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

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By

Dhiraj Sharma

Graduate Program in Agricultural, Environmental and Development Economics

The Ohio State University

2011

Dissertation Committee:

Professor Brian Roe, Advisor

Professor Claudio Gonzalez-Vega

Professor Joyce Chen

Copyrighted by

Dhiraj Sharma

2011

Abstract

This dissertation examines the impact of financial incentives on students' academic achievement and households' response to these incentives. Economic theory suggests that financial incentives can improve students' academic performance by acting as a price subsidy to 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. On the other hand, incentives may lead to a worse performance due to a loss in students' intrinsic motivation, that is, the innate desire to learn.

I conducted a randomized field experiment among students in public schools in Lalitpur district in Nepal. From a pool of 33 schools, 11 were randomly chosen to be in the treatment group while the remaining 22 constituted the control group. Grade eight students in the treatment schools were offered cash incentives that increased monotonically with their scores (e.g., piece-rate incentives) in each of the two semester exams and the end-of-the-year district level exam. To encourage students to pass, those who passed all subjects received twice the reward per score compared to those who failed one or more subjects.

Despite several experimental and institutional factors making it less likely of finding a positive treatment effect, I find that the incentives increased the average

aggregate scores by approximately 0.14 standard deviations. There is no noticeable difference in gains between males and females and incentives had a relatively higher impact on students from higher socioeconomic strata and lower academic performance quartiles. This study contributes to the growing literature on the short-term impact of academic incentives by recording households' response to the incentives. I show that the increase in aggregate scores is likely due to the increased availability of help with schoolwork at home, either from a hired tutor or a household member. Finally, financial rewards did not have an adverse impact on students' intrinsic motivation to learn.

Policy implications of this study include the potential use of piece-rates incentive schemes in merit scholarships and cash transfer programs to encourage students with low ability. It also highlights the importance of improving students' learning environment at home by increasing the involvement of parents in their children's education and schooling.

To *mami*, *bua*, *dada*, and *bhauju*,
for this is the fruit of their sacrifice.

Acknowledgments

I am deeply indebted to my advisor Dr. Brian Roe for his strong and unwavering support during this process and for the autonomy he gave me to pursue my passion.

I am fortunate to be a student of Dr. Claudio Gonzalez-Vega from whom I never stop learning.

I am thankful to Dr. Joyce Chen for her valuable advice, comments, and suggestions.

I want to thank my classmates Emilio, Jose, Kora, Matt, Mike, Nick, Shibalee, and Shigeharu. It would have been that much harder to survive the first year and beyond without you in the trenches.

Many thanks to the Nepali community in Columbus – Amulya, Aruna, Balaram, Dibu, Enna, Isha, Keshav, Nagendra, Nripesh, Pradeep, Prakrit, Rabi, Rakshya, Ramita, Rashmi, Rhitom, Shruti *didi*, Suresh *dai*, Tripti, Umakanta *dai*, Unnat, and Usha – for a vibrant social life.

Above all, my gratitude to *sana* who has endured much in this process. You give my life its purpose.

Vita

April 3, 1982	Born Kathmandu, Nepal
2005.....	B.A. Economics and Mathematics, Lafayette College, Easton, Pennsylvania
2005-2006 and 2009-10	University Distinguished Fellowship, The Ohio State University, Columbus, Ohio
2006.....	M.A. Economics, The Ohio State University
2006-2007	Ohio Agricultural Research and Development Center Director's Associateship Award, The Ohio State University, Columbus, Ohio
2007-2011	Graduate Teaching Associate, The Ohio State University, Columbus, Ohio

Fields of Study

Major Field: Agricultural, Environmental and Development Economics

Minor Field: Development Economics

Table of Contents

Abstract	ii
Acknowledgments.....	v
Vita.....	vi
Table of Contents	vii
List of Tables	ix
List of Figures	xi
Chapter 1: Introduction	1
1.1 Methodology and Experiment Design.....	5
1.2 Contribution to the Literature.....	6
1.3 Organization of the Dissertation	7
Chapter 2: Literature Review	9
2.1 Education Production Functions and Inputs-Based Policies.....	9
2.2 Incentives-Based Policies.....	17
2.3 Returns to Education	30
Chapter 3: Methodology and Experiment Design	34
3.1 Methodology	35

3.2	Experimental Design and Program Structure.....	37
Chapter 4: Theoretical Model and Empirical Results.....		44
4.1	Theoretical Model	44
4.2	Data Description.....	46
4.3	Impact of the Incentives on Academic Outcomes.....	54
4.4	Impact of the Incentives on Passing Rate.....	68
4.5	Households' Response to the Incentives.....	71
4.6	Impact of Incentives on Intrinsic Motivation.....	75
Chapter 5: Discussion and Conclusion		77
5.1	Cost Effectiveness of Cash Incentives	77
5.2	Discussion	78
5.3	Conclusion.....	81
References.....		83
Appendix A: Tables		92

List of Tables

Table 1: Core and Peripheral Subjects	38
Table 2: Number of Schools Providing Instruction on Peripheral Subjects	41
Table 3: Timeline of the Incentives Program, 2009 – 2010	43
Table 4: Summary Statistics of School Variables.....	51
Table 5: Summary Statistics of Socioeconomic Variables	52
Table 6: Attrition Pattern	53
Table 7: Estimate of the Impact of Cash Incentives on Average Aggregate Scores	57
Table 8: Estimate of the Impact of Cash Incentives on Average Aggregate Scores (Three Period Panel Data)	59
Table 9: Impact of Incentives on Average Aggregate Scores Using Grade Seven Scores as Baseline	60
Table 10: Fixed Effects Estimates of the Impact of Cash Incentives on District Exams Scores (by Individual Subjects)	61
Table 11: Fixed Effects Estimates of the Impact of Cash Incentives on District Exams Scores (by Gender)	63
Table 12: Fixed Effects Estimates of the Impact of Cash Incentives on District Exams Scores (by Socioeconomic Status).....	64

Table 13: Fixed Effects Estimates of the Impact of Cash Incentives on District Exams Scores (by Baseline Ability)	66
Table 14: Fixed Effects Estimates of the Impact of Cash Incentives on Second Semester Scores (by Baseline Ability)	67
Table 15: Fixed Effects Estimates of the Impact of Cash Incentives on the Probability of Passing	69
Table 16: Fixed Effects Estimates of the Impact of Cash Incentives on the Probability of Passing (by Subgroups).....	70
Table 17: Fixed Effects Estimates of the Impact of Incentives on Academic Reinforcement at Home	73
Table 18: Fixed Effects Estimates of the Impact of Incentives on Academic Reinforcement at Home (by Subgroups)	74
Table 19: Ordered-Probit Analysis of the Impact of Cash Incentives on Intrinsic Motivation.....	76
Table 20: Estimate of the Impact of Incentives on Average Aggregate Scores (Second Semester).....	92
Table 21: Estimate of the Impact of Incentives on Average Aggregate Scores (District Exams)	93

List of Figures

Figure 1: Reward Structure.....	40
Figure 2: Histogram of Mathematics Scores	47
Figure 3: Proportion of Students Passing Each Subject.....	48

Chapter 1: Introduction

In recent years, there has been a substantial increase in the school enrollment rate in developing countries. For example, the net primary enrollment rate in Nepal increased from 57 to 72 percent between 1995/96 and 2003/04 alone (Central Bureau of Statistics 2004). Gender and regional disparities in education are narrowing at an equally encouraging rate. This has been possible because national governments, and bilateral and multilateral donor agencies have invested heavily in education, and education has also been the priority of the global community as exemplified by commitments like Education for All (EFA) and the Millennium Development Goals (MDGs).

However, despite, or perhaps because of, the substantial increase in school enrollment rates, schooling quality remains low in these focal countries. Widely used measures of quality like promotion rate, repetition rate, and dropout rate show that much remains to be done to improve the quality of schools in the least developed countries. For instance, in the academic year 2007/08 less than 55 percent of students enrolled in grade one in public schools in Nepal advanced to grade two while 16 percent dropped out of school altogether (Department of Education 2008). Direct measures of student's learning paint a similar picture. Banerjee, Cole, Duflo, and Linden (2007) report that even in a major Indian city, where school attendance is high and teacher absenteeism is low, only 19.5 percent of students enrolled in grade three show grade one equivalent

math competencies. Similarly, Glewwe (1999) finds that among grade six students in Ghana, the mean score on a simple multiple choice test is 25 percent, the score expected if the students were guessing randomly, and Lockheed and Verspoor (1991) find evidence of poor learning levels in many parts of the developing world.

It is important that students acquire cognitive skills in school because research has shown that not just the years of schooling but also the cognitive skills acquired in school are important determinants of economic and noneconomic outcomes like wage, farm earnings, personal health, children's schooling, health, nutrition, and women's fertility (Glewwe 2002, Oreopoulos and Salvanes 2011). This result has been substantiated from research conducted in a number of countries including the United States (Blackburn and Neumark 1993, Murnane, Willett and Levy 1995), Ghana (Glewwe 1999), Pakistan (Alderman, Behrman, Ross and Sabot 1996), and Kenya and Tanzania (Boissiere, Knight and Sabot 1985).

The current study is motivated by these twin observations – that students in developing countries do not acquire necessary cognitive skills in school and that cognitive skills are important in determining a host of economic and noneconomic outcomes. Therefore, to lift people out of poverty, it is imperative that efficient and effective means to improve the quality of schooling be identified for the developing world.

Over the years, economists and policy makers have tried a number of approaches to improve the educational outcomes of students. Although they are not necessarily mutually exclusive, such approaches can be classified into two categories: “inputs-based

policies” and “incentives-based policies.” Inputs-based policies are policies that seek to improve school quality by investing additional resources in schools. An assumption that is often implicit in these policies is that greater funding is a prerequisite to and a cause of improved quality (Hanushek 2003). Theoretical justification for inputs-based policies is derived from the application of neoclassical production function to education. This approach treats schools as firms that produce educational outcomes by using a mix of inputs in the most efficient manner (Pritchett and Filmer 1999). While a school cannot function without a minimum level of resources like physical infrastructure, qualified and well-trained teachers, and availability of instructional materials, research has shown that beyond a certain level, additional resources do not necessarily translate to improvements in school quality (Hanushek 1986, Hanushek 1997).

In response to the inadequacy of inputs to explain school quality, researchers have turned to the role of incentives in “education production.” This approach extends the agency problems inherent in many economic transactions to education. From this point of view, parents or taxpayers are principals who hire teachers, the agents, to perform a task. If teachers’ preferences diverge from that of the principals’, the agency problem arises. Therefore, the proponents of incentives-based policies claim that the exclusive focus on inputs is inherently flawed and the correct approach to improve educational outcomes is to align the teachers’ incentives with that of the students, the parents, and society. Such reform initiatives may include “...decentralization of budget authority, to strengthening links between teacher pay and performance, to vouchers and school choice” (Kremer 2003).

In recent years, economists and policy makers have experimented with financial incentives to improve students' academic performance which include providing incentives to schools, teachers, parents, or students. An example of financial incentives to schools is the voucher program where vouchers that subsidize the cost of attending a school are provided to families. Proponents of the voucher program argue that because of the vouchers, families can “shop” for the best schools, thus increasing competition among schools and improving their quality. Studies show that vouchers can improve students' outcomes like test scores, high school graduation rate, and college enrollment rate (Angrist, Bettinger, Bloom, King and Kremer 2002, Angrist, Bettinger and Kremer 2006, Rouse 1998).

Recently, there has also been a renewed focus on the relevance of teachers' incentives in educational outcomes. Some argue that it is important to make teachers accountable for their students' performance by making the teachers' pay contingent on their students' performance. In the US, teachers' salary, bonus, promotion, and tenure are increasingly tied to their “value-added” ratings, the progress made by their students during the year. Researchers have also investigated the impact of teacher incentives on students' test scores in developing countries (Glewwe, Ilias and Kremer 2010, Muralidharan and Sundararaman 2009).

The most prominent example of financial incentives to parents is the conditional cash transfer (CCT) programs that have been implemented in a number of developing countries. CCT programs provide cash to poor households, usually on a monthly basis, conditional on their children attending school regularly and receiving regular health

checkups. The programs have had varying levels of success and they have become a staple of international economic development policy (Rawlings and Rubio 2005).

Finally, one class of intervention provides financial incentives directly to students based on their academic performance. After all, student motivation and effort are perhaps the most important determinants of academic achievement. A number of recent studies have evaluated the impact the financial incentives have on students' academic outcomes when the incentives are provided to the students themselves (Angrist, Lang and Oreopoulos 2009, Angrist and Lavy 2009, Bettinger 2010, Fryer 2010, Kremer, Miguel and Thornton 2009, Leuven, Oosterbeek and van der Klaauw 2010). This study falls under this class of financial incentives.

1.1 Methodology and Experiment Design

The question of causality has been a central issue in econometrics. Causal inference questions like “what is the effect of financial incentives on academic outcomes” requires answering a counterfactual question like “what would have been the outcome of the students who received financial incentives in absence of the incentives” (Duflo, Glennerster and Kremer 2007). This requires a comparison group that is similar in all dimensions to the treatment group except one – the treatment status itself. However, when a program is not allocated randomly, individuals who participate in the program are likely to be systematically different from those who do not. Random allocation of students into treatment and control groups ensures in expectation that all the observed, unobserved, and unobservable characteristics of schools, teachers, parents, and students

are orthogonal to the treatment status. Thus randomization can overcome the “selection bias” inherent in retrospective studies.

