DISCUSSION PAPER SERIES

No. 6682

MONITORING WORKS: GETTING TEACHERS TO COME TO SCHOOL

Esther Duflo, Rema Hanna and Stephen Ryan

DEVELOPMENT ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP6682.asp**

ISSN 0265-8003

MONITORING WORKS: GETTING TEACHERS TO COME TO SCHOOL

Esther Duflo, Massachusetts Institute of Technology (MIT), Paris School of Economics and CEPR
Rema Hanna, New York University and J-PAL
Stephen Ryan, Massachusetts Institute of Technology (MIT)

Discussion Paper No. 6682
February 2008

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 6682

February 2008

ABSTRACT

Monitoring Works: Getting Teachers to Come to School*

This paper combines a randomized experiment and a structural model to test whether monitoring and financial incentives can reduce teacher absence and increase learning. In 57 schools in India, randomly chosen out of 113, a teacher's daily attendance was verified through photographs with time and date stamps, and his salary was made a non-linear function of his attendance. The teacher absence rate changed from 42 percent in the comparison schools to 21 percent in the treatment schools. To separate the effects of the monitoring and the financial incentives, we estimate a structural dynamic labour supply model that allows for heterogeneity in preferences and auto-correlation of external shocks. The teacher response was almost entirely due to the financial incentives. The estimated elasticity of labour with respect to the incentive is 0.306. Our model accurately predicts teacher attendance in two out-of-sample tests on the comparison group and a treatment group that received different financial incentives. The program improved child learning: test scores in the treatment schools were 0.17 standard deviations higher than in the comparison schools.

JEL Classification: I20, I21, J13, J30 and O10
Keywords: education, financial incentives and India

Esther Duflo
Department of Economics, E25-252
Massachusetts Institute of Technology
50 Memorial Drive
Cambridge MA 02139
USA
Email: eduflo@mit.edu

Rema Hanna
Wagner Graduate School of Public Service
295 Lafayette Street
Room 3046
New York, NY 10012
USA
Email: rema.hanna@nyu.edu

Stephen Ryan
Department of Economics, E25-252
Massachusetts Institute of
Technology
50 Memorial Drive
Cambridge MA 02139
USA
Email: sryan@mit.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=167704

* This project is a collaborative exercise involving many people. Foremost, we are deeply indebted to Seva Mandir, and especially to Neelima Khetan and Priyanka Singh, who made this evaluation possible. We thank Ritwik Sakar and Ashwin Vasan for their excellent work coordinating the fieldwork. Greg Fischer, Shehla Imran, Callie Scott, Konrad Menzel, and Kudzaishe Takavarasha provided superb research assistance. For their helpful comments, we thank Abhijit Banerjee, Rachel Glennerster, Michael Kremer and Sendhil Mullainathan. For financial support, we thank the John D. and Catherine T. MacArthur Foundation.

Submitted 20th of December 2007

1 Introduction

Over the past decade, many developing countries have expanded primary school access. This expansion has been energized by initiatives such as the United Nations Millennium Development Goals, which call for achieving universal primary education by 2015. However, these improvements in school access have not been accompanied by improvements in school quality. For example, in India, a nationwide survey found that 65 percent of children enrolled in grades 2 through 5 in government primary schools could not read a simple paragraph, and 50 percent could not do simple subtraction or division (Pratham, 2006). These poor learning outcomes may be due, in part, to high absence rates among teachers. Using unannounced visits to measure teacher attendance, a nationally representative survey found that 24 percent of teachers in India were absent from the classroom during normal school hours (Chaudhury, et al., 2005a, b).[1] Improving attendance rates may be the first step needed to make "universal primary education" a meaningful term.

Solving the absentee problem poses a significant challenge (see Banerjee and Duflo (2005) for a review). One solution—championed by many, including the 2004 World Development Report—is to involve the community in teacher oversight, including the decisions to hire and fire teachers. However, in many developing countries, teachers are a powerful political force, and may resist attempts to curb their influence. As such, many governments have begun to shift from hiring government teachers to instead hiring "para-teachers." Para-teachers are teachers that are hired on short, flexible contracts to work in primary schools and in non-formal education centers (NFEs) run by NGOs and local governments (that are often financed by central governments). In some countries, para-teachers account for most of the growth in the teaching staff over the last few years. In India alone, 21 million children—mainly poor children in rural areas—attend NFEs.[2]

The evidence that para-teachers are more motivated than other teachers is mixed, however. In India, Chaudhury et al. (2005b) found that locally hired para-teachers had absence rates significantly higher than those of government teachers. In contrast, Duflo, Dupas and Kremer (2007) found that para-teachers in Kenya were more likely to be present than regular teachers. One possible explanation for this difference is that para-teachers in Kenya face stronger incentives than those in India, since the Kenyan para-teachers had the opportunity to be regularized (and thereby

[1] Although teachers do have some official non-teaching duties, this absence rate is too high to be fully explained by this particular story.

[2] This is a large number of children: for comparison, in the US, enrollment in all public schools from kindergarten to grade 8 was just 33.6 million in 2004. In Gujarat, one of India's largest states, para-workers comprise 43 percent of the teaching staff in rural areas (Duthilleul, 2004).

earn higher salaries) if they performed well.[3] Unlike government teachers, it may be feasible to implement such incentives for para-teachers since they do not form a entrenched constituency, they are already subject to yearly renewal of their contract, and there is a long queue of qualified job applicants. Thus, providing para-teachers with strong incentives may be an effective way to improve the quality of education, provided that para-teachers can teach effectively.[4]

In this paper, we combine a randomized experiment and structural methods to empirically test whether the direct monitoring of the attendance of para-teachers (referred to simply as teachers in the rest of the paper), coupled with high-powered incentives based on their attendance, improves school quality. We ask three main questions: If teachers are given incentives to attend school, will they actually attend school more? If they attend school more, will they teach more? Finally, if teacher absenteeism is reduced, will children learn more?

The effect of incentives based on presence is theoretically ambiguous. While simple labor supply models would predict that incentives should increase effort, the incentives could fail to improve attendance for a variety of reasons. First, teachers may be unable to take advantage of the incentives if they must participate in village meetings, training sessions, election or census duty. These pressures may be particularly high on para-teachers, who are often among the few literate individuals in the village. Second, the incentive schemes may crowd out the teacher's intrinsic motivation to attend school (Benabou and Tirole, 2006). Finally, some teachers, who previously believed that they were required to work every day, may decide to stop working once they have reached their target income for the month (Fehr and Gotte, 2002).

