	0.00 (0.00)	0.02 (0.01)	0.03 (0.03)	-0.05 (0.05)	993
Population and Environment	3.26	3.32	0.00 (0.00)	0.00 (0.02)	0.02 (0.03)	-0.03 (0.05)	990
Moral Education	3.22	3.21	-0.00 (0.02)	-0.01 (0.04)	-0.01 (0.04)	0.03 (0.10)	589
Vocational Studies	3.27	3.24	-0.00 (0.00)	-0.00 (0.02)	-0.00 (0.03)	0.00 (0.05)	764

"Marginal Effects of Treatment" refers to the marginal change in probability when the dummy variable for treatment changes from 0 to 1. Ordered probit regressions are conducted on the following four categories: (1) not interested at all; (2) not very interested; (3) moderately interested; and (4) highly interested. Probabilities may not sum to 1 due to rounding. All regressions control for subject-specific first semester scores and include fixed effects for the level of school (lower secondary, secondary, and higher secondary). Standard errors are clustered at the school level.