My study is a randomized field experiment conducted in public schools in the Lalitpur district in Nepal. From a pool of 33 schools, 11 were randomly placed into the treatment group while the remaining 22 constituted the control group. Grade eight students in the treatment schools were offered cash rewards that increased monotonically with their scores in each of the two semester exams and the end-of-the-year district level exam. To encourage students to pass, those who passed all subjects received reward at twice the rate per score compared to those who failed one or more subjects. Furthermore, those who failed one or more subjects in the earlier semesters could earn back the “withheld” amount if they passed all subjects in the later semesters. This stacked incentives structure was adopted to encourage those who failed in the earlier semesters to pass all subjects in the later semesters. Chapter three discusses the methodology and the experiment design in more detail.

1.2 Contribution to the Literature

This dissertation makes a number of contributions to the literature on the impact of short-term financial incentives on the academic outcomes of students. First, it measures the impact of piece-wise linear incentives. Most of the incentives in education are structured as a tournament where students have to score in the top percentiles or cross a pre-specified threshold to be eligible for the rewards. Such incentive structures may not affect the effort and the performance of the most able and the 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. Piece-rate payment schemes can be a simple alternative to tournaments where students are paid in proportion to their grades or test scores. This structure is likely to motivate all students to increase their effort at the margin.

The second contribution of this dissertation is to increase our understanding of how households respond to the incentives. Most of the studies in the literature measure the impact of financial incentives on students' outcomes like test scores, school attendance rates, graduation rates, and re-enrollment rates, while only a few study household response to the incentives (Kremer, Miguel and Thornton 2009). If home inputs and student effort are complements in education production, incentives to students might be a cost-effective means to improve academic outcomes. On the other hand, if home inputs and student effort are substitutes, the net impact of financial incentives may be small.

Finally, some are concerned that financial incentives may have a negative impact on students' intrinsic motivation to learn, thus weakening the net impact of the incentives. I analyze the impact of the incentives on students' intrinsic motivation by eliciting their response directly using a survey.

1.3 Organization of the Dissertation

The next chapter summarizes the literature on “education production” and the debates surrounding inputs-based policies vs. incentives-based policies. It also reviews

studies from two other threads of literature that are relevant to the current discussion.

The first is the literature on agency or contract theory that analyzes various compensation mechanisms that can be used to optimally modify agent behavior in favor of a principal's interest. The second is the literature on returns to human capital and the impact of human capital on economic and noneconomic outcomes.

Chapter three describes the methodology and experimental design in detail.

Chapter four sketches a simple theoretical model of education production and its interaction with financial incentives and intrinsic motivation. It then presents the econometric results of the impact of incentives on academic outcomes of students, on households' response, and on students' intrinsic motivation. Finally, chapter five concludes with a discussion of the cost effectiveness of the incentives scheme, the main results, its policy implications, and possible future directions of the study.

Chapter 2: Literature Review

Economists, educators, and policymakers have long sought to understand the determinants of academic achievement and the means to improve the academic performance of students. Although not necessarily mutually exclusive, such efforts can be classified into two broad categories: inputs-based policies and incentives-based policies. This chapter summarizes the relevant studies and situates the current study in the literature. It also discusses the conceptual underpinnings of inputs- and incentives-based policies. Finally, the chapter reviews studies from the “returns to education” literature to justify an investment in improving academic outcomes of students.

2.1 Education Production Functions and Inputs-Based Policies

One of the earliest studies to evaluate the determinants of student achievement is the report on educational equality in the United States commissioned by Department of Health, Education, and Welfare (Coleman 1966). The report, based on extensive surveys of students and teachers across America, showed that variations in school characteristics explained little variation in students’ educational achievement. On the other hand, the socioeconomic background of a student and her peers were strongly related to the student’s educational attainment. The controversial findings of the report led to a spate

of research in the field (Heyneman 1976) and the debate on the relative importance of home vs. school variables in educational outcomes is alive to this day.

Research on the topic picked up pace in developing countries in the 1970s as these countries increased their investment in public schooling. One of the earliest available studies on the determinants of academic achievement is a review of studies from 12 less-developed countries (Heyneman, Farrell and Sepulveda 1981). The authors find that among a number of variables, including teacher training and school size, only the availability of textbooks in classroom is consistently positively correlated to academic achievement. Schiefelbein and Simmons (1981) conduct similar review of 26 production function studies conducted in more than 20 developing countries. The studies treat cognitive achievement as an output that is produced using labor and capital inputs like teacher education and textbooks. The researchers isolate 10 variables that are consistently correlated with educational outcomes in the expected direction, including textbook availability and the use of homework assignments. Surprisingly, variables like educational outcomes, school facilities, class size, and expenditure per pupil are not consistently related to academic outcomes in the expected direction.

Studies conducted in the 1980s continued to show that inputs do not consistently improve educational outcomes. Harbison and Hanushek (1992) find ambiguous results in their evaluation of 96 studies on the impact of resources on educational outcomes. Inputs such as teachers' education, experience, and salary, expenditure per pupil, and teacher-pupil ratio did not always have a predictable impact on student performance. Velez, Schiefelbein, and Valenzuela (1993), Fuller (1987), and Fuller and Clarke (1994) reach a

similar conclusion in their evaluation of studies on the impact of school inputs on academic outcomes.

At around the same time, a closely related line of inquiry was being pioneered by Erik Hanushek in the context of public schools in the US. Hanushek contended that differences in school quality were not related to variations in expenditure per pupil and other variables commonly thought to be inputs to educational outcomes like class size, teacher experience, and teacher qualification (Hanushek 1981, Hanushek 1986). In a series of subsequent studies, Hanushek asserted that “there is not a strong or consistent relationship between student performance and school resources” (Hanushek 1986, Hanushek 2003, Hanushek 1995, Hanushek 1997).

The persistent inconsistencies could have been an artifact of the studies under consideration. On at least one occasion Hanushek acknowledges that,

“...the standards of data collection and analysis are so variable that the results from this work are subject to considerable uncertainty. Much of the analysis of input-output relationships for developing countries is not published in standard academic journals, and thus it lacks that basic level of quality control. Even more important, the data for many of these studies do not come from regular collection schemes, are difficult to check for quality, and miss key elements of the educational process. Therefore, even if the analytical approaches are state of the art, many questions remain” (Hanushek 1995).

Kreuger (2003) raises other issues with Hanushek’s studies. He contends that Hanushek’s analysis provides equal weight to every estimate, thus providing proportionately more weight to studies where multiple estimates are calculated from small subsamples than to studies that utilize the full sample but have fewer estimates. This methodology is likely to yield weaker results because estimates drawn from smaller

samples are not very consistent. Krueger also points out that the econometric specifications used in the bulk of the estimates are faulty, hence unable to capture a class-size effect even if there was one. The author finds a more consistent relationship between class size and student achievement after correcting for these deficiencies.

Another reason for the apparently inconsistent results could be inadequate consideration of the level of inputs in “returns to input” analyses. If inputs have diminishing returns in the production of cognitive achievement, it is not very informative to ask what the marginal effect of an additional unit of input is unless the level at which the input is applied is made explicit. As discussed by Pritchett and Filmer (1999), a failure to reject zero marginal effect of additional resources may simply reflect the level at which the resources are applied. It is not always clear if “returns to input” studies take into account the level of inputs.

Indirect evidence of the diminishing marginal returns of educational inputs comes from the literature on the sociology of education. In response to the findings of the Coleman Report (1966), Heyneman (1976) investigated if the conclusions held for less-developed societies, i.e., if home variables had more influence on students’ academic outcomes than school variables in developing countries. He found that school variables like availability of school infrastructure, teachers’ years of schooling, and teacher experience explained far more variation in academic outcomes in Uganda than in industrialized countries like the United States. This suggests that at low levels of application, school inputs are relatively more important in determining academic outcomes.

Perhaps the most important reason for the lack of consistency in returns to input is the methodology employed by the studies. Historically, economists have relied on a neoclassical production function approach to study educational outcomes. In the context of the firm, a production function describes the maximum level of output that can be produced with different combinations of inputs. Firms are assumed to be profit maximizers, thus it is assumed that they produce a given level of output with the most efficient mix of inputs. Firms are also assumed to be sensitive to changes in technology, input costs, the price of outputs, etc. so that any change in any of the parameters leads to an optimal reallocation of the inputs and the production process. In the context of education, output is usually measured as scores in achievement tests, whereas inputs consist of a vector of school and teacher characteristics, pedagogical practices, child and peer attributes, and household and community level variables (Hanushek 1986).

It is not clear, however, if a “production function” is the appropriate modeling tool to apply to educational outcomes. At one level, the processes of education production may be fundamentally different from the processes of material production so that the production-function approach may not be applicable at all. For example, production function approach assumes that all inputs can be varied without restriction. However, inputs in educational production include variables like socioeconomic status and peer attributes that cannot be varied easily.

At another level, the assumption that schools operate at a maximum efficiency by choosing the levels of inputs by equating their marginal products to their price is also not tenable. There is evidence of severe distortion in the budget allocation in the education

sector for political economy reasons (Kremer 2003, Kremer, Moulin and Namunyu 2003). For example, Pritchett and Filmer (1999) show that in Brazil, the marginal product per dollar of inputs like textbooks and writing materials that are not directly valued by teachers are at least an order of magnitude higher than that of inputs valued by teachers like teachers' salary.

Even if a production-function approach is a reasonable method to study educational outcomes, numerous practical difficulties are encountered in estimating the marginal impact of inputs. The production function may be unknown to schools, researchers, and decision-makers and it may have to be inferred from often imperfect and inaccurate data. A number of estimation problems are also encountered when using retrospective data for empirical estimation of the production function that include specification bias, unobserved and omitted variable bias, measurement error, attrition bias, endogeneity of school quality, and selection bias (Glewwe 2002, Glewwe and Kremer 2006). These problems seem so formidable that some scholars are skeptical that they could ever be overcome (Hanushek 1995).

While some scholars focus on improving the theoretical models and collecting better data to overcome the problems encountered in retrospective studies, others argue for well-executed randomized experiments (Glewwe and Kremer 2006, Kremer 2003). Random assignment of educational units (students, classrooms, or entire schools) to program and control groups ensures that the only systematic difference between the two groups is the treatment status. Any difference in the outcome between the groups can then be attributed to the treatment with reasonable confidence. As randomized

evaluations have become more common in economics, fresh insight has been available on the “returns to input” debate.

One of the largest and well-studied experiments in the United States is the Student/Teacher Achievement Ratio (STAR) class-size experiment in Tennessee. Under the STAR program, kindergarten students were randomly assigned to small or normal classes where small classes had 13-17 students per teacher whereas normal classes had 22-25 students in a class (Krueger 1999). After the initial assignment, students remained in the same class type for four years. Krueger finds that at the end of the first year of the program, the average score on a standardized test in small classes was four percentile points higher than in normal classes and this advantage in test scores increased by 1 percentile point in every subsequent year. He also finds that the lower class size had a relatively higher effect on minority and poor students.

In the context of developing countries, one of the earliest experimental studies on the impact of inputs is the Nicaraguan program where textbook and radio-based instruction were provided to a set of randomly chosen students (Jamison, Galda, Searle and Heyneman 1981). The authors find that availability of textbooks increased students’ post-test scores by 0.33 standard deviations and the radio-based instructional program increased it by 1.5 standard deviations. These are remarkable gains, particularly when compared to the gains observed in more recent experiments. Similarly, Heyneman et al. (1984) evaluated the impact of a nationwide distribution of textbooks in the Philippines. Randomly selected students from the year before the textbook distribution formed the control group while randomly selected students after the distribution of textbooks formed

the treatment group. The authors find that textbooks had a strong impact on test scores in science, mathematics, and Filipino, the subjects for which textbooks were distributed, ranging from 0.32 standard deviations in math to 0.51 standard deviations in science.

More recently, Banerjee, Cole, Duflo, and Linden (2007) evaluate a remedial education program in India where young women from the community who had completed secondary schooling were assigned to schools as remedial tutors. The weakest students in these schools were pulled out of regular classrooms for two hours a day to work in the remedial program. The program raised average test scores of all students in treatment schools by 0.14 standard deviations in the first year and 0.28 standard deviations in the second year. Moreover, the study finds that the weaker students gained the most from the program. Another intervention that allowed students in the fourth grade to spend two hours every week playing games and solving math problems on the computer increased the participants' average test scores by 0.34 standard deviations in the first year and 0.47 standard deviations in the second year. Taken together, these studies suggest that inputs can have a substantial impact on school outcomes.

However, other experimental studies find that additional school inputs do not necessarily improve students' test scores. Glewwe, Kremer and Moulin (2009) evaluate a program in Kenya where a Dutch non-governmental organization (NGO) provided textbooks to randomly chosen primary schools. The authors find that the intervention raised test scores of only the better-prepared students. In another randomized evaluation, Glewwe, Kremer, Moulin, and Zitzewitz (2004) show that flip charts have no effect on

test scores in rural Kenya. This study is particularly illuminating because it reverses earlier findings from retrospective studies of positive impact of flip charts on test scores.

2.2 Incentives-Based Policies

Conflicting results, and apparent ineffectiveness, of input-based policies have led some scholars to advocate for a new approach altogether. They argue that the exclusive focus on inputs-based policies is inherently flawed because it ignores the role of incentives faced by schools and teachers. Pay, promotion, and job retention of teachers are rarely tied to their performance, so it is not surprising that additional resources do not translate into improved performance. Therefore, the scholars insist that resources alone cannot be counted upon to improve academic standards unless there is a fundamental restructuring and reform of school incentives (Hanushek 2003, Hanushek 1995). Such institutional reforms may include providing vouchers to parents, thus giving schools an incentive to supply high quality services, merit pay system that rewards teachers based on students' learning, increased competition between public and charter schools, and, especially in the context of developing countries, providing local authorities more autonomy to make decision regarding teacher hiring, budget allocation, etc. (Kremer 2003).