Even if incentives increase teacher attendance, it is unclear whether child learning levels will actually increase. Teachers may multitask (Holmstrom and Milgrom, 1991), reducing their efforts along other dimensions.[5] Such schemes may also demoralize teachers, resulting in less effort (Fehr and Schmidt, 2004), or may harm teachers' intrinsic motivation to teach (Kreps, 1997). On the other hand, incentives can improve learning levels if the main cost of working is the opportunity cost of attending school and, once in school, the marginal cost of teaching is low. In this case, an incentive system that directly rewards presence would stand a good chance of increasing child

[3]Another possibility is that the schools where para-teachers teach in India are very different from the schools where most regular teachers teach, whereas they are the same in Kenya.

[4]This ability to discipline para-teachers is another reason why the Indian government has favored hiring contract teachers over regular teachers, even though in practice, there has been little effort to put such systems in place (Duthilleul, 2004).

[5]This is a legitimate concern as other incentive programs (based on test scores) have been subject to multitasking (Glewwe, Ilias and Kremer, 2003), manipulation (e.g. Figlio and Winicki, 2002; Figlio and Getzler, 2002) or outright cheating (Jacob and Levitt, 2003). On the other hand Lavy (2002) and Mulharidharan and Sundaraman (2006) find very positive effect of similar programs.

learning. Thus, whether or not the incentives can improve school quality is ultimately an empirical question.

We study a teacher incentive program run by the NGO Seva Mandir. Seva Mandir runs single-teacher NFEs in the rural villages of Rajasthan, India. As in many rural areas, teacher absenteeism is high, despite the threat of dismissal for repeated absence. In our baseline study (August 2003), the absence rate was 44 percent.

Faced with such high absenteeism, Seva Mandir implemented an innovative monitoring and incentive program in September 2003. In 57 randomly selected program schools, Seva Mandir gave teachers a camera, along with instructions to have one of the students take a picture of the teacher and the other students at the start and close of each school day. The cameras had tamper-proof date and time functions, allowing for the collection of precise data on teacher attendance that could be used to calculate teachers' salaries. Each teacher was then paid according to a non-linear function of the number of valid school days for which they were actually present, where a "valid" day was defined as one for which the opening and closing photographs were separated by at least five hours and both photographs showed at least eight children. Specifically, they received Rs 500 if they attended fewer than 10 days in a given month, and Rs 50 for any additional day (up to a maximum of 25 or 26 days depending on the month). In the 56 comparison schools, teachers were paid a fixed rate for the month, and were told (as usual) that they could be dismissed for repeated, unexcused absences.

The program resulted in an immediate and long lasting improvement in teacher attendance rates in treatment schools, as measured through monthly unannounced visits in both treatment and comparison schools. Over the 30 months in which attendance was tracked, teachers at program schools had an absence rate of 21 percent, compared to 44 percent baseline and the 42 percent in the comparison schools. Absence rates stayed low after the end of the proper evaluation phase (the first fourteen months of the program), suggesting that teachers did not change their behavior simply for the evaluation.

While the reduced form results inform us that this program was effective in reducing absenteeism, it does not tell us what the effect of another scheme with a different payment structure would be. Moreover, it does not allow us to disentangle the effect of the monitoring from the effect of the incentives, since the comparison school teachers were not monitored.

To answer these questions, we exploit the non-linear nature of the incentive scheme to estimate a dynamic labor supply model using the daily attendance data in the treatment schools. The identification exploits the fact that the incentive for a teacher to attend school on a single day

changes as a function of the number of days they attend school in the month, and the number of days left in the month. This is because they have to attend at least 10 days in a month to begin to receive the incentive (by working in the beginning of the month, the teacher builds up the option to work for Rs 50 per day at the end of the month). Indeed, regression discontinuity design estimates show that teachers work significantly more at the beginning of the month than at the end of the previous month, when they had not accumulated at least 10 days of work in that month.

We use this fact to estimate the teachers' marginal utility of money. We allow serial correlation in the opportunity cost of attending school and heterogeneity in teachers' outside option, and we use the method of simulated moments to estimate the parameters. Allowing for serial correlation and heterogeneity considerably complicates the estimation procedure, but we show that these features are very important in this application. To our knowledge, our paper is one of the few papers to estimate dynamic labor supply decisions with unobserved heterogeneity and a serially correlated error structure.[6] Our approach is similar in spirit to that in Card and Hyslop (2005), though the set up and methods employed are quite different. They also combined a randomized evaluation and parametric assumptions to estimate a dynamic labor supply model in the context of the Canadian Self Sufficiency Program (SSP). In this program, participants need to find work within twelves month of assignment to the treatment group to establish eligibility for a negative income tax for the next three years. Card and Hyslop (2005) estimate a dynamic labor supply model to distinguish the "establishment effect" (the incentive to work in order to establish eligibility) from the incentive to work due to the negative income tax. Like us, they find that individuals respond to dynamic incentives in a manner consistent with the prediction of a simple dynamic labor supply model.[7] Other papers combining randomized experiments and structural models are Todd and Wolpin (2007) and Attanazio, Meghir, and Santiago (2006).

We find that teachers are very responsive to the financial incentives: our preferred estimates suggest that the elasticity of labor supply with respect to the level of the financial bonus is 0.306. Furthermore, *decreasing* the number of days that workers must work until they are eligible for the incentive by a single day *increases* the expected number of days worked by about 1.29 percent. An unusual feature of this application is the ability to carry out convincing out-of-sample tests based on the randomized evaluation (as in Todd and Wolpin (2007)). When allowing for serial correlation and heterogeneity, we find that our model accurately predicts the difference in attendance in the treatment and the control group, as well as the number of days worked under a new incentive

[6] The only other paper we know of that allows for both of these factors is the working paper of Bound, Stinebrickner, and Waidman (2005), which examines the retirement decisions of older men.

[7] They deal with auto-correlated shocks by introducing a second order state dependance in the probability to work.

system initiated by Seva Mandir after the experiment.

Although we find that teachers are sensitive to the financial incentives, we see no evidence of multitasking. When the school was open, teachers were as likely to be teaching in treatment as in comparison schools, suggesting that the marginal costs of teaching are low conditional on attendance. Student attendance when the school was open was similar in both groups, so student in treatment group received more days of instruction. A year into the program, test scores in the treatment schools were 0.17 standard deviations higher than in the comparison schools. Two and a half years into the program, children from the treatment schools were also 10 percentage points (or 62 percent) more likely to transfer to formal primary schools, which requires passing a competency test. The program's impact and cost are similar to other successful education programs.

The paper is organized as follows. Section 2 describes the program and evaluation strategy. The results on teacher attendance, including the estimates from the dynamic labor supply model, are presented in Section 3. Section 4 presents the results on other dimensions of teacher effort, as well as student outcomes. Section 5 concludes.

2 Experimental Design and Data Collection

2.1 Non-formal Education Centers

Since the enactment of the National Policy on Education of 1986, non-formal education centers (NFEs) have played an increasingly important role in India's drive towards universal primary education. The NFEs serve two main purposes. First, since they are easier to establish and cheaper to run, they have been the primary instrument for expanding school access to children in remote and rural areas. The government of the Indian state of Madhya Pradesh, for example, mandated that NFEs be established for all communities where there were no schools within one kilometer. Second, since NFEs are subject to fewer regulations than government schools, they can tailor their hours and curricula to meet the diverse needs of children. Thus, they have been used to ease children who may otherwise not attend school into a government school at the age-appropriate grade level. As of 1997, 21 million children were enrolled in NFEs across India (Education for All Forum, 2000), and similar informal schools operate throughout most of the developing world (Bangladesh, Togo, Kenya, etc.).

Children of all ages may attend, though, in our sample, most are between 7-10 years of age. Nearly all of the children are illiterate when they enroll. In the setting of our study, the NFEs are open six hours a day and have about 20 students each. All students are taught in one classroom

by one teacher, who is recruited from the local community and who has, on average, a 10th grade education. Instruction focuses on basic Hindi and math skills. The schools only have one teacher; thus, when the teacher is absent, the school is closed.

2.2 The Incentive Program

Seva Mandir runs about 150 NFEs in the tribal villages of Udaipur, Rajasthan. Udaipur is a sparsely populated, arid and hilly region, where villages are remote and access is difficult. Thus, it is difficult for Seva Mandir to regularly monitor the NFEs. Absenteeism is high, despite the organization's policy calling for the dismissal of truant teachers. A 1995 study (Banerjee et al., 2005) found that the absence rate was 40 percent, while our baseline (in August 2003) found that the rate was 44 percent.

Seva Mandir was, therefore, motivated to identify ways in which to reduce teacher absenteeism. To this end, they implemented an external monitoring and incentive program in September 2003. They chose 120 schools to participate, with 60 randomly selected schools serving as the treatment group and the remaining 60 as the comparison group.[8] Prior to the announcement of the program, 7 of these schools closed or failed to open; these closure were equally distributed amount the treatment and controls schools, and were not due to the program.

In the 57 treatment schools, Seva Mandir gave each teacher a camera, along with instructions for one of the students to take a photograph of the teacher and the other students at the start and end of each school day. The cameras had a tamper-proof date and time function that made it possible to precisely track each school's openings and closings.[9] Figure 1 displays two typical pictures; the day of the month and the time of day appear in the lower-righthand corner; there is no ambiguity about the month since the rolls were changed every month. Camera upkeep (replacing batteries, and changing and collecting the film) was conducted monthly at regularly scheduled teacher meetings. If a camera malfunctioned, teachers were instructed to call the program hotline within 48 hours. Someone was then dispatched to replace the camera, and teachers were credited for the missing day.[10]

At the start of the program, Seva Mandir's monthly base salary for teachers was Rs 1000 ($23

[8] Seva Mandir operates in five blocks in the district. Stratified random sampling was conducted within block.

[9] The time and data buttons on the cameras were covered with heavy tape, and each had a seal that would indicate if it had been tampered with. Fines would have been imposed if cameras had been tampered with (this did not happen) or if they had been used for another purpose (this happened in one case, when a teacher photographed his family).

[10] Teachers were given the 48-hour leeway to report malfunctioning cameras because not all villages have a working phone and phone services are not always reliable.

at the real exchange rate, or about $160 at PPP) for at least 20 days of work per month. In the treatment schools, teachers received a Rs 50 ($1.15) bonus for each additional day they attended in excess of the 20 days, and they received a Rs 50 fine for each day of the 20 days they skipped work. Seva Mandir defined a "valid" day as one in which the opening and closing photographs were separated by at least five hours and at least eight children were present in both photos to indicate that the school was actually functioning. Due to ethical and political concerns, Seva Mandir capped the fine at Rs 500. Thus, salaries ranged from Rs 500 to Rs 1,300 (or $11.50 to $29.50). In the 56 comparison schools, teachers were paid the flat rate of Rs 1,000, and were informed that they could be dismissed for poor attendance. However, this happens very rarely, and did not happen during the span of the evaluation.[11]

Seva Mandir pays its teachers every two months. In each two-month period, they collected the last roll of film a few days before the salary payment, so that payment could be made immediately at the end of the relevant time period. To reinforce the understanding of the program, Seva Mandir showed treatment teachers a detailed breakdown of how their salary was calculated after the first payment.

2.3 Data Collection

An independent evaluation team led by Vidhya Bhawan (a Udaipur-based consortium of schools and teacher training institutes) and the researchers collected data to answer three basic questions: If teachers are provided with high-powered incentives to attend school that are based on external monitoring, will they attend more? If they do attend school more, will teaching time increase? Finally, will the students learn more?

We have two sources of attendance data. First, we collected data on teacher attendance through one random unannounced visit per month in all schools. By comparing the absence rates obtained from the random checks across the two types of schools, we can determine the program's effect on absenteeism.[12] Second, Seva Mandir provided access to all of the camera and payment data for the treatment schools. This allows us to verify whether the random checks provide a good estimate of actual attendance rates (as measured by the cameras), and also allows us to check whether teachers were simply coming to school in for the photos, rather than attending the entire school

[11] Teachers in the control schools knew that the camera program was occurring, and that some teachers were randomly selected to be part of the pilot program.

[12] The random checks were not linked with any incentives, and teachers were aware of that fact. We cannot rule out the fact that the random check could have increased attendance in comparison schools. However, we have no reason to believe that the random checks would differentially affect the attendance of comparison and treatment teachers.

day. Furthermore, we exploit the daily camera data to estimate a structural model of dynamic labor supply.

We collected data on teacher and student activity during the random check to allow us to determine whether teachers taught more as a result of the program. For schools that were open during the visit, the enumerator noted the school activities: how many children were sitting in the classroom, whether anything written on the blackboard, and whether the teacher was talking to the children. While these are crude measures of teacher performance, they were chosen because each could be easily observed before the teacher and students could adjust their behavior: for example, the enumerator could see the blackboard the instant he entered the school. Since the schools have only one classroom and one teacher, teachers could not be warned that the enumerator was at the school and change their behavior.