The best case study for the need of an appropriate institutional framework to complement additional resources comes from a study conducted in western Kenya. Duflo, Dupas, and Kremer (2007) report from a randomized field experiment where a set of randomly selected schools were provided funds to hire an extra teacher on a short-term

contract. The intervention effectively halved the average pupil-teacher ratio, reducing it from 80 to 46. The authors find that students who were taught by contract teachers scored significantly higher in test scores relative to their peers in small classes that were taught by civil service teachers and also relative to students in large classes. Contract teachers were more likely to be present in school and be teaching in class compared to civil service teachers. This result can be best understood in light of the stronger incentives faced by contract teachers who are hired for a short term on a temporary basis. It also highlights the need to reform institutional structures before committing additional resources in education.

A number of studies have begun to shed more light into the impact of incentives and institutional reform on educational outcomes. One class of such studies examines the impact of school vouchers on outcomes like test scores, high school graduation rate, college enrollment, and labor market earnings. Rouse (1998) uses a quasi-randomized design to evaluate the Milwaukee Parental Choice Program where low-income families in Milwaukee were provided vouchers to enroll their children in private schools. She finds that students who attended private schools achieved significantly higher scores in math compared to those who applied for voucher but did not receive it. However, she finds no such impact on reading scores.

Similarly, Howell et al. (2002) report the impact of voucher programs on students' test scores in New York City, Washington, D.C., and Dayton, Ohio. The authors find that private schools improved the test scores of African-American students but not of students from other ethnic groups. Studies that review evidence from the

United States and other developed countries conclude that any gains in student achievement from a large voucher program are likely to be small and that there is little evidence of an increase in competition among public schools, one of the main justifications for providing school vouchers (Ladd 2002, Rouse and Barrow 2009). The authors also caution that any observed impact of school vouchers may vanish in a scaled-up version due to general equilibrium effects.

The disparity between public and private schools is sharper in developing countries, and school voucher programs have become popular in these countries as a tool to foster competition between public and private schools. One of the largest school voucher programs was carried out in Colombia in the 1990s where 125,000 students were provided vouchers worth slightly more than half the cost of private secondary schools (Angrist, Bettinger, Bloom, King and Kremer 2002). Many vouchers were distributed by lottery so the program is comparable to a randomized trial. The authors find that three years after the assignment, lottery winners had more years of schooling, were more likely to have completed the eighth grade, and had higher scores on achievement tests relative to the losers. In a follow up study, Angrist, Bettinger, and Kremer (2006) provide one of the few studies of the long term-impact of vouchers. They show that winning the voucher led to an increase in high school graduation rate and improvement in learning levels seven years after the lottery. These results stand in contrast to the findings in developed countries.

Another class of financial incentives includes the pay-for-performance programs under which teachers' compensation is partially based on student outcomes.

Muralidharan and Sundararaman (2009) report from an experimental study in India where teachers' bonuses were tied to their students' test scores. The authors find that the pay-for-performance program had a significant impact on math and language test scores at the end of two years of the program. They also find that students in program schools performed better in both mechanical and conceptual components of the test and in subjects that were not incentivized. The program did not decrease teacher absenteeism but, conditional on being present, it did prompt teachers to exert more effort.

Lavy (2002) presents results from two programs in Israel that provided monetary incentives to teachers. The first program, conducted in 1996 and 1997, awarded monetary incentives based on improvement in students' annual achievement to the top third performers (out of 62 schools). The sum that schools received was distributed to teachers and staff as merit pay and to improve general working conditions of teachers. Lavy finds that monetary incentives had some effect on religious schools in the first year of the program and significant impact on both religious and secular schools in the second year. The program caused an increase in average test scores and in the number of credits with which students graduated high school. It also led to a decrease in dropout rates from middle to high school and an increase in the proportion of students who obtained a matriculation certificate. The author does not report teacher behavior so it is not clear if the effect is due to fundamental improvements in learning or due to "gaming" by teachers.

The second program is a pay-for-performance scheme based on individual performance. English, Hebrew, Arabic, and mathematics teachers who taught grades ten

to twelve were provided a bonus based on the mean class performance and the passing rate. Lavy (2009) finds that teachers' incentives led to an increase in test taking rates, conditional pass rates, and average test scores of students. He also finds evidence for teachers tracking their student by ability, adding time of instruction during the year and before national exams, and paying more attention to weak students. There is no evidence for manipulation of test scores by teachers.

These studies suggest that performance-based pay schemes for teachers can motivate teachers and improve student outcomes. However, when an agent's task is multidimensional and a contract cannot be drawn to specify all relevant aspects of the task, the agent can reallocate his effort away from activities that are not compensated (Holmstrom and Milgrom 1991). Pay-for-performance contracts to teachers can give rise to dysfunctional behavioral responses like artificial grade inflation, "teaching to the test," or outright cheating or grade manipulation by teachers. Indeed, a randomized study from Kenya finds that in program schools where teachers' compensation was based on test scores, there was an increase in test preparation sessions and students had higher scores, but there was no evidence of an increase in teacher attendance, homework assignment, or change in pedagogy (Glewwe, Ilias and Kremer 2010). Moreover, the gains in scores vanished after the incentive program was phased out, which further suggests that teachers did little to fundamentally improve the long-run learning by children. Therefore, caution should be exercised in implementing pay-for-performance programs. In light of the conflicting results, Lavy (2007) recommends that any pay-for-performance scheme

should be based on outcomes broad enough to prevent potential “gaming” by teachers but narrow enough to be easy to measure, monitor, and have little random variation.

Two of the studies discussed above also compare the impact of the incentives program to that of providing additional resources to schools without an active incentive component. Muralidharan and Sundararaman (2009) compare the impact of incentives to that of an extra contract teacher and a block grant provided to schools to procure non-teacher inputs used by students. The authors find that “incentive schools” significantly outperformed “resource schools” at the end of two years of the program. Lavy (2002) compares the impact of teacher incentives to a program where schools were endowed with conventional resources like additional teaching time and on-the-job teacher training. The author finds that although both the programs had a significant impact on student performance, teacher incentives were more cost effective. These results show that incentives-based policies can be more cost effective when implemented properly.

Many developing countries have already experimented with programs aimed at increasing school enrollment and attendance among the poor by providing financial incentives to parents. The programs, known as conditional cash transfer (CCT) programs, provide monthly cash compensation to parents conditional on their children attending schools and receiving health checkups on a regular basis. The programs have attained their goals with varying levels of success, and they have become a staple of international economic development policy (Rawlings and Rubio 2005).

However, a number of studies have shown that CCT programs have had little, qualified, or no impact on school *performance* of students. Behrman, Sengupta, and

Todd (2000) evaluate the impact of PROGRESA, the Mexican conditional cash transfer program, on achievement scores. They find that exposure to PROGRESA does not have a significantly positive effect on achievement of students in the short run. In a follow up study, Behrman, Parker, and Todd (2005) find that more than five years after the first round of benefits was distributed, youth in the treatment group have 0.2 more years of schooling than youth in control group but there is no difference in achievement levels. Similarly, studies by Duryea and Morrison (2004) of the Supermonos program in Costa Rica and by Ponce and Bedi (2010) of a cash transfer program in Ecuador reveal that conditional cash transfer programs do not have a significant impact on academic achievement. A report by UNESCO Institute for Statistics also concludes that “there is very limited support for the conclusion that CCTs are effective educational instruments, in particular with regards to their ability to increase learning” (Reimers, DeShano da Silva and Trevino 2006).

The general lack of evidence for CCT programs to improve educational outcomes of students is perhaps simply due to the objective of the program. Under most CCT programs, cash payments are contingent on students’ school attendance and not their outcome. Other programs that aim to increase school attendance have similarly had no impact on actual learning of students. An example is the study by Miguel and Kremer (2004) where a deworming program in Kenya reduced absenteeism in treatment and untreated schools dramatically but had no impact on academic test scores. If we observe positive results from students’ incentives program, it perhaps gives some scope for an amendment of conditional cash transfer programs. To the extent that school attainment is

correlated to academic achievement, as some studies suggest, making cash contingent on students' outcomes may attain the goals of increasing schooling and learning levels simultaneously.

A more direct approach to improve students' outcomes would be to provide incentives to students themselves. Student effort is arguably the most important input in education production, and 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. Providing incentives to students as opposed to teachers or schools also divorces the agency problem discussed previously.

In the context of the United States, a number of states have resorted to providing merit scholarships to college students in an attempt to boost college enrollment rates and student achievement. Georgia's Helping Outstanding Pupils Educationally (HOPE) and New Mexico's NM Success scholarship programs are two examples of such initiatives. The HOPE program awards full tuition cost, other fees and book expenses to students who graduate high school with an average of B and 3.0 GPA in college. Cornwell, Mustard, and Sridhar (2006) find that students took fewer courses and withdrew from courses with higher frequency under the HOPE program and this behavior was concentrated among students who were at the margin of eligibility for the scholarship. Under the NM Success program, high school graduates who enroll in college, take a full course load in the first semester, and maintain a GPA of 2.5 are eligible for scholarship in

their second semester. Binder, Ganderton, and Hutchens (2002) find that the program increased the proportion of white students or student from high income families while the proportion of students from low income declined. The program also increased the attrition rate of first year students, probably because it attracted ill-prepared students to college.

Angrist, Lang, and Oreopoulos (2009) evaluate a program aimed at improving student achievement and retention at a large Canadian university. First-year students were randomly assigned to one of three treatment groups or a control group. In the first treatment group, students received support services, including mentoring by senior students. Students in the second treatment group received a substantial cash reward for meeting a target GPA while students in the third treatment group received both support services and cash incentive. The authors find that the incentives did not have an impact on the overall sample, but it had a significant effect on females, with the effect concentrated on the subgroup that intended to fulfill the minimum eligibility requirement for the scholarship.

Angrist and Lavy (2009) analyze a program from Israel that provided substantial cash incentives to low achieving high school students if they collected enough credit to obtain high school certification. The authors find that the achievement awards program had no effect on the aggregate matriculation rates, but it did have a positive impact on the certification rates of girls. Moreover, this increase was driven by the girls who were at the margin of completing the certification requirements.

Kremer, Miguel, and Thornton (2009) evaluate the impact of a merit scholarship program in Kenya where girls who ranked in top 15% of class were awarded a grant to cover tuition fees and school supplies. The authors find that the scholarship program had a large positive impact on the test scores of girls. Even girls with low pretest scores showed significant improvement, and test scores gains was also observed among boys who were not eligible for scholarship at all.

The works that are most similar to mine include Bettinger (2010), Fryer (2010), and Leuven, Oosterbeek, and van der Klaauw (2010). Bettinger (2010) presents evidence from a pay-for-performance program among students in grade three through six in the Coshocton school district in Ohio. The students in a randomly chosen subset of schools received cash payments if they pass district- and state-mandated standardized exams in four areas: Math, Reading, Science, and Social Science. The author finds that incentives had a positive impact on math scores but not on other subjects.

Leuven, Oosterbeek, and van der Klaauw (2010) report the results of an experiment in the Netherlands where economics and business students were awarded cash incentives for fulfilling all the first-year requirements before starting the second year. They find no significant difference in the passing rate and the number of credit points collected across high reward, low reward, or no reward groups. Financial incentives have a positive impact on the achievement of the high-ability group while they have a negative impact on the low-ability students.

Finally, Fryer (2010) conducts multiple studies in urban schools districts in the United States and finds weak and inconsistent impacts of the incentives. He 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. The author hypothesizes that this could be because of students' lack of understanding of the educational production function.

Theoretical justifications for incentives-based policies are derived from the literature on agency theory. The central problem of principal-agent theory is how the principal can best motivate the agent to perform as the principal would prefer. This problem arises because of a number of sources of friction in the contractual agreement: risk aversion of the agent, asymmetry of pre-contractual beliefs on the probabilities of external shocks that affect outcomes, limits on commitment liabilities, and unobservability and unverifiability of the agent's effort or final outcomes (Sappington 1991).

To reduce the premium that the principal has to pay to the agent for his effort, the principal may monitor agent's actions, or he may seek to design contracts that elicit private information the agent has about his working environment. This is particularly salient when there are multiple agents because information relevant to one agent can be solicited from the actions and outcomes of other agents.

Tournament contracts fall in this class of contractual mechanisms. A rank-order tournament contract is a special case of a tournament that provides compensation based on relative performance alone, and not on absolute levels of outcome. In a seminal work, Lazear and Rosen (1981) showed that under the assumption that two contestants have

symmetric information and are equally able, both contestants exert equal effort in a Nash equilibrium. Since then, circumstances under which payments based on relative outcomes are superior to those based on individual outcomes have been investigated by economists (Green and Stokey 1983, Holmstrom 1982, Nalebuff and Stiglitz 1983).

When each agent's output does not depend on random shocks or the shocks are uncorrelated across agents, it is efficient to design a contract where agents are held responsible for their own output. However, it is more often the case that agents face correlated shocks. In such a scenario, a payment scheme that is dependent on outputs of all agents is more efficient than individualistic rewards because the vector of outputs is informative of the common productivity shock experienced by agents. Holmstrom (1982) also shows that tournaments are "consistent" in the sense that as the number of agents grows large, all common uncertainty can be detected and can be eliminated from agents' responsibility. The point is further reinforced by Green and Stokey (1983) who show that in absence of common shocks, optimal independent contracts dominate optimal tournament contract whereas if agents face common productivity shocks and the variance of the shock is sufficiently high, optimal tournaments dominate optimal independent contracts. Furthermore, for any given variance of the common shock, optimal tournaments dominate optimal independent contracts for a sufficiently large number of agents.

These theoretical predictions, however, are based on the assumption that agents are of equal ability. They do not hold when one agent is "disadvantaged," i.e., when the marginal cost of effort for one is higher than that for the other (Okeeffe, Viscusi and

Zeckhauser 1984). In such a case, the disadvantaged agent exerts less effort than the advantaged agent in the Nash equilibrium and the advantaged agent exerts less effort than when the abilities were symmetric (Bull, Schotter and Weigelt 1987).

Agency theory has identified that continuous payment schemes offer better incentives than tournaments when agents are of heterogeneous ability (Lazear 2000a, Lazear 2000b). Pay-for-performance or piece-rate incentive systems reward agents on the basis of some measure of output and the pay is not dependent on other agents' actions or outcomes. The current study rewards students on a piece-rate basis where a student's payment depends on his or her absolute performance in a piece-wise linear fashion. I recognize that this "continuous" incentive scheme has not been tried in education in the context of developing countries. My design also has a weak version of fixed performance standards that leads to the piece-wise linear structure mentioned above; students scoring less than the passing grade receive half the price per score that those who score more than the threshold.

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, the intrinsic joy of performing a task is a powerful driver of human action (Fehr and Falk 2002, Frey and Jegen 2001, Frey and OberholzerGee 1997).

A rich literature in psychology examines the relationship between extrinsic rewards and intrinsic motivation. One is said to be intrinsically motivated to perform a task if there are no obvious or apparent external reasons to do so. In contrast,

extrinsically motivated behaviors are those that are controlled by promises of external rewards. In the late 1960s, psychologists began to speculate that extrinsic rewards could undermine intrinsic motivation (DeCharms 1968, Festinger, Ono, Burnham and Bamber 1967). The first and the most cited laboratory experiments on the impact of external rewards on intrinsic motivation conducted in the early 1970s showed that incentives did have a negative impact on intrinsic motivation (Deci 1971, Deci 1972a, Deci 1972b).

Later scholars have critiqued the early experiments and offered a more nuanced view of the interaction of external incentives with intrinsic motivation (Cameron and Pierce 1994, Eisenberger, Pierce and Cameron 1999). From a meta-analysis of 96 experimental studies, the authors find that rewards are detrimental to intrinsic motivation only under a limited and unrealistic set of circumstances. They conclude that, as suggested by some studies, when incentives are contingent on exhibiting competence, they do not have a harmful effect on intrinsic motivation (Arkes 1979). Given this result, it is unlikely that providing financial incentives to students conditional on grades are detrimental to their intrinsic motivation. To check for the impact of incentives on intrinsic motivation, my year-end survey consists of a section that explicitly elicits students' interest in learning.

2.3 Returns to Education

Before conducting a program intended to improve education outcomes, it is instructive to consider if such an investment is justified at all. Ever since the pioneering works of Becker (1962) and Mincer (1974), there have been numerous studies on the

rates of returns to investment in education. In a series of empirical studies in developing countries, Psacharopoulos (1981, 1985, 1994) finds a number of consistent results. First, investment in primary education has a higher rate of return than investment in secondary education. Second, the marginal return to education is decreasing with the level of economic development. Finally, the private rate of return is higher than the social rate. These results have often been used to justify investments in education in developing countries.

While the early literature took education to be simply the number of years of schooling completed, later studies have also considered returns to quality of schooling (Behrman and Birdsall 1983, Johnson and Stafford 1973). The motivation for this innovation is twofold: to see if quality of schooling is important in its own right and to correct for potential bias in estimates of rates of return to “quantity” of schooling due to the omission of quality indicators. When Behrman and Birdsall (1983) include quality of schooling, as measured by average schooling of teachers, in the traditional Mincerian equation, they find that private returns to years of schooling is only half that of the estimate obtained from traditional specifications. They also find that the social return to quality is substantially higher than the social return to quantity. Johnson and Stafford (1973) conclude that the influence of the quality of schooling on earnings potential and rates of return is much higher than the influence of the quantity of schooling, with schooling quality measured by expenditure per pupil. More recently, Card and Krueger (1992) reach similar conclusions in their evaluation of earnings data from the United States. These findings have important implications for education policy in developing

countries. If quality of schooling does not matter, then the optimal policy would be to build more schools and increase enrollment rates; if quality does matter, it would be more efficient to improve the quality of existing schools.

More pertinent to our current discussion however is the return to students' learning levels. In the context of the United States, two early studies stand in favor of the hypothesis that there are considerable returns to academic achievement. Murnane, Willett, and Levy (1995) conclude that as much as 38% of the increase in return to education between late the 1970s and mid 1980s for 24-year-old males was due to improvement in ability and Blackburn and Neumark (1993) report that the increase in returns to education in the 1980s was concentrated among those with relatively high test scores.

Boissiere, Knight, and Sabot (1985) parse the influence of cognitive achievement, innate ability, and years of education on earnings of urban wage laborers in Kenya and Tanzania. The authors find that the labor market returns to cognitive skills, measured as literacy and numeracy, are high. On the other hand, they find that the direct returns to innate ability are small and the returns to years of schooling are moderate. The authors also suggest that the main effects of innate ability and years of education on earnings operate indirectly through the development of cognitive skills as those who are more able and those who stay in school longer are likely to be more literate and numerate.

In a similar study, Glewwe and Jacoby (1994) estimate returns to cognitive skills using nationally representative data from Ghana. They find that acquired skills and not years of schooling or innate ability determine wages in the private sector. Alderman et

al. (1996) also find that cognitive achievement, and not years of schooling or innate ability, explains wage differentials in the rural labor market in Pakistan. These findings provide justification for investments to boost students' learning in school.

Chapter 3: Methodology and Experiment Design

Since at least the early 1980s, there has been a spirited debate among economists on the role of randomized experiments in economics (Leamer 1983). The debate is motivated by two observations on the nature of econometric practice. First, to draw inferences from the data, unrealistic assumptions often have to be imposed on the data. Second, any inference drawn from the data is very sensitive to the underlying assumptions, thus very unstable and fragile. As a remedy to the problem, scholars proposed that at a minimum regression analyses be subjected to sensitivity analysis (Leamer 1983, Sims 1980).

The debate gained further momentum when studies in the mid 1980s showed that estimates from an econometric evaluation differed substantially from that of a randomized trial (Lalonde 1986). As a result, applied econometricians began to use quasi-experimental designs to estimate the causal impact of various social programs (Angrist 1990, Angrist and Krueger 1991). More recently, randomized evaluations have become a popular tool in economics, especially in the field of development economics (Duflo 2006).

3.1 Methodology

Randomized controlled trials can be best understood under Rubin's Causal Model (RCM) framework (Duflo, Glennerster and Kremer 2007, Imbens and Wooldridge 2009). Let a program affect n individuals indexed by $i = 1, 2, \dots, n$. Let $Y_i(C)$ be the outcome for individual i if she does not participate in a program and let $Y_i(T)$ be the outcome for the same individual if she participates in the program. Many empirical questions in the social sciences involve answering the question $Y_i(T) - Y_i(C) = ?$, i.e., what is the impact of a program on individual i ? The difficulty with this question is immediately apparent because we do not observe both $Y_i(C)$ and $Y_i(T)$ for the same individual – she either participates in the program or she does not. However, if we have a sample of n individuals, some of whom participate in the program and others do not, we can obtain the following:

$$E[Y_i(T)|T] - E[Y_i(C)|C].$$

The expression gives the difference in average outcomes among participants and nonparticipants. As the sample size increases, the sample averages converge to the population value. However, when participation in a program is nonrandom, a simple comparison of those who participate in a program and those who do not may provide a biased estimate of the treatment effect. To demonstrate this, let us add and subtract $E[Y_i(C)|T]$, the potential outcome for program participants had they not been exposed to the program, to the expression:

$$E[Y_i(T)|T] - E[Y_i(C)|C] - E[Y_i(C)|T] + E[Y_i(C)|T]$$

After a slight manipulation, the expression can be written as:

$$\begin{aligned}
& E[Y_i(T)|T] - E[Y_i(C)|T] - E[Y_i(C)|C] - E[Y_i(C)|T] \\
& = E[Y_i(T) - Y_i(C) |T] + \{E[Y_i(C)|T] - E[Y_i(C)|C]\}
\end{aligned}$$

Thus the naïve difference in average outcomes can be written as a sum of two components. The first term $E[Y_i(T) - Y_i(C) |T]$ gives the average treatment effect on the treated (ATT). The term $E[Y_i(C)|T] - E[Y_i(C)|C]$ represents the selection bias. $E[Y_i(C)|T]$ is the potential outcome of program participants had they not participated in the program. When program participation is nonrandom, it is likely that $E[Y_i(C)|C]$, the outcome for program nonparticipants when they are not exposed to the program, is not a good approximation of $E[Y_i(C)|T]$.

Random assignment of individuals into treatment and control groups overcomes the selection bias. The mechanism ensures that the potential outcomes of the program participants is equal, in expectation, to the potential outcomes of the nonparticipants, i.e., the potential outcomes is independent of the treatment status:

$$E[Y_i(C)|T] - E[Y_i(C)|C] = 0$$

Therefore,

$$E[Y_i(T)|T] - E[Y_i(C)|C] = E[Y_i(T) - Y_i(C) |T]$$

In terms of regression analysis, this expression is equivalent to the parameter β_1 in the following regression equation:

$$Y_i = \beta_0 + \beta_1 * T_i + \epsilon_i$$

where T is the dummy variable for treatment status. Random assignment of elements into treatment and control groups ensures that the error term is uncorrelated to T .

Even when randomization splits the sample into two similar groups, the estimate of the impact of the incentives will be biased if the attrition pattern is correlated with the program. For example, those who receive a job training program are more likely to be in touch with the researchers at the end of a period than their counterparts who were unemployed but did not receive the job training. If the purpose of the study is to evaluate the effectiveness of the job training program, this will lead to an underestimate of the impact.

3.2 Experimental Design and Program Structure

The experiment was conducted among grade eight students in 33 public schools in the 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, I visited each of the 33 sample schools and explained the incentive program to school 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.¹ One lower secondary school originally assigned to the control group had to be dropped from the sample because no student enrolled in grade eight for the academic year.²

There are nine officially mandated courses for students in grade eight. The core courses include Nepali, English, Math, Science, and Social Studies and the peripheral 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. This information is summarized in table 1.

Core Subjects	Full Score	Passing Score	Peripheral Subjects	Full Score	Passing Score
Nepali	100	32	Health and Physical Education	50	16
English	100	32	Population and Environment	50	16
Math	100	32	Moral Education	50	16
Science	100	32	Vocational Studies	50	16
Social Studies	100	32			

Table 1: Core and Peripheral Subjects

¹ 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 this study is 1:2.

² Zero enrollment for the school was not due to its control status. The program was announced one month into the academic year 2009/10, and no students were enrolled in the school even prior to the date.

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 secured an average of 40 percent ($280 = 40\%$ of 700). Under the scheme, if a student passed all courses, she received 5 rupees (~ 7 cents) per percent; if she failed any course, the price per percent was 2.5 rupees (~ 3.5 cents). For instance, if the student passed all courses while securing 40 percent, she received 200 rupees ($Rs. 5.00 \times 40 = Rs. 200$) but if she failed one or more courses, she only received 100 rupees ($Rs. 2.50 \times 40 = Rs. 100$). 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 but passed all classed, 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). Figure 1 provides a graphical illustration of the rewards structure.

Figure 1: Reward Structure

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). Table 2 summarizes this information. To keep the comparison valid, scores of only the nine officially mandated subjects are compared and the average aggregate score is computed by averaging the scores over an appropriate total.

Subject	Treatment Schools	Control Schools	Total Schools
Health and Physical Education	11	21	32
Population and Environment	11	21	32
Moral Education	5	10	15
Vocational Studies	9	16	25

Table 2: Number of Schools Providing Instruction on Peripheral Subjects

The average amount of incentives received by students over the course of the program was Rs. 590 (~\$7.50). This sum is not significant relative to household income or expenditure in the semi urban areas of Kathmandu.³ However, the amount is a significant portion of annual educational expenditure as school fees are about 600 rupees (~\$6.50 – 8.00) per year. Anecdotal evidence also suggests that students received a daily stipend of about Rs. 10 for afternoon snacks. Therefore, an average student earned about two months' worth of snack stipend over the course of the year. When seen from this perspective, the incentives are not as insignificant as we might think.

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.⁴ I explained the incentive scheme to students in the classroom, and worked through a number of examples on the blackboard to cover all the possible payment scenarios. In

³ Nepal Living Standard Survey 2003/04 (NLSS II) found the 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.

⁴ Our data shows no transfers of students between schools.

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, I 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.⁵ 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.

As shown in table 3, 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). 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 schools

⁵ However, students were eager to compare their rewards with each other. Therefore, to a large extent, students knew their performance relative to their peers.

were also asked to answer questions intended to measure their understanding of the incentive scheme.⁶

Date	Activity
<u>2009</u>	
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
<u>2010</u>	
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 3: Timeline of the Incentives Program, 2009 – 2010

⁶ 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.

Chapter 4: Theoretical Model and Empirical Results

4.1 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 the agents' inference about the nature of the task. 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 a student, and

e = student's effort (e.g. school attendance, study time, etc.).