During the random check, the enumerator also conducted a roll call to document which children on the evaluation roster were present.[13] They noted whether any of the absent children had left school or had enrolled in a government school, and then updated the evaluation roster to include new children.

To determine whether child learning increased as a result of the program,the evaluation team, in collaboration with Seva Mandir, administered three basic competency exams to all children enrolled in the NFEs in August 2003: a pre-test in August 2003, a mid-test in April 2004, and a post-test in September 2004. The pre-test followed Seva Mandir's usual testing protocol. Children were given either a written exam (for those who could write) or an oral exam (for those who could not). For the mid-test and post-test, all children were given both the oral exam and the written exam; those unable to write, of course, earned a zero on the written section. The oral exam tested simple math skills (counting, one-digit addition, simple division) and basic Hindi vocabulary skills, while the written exam tested for these competencies plus more complex math skills (two-digit addition and subtraction, multiplication and division), the ability to construct sentences, and reading comprehension. Thus, the written exam tested both a child's ability to write and his ability to handle material requiring higher levels of competency relative to the oral exam.

Finally, we collected detailed data on teachers' characteristics. First, to determine whether the effect on learning depended upon a teacher's academic ability, Seva Mandir administered a competency exam to each teacher prior to the program. Second, two months into the program, the evaluation team observed each school for a whole day in order to assess whether the program

[13] Evaluation rosters were different from the school roster in that they included all children enrolled at the beginning of the experiment and all children enrolled subsequently.

impact depended on the pedagogy employed by the teachers.[14]

2.4 Baseline and Experiment Integrity

Given that schools were randomly allocated to the treatment and comparison groups, we expected school quality to be similar across groups prior to the program onset. Before the program was announced in August 2003, the evaluators were able to randomly visit 41 schools in the treatment group and 39 in the comparison.[15] Panel A of Table 1 shows that the attendance rates were 66 percent and 64 percent, respectively. This difference is not statistically significant. Other measures of school quality were also similar prior to the program: in all dimensions shown in Table 1, the treatment schools appear to be slightly better than comparison schools, but the differences are always small and never significant. Finally, to determine the joint significance of the treatment variable on all of the outcomes listed in Panels B through E, we estimated a SUR model. The results are listed in the final row of Table 1. The F-statistic is 1.21, with a p-value of 0.27, implying that the comparison and treatment schools were similar to one another at the program's inception.

Baseline academic achievement and preparedness were the same for students across the two types of schools. Table 2 presents the results of the pre-test (August 2003). Panel A shows the percentage of children who could write. In Panels B and C, we report the results of the oral and written tests, respectively. On average, students in both groups were at the same level of preparedness before the program. Seventeen percent of the children in the treatment schools and 19 percent in the comparison schools took the written exam. This difference is not significant. Those who took the oral exam in the treatment schools performed somewhat better than those who took the oral exam in the comparison schools, while those who took the written exam in the treatment school performed somewhat better than in the comparison schools. These differences are not significant.

[14] Note that unlike the crude measures of teacher performance collected at the random checks, teachers may have changed their behavior as a result of the observations.

[15] Due to time constraints, only 80 randomly selected schools of the 113 were visited prior to the program. There was no significant (or perceivable) difference in the characteristics of the schools that were not observed before the program. Moreover, the conclusion of the paper remains unchanged when we restrict all the subsequent analysis to the 80 schools that could be observed before the program was started.

3 Results: Teacher attendance

3.1 Reduced form results: Teacher Behavior

The effect on teacher absence was both immediate and long lasting. Figure 2 shows the fraction of schools found open on the day of the random visit, by month. Between August and September 2003, teacher attendance increased in treatment schools relative to the comparison schools. Over the next two and a half years, the attendance rates in both types of schools followed similar seasonal fluctuations, with treatment school attendance systematically higher than comparison school attendance.

As Figure 2 shows, the treatment effect remained strong even after the post-test, which marked the end of the formal evaluation. Since the program had been very effective, Seva Mandir maintained it. However, they only had enough resources to keep the program operating in the treatment schools (expansion to all schools is planned, but Seva Mandir is currently performing other experiments to chose the most effective way to improve attendance). The random checks conducted after the post-test showed that the higher attendance rates persisted at treatment schools even after the teachers knew that the program was permanent, suggesting that teachers did not alter their behavior simply for the duration of the evaluation.

Table 3 presents a detailed breakdown of the program effect on absence rates.[16] Columns 1 and 2 report the means for the treatment and comparison schools, respectively, for the period September 2003 to February 2006. Column 3 presents the difference between the treatment and comparison schools for this period, while Columns 4 through 6 respectively present this difference until the mid-test, between the mid-test and post-test, and after the post-test. On average, the teacher absence rate was 21 percentage points lower in the treatment than in the comparison schools (Panel A). Thus, the program halved the absence rate.[17]

The effects on teacher attendance were pervasive—teacher attendance increased for both low and high quality teachers. Panel B reports the impact for teachers with above median test scores on the teacher skills exam conducted prior to the program, while Panel C shows the impact for

[16] As Appendix Table 1 shows, roughly equal numbers of schools across the treatment and control group either closed or experienced a change in teacher. Very few schools experienced a change in teacher, particularly during the formal evaluation period. Seva Mandir gave a competency exam to teachers in December 2006. New teachers had slightly higher test scores than existing teachers on the exam, but the difference between new teachers in treatment and control schools was minimal.

[17] This reduction in school closures was comparable to that of a previous Seva Mandir program which tried to reduce school closures by hiring a second teacher for the NFEs. In that program, school closure only fell by 15 percentage points (Banerjee, Jacob and Kremer, 2005), both because individual teacher absenteeism remained high and because teachers did not coordinate to come on different days.

teachers with below median scores.[18] The program impact on attendance was larger for below median teachers (a 24 percentage point increase versus a 15 percentage point increase). However, this was due to the fact that the program brought below median teachers to the same level of attendance as above median teachers (78 percent).

The program reduced absence everywhere in the distribution. Figure 3 plots the observed density of absence rates in the treatment and comparison schools for the 25 random checks. The figure clearly shows that the incentive program shifted the entire distribution of absence for treatment teachers.[19] Not one of the teachers in the comparison schools was present during all 25 observations. Almost 25 percent of teachers were absent more than half the time. In contrast, 5 of the treatment teachers were present on all days, 47 percent of teachers were present on 21 days or more, and all teachers were present at least half the time. Therefore, the program was effective on two margins: it eliminated extremely delinquent behavior (less than 50 percent attendance), and increased the number of teachers with perfect or very high attendance records.