I assume that f is strictly concave in e i.e. $f_e > 0$ and $f_{ee} < 0$. Private monetary rewards to learning are given by wy where w can be interpreted as the present discounted

value of lifetime after-tax earnings (Bishop 2006). Students also derive satisfaction from learning itself, regardless of the monetary rewards. Nonmonetary rewards to learning are indexed by $j(e)$ where $j_e(e) > 0$, $j_{ee}(e) < 0$. 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 $wy + j(e) - c(e)$. Students choose effort e^* that solves the first-order condition:

$$wf_e + j_e - c_e = 0$$

Here I have assumed that there are no explicit rewards to learning other than the labor market earnings. Now suppose students are offered monetary incentives in the current period as a linear function of their level of learning y . For now, we will assume that there are no interactions between the joy of learning and external rewards. Students' new net payoff function is $(w + p)y + j(e) - c(e)$ where p is the rate of reward per scoring unit. Under this new condition, students choose effort e^{**} to solve the first order condition:

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

The effort level under the new regime is higher than when there were no incentives. The interesting case arises when nonmonetary rewards of learning are a function of monetary incentives. Let nonmonetary benefits depend both on level of effort and external rewards i.e. $j(e, p)$. The net payoff function for students is $(w + p)y + j(e) - c(e)$ and they choose e^{***} that solves:

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

To see how the optimal level of effort changes with external rewards, I compute the total derivative of e^* with respect to p :

$$\begin{aligned}\frac{de^*}{dp} &= -\frac{\frac{\partial}{\partial p}[(w+p)f_e + j_e(e,p) - c_e(e)]}{\frac{\partial}{\partial e}[(w+p)f_e + j_e(e,p) - c_e(e)]} \\ &= \frac{f_e + j_{ep}(e,p)}{-[(w+p)f_{ee} + j_{ee}(e,p) - c_{ee}(e)]}\end{aligned}$$

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 the incentives, the expression is unambiguously 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 depending on the curvature of $j(e,p)$.

4.2 Data Description

At the time the experiment was designed, final semester scores from grade seven were intended to be the baseline scores and grade eight district level-exam scores were meant to be the final measure of achievement. Several patterns in the data led me to question the suitability of these scores for my analysis. First, grade seven scores are

Figure 2: Histogram of Mathematics Scores

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 for these students. Third, tests and curricula are not uniform across schools in grade seven. Fourth, I 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 2. There is a sharp spike in density around 32, the passing grade, in the end-of-the-year scores while the spike is muted in the semester exams.

Figure 3 presents the proportion of students who passed each subject in grade seven and eight final exams, and grade eight semester exams. The proportions are higher for the year-end exams compared to the semester exams for all subjects and the difference is especially pronounced in English, Math, and Science.

Figure 3: Proportion of Students Passing Each Subject

There are a number of possible reasons for grade inflation in the year-end exams. Schools have an incentive to nudge a marginal student over the threshold because the funding the schools receive from the district education office is partly based on the

number of students who 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, I conduct two sets of analyses, the first of which treats grade eight first semester scores as the baseline measure and second semester scores as the final measure of achievement in the analyses. This severely stacks the deck *against* finding a positive treatment effect for several reasons. First, the incentives program was announced before the first semester exams so to the extent that incentives had a positive impact on first semester scores, I am less likely to find positive treatment effect.⁷ 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. Finally, to the extent that students anticipate grade inflation in the end-of-the-year exams, they are less motivated by the incentives for the semester exams.

The second set of analyses treats grade eight first semester scores as the baseline measure of ability and uses scores from the year-end district exams as the final measure

⁷ 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.

of achievement. To reduce the possibly spurious influence of grade inflation, I drop observations if the scores are close to the passing threshold. In particular, I drop observations if the scores in the core subjects are 32 or 33 or if the scores in the peripheral subjects are 16 or 17.

I cannot rule out grade inflation in the semester exams, though I believe the incentives to do so are substantially weaker. It is conceivable that 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 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 semester exam cash rewards in return for higher scores. This is also unlikely to occur in this experiment because the marginal increase in cash rewards due to grade inflation in just one subject would be insignificant as 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 well established in the 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 the year-end scores. Thus I am confident that the semester exams are appropriate measures of students' learning.

Tables 4 and 5 report the summary statistics of key school-level variables and socioeconomic variables. Schools are balanced in terms of pupil-teacher ratio and availability of school infrastructure, the two largest areas of school expenditure, but I find that average years of schooling of teachers qualified to teach at the lower secondary level is slightly lower in treatment schools.

Variable	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
Boys	0.45	0.43	0.02	1479
Grade eight repeaters (proportion)	0.10	0.09	0.01	1402
Notes: Information on the number of teachers is unavailable for 1 school. Information on teachers' education is unavailable for 2 schools. The infrastructure index is obtained by summing seven binary variables: availability of sufficient drinking water, toilets, urinals, playgrounds, electricity, school compound fences, and computer facilities. Difference in group means is tested using a <i>t</i> -test. Standard errors used in testing the difference between group means are clustered at the school level.				
*Significant at 10% level **Significant at 5% level ***Significant at 1% level				

Table 4: Summary Statistics of School Variables

In terms of home variables, there is little difference in home ownership, a key indicator of socioeconomic status. However, indicators for households' tastes for education, like parents with secondary education, availability of a tutor for students to

study with after school, and family members helping children with schoolwork, suggest that students in treatment schools are likely to come from relatively disadvantaged homes.

Variable	Treatment	Control	Difference	Obs.
Age of students	14.49	14.35	0.14	1227
Students living with parents	0.92	0.90	0.02	1228
Family owns the place of residence	0.77	0.72	0.05	1194
Number of siblings	2.18	2.26	-0.08	1228
Guardians illiterate	0.18	0.16	0.02	1227
Guardians have primary education	0.34	0.30	0.04	1227
Guardians have secondary education	0.36	0.44	-0.08*	1227
Guardians' education missing	0.11	0.10	0.01	1227
Tutor helps with schoolwork	0.03	0.11	-0.08***	1227
Household member helps with schoolwork	0.38	0.52	-0.14***	1208
Notes: Difference between group means is tested using a <i>t</i> -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 5: Summary Statistics of Socioeconomic Variables

Estimates of the impact of incentives on scores could be biased if the attrition rate is different between treatment and control groups. Table 5 shows that there is little difference in the proportion of students whose scores were available in each of the three

exams.⁸ The estimates could still be biased if the composition of the dropouts is different across treatment and control groups. To check for the difference in the composition of the dropouts, I 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.⁹ As shown in table 6, I also find little difference in the composition of the dropouts. Therefore the estimates are not likely to be biased due to nonrandom attrition across treatment arms.

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

Notes: The figures in the third column 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 fourth column refer to the pre-program percentile rank of the dropouts. Students' roll numbers are used as an approximate measure of pre-program class rank. Difference between group means is tested using a *t*-test. Standard errors are clustered at the school level and reported in parentheses.

Table 6: Attrition Pattern

⁸ The dropout rate observed in our sample is also similar to the district-wide dropout rates in Lalitpur district (Department of Education, 2010).

⁹ 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).

4.3 Impact of the Incentives on Academic Outcomes

Given two periods of data, the causal effect of the program can be estimated using a lagged dependent variable model, a difference-in-difference specification, or an individual fixed effects model under different identifying assumptions (Angrist and Pischke 2009). If the randomization was successful, we would expect to find broadly similar results from alternative specifications. Therefore, I first check the robustness of the results using the three models.

The lagged dependent variable model includes a lagged measure of achievement as a right-hand-side variable. The specification measures the gains in scores for treated students relative to that of control students with the same baseline achievement level. The identifying assumption for this specification is that of conditional independence. Conceptually, the baseline achievement measure is assumed to be a sufficient statistic for the entire history of family and school inputs and the unobserved innate ability of students. The lagged dependent variable model is given by:

$$y_{ij,t=2 \text{ or } 3} = \beta_0 + \beta_1 * y_{ij,1} + \beta_2 * incentives + X'_{ijt}\delta + \varepsilon_j + \varepsilon_{ij}$$

where y_{ijt} is the normalized score of student i in school j in time t ($y_{ij,1}$ corresponds to the first semester scores, $y_{ij,2}$ corresponds to second semester scores and $y_{ij,3}$ represent district exams scores), *incentives* is the dummy variable for treatment schools, X_{ijt} is a vector of variables that vary over time, and errors are clustered at the school level to allow for intra-cluster correlation. As discussed previously, I conduct two sets of

analyses, one using second semester scores and the other using district exams scores as the final measures of outcome. To reduce the spurious effect of grade inflation, I drop observations with scores close to the passing threshold (32 or 33 for core subjects and 16 or 17 for peripheral subjects). The covariates include dummy variables for the level of school (lower secondary, secondary, and upper secondary), pupil-teacher ratio, infrastructure index, years of teachers' education, number of siblings, highest level of education completed by guardians, and availability of academic reinforcement at home (either from a tutor or a household member). 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 standard deviation (s.d.) gains in scores.

The identifying assumption for the difference-in-difference specification is that the average scores would have followed the same trend in treatment and control groups in the absence of the treatment. Thus, any observed difference in trends can be attributed to the incentives. The difference-in-difference equation is estimated as following:

$$y_{ijt} = \beta_0 + \beta_1 * incentives + \beta_2 d_t + \beta_3 * incentives * d_t + X'_{ijt} \delta + \varepsilon_{jt} + \varepsilon_{ijt}$$

where d_t is the dummy variable for the second semester ($t = 2$) and the end-of-the-year ($t = 3$), and β_3 represents the between-group difference in within-group differences in pre- and post-test scores.

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. In this study, it is also important to account for the difference in the difficulty level of the semester exams and the grading scheme within schools that are likely to remain unchanged across exams. Formally, the fixed effects specification is:

$$y_{ijt} = \beta_0 + \beta_1 d_t + \beta_2 * incentives * d_t + X'_{ijt} \delta + \mu_i + \varepsilon_{jt} + \varepsilon_{ijt}$$

where d_t is the dummy variable for the second semester ($t = 2$) and the end-of-the-year ($t = 3$), and μ_i is the time-invariant individual fixed effect for student i .

Estimates of the impact of incentives for all specifications and for both second semester and year-end scores are reported in table 7.¹⁰ When the second semester scores are used as the final measure of achievement, the point estimate of the parameter of interest hovers around 0.03 and 0.04 in all specifications and it does not change significantly with the inclusion of school controls (dummy variables for the type of school, pupil-teacher ratio, infrastructure index, and average years of teachers' education), and home controls (number of siblings, highest level of education completed by parents, availability of help at home from a tutor, and availability of help at home from a household member).¹¹ The parameter estimates are about four times higher (although not significant) for all specifications when the scores from the district exams are used as the final measure of achievement.¹² This shows that the impact of the incentives increases with an increase in the duration of students' exposure to the program, at least for students outside a narrow band of performance levels clustered around the

¹⁰ The coefficients for school and home controls are reported in the tables in the appendix.

¹¹ Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods.

¹² The sample size decreases significantly when students with scores close to the passing threshold are dropped from analysis.

threshold of passing scores. Given that incentives are sharpest near these thresholds, and that I am removing these students from this sample, it further suggests that incentives increased test scores in this trial.

	Lagged Model	Lagged Model	Diff-in- Diff	Diff-in- Diff	Fixed Effects	Fixed Effects
Second Semester Scores						
Causal Parameter	0.04 (0.07)	0.05 (0.09)	0.03 (0.07)	0.04 (0.08)	0.03 (0.07)	0.04 (0.08)
School Controls		X		X		X
Home Controls		X		X		X
Obs.	1312	925	2624	1850	2624	2258
District Exams Scores						
Causal Parameter	0.16 (0.14)	0.18 (0.13)	0.13 (0.25)	0.14 (0.26)	0.13 (0.25)	0.14 (0.24)
School Controls		X		X		X
Home Controls		X		X		X
Obs.	668	493	1336	986	1336	1180
Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of first semester scores in comparison schools). School controls include dummy variables for the type of school (lower secondary, secondary, higher secondary), pupil-teacher ratio, infrastructure index, and years of teachers' education. Home controls include number of siblings, highest level of education completed by parents, and availability of academic reinforcement at home (either from a tutor or a household member). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. District exam scores data omits any student who received the minimum score necessary to pass or the minimum score plus one to control for grade inflation discussed in the text. 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: Estimate of the Impact of Cash Incentives on Average Aggregate Scores

To put estimates from the district exams in perspective, I compare my 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 (Kremer, Miguel and Thornton 2009). At the end of the first year of its implementation, teacher incentives in Kenya had an effect of 0.05 s.d. on test scores (Glewwe, Ilias and Kremer 2010) and textbook provision in Kenyan schools had an impact 0.04 s.d (Glewwe, Kremer and Moulin 2009). 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 Sundararaman 2009) while 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 (Banerjee, Cole, Duflo and Linden 2007). Finally, in the works most similar to mine, Bettinger (2010) finds that incentives increased math scores of the recipients by 0.15 s.d., while Fryer (2010) finds that when incentives are conditioned on grades, students gained between 0.09 and 0.13 s.d.¹³ Thus, my result based on district exams scores is on par with many of the school interventions experimented with in the last decade, especially given that the use of first semester scores as the baseline measure of the outcome likely understates the total impact.

I also estimate the following individual fixed effects model by stacking three periods of data (first semester, second semester, and district exams):

¹³ The estimates depend on the parameter of choice (Average Treatment Effect or Average Treatment Effect on Treated). The study finds no impact on standardized test scores when the incentives are conditioned on grades.

$$y_{ijt} = \beta_0 + \beta_1 d_2 + \beta_2 d_3 + \beta_3 * incentives * d_2 + \beta_4 * incentives * d_3 + X'_{ijt} \delta + \mu_i + \varepsilon_{jt} + \varepsilon_{ijt}, t = 1, 2, 3$$

where d_2 and d_3 are dummy variables for second semester and end-of-the-year respectively. The results are reported in table 8.