A comparison of the random check data and the camera data suggests that, for the most part, teachers did not "game" the system. A comparison of the random check data and the camera data provides direct evidence of this. Table 4 shows that for the treatment schools, the camera data tend to match the random check data quite closely. Out of the 1337 cases, 80 percent matched perfectly; that is, the school was open and the photos were valid or the school was closed and the photos were not valid. In 13 percent of the cases, the school was found open during the random check, but the photos indicated that the day was not considered "valid" (which is not considered an instance of "gaming"). There are 88 cases (7 percent) in which the school was closed and the photos were valid, but only 54 (4 percent of the total) of these were due to teachers being absent in the middle of the day during the random check and shown as present both before and after. In the other cases, the data did not match because the random check was completed after the school had closed for the day, or there were missing data on the time of the random check or photo (Table 4, Panel C). Overall, the fact that gaming occurred only 4 percent of the time suggests that the program was quite robust.

Of the 179 cases (13 percent) where the school was open but the photos were invalid, it was primarily because there was only one photo (90 cases) or because the school was open for less than

[18] Teacher test scores and teacher attendance are correlated: In the control group, below median teachers came to school 53 percent of the time, while above median teachers came to school 63 percent of the time.

[19] We also graphed the estimated cumulative density function of the frequency of attendance, assuming that the distribution of absence follows a beta-binomial distribution (not shown for brevity). The results are similar to that of Table 3.

the full five hours (43 cases). This suggests that for a small number of cases, the random check may have designated a comparison school as open for the day, even though it was open for only part of the school day. Therefore, since the program may also have affected the length of each day, the random check data may, if anything, underestimate the effect of the program on the total teaching time a child received. Figure 4, which plots the difference in average teacher attendance rates for treatment and comparison schools by the time of the random check, provides support for this hypothesis. The difference in the attendance rate was larger at the start and end of the day, suggesting that teachers in treatment schools not only attended more often, but also kept the schools open for more hours.

3.2 Monitoring or incentives: Preliminary Evidence

The program had two components: the daily monitoring of teacher attendance and an incentive that was linked to attendance. It is possible that the daily monitoring itself could crowd out the teachers' intrinsic motivation to attend school, but that the financial incentive is strong enough to overcome this effect. Benabou and Tirole (2006) formally model this effect: in their model, when incentives are introduced to motivate a "pro-social" behavior, small incentives may have a discouraging effect, by destroying a person's motivation to do the task in terms of her self image. If the incentive is large enough, however, the usual effect of the monetary incentive dominates.

Ideally, to disentangle these effects, we would have provided different monitoring and incentive schemes in different, randomly selected schools. Some teachers could have been monitored, but without receiving incentives. Some could have received a small incentive, while others could have received a larger one. This was not feasible for a variety of reasons. However, combining a "natural experiment" in the data with several structural assumptions can shed some light on these questions.[20]

In particular, the non-linear nature of the incentive scheme provides a promising source of identification. Consider a teacher who, because he was ill, was unable to attend school on most of the first 20 days of the 26 days of the month. By day 21, assuming he has attended only 5 days so far, he knows that, if he works every single day remaining in the month, he will have worked only 10 days. Thus, he will earn Rs 500, i.e. the same amount he would earn if he did not work any other days that month. Although he is monitored, his monetary incentive to work in these last few days is zero. At the start of the next month, the clock is re-set. He now has incentive to start

[20] See Attanasio, Meghir and Santiago (2006) and Todd and Wolpin (forthcoming) for two other applications of this method of combining structural estimation and a randomized experiment.

attending school again, since by attending at the beginning of the month he can hope to be "in the money" by the end of the month, thereby benefiting from the incentive. Consider another teacher of the same type who has worked 10 days by the 21st day of the month. For everyday he works in the five remaining days, he earns Rs 50. By the beginning of the next month, his incentive to work is no higher. In fact, it could even be somewhat lower since he may not benefit from the work done the first day of the month if he does not work at least 10 days in that month.

In the data, 65 percent of the teachers who had worked less than 10 days in a given month (and thus were not eligible for the incentives at the end of the month) worked at least 10 days the next month. Thus, this suggests a sharp increase in incentives at the beginning of the next month for a teacher who was delinquent at the end of the previous month. On the other hand, 95 percent of the teachers who had worked at least 10 days in a month also worked at least 10 days the following month, suggesting that those teachers face essentially the same incentive to work at the end of a given month and the start of the next month.

This leads to a simple test for whether financial incentives matter. For teachers in the treatment group, we created a dataset that contains their attendance records for the last and the first day of each month. The last day of each month and the next day of the following month form a pair, indexed by m. We run the following equation, where $Work_{it}$ is a dummy variable equal to 1 if teacher i in school s works in day t in the pair of days m.

$$Work_{itm} = \alpha + \beta 1_m(d > 10) + \gamma Firstday + \lambda 1_m(d > 10) * Firstday + \upsilon_i + \mu_m \epsilon_{is}, \qquad (1)$$

where $1_m(d > 10)$ is a dummy equal to 1 for both days in the pair m if the teacher had worked more than 10 days in the month of the first day of the pair, and 0 otherwise. We estimate this equation treating υ_i and μ_m as either fixed effects or random effects.

If the teachers are sensitive to financial incentives, we expect β to be positive (teachers should work more when they are in the money than out of the money), γ to be positive (a teacher who is out of the money in a given month should work more in the first day of the following month) and λ to be negative and as large as γ (there is no increase in incentive for teachers who had worked at least 10 days before). Even with teacher fixed effects, β may not have a causal interpretation, because shocks may be auto-correlated. For example, a teacher who has been sick the entire month, and thus has worked less than 10 days, may also be less likely to work the first day of the next month. However, because when a month starts and finishes is arbitrary and should not be related to the underlying structure of shocks, a positive γ strongly indicates that teachers are sensitive

to financial incentives, unless there is a common "first day of the month" effect unrelated to the incentives. A negative λ will be robust even to this effect.

Table 5 presents these results, without fixed effects in column 1 and with teacher and pair fixed effects in column 2. These results clearly show that teachers are more likely (37 percentage points in the fixed effect specification) to work on the last day of the month when they are "in the money" than when they are not. More importantly, teachers who were out of the money in one month are 12 percentage points more likely to work in the first day of the next month, while teachers who are in the money in a given month work about as much the first day of the following month as the last day of that month (the interaction between "in the money" and "first day of the month" is negative and of the same absolute magnitude as the coefficient of "first day of the month"). This is exactly what would be expected if teachers respond to the financial incentives.