	Fixed Effects Coefficients	Fixed Effects Coefficients
Time 2	0.11 (0.08)	0.10 (0.08)
Time 3	0.51** (0.20)	0.49** (0.20)
Incentives * Time 2	-0.03 (0.09)	-0.02 (0.10)
Incentives * Time 3	0.15 (0.24)	0.16 (0.24)
Academic Reinforcement at Home		0.05 (0.03)
Obs.	1950	1725
Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of first semester scores in comparison schools). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies across time periods. The data omits any student who received the minimum score necessary to pass or the minimum score plus one to control for grade inflation discussed in the text. 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: Estimate of the Impact of Cash Incentives on Average Aggregate Scores (Three Period Panel Data)

The estimates on the interaction terms of the treatment status with district exams are positive (although not significant) and they are comparable to the results obtained in table 7. Therefore, the results for the fixed effects model in table 7 are robust to analysis with a three-period model. I find that the estimates on the interaction terms of the

treatment status with second semester exams are negative but both practically and statistically insignificant.

To further test the robustness of my results, I use grade seven year-end scores and grade eight year-end scores as the baseline and final measures of achievement respectively. The analysis drops students with scores close to the passing threshold in both grade seven and grade eight year-end exams. Table 9 reports the results when only two periods of data are used.

	Fixed Effects Coefficients	Fixed Effects Coefficients
Time 3	-0.10 (0.20)	-0.03 (0.20)
Incentives * Time 3	0.06 (0.29)	-0.03 (0.29)
Academic Reinforcement at Home		0.21** (0.10)
Obs.	708	644
Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of grade seven year-end scores in comparison schools). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. The data omits any student who received the minimum score necessary to pass or the minimum score plus one to control for grade inflation discussed in the text. 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: Impact of Incentives on Average Aggregate Scores Using Grade Seven Scores as Baseline

I find that students in treatment schools do not gain more relatively to students in control schools between grade seven and grade eight year-end exams. However, it

should be recognized that many variables are likely to have changed between the two time periods so an individual fixed effects model is unable to account for those changes.

	Incentives * District Exams	Obs.	Raw Baseline Score/Maximum
Aggregate (All Subjects)	0.14 (0.24)	1180	37.7/100
Nepali	-0.00 (0.22)	2138	40.1/100
English	0.03 (0.09)	1904	36.6/100
Math	0.17 (0.21)	1692	26.0/100
Science	-0.27 (0.26)	1808	34.2/100
Social Studies	-0.18 (0.20)	2010	41.6/100
Health and Physical Education	0.15 (0.20)	2080	19.5/50
Population and Environment	0.08 (0.14)	1978	21.2/50
Moral Education	-0.06 (0.30)	1204	24.2/50
Vocational Studies	-0.10 (0.23)	1454	21.7/50

Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of first semester scores in comparison schools). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. The data omits any student who received the minimum score necessary to pass or the minimum score plus one to control for grade inflation discussed in the text. 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: Fixed Effects Estimates of the Impact of Cash Incentives on District Exams Scores (by Individual Subjects)

Next, I disaggregate the impact of incentives on year-end scores for individual subjects using a fixed effects model with two periods of data (first-semester and year-end district exam scores). For each subject, observations are dropped if the year-end score is

close to the passing threshold (32 or 33 for core subjects and 16 and 17 for peripheral subjects). The results are reported in table 10.

I observe that gains in Math, Health and Physical Education, and Population and Environment constitute the bulk of the aggregate gains in the year-end exams. Several subjects feature lower scores under the treatment (Science, Social Studies, Moral Education, and Vocational Studies), though neither the gains nor losses reach statistical significance.

Using the same scores, I next investigate if there are heterogeneous treatment effects across subsamples of gender, socioeconomic status, and baseline achievement levels.¹⁴ Socioeconomic strata were constructed on the basis of students' report of their guardians' education.¹⁵ 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. To analyze if treatment effects differed by initial academic performance, students were placed into one of four quartiles based on aggregate first semester scores.¹⁶

The differential impact of incentives on district exam scores on males is estimated using the following regression:

¹⁴ Information on gender was filled out by researchers for students who did not complete the baseline survey but had unambiguous male or female names.

¹⁵ 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.

¹⁶ To the extent that incentives affected the distribution of scores in the first semester, the quartiles may not correspond to the "true" baseline quartiles.

$$y_{ijt} = \beta_0 + \beta_1 d_3 + \beta_2 * incentives * d_3 + \beta_3 * incentives * d_3 * male + X'_{ijt} \delta + \mu_i + \varepsilon_{jt} + \varepsilon_{ijt}, t = 1, 3$$

where *male* is the dummy variable indicating male student and β_3 is the differential impact of the incentives on male students. The estimates are reported in table 11. Under both sets of analyses, males do not gain significantly more relative to females. This finding contrasts with the pattern observed in the literature where girls respond relatively strongly to incentives (Angrist and Lavy 2009).

	Fixed Effects Coefficients	Fixed Effects Coefficients
Time 3	0.50** (0.20)	0.51** (0.20)
Incentives * Time 3	0.15 (0.26)	0.13 (0.26)
Incentives * Time 3 * Male	0.02 (0.10)	0.00 (0.10)
Academic Reinforcement at Home		0.09* (0.04)
Obs.	1324	1180
Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of first semester scores in comparison schools). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. The data omit any student who received the minimum score necessary to pass or the minimum score plus one to control for grade inflation discussed in the text. 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: Fixed Effects Estimates of the Impact of Cash Incentives on District Exams Scores (by Gender)

The interaction effects of incentives with socioeconomic status are estimated using the following regression:

$$y_{ijt} = \beta_0 + \beta_1 d_3 + \beta_2 * incentives * d_3 + \beta_3 * incentives * d_3 * prim + \beta_4 * incentives * d_3 * sec + X'_{ijt} \delta + \mu_i + \varepsilon_{jt} + \varepsilon_{ijt}, t = 1, 3$$

where *prim* and *sec* are dummy variables that indicate if the students' guardians have primary and secondary education respectively, and β_3 and β_4 represent the differential impact of the incentives on these students relative to students whose parents are illiterate.

	Fixed Effects Coefficients	Fixed Effects Coefficients
Time 3	0.49** (0.21)	0.49** (0.21)
Incentives * Time 3	0.07 (0.27)	0.06 (0.27)
Incentives * Time 3 * Primary	0.06 (0.15)	0.06 (0.14)
Incentives * Time 3 * Secondary	0.14 (0.11)	0.15 (0.10)
Academic Reinforcement at Home		0.06 (0.04)
Obs.	1068	1068
Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of first semester scores in comparison schools). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. The data omit any student who received the minimum score necessary to pass or the minimum score plus one to control for grade inflation discussed in the text. 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: Fixed Effects Estimates of the Impact of Cash Incentives on District Exams Scores (by Socioeconomic Status)

The results, reported in table 12, show that the impact of the incentives increase monotonically as we move up the socioeconomic strata, although the coefficients are not

statistically significant and the null hypothesis that treatment effects are equal for students from all strata cannot be rejected (F -test, p -value = 0.55). These findings are significant because it suggests the *channel* through which the incentives affect students' outcomes. It is possible that there exist multiple equilibria in the home learning environment. In the absence of incentives, students and families are locked in a lower equilibrium because of their mutual expectations: 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 was at work, I would expect students with better educated parents to show the most gains because the parents are likely to be more interested in their children's schooling. My analysis of the impact of incentives on passing rate and academic reinforcement at home, which appears in the next section, provides further support for this hypothesis.

The following regression is used to estimate the heterogeneous treatment effects across students of different ability:

$$\begin{aligned}
 y_{ijt} = & \beta_0 + \beta_1 d_3 + \beta_2 * incentives * d_3 + \beta_3 * incentives * d_3 * Q_2 \\
 & + \beta_4 * incentives * d_3 * Q_3 + \beta_5 * incentives * d_3 * Q_4 + X'_{ijt} \delta + \mu_i + \varepsilon_{jt} + \varepsilon_{ijt} \\
 & t = 1, 3
 \end{aligned}$$

where Q_2 , Q_3 , and Q_4 are the dummy variables for students in the second quartile, the third quartile, and the bottom quartile respectively and β_3 , β_4 , and β_5 are the differential

impacts of the incentives relative to the students in the top quartile. The results are reported in table 13.

	Fixed Effects Coefficients	Fixed Effects Coefficients
Time 3	0.53** (0.21)	0.51** (0.20)
Incentives * Time 3	-0.17 (0.23)	-0.16 (0.22)
Incentives * Time 3 * 2 nd Quartile	0.32*** (0.05)	0.34*** (0.05)
Incentives * Time 3 * 3 rd Quartile	0.62*** (0.07)	0.62*** (0.08)
Incentives * Time 3 * Bottom Quartile	0.58*** (0.17)	0.55*** (0.18)
Academic Reinforcement at Home		0.08* (0.04)
Obs.	1336	1180
Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of first semester scores in comparison schools). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. The data omit any student who received the minimum score necessary to pass or the minimum score plus one to control for grade inflation discussed in the text. 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 13: Fixed Effects Estimates of the Impact of Cash Incentives on District Exams Scores (by Baseline Ability)

The interaction effects for students in the second, third, and bottom quartiles are practically large and statistically significant and the null hypothesis that treatment effects are equal for students from all quartiles can be rejected (F -test, p -value = 0.00).

However, the point estimates could be an artifact of the data because my analysis of year-end scores drops students whose scores are close to the passing threshold. To check for

the robustness of the results, I conduct the analysis using second semester scores as my final measure of outcome. The results are reported in table 14.

	Fixed Effects Coefficients	Fixed Effects Coefficients
Time 3	0.05 (0.06)	0.05 (0.06)
Incentives * Time 3	-0.02 (0.07)	-0.03 (0.08)
Incentives * Time 3 * 2 nd Quartile	0.05 (0.05)	0.06 (0.05)
Incentives * Time 3 * 3 rd Quartile	0.07 (0.05)	0.08 (0.05)
Incentives * Time 3 * Bottom Quartile	0.11** (0.04)	0.11** (0.05)
Academic Reinforcement at Home		0.04 (0.03)
Obs.	2624	2258
Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of first semester scores in comparison schools). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. 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 14: Fixed Effects Estimates of the Impact of Cash Incentives on Second Semester Scores (by Baseline Ability)

Table 14 shows that the interaction effects increase monotonically for students in lower quartiles, and the estimate on students in the bottom quartile is statistically significant (0.11; p -value 0.04) although we cannot reject the null hypothesis that the interaction effects are equal (F -test, p -value 0.46). This finding can be attributed to the piece-wise linear incentive scheme because it provides incentives to even the least able students to put in more effort at the margin. 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 the low-ability students in the second semester and year-end exams 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 if they exert high level of effort and if there are diminishing marginal returns to effort.

4.4 Impact of the Incentives on Passing Rate

According to the rewards structure, students who passed all subjects received rewards at twice the rate relative to students who failed one or more subjects. To see if the incentives had an impact on the passing rate, I compute the marginal increase in the probability of passing all subjects for students in the treatment group. To be consistent with earlier analyses, I estimate a linear probability model with individual fixed effects. Formally, the fixed effects specification is given by:

$$pass_{ijt} = \beta_0 + \beta_1 d_t + \beta_2 * (incentives) * d_t + X'_{ijt} \delta + u_i + \varepsilon_{jt} + \varepsilon_{ijt}$$

where $pass_{ijt}$ is the indicator variable for whether a student passed all subjects in period t , d_2 and d_3 are the dummy variables indicating second semester and the end-of-the-year exams respectively and u_i is the individual fixed effect. Errors are clustered at the school level to allow for intra-cluster correlation. The estimates are reported in table 15.

	Fixed Effects	Fixed Effects	Fixed Effects	Fixed Effects
Time 2	-0.00 (0.02)	0.00 (0.02)		
Incentives * Time 2	0.04 (0.03)	0.03 (0.03)		
Time 3			0.47*** (0.09)	0.44*** (0.09)
Incentives * Time 3			0.06 (0.11)	0.09 (0.12)
Academic Reinforcement at Home		0.02 (0.02)		-0.02 (0.03)
Obs.	2624	2258	1336	1180

Notes: The dependent variable is the dummy variable indicating if a student passed all subjects. Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between the two periods. The district exams data omit any student who received the minimum score necessary to pass or the minimum score plus one to control for potential spurious effect. 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 15: Fixed Effects Estimates of the Impact of Cash Incentives on the Probability of Passing

I find that in the second semester, students in treatment schools were about 3 percentage points more likely to pass all subjects in the second semester compared to students in the control group although the point estimate is not significant. The comparable figure for the district level exams is 6 to 9 percent where my analysis drops observations that are most likely to have been artificially pushed over the passing threshold. These results are consistent with the results in table 7 where the impact of the incentives are larger for the year-end exams, possibly due to longer duration of exposure to the program.

	Fixed Effects Coefficient	Obs.
<u>Second Semester</u>		
Incentives * Time 2 * Male	0.02 (0.04)	2258
Incentives * Time 2 * Primary	-0.00 (0.04)	2032
Incentives * Time 2 * Secondary	0.04 (0.06)	2032
Incentives * Time 2 * 2 nd Quartile	0.13** (0.06)	2258
Incentives * Time 2 * 3 rd Quartile	0.11 (0.08)	2258
Incentives * Time 2 * Bottom Quartile	0.04 (0.06)	2258
<u>District Exams</u>		
Incentives * Time 3 * Male	-0.05 (0.09)	1180
Incentives * Time 3 * Primary	0.04 (0.11)	1068
Incentives * Time 3 * Secondary	0.13 (0.06)	1068
Incentives * Time 3 * 2 nd Quartile	0.56*** (0.06)	1180
Incentives * Time 3 * 3 rd Quartile	0.69*** (0.09)	1180
Incentives * Time 3 * Bottom Quartile	0.44*** (0.15)	1180
<p>Notes: The dependent variable is the dummy variable indicating if a student passed all subjects. Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. The district exam data omit any student who received the minimum score necessary to pass or the minimum score plus one to control for potential spurious effect. Standard errors are clustered at the school level and reported in parentheses.</p> <p>*** Significant at 1% level; **Significant at 5% level; *Significant at 10% level</p>		

Table 16: Fixed Effects Estimates of the Impact of Cash Incentives on the Probability of Passing (by Subgroups)

Next, I analyze the impact of incentives on the passing rate for subgroups of students (table 16).¹⁷ In both the second semester and year-end exams, incentives had a relatively larger (but insignificant) impact on the passing probability of students from the highest socioeconomic stratum. Although in both the second semester (F -test, p -value 0.49) and the year-end exams (F -test, p -value 0.28) I cannot reject the null hypothesis that the interaction effects are equal, the coefficients are suggestive of the complementary interaction between students' motivation and a supportive home environment in the production of education.

It is 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 during the semester exams. If the incentives did have an impact on the passing rate, we would expect it to be concentrated among the marginal students who failed one or two subjects and just fell short of the higher price regime. Consistent with this expectation, the coefficients are relatively high for students in the second and the third quartiles of the distribution for both the second semester and year-end exams.

4.5 Households' Response to the Incentives

One of the original contributions of this study is the analysis of student and household response to cash incentives. The exit survey allows us to estimate if there were any changes in students' daily time use at home. I find that students in the treatment schools spend about 2 and 6 minutes less in schoolwork in the morning and

¹⁷ The coefficients reported in table 16 are from regression specifications that control for academic reinforcement at home.

evening respectively (not shown).¹⁸ 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 I 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 study time, but I cannot verify this claim. In any event, both the magnitudes are practically and statistically insignificant for it to be of concern.

To estimate if the incentives had an impact on the availability of help with schoolwork at home, either from a tutor or a household member, I estimate a linear probability model with individual fixed effects.

$$help_{ijt} = \beta_0 + \beta_0 + \beta_1 d_3 + \beta_2 * (incentives) * d_3 + X'_{ijt} \delta + u_i + \varepsilon_{jt} + \varepsilon_{ijt}, t = 1 \text{ and } 3$$

where $help_{ijt}$ is the indicator variable for the help received by student i in school j in time period t . The estimates are presented in table 17.

The results show that students in treatment schools were 8 percent more likely to receive academic reinforcement at home at the end of the year. This suggests that incentives had a favorable impact on the learning environment at home.

¹⁸ 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).

	Fixed Effects Coefficients
Time 3	-0.00*** (0.03)
Incentives * Time 3	0.08* (0.04)
Obs.	1990
Notes: Dependent variable is the dummy variable indicating if a student received academic reinforcement at home, either from a tutor or a household member. 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 17: Fixed Effects Estimates of the Impact of Incentives on Academic Reinforcement at Home

Table 18 shows the impact of incentives on different subgroups of students. I find that although boys were significantly less likely to receive help at home at the end of the year relative to the girls, they were still more likely to receive help than students in control schools. Although the point estimates are not significant, students whose guardians have primary and secondary education were relatively more likely to receive academic reinforcement at home. These observations provide further affirmation of the hypothesis that households with relatively more education were more responsive to the incentives. Finally, when the sample is disaggregated according to baseline ability, I find no significant difference in the availability of help at home.

	Fixed Effects Coefficient	Obs.
Incentives * Time 3 * Male	-0.07** (0.03)	1990
Incentives * Time 3 * Primary	0.06 (0.05)	1780
Incentives * Time 3 * Secondary	0.08 (0.05)	1780
Incentives * Time 3 * 2 nd Quartile	-0.05 (0.04)	1962
Incentives * Time 3 * 3 rd Quartile	-0.01 (0.03)	1962
Incentives * Time 3 * Bottom Quartile	0.00 (0.06)	1962

Notes: The dependent variable is the dummy variable indicating if a student received academic reinforcement at home, either from a tutor or a household member. 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 18: Fixed Effects Estimates of the Impact of Incentives on Academic Reinforcement at Home (by Subgroups)

To gauge the impact of incentives on the attendance rate, I use the number of students who filled out the baseline and exit surveys as measures of attendance.¹⁹ Difference-in-difference estimates show no significant increase in the attendance rate at treated schools (0.03; p -value 0.44).²⁰ 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

¹⁹ This method is adopted because the official attendance information available from school register is not always reliable in developing countries. When conducting surveys, I generally contacted the principals for an appointment either the previous evening or the same morning and my requests were mostly met so the visits are “unannounced” for all practical purposes. Attendance rates might be nonrandom in the few occasions when the request was not met and I was asked to visit at a later time.

²⁰ This finding is reasonable because of high pre-program attendance rates in sample schools. The average attendance rate at baseline was 78 percent. I use a difference-in-difference specification because attendance information is available only at the school level and not at the individual level.

so surprising because educational materials were sufficiently available at baseline in both treatment and control schools and marginal returns to additional materials is likely to be sharply diminishing.²¹

4.6 Impact of Incentives on Intrinsic Motivation

Finally, 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.²² Table 19 reports the results from an ordered-probit analysis.²³ I find that the treatment status does not significantly predict the probability of choosing any option for any subjects. This analysis suggests that financial rewards had no noticeable adverse impact on students’ motivation to learn.

²¹ 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.

²² 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.

²³ The ordered probit analysis requires larger values to represent “higher” outcomes. Therefore, I recoded the survey responses as following: (1) highly uninterested; (2) uninterested; (3) interested; and (4) highly interested.

	Mean (Treatment)	Mean (Control)	Marginal Effects of Treatment				Obs.
			Highly Uninterested	Moderately Uninterested	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
76 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) highly uninterested; (2) uninterested; (3) interested; and (4) highly interested. Probabilities may not sum to 1 due to rounding. Standard errors are clustered at the school level.

Table 19: Ordered-Probit Analysis of the Impact of Cash Incentives on Intrinsic Motivation

Chapter 5: Discussion and Conclusion

5.1 Cost Effectiveness of Cash Incentives

As researchers experiment with various school interventions, programs can be 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, I adopt the cost effectiveness methodology developed by researchers at the Poverty Action Lab (2010).

After I drop students who have scores close to the passing threshold who were excluded from the regression analysis of the district exam scores, the sample consists of 190 students from treatment schools. The students were eligible to receive Rs. 43,888.41 (~\$558.37) in the second semester and Rs. 46,268 in the final semester (~\$588.60). Assuming that students in treatment schools gained 0.14 standard deviations between the first semester and the end-of-the-year exams, the cost per 0.1 standard deviation gain is \$4.02.²⁴ This figure is more than the figure for other comparable programs like textbook provision and teacher incentives in Kenya (\$4.01 and \$3.41 respectively), girls' scholarship program in (\$3.53), and remedial education in India (\$0.67 - \$1.77 depending on the year). However, there are two caveats to this comparison. Costs reported in Kremer et al. (2004) and Duflo et al. (2005) are in nominal US dollars for the year in

²⁴ The administrative costs up to the final round of incentives distribution is Rs. 100,892 (~\$1283.61). Therefore, including the administrative costs, the cost per pupil per standard deviation gain is \$9.14.

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. Furthermore, the gains in scores for my study represent the gains between the first semester and year-end exams where the first semester scores were incentivized. Therefore, the cost effectiveness of financial incentives is likely to fare better when these factors are taken into consideration.

There is also a discussion in the literature of whether to treat financial rewards as costs or transfers; they are costs from the point of view of the one who is financing the program, but they are mere transfers for society. If I consider gains in district exams scores, the cost per pupil per 0.1 standard deviation gain is virtually zero excluding the administrative costs.²⁵ Therefore, a design like mine that utilizes test score data from schools is cheaper to implement. At the same time, I also recognize that the scores are more amenable to manipulation.

For a more rigorous analysis the of cost effectiveness of incentives, we need to understand the consequences of incentives on long term variables of interest like years of schooling, grade transition and re-enrollment rates, and lifetime earnings. However, such an analysis is outside the scope of this study

5.2 Discussion

This study has highlighted a number of interesting facets of incentives and education. This study finds that when incentives are structured as piece rates, low ability

²⁵ The cost per pupil per 0.1 standard deviation gain is \$4.8 including the administrative costs.

students gain substantially more relative to high ability students. I am 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, Miguel and Thornton 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 this experiment allows me to evaluate the impact of incentives both at the extensive (passing all subjects) and intensive (improving aggregate scores) margins. I 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 outcomes of the low ability students who had little chance of crossing over to the higher regime.

I also document the incidence of grade inflation in year-end exams. 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 in developing countries (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, this 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 the experiment did not allow for differential levels of incentives and therefore I am unable to provide an estimate of the price elasticity of

achievement gains. Under an ideal experiment, the 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 paying too little may actually be counterproductive (Gneezy and Rustichini 2000).

The finding concerning the behavioral response of households is also revealing of the interaction between incentives and households' actions. The probability that incentives increased the availability of academic reinforcement was the highest for students from the uppermost socioeconomic stratum. 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 guardian 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.²⁶

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 a change is temporary and incentives may lead to poorer outcomes in the long-term (Kohn 1993).²⁷ I plan to measure the long-term impact of incentives using scores from national level exams students take at the end of grade ten.

²⁶ 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.

²⁷ If we do observe such fade-out effect, it would not be unique to our study.

5.3 Conclusion

In recent years, researchers have experimented with innovative school interventions to increase school attendance and academic outcomes of students. This dissertation reports the results of a piece-rate cash incentives program where students were provided monetary rewards in proportion to their scores in academic exams. I find that monetary incentives raise academic achievement, with the weakest students and students from higher socioeconomic strata exhibiting the greatest gains. My 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 my 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, Miguel and Thornton 2009).

I also recognize that there is scope for the implementation of cash incentive programs like this in developing countries because of its cost effectiveness. Incentive 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. The relative superiority of input versus

output based incentives is therefore not obvious *a priori* and should be the subject of further inquiry.

This study adds to the new generation of “determinants of achievement” studies that has begun to shed more light into the “black box” of the 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, Deaton 2010b). Future studies should be designed based on richer and more nuanced behavioral models of cognitive achievement.

References

- Alderman, Harold, Jere Behrman, David R. Ross, and Richard Sabot, "The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market," *Oxford Bulletin of Economics and Statistics*, 58 (1996), 29-55.
- Angrist, Joshua, "Lifetime Earnings and the Vietnam Era Draft Lottery - Evidence from Social-Security Administrative Records," *American Economic Review*, 80 (1990), 313-336.
- Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael. Kremer, "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment," *American Economic Review*, 92 (2002), 1535-1558.
- Angrist, Joshua, Eric Bettinger, and Michael Kremer, "Long-term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia," *American Economic Review*, 96 (2006), 847-862.
- Angrist, Joshua, and Alan B. Krueger, "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, 106 (1991), 979-1014.
- Angrist, Joshua, Daniel Lang, and Philip Oreopoulos, "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal - Applied Economics*, 1 (2009), 136-163.
- Angrist, Joshua, and Victor Lavy, "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 99 (2009), 1384-1414.
- Angrist, Joshua, and Jorn-Steffen Pischke, *Mostly Harmless Econometrics* (Princeton, New Jersey: Princeton University Press, 2009).
- Arkes, Hal R., "Competence and the Overjustification Effect," *Motivation and Emotion*, 3 (1979), 143-150.
- Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden, "Remedying Education: Evidence from Two Randomized Experiments in India," *National Bureau of Economic Research Working Paper Series*, No. 11904 (2005).

---, "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics*, 122 (2007), 1235-1264.

Becker, Gary S., "Investment in Human Capital: A Theoretical Analysis," *The Journal of Political Economy*, 70 (1962), 9-49.

Behrman, Jere R., and Nancy Birdsall, "The Quality of Schooling: Quantity Alone is Misleading," *American Economic Review*, 73 (1983), 928-946.

Behrman, Jere R., Todd Petra, and Piyali Sengupta, "The Impact of PROGRESA on Achievement Test Scores in the First Year," (2000).

Behrman, Jere R., Piyali Sengupta, and Petra Todd, "Progressing Through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico," *Economic Development and Cultural Change*, 54 (2005), 237-275.

Benabou, Roland, and Jean Tirole, "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70 (2003), 489-520.

Bettinger, Eric P., "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores," *National Bureau of Economic Research Working Paper Series*, No. 16333 (2010).

Binder, Melissa, Philip T. Ganderton, and Kristin Hutchens, "Who Benefits from a Lottery-Funded College Subsidy? Evidence from the New Mexico Success Scholarship," (2002).

Bishop, John, "Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn - Externalities, Information Problems, and Peer Pressure," in *Handbook of the Economics of Education*, Erik Hanushek, and Finis Welch, eds. (Amsterdam: North-Holland, 2006).

Blackburn, McKinley L., and David Neumark, "Omitted-Ability Bias and the Increase in the Return to Schooling," *Journal of Labor Economics*, 11 (1993), 521-544.

Boissiere, M., J. B. Knight, and R. H. Sabot, "Earnings, Schooling, Ability, and Cognitive Skills," *American Economic Review*, 75 (1985), 1016-1030.