More generally, the structure of the incentives lends itself to a regression design estimation, with the calendar days as the running variable which determines changes in treatment. The set up generates may different discontinuities (one per month), and two groups of teachers for whom the discontinuity has different implications, which is rare in regression discontinuity designs settings. Figure 5 gives a graphical representation of the approach (Table 5, columns 1 and 2 provide the regression estimates). It shows a regression of the probability that a teacher works if she is in the money by day 21 of the month (with 4 days left), in the last 10 days of that month and the first 10 days of the next month. We fit a third order polynomial on the left and the right of the change in month. The figure shows a jump up for teacher who were not in the money, and no jump for those who were in the money. This is exactly what we would expect: the change in incentive at the beginning of a month is important for teachers who were not in the money. Intuitively, they have 65 percent chance to be in the money the following month, so they have a 65 percent chance that this day will "count." They also value the fact that working early on increases the chance that they reach the 10 days threshold. The teachers who were in the money, however, have a 95 percent chance to be in the money again. In addition, these teachers value the fact that the first days worked help them work towards the 10 days threshold. Therefore, we do not see a sharp drop in incentives for the teachers who had been in the money the previous month.

These results imply that teachers are responsive to the financial incentives, and not only to the monitoring system. However, without more structure, it is not possible to conclude whether all of the effect of the program is due to incentives. To do so, we need to estimate the marginal utility of the incentive to the teacher, and therefore the marginal value of working on the first day of the month. To analyze this problem, we set up a dynamic labor supply model and we use the

additional restrictions that the model provides to estimate its parameters.

3.3 A Dynamic Model of Labor Supply

In this section, we propose and estimate a simple model of dynamic labor supply over the month, which incorporates the teacher response to the varying incentives over the month.

3.3.1 The Model

Each day, a teacher chooses whether or not to attend school, by comparing the value of attending school to that of staying home or doing something else. His payoff to attending school is realized at the end of the month, when his number of days worked are counted and his salary is calculated. We assume no discounting within the month.

We denote the state space by s, where $s = (t, d)$, where t is the current time and d is the days worked previously in the current month. We are interested in the response to dynamic incentives within each month, so we allow time to take on values between 1 and $T = 26$, which is the maximal number of days that a teacher can work in a month.

The transition process between states is trivial. At the end of each day, t increases by one, unless $t = T$, in which case it resets to $t = 1$. If a teacher has worked in that period d increases by one, otherwise it remains constant.

The payoffs are as follows. In each period, the teacher has the choice to attend school or not. If the teacher does not attend school, he or she receives a utility of $\mu + \epsilon_t$, where μ is the average utility of not working and ϵ is a shock drawn from the standard normal distribution. In some specifications, we will allow μ to vary across teachers.

At the conclusion of time T, the teacher will also receive a payoff which is a function of the number of days worked in the current month:

$$\pi(d) = 500 + \max\{0, d - 10\} \cdot 50. \tag{2}$$

Each teacher gets at least Rs 500, and for every day over 10 that a teacher works in the current month, he or she receives a bonus of Rs 50.

Given this payoff structure, for $t < T$, we can write the value function for each teacher as follows:

$$V(t, d) = \max\{\mu + \epsilon_t + EV(t + 1, d), EV(t + 1, d + 1)\}. \tag{3}$$

At time T, we have:

$$V(T,d) = \max\{\mu + \epsilon_T + \beta\pi(d) + EV(1,0), \beta\pi(d+1) + EV(1,0)\}, \tag{4}$$

where β is marginal utility of income. Note that the term $EV(1,0)$ enters into both sides of the maximum operator in Equation 4. Since the expectation of this term is independent of any action taken today, without loss of generality we can ignore any dynamic considerations that arise in the next month when making decisions in the current month. This is useful since we can think about solving the value function by starting at time T and working backward, which breaks an infinite-horizon dynamic program into a repeated series of independent finite-time horizon dynamic programs. Equation 4 also motivates the location normalization of zero utility for attending school. The choice between working and not working only depends on the difference in utilities; thus, we can add any constant to both sides and obtain the same set of policy functions. Therefore, without loss of generality, we set the payoff to attending school to be equal to zero.

3.3.2 Identification

As in other finite-horizon games, identification is constructive, and based on partitions of the state space. At time T, the agent faces a static decision; he or she will work if:

$$\mu + \epsilon_T + \beta\pi(d) > \beta\pi(d+1). \tag{5}$$

The probability of this event is:

$$\begin{aligned} Pr(work|d,\theta) &= Pr(\epsilon_T > \beta(\pi(d+1) - \pi(d)) - \mu) && (6) \\ &= 1.0 - \Phi(\beta(\pi(d+1) - \pi(d)) - \mu), && (7) \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. We denote the vector of unknown parameters governing that probability by θ. The non-linearity in $\pi(d)$ allows identification of μ and β. When $d < 10$, the difference between $\pi(d+1)$ and $\pi(d)$ is zero, and β does not enter the equation. The resulting equation is:

$$Pr(work|d,\theta) = 1 - \Phi(\mu), \tag{8}$$

which is a simple probit. The econometrician observes the left-hand side in the data, and searches for the value of μ that matches that empirical probability as closely as possible. The monotonicity

of Φ ensures that this value is unique. Once μ has been consistently estimated from observations on the behavior of teachers who are "out of the money," we can uniquely identify β in Equation 7 by a similar argument applied to the observations of teachers with $d \geq 10$. Equation 8 also motivates the scale normalization on ϵ: any choice of variance for ϵ can be counteracted by a scaling of μ. Therefore, without loss of generality, we set the variance of ϵ to be equal to one.

If teacher have different μ, the model cannot be identified by simply comparing a cross-section of teachers. However, it remains identified by the fact there is variation in whether teachers are in or out of the money in the last day. Equation 8 can be estimated allowing for a different μ for each teacher.

If ϵ is serially correlated, identification is more complicated. Suppose that the shock follows an AR(1) process:

$$\epsilon_t = \rho\epsilon_{t-1} + \nu_t, \tag{9}$$

where ρ is the persistence parameter and ν_t is a draw from the standard normal distribution. Autocorrelation could be either positive or negative: it would be positive if the teachers stopped coming, for example, because they are ill, and the illness lasts more than one day. It would be negative if teachers do not attend school because they have to accomplish a task (e.g. harvest their fields), but once the task is accomplished, they are done with it for a while.

Irrespective of whether ρ is positive or negative, we can no longer directly apply the partitioning argument above. The reason is that the ϵ_T will be correlated with d, as teachers with very high draws on ϵ_T are more likely to be in the region where $d < 10$ if ρ is positive (the converse will be true if ρ is negative). In this case, the expectation that $\epsilon = 0$ is invalid, and will bias our estimates of α and subsequently β.

Our solution to this problem is to integrate out over the unknown distribution of ϵ when solving the dynamic program. Our identifying assumption, which solves the "incidental parameters" problem, is that at time $t = 1$ in the first month all agents receive an idiosyncratic draw from the limiting distribution of Equation 9, which is given by $\bar{F} = N(0, 1/(1-\rho^2))$. We then solve the dynamic program defined by following:

$$V(t, d, \epsilon_t) = \max\{\mu + \epsilon_t + EV(t+1, d), EV(t+1, d+1)\}. \tag{10}$$

At time T, we have:

$$V(T, d, \epsilon_T) = \max\{\mu + \epsilon_T + \beta\pi(d) + EV(1, 0), \beta\pi(d+1) + EV(1, 0)\}, \tag{11}$$

The difference in these value functions as compared to the iid case is that we now explicitly account for ϵ_t as a state variable. The difference between this dynamic program and Equations 3 and 4 is that the expectation about ϵ_{t+1} depends on ϵ_t through the persistence parameter ρ. We discretize ϵ into 200 states.[21] Conditioning on the initial draw from $\bar{F}$, we can simulate the probabilities of ending up in the regions where $d < 10$ and $d \geq 10$. Once we have used the model to condition on being in these partitions of the state space, we can apply the same logic to identify μ and β as used in the idiosyncratic error case. Intuitively, runs of days (not) worked identify the autocorrelation parameter, and the average number of days worked identifies the utility of the outside option.

3.3.3 Estimation

We estimate several different models of the general dynamic program described by Equations 3 and 4.

Models with i.i.d. Errors The simplest model that we estimate is one where all agents share the same marginal utility of income and average outside option of not working, and the shocks to the utility of not working are i.i.d. The discussion about identification suggests an estimation approach for recovering the unknown parameters in practice.

We use all of the days in the month in our estimation by utilizing the empirical counterpart of Equation 3 for $t < T$:

$$\begin{aligned} Pr(\text{work}|t, d, \theta) &= Pr(\mu + \epsilon_t + EV(t+1, d) < EV(t+1, d+1)) \\ &= Pr(\epsilon_t < EV(t+1, d+1) - EV(t+1, d) - \mu) \\ &= \Phi(EV(t+1, d+1) - EV(t+1, d) - \mu), \end{aligned} \tag{12}$$

where each of the value functions in Equation 12 is computed using backward recursion from period T. The log-likelihood function for the model without serial correlation in the error terms is then:

$$LLH(\theta) = \sum_{i=1}^{N} \sum_{m=1}^{M_i} \sum_{t=1}^{T_m} [1(\text{work}) Pr(\text{work}|t, d, \theta) + 1(\text{not work})(1 - Pr(\text{work}|t, d, \theta)], \tag{13}$$

where each agent is indexed by i, the months they work are indexed by $m = \{1, \ldots, M_i\}$, and the days within each of those months are indexed by $t = \{1, \ldots, T_m\}$. This likelihood is well-behaved and can be evaluated quickly since numerical integration is not necessary.

[21] We arrived at this number by increasing the number of points in the discretization of ϵ until there was no change in the expected distribution of outcomes. For alternative approaches to estimating dynamic discrete choice models with serially-correlated errors, see Keane and Wolpin (1994) and Stinebrickner (2000).

This framework also allows us to relax the assumption that teachers all have the same μ. It is natural to assume that different teachers face different average outside options. We can use the panel structure of the data to identify and estimate the variation in μ across teachers in our sample population using either fixed effects or a random coefficients approach.

Models with Serial Correlation When we allow for serial correlation in the unobserved error term, we have to use the method of simulated moments (MSM) to estimate our unknown parameters. This is considerably more involved, and requires a number of steps.

First, we solve the dynamic programming problem defined by Equations 10 and 11 for an initial guess of our parameters. We then simulate many work histories from this model, forming an unbiased estimate of the distribution of *sequences* of days worked at the beginning of each month. We simulate the model by drawing sequences $\epsilon = \{\epsilon_0, \ldots, \epsilon_t\}$ and following the optimal policy prescribed the dynamic program. For different sequences of ϵ's we obtain different work histories. Repeating this process many times results in unbiased estimates of the probability one observes a given sequence. We then match the model's predicted set of probabilities over these sequences against their empirical counterparts. Denoting a sequence of days worked as A, we form a vector of moment conditions:

$$E[Pr(A; X) - \hat{Pr}(A; X, \hat{\theta})] = 0, \tag{14}$$

where $Pr(A; X)$ is the empirical probability of observing a sequence of days worked conditioning on X, a vector containing the number of holidays and the maximum number of days in that month an agent could potentially work in that month.[22] We form $\hat{Pr}(A)$ through Monte Carlo simulation:

$$\hat{Pr}(A; X, \hat{\theta}) = \frac{1}{M \cdot NS} \sum_{m=1}^{M} \sum_{i=1}^{NS} 1(A_i = A; X_m, \hat{\theta}), \tag{15}$$

where A_i is the simulated work history associated with simulation i, as derived from a dynamic program constructed in accordance with the parameters $\hat{\theta}$ and the characteristics of the month. The number of simulations used to form the expected probability of observing a sequence of days worked is denoted as NS. In all of our estimations we use $NS = 200,000$. Note that we are also drawing ϵ_1 anew from the distribution $\bar{F}$ for every simulated path, where we keep track of the seeding values in the random number generator, as to ensure that the function value is always the

[22] This is necessary since the maximum payoff a teacher could obtain varies across months with the length of the month and the number of holidays in that month, which count as a day worked in the bonus payoff function if they fall on a workday.

same for a given $\hat{\theta}$.[23] The objective function under the method of simulated moments is:

$$\min_{\theta} \left[\sum_{i=1}^{n} g(X_i, \theta) \right]' \Omega^{-1} \left[\sum_{i=1}^{n} g(X_i, \theta) \right], \tag{16}$$

where $g(X_i, \theta)$ is the vector of moments formed by stacking the $2^N - 1$ moments defined by Equation 14, and Ω^{-1} is the standard two-step optimal weighting matrix. For more details concerning the implementation and asymptotic theory of simulation estimators, see McFadden (1989) and Pakes and Pollard (1989).