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

Cameron, Judy, and W. David. Pierce, "Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis," *Review of Educational Research*, 64 (1994), 363-423.

Card, David, and Alan B. Krueger, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United-State," *Journal of Political Economy*, 100 (1992), 1-40.

Central Bureau of Statistics, "Nepali Living Standard Survey 2003/04 Statistical Survey," (2004)

Coleman, James S., "Equality of Educational Opportunity (COLEMAN) Study (EEOS), 1966," (Inter-university Consortium for Political and Social Research (ICPSR) [distributor], 1966).

Cornwell, Christopher, David B. Mustard, and Deepa J. Sridhar, "The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Program," *Journal of Labor Economics*, 24 (2006), 761-786.

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

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

DeCharms, Richard, *Personal Causation: The Internal Affective Determinants of Behavior* (New York: Academic Press, 1968).

Deci, Edward L., "Effects of Externally Mediated Rewards on Intrinsic Motivation," *Journal of Personality and Social Psychology*, 18 (1971), 105-115.

---, "Effects of Contingent and Noncontingent Rewards and Controls on Intrinsic Motivation," *Organizational Behavior and Human Performance*, 8 (1972a), 217-229.

---, "Intrinsic Motivation, Extrinsic Reinforcement, and Inequity," *Journal of Personality and Social Psychology*, 22 (1972b), 113-120.

Department of Education, "Flash I Report 2007/08," (2008).

Duflo, Esther, "Field Experiments in Development Economics," (2006).

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

Duflo, Esther, Rachel Glennerster, and Michael Kremer, "Using Randomization in Development Economics Research: A Toolkit," in *Handbook of Development Economics*, T. Paul Schultz and John A. Strauss, ed. (Elsevier, 2007).

Duryea, Suzanne, and Andrew Morrison, "The Effect of Conditional Transfers on School Performance and Child Labor: Evidence from an Ex-Post Impact Evaluation in Costa Rica," (2004).

Easterly, William R., *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics* (MIT Press, 2002).

Eisenberger, Robert, W. David. Pierce, and Judy Cameron, "Effects of Reward on Intrinsic Motivation - Negative, Neutral, and Positive: Comment on Deci, Koestner, and Ryan (1999)," *Psychological Bulletin*, 125 (1999), 677-691.

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

Festinger, Leon, Hiroshi Ono, Clarke A. Burnham, and Donald Bamber, "Efference and Conscious Experience of Perception," *Journal of Experimental Psychology*, 74 (1967), 1-36.

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

Frey, Bruno S., and Felix OberholzerGee, "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out," *American Economic Review*, 87 (1997), 746-755.

Fryer, Roland G., Jr., "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *National Bureau of Economic Research Working Paper Series*, No. 15898 (2010).

Fuller, Bruce, "What School Factors Raise Achievement in the Third-World," *Review of Educational Research*, 57 (1987), 255-292.

Fuller, Bruce, and Prema Clarke, "Raising School Effects while Ignoring Culture? Local Conditions and the Influence of Classroom Tools, Rules, and Pedagogy," *Review of Educational Research*, 64 (1994), 119-157.

Glewwe, Paul, *The Economics of School Quality Investments in Developing Countries: An Empirical Study of Ghana* (St. Martin's Press, 1999).

---, "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes," *Journal of Economic Literature*, 40 (2002), 436-482.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics*, 2 (2010), 205-227.

Glewwe, Paul, and Hanan Jacoby, "Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana," *The Journal of Human Resources*, 29 (1994), 843-864.

Glewwe, Paul, and Michael Kremer, "Schools, Teachers, and Education Outcomes in Developing Countries," in *Handbook of the Economics of Education*, E. Hanushek and F. Welch, ed. (Elsevier, 2006).

Glewwe, Paul, Michael Kremer, and Sylvia Moulin, "Many Children Left Behind? Textbooks and Test Scores in Kenya," *American Economic Journal-Applied Economics*, 1 (2009), 112-135.

Glewwe, Paul, Michael Kremer, Sylvia Moulin, and Eric Zitzewitz, "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya," *Journal of Development Economics*, 74 (2004), 251-268.

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

Green, Jerry R., and Nancy L. Stokey, "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91 (1983), 349-364.

Hanushek, Eric A., "Throwing Money At Schools," *Education Week*, 1 (1981), 24-24.

---, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 24 (1986), 1141-1177.

---, "The Failure of Input-Based Schooling Policies," *Economic Journal*, 113 (2003), F64-F98.

Hanushek, Eric. A., "Interpreting Recent Research on Schooling in Developing Countries," *World Bank Research Observer*, 10 (1995), 227-246.

---, "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis*, 19 (1997), 141-164.

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

Heyneman, Stephen P., "Influences on Academic Achievement: A Comparison of Results from Uganda and More Industrialized Societies," *Sociology of Education*, 49 (1976), 200-211.

Heyneman, Stephen P., J. P. Farrell, and M. A. Sepulvedaastuardo, "Textbooks and Achievement in Developing Countries: What We Know," *Journal of Curriculum Studies*, 13 (1981), 227-246.

Heyneman, Stephen P., Dean T. Jamison, and Xenia Montenegro, "Textbooks in the Philippines: Evaluation of the Pedagogical Impact of a Nationwide Investment," *Educational Evaluation & Policy Analysis*, 6 (1984), 139-150.

Holmstrom, Bengt, "Moral Hazard in Teams," *Bell Journal of Economics*, 13 (1982), 324-340.

Holmstrom, Bengt, and Paul Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law Economics & Organization*, 7 (1991), 24-52.

Howell, William G., Patrick J. Wolf, David E. Campbell, and Paul E. Peterson, "School Vouchers and Academic Performance: Results from Three Randomized Field Trials," *Journal of Policy Analysis and Management*, 21 (2002), 191-217.

Imbens, Guido, and Jeffrey Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47 (2009), 5-86.

Jamison, Dean T., Klaus Galda, Barbara Searle, and Stephen P Heyneman, "Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement," *Journal of Educational Psychology*, 73 (1981), 556-567.

Johnson, George E., and Frank P. Stafford, "Social Returns to Quantity and Quality of Schooling," *Journal of Human Resources*, 8 (1973), 139-155.

Kohn, Alfie, *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes* (New York: Houghton Mifflin Co., 1993).

Kremer, Michael, "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons," *American Economic Review*, 93 (2003), 102-106.

Kremer, Michael, Edward Miguel, and Rebecca Thornton, "Incentives to Learn," *National Bureau of Economic Research Working Paper Series*, No. 10971 (2004).

---, "Incentives to Learn," *Review of Economics and Statistics*, 91 (2009), 437-456.

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

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

---, "Economic Considerations and Class Size," *Economic Journal*, 113 (2003), F34-F63.

Ladd, Helen F., "School Vouchers: A Critical View," *Journal of Economic Perspectives*, 16 (2002), 3-24.

Lalonde, Robert J., "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76 (1986), 604-620.

Lavy, Victor, "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement," *Journal of Political Economy*, 110 (2002), 1286-1317.

---, "Using Performance-Based Pay to Improve the Quality of Teachers," *Future of Children*, 17 (2007), 87-109.

---, "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics," *American Economic Review*, 99 (2009), 1979-2011.

Lazear, Edward P., "Performance Pay and Productivity," *American Economic Review*, 90 (2000a), 1346-1361.

---, "The Power of Incentives," in *112th Annual Meeting of the American-Economic-Association*, (Boston, Massachusetts: American Economic Review, 2000b).

Lazear, Edward P., and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89 (1981), 841-864.

Leamer, Edward E., "Let's Take the Con Out of Econometrics," *American Economic Review*, 73 (1983), 31-43.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw, "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment," *Journal of the European Economic Association*, 8 (2010), 1243-1265.

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

Miguel, Edward, and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72 (2004), 159-217.

Mincer, Jacob, *Schooling, Experiences, and Learning* (Columbia University Press, 1974).

Muralidharan, Karthik, and Venkatesh Sundararaman, "Teacher Performance Pay: Experimental Evidence from India," National Bureau of Economic Research Working Paper Series, No. 15323 (2009).

Murnane, Richard J., John B. Willett, and Frank Levy, "The Growing Importance of Cognitive Skills in Wage Determination," *Review of Economics and Statistics*, 77 (1995), 251-266.

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

Okeeffe, Mary, W. Kip Viscusi, and Richard J. Zeckhauser, "Economic Contests: Comparative Reward Schemes," *Journal of Labor Economics*, 2 (1984), 27-56.

Oreopoulos, Philip, and Kjell G. Salvanes, "Priceless: The Nonpecuniary Benefits of Schooling," *Journal of Economic Perspectives*, 25 (2011), 159-184.

Ponce, Juan, and Arjun Bedi, "The Impact of a Cash Transfer Program on Cognitive Achievement: The Bono de Desarrollo Humano of Ecuador," *Economics of Education Review*, 29 (2010), 116-125.

Poverty Action Lab, "J-PAL Cost Effectiveness Methodology," (2010)

Pritchett, Lant, "Where Has All the Education Gone?," *World Bank Economic Review*, 15 (2001), 367-391.

Pritchett, Lant, and Deon Filmer, "What Education Production Functions Really Show: A Positive Theory of Education Expenditures," *Economics of Education Review*, 18 (1999), 223-239.

Psacharopoulos, George, "Returns to Education: An Updated International Comparison," *Comparative Education*, 17 (1981), 321-341.

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

---, "Returns to Investment in Education: A Global Update," *World Development*, 22 (1994), 1325-1343.

Rawlings, Laura B., and Gloria M. Rubio, "Evaluating the Impact of Conditional Cash Transfer Programs," *The World Bank Research Observer*, 20 (2005), 29-55.

Reimers, Fernando, Carol DeShano da Silva, and Ernesto Trevino, "Where is the "Education" in Conditional Cash Transfers in Education?," (UNESCO Institute for Statistics, 2006).

Rouse, Cecilia Elena, "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," *The Quarterly Journal of Economics*, 113 (1998), 553-602.

Rouse, Cecilia Elena, and Lisa Barrow, "School Vouchers and Student Achievement: Recent Evidence and Remaining Questions," *Annual Review of Economics*, 1 (2009), 17-42.

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

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

Sims, Chris, "Macroeconomics and Reality," *Econometrica*, 48 (1980), 1-48.

Velez, Eduardo, Ernesto Schiefelbein, and Jorge Valenzuela, "Factors Affecting Achievement in Primary Education," in *Human Resources Development and Operations Policy*, (The World Bank, 1993).

Appendix A: Tables

	Lagged Model	Diff-in- Diff	Fixed Effects
Causal Parameter	0.05 (0.09)	0.04 (0.08)	0.04 (0.08)
Lagged Score	0.93*** (0.03)	.	.
2 nd Semester	.	0.04 (0.07)	0.05 (0.06)
Secondary	-0.16 (0.11)	-0.05 (0.15)	.
Lower Secondary	-0.24 (0.21)	0.00 (0.22)	.
Pupil-teacher Ratio	0.00 (0.00)	-0.00 (0.00)	.
Infrastructure Index	0.01 (0.04)	0.06 (0.05)	.
Teachers' Education	-0.08 (0.08)	0.08 (0.10)	.
Number of Siblings	0.02 (0.01)	-0.04 (0.04)	.
Highest Level of Parents' Education	-0.00 (0.00)	0.02 (0.02)	.
Academic Enforcement at Home	-0.03 (0.03)	-0.13 (0.08)	0.04 (0.03)
Obs.	925	1850	2278

Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of first semester scores in comparison schools). School controls include dummy variables for the type of school (lower secondary, secondary, higher secondary), pupil-teacher ratio, infrastructure index, and years of teachers' education. Home controls include number of siblings, highest level of education completed by parents, and availability of academic reinforcement at home (either from a tutor or a household member). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. District exam scores data omits any student who received the minimum score necessary to pass or the minimum score plus one to control for grade inflation discussed in the text. 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 20: Estimate of the Impact of Incentives on Average Aggregate Scores (Second Semester)

	Lagged Model	Diff-in- Diff	Fixed Effects
Causal Parameter	0.18 (0.13)	0.14 (0.26)	0.14 (0.24)
Lagged Score	0.67*** (0.07)	.	.
District Exams	.	0.50*** (0.22)	0.51** (0.22)
Secondary	0.27 (0.18)	0.09 (0.18)	.
Lower Secondary	0.18 (0.28)	-0.09 (0.25)	.
Pupil-teacher Ratio	-0.00 (0.01)	-0.00 (0.01)	.
Infrastructure Index	0.10 (0.08)	0.17** (0.07)	.
Teachers' Education	0.11 (0.12)	-0.02 (0.12)	.
Number of Siblings	0.04** (0.02)	-0.00 (0.04)	.
Highest Level of Parents' Education	-0.00 (0.02)	0.02 (0.03)	.
Academic Enforcement at Home	0.04 (0.05)	-0.16** (0.08)	0.09* (0.04)
Obs.	493	986	1180
Notes: Dependent variables are normalized scores (with respect to the mean and the standard deviation of first semester scores in comparison schools). School controls include dummy variables for the type of school (lower secondary, secondary, higher secondary), pupil-teacher ratio, infrastructure index, and years of teachers' education. Home controls include number of siblings, highest level of education completed by parents, and availability of academic reinforcement at home (either from a tutor or a household member). Academic reinforcement at home is the only control variable salient for the fixed effects models because it alone varies between two periods. District exam scores data omits any student who received the minimum score necessary to pass or the minimum score plus one to control for grade inflation discussed in the text. 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 21: Estimate of the Impact of Incentives on Average Aggregate Scores (District Exams)