Matching sequences of days worked from the first N days in each month produces $2^N - 1$ linearly independent moments, where we subtract one to correct for the fact that the probabilities have to sum to one. In our estimation, we match sequences of length $N = 5$, which generates 31 moments. Experimentation with shorter and longer sequences of days worked did not result in significant changes to the coefficients.

We also relax the assumption that the outside option is equal across all agents in the model with autocorrelated errors. We model the distribution of mean levels of the outside option to each agent in each month, μ_{it}, as being drawn from an unknown distribution denoted by $G(\mu)$. When forming moments in the MSM estimator in Equation 16 we need to integrate out this unobserved heterogeneity. The modification to the expected probability of observing a sequence of actions, A_i, is then:

$$\hat{Pr}(A; X, \hat{\theta}) = \frac{1}{M \cdot NS \cdot U} \sum_{m=1}^{M} \sum_{i=1}^{NS} \sum_{u=1}^{U} 1(A_i = A; X_m, \hat{\theta}_1, u), \tag{17}$$

where u is a draw of the mean level of the outside option from $G(\hat{\theta}_2)$, the unknown distribution of heterogeneity in the population. In practice we set $U = 200$. For clarity, we have partitioned the set of unknown parameters into $\hat{\theta}_1 = \{\beta, \rho\}$ and $\hat{\theta}_2$, the set of parameters governing the distribution of unobserved heterogeneity. Note that this model is slightly different than the fixed-effects model considered in the iid case above, as it allows the draw of the outside option to vary across both months and agents. On the other hand, the distribution of unknown heterogeneity in the fixed effects model is estimated nonparametrically.

We estimate three different models with unobserved heterogeneity which differ through the specification of $G(\hat{\theta}_2)$. In the first model $G(\hat{\theta}_2)$ is distributed normally with mean and variance

[23] The conditional distribution of ϵ in future months still depends on the actions taken today, the crux of the "incidental parameters" problem. However, the dependence across the first five days of one month to the first five days of the following month is quite weak, even with high values of ρ. In principle, we can test this restriction by estimating the model on just the first month's worth of work sequences.

$\hat{\theta} = \{\mu_1, \sigma_1^2\}$. In the second model, our preferred specification, we allow for a mixture of two types, where each type is distributed normally with proportion p and $(1-p)$ in the population. The third model is identical to the second model except that we remove the autocorrelation in the error structure for comparison.

3.3.4 Results

We present these results in Table 6. We present the main parameters of the model (β, μ, σ, and other parameters as relevant), as well as the implied labor supply elasticity (percentage increase in days worked caused by a one percent increase in the value of the bonus) and the semi-elasticity with respect to the bonus cut-off (percentage increase in days worked in response to an increase in one day in the minimum number of days necessary for a bonus).

The parameters in column 1 (Model I) correspond the simplest iid model where all teachers have the same marginal utility of income, β, and mean opportunity cost of working, μ. We precisely estimate a β of 0.049 and a μ of 1.550. The combination of these parameters leads the teachers to be very sensitive to changes in the financial incentives: the labor supply elasticity with respect to the bonus is 3.52%. Even more stark is the response to a one-day increase in the number of days before the incentives kick in; increasing the cutoff to eleven days instead of ten days reduces the number of days worked by 75.49 percent. These parameters would however be biased in the presence of teacher heterogeneity or auto-correlated shocks.

Model II, which introduce teacher fixed effects to the basic framework, suggest that this may be the case. The estimate of the marginal utility of income is much lower (0.024), but remains highly significant. The implied elasticity of labor supply with respect to the size of the bonus is now 1.687, which is slightly lower than it was in Column 1, although it is still high. The estimated responsiveness to an increase in the minimum number of days (-16%) is much lower, suggesting once again that agent heterogeneity is an important part of the story.

As discussed above, one may be concerned that the shocks to the opportunity cost of going to school may be serially correlated. In Model III, we introduce autocorrelation into the structure of Model I, without allowing for any heterogeneity in the teacher population. We precisely estimate a marginal utility of income of 0.059, which is slightly higher than Model I. The mean level of the outside option is estimated to be 2.315, also higher than Model I. We estimate a strong positive autocorrelation in the shock to that outside option, $\rho = 0.682$. These parameters suggest estimated elasticities of 6.225 and -50% with respect to the bonus and bonus cutoff, respectively. These elasticities are implausibly high, most likely because we have not taken heterogeneity into account.

The contrast between Models II and III illustrates the differences between teachers with highly autocorrelated errors and teachers with persistently high or low outside options. Either factor can result in low or high average numbers of days worked in a month. However, with highly autocorrelated errors, teachers tend to work or not work in long bunches. Model II attributes all the persistence to differences among teachers, while model III attribute all the persistence to auto-correlation. To further disentangle the effect of unobserved heterogeneity from autocorrelation, Models IV and V (Columns 4 and 5 respectively) add unobserved heterogeneity in the outside option, in a model with auto-correlation.

In Model IV, we specified that teachers receive a draw of their outside option for the month from a normal distribution with unknown mean and variance. We estimate a low variance of the outside option ($\sigma_1^2 = 0.001$), although this is imprecisely estimated. Not surprisingly then, within sampling and numerical error, Model III and Model IV appear very similar. Both models produce similar elasticities. The high variance in the estimated σ_1 suggests that heterogeneity may not have been well taken into account by drawing the outside option from the same normal distribution. Furthermore, the model (as well as the previous one we estimated, except for the fixed effect model) fails to capture an important feature of the data: there is a non-negligible share (about 2%) of the teacher-month observation where the teacher has worked exactly zero days.

Thus, in Model V, we extend Model IV to allow for two types of heterogeneity. By allowing the distribution of outside options in the data to be distributed as a mixture of normals, we can empirically investigate the hypothesis that there are "slackers" as well as hard working teachers. The results of Model V strongly suggest that there are (at least) two types of workers in the data. The estimated marginal utility of income is significantly lower than the previous models, $\beta = 0.014$. While there is still strong evidence of auto-correlation in the data, the estimated ρ is a little lower (0.461).

The average level of the outside option for the first group is much lower than that of Models III and IV, $\mu_1 = -0.107$, with a slightly higher variance of $\sigma_1^2 = 0.153$. Thus, we estimate that a majority of teachers derive some utility of going to school in most months (they have a negative outside option). On the other hand, we estimate that 4.7 percent (p) of the population draws an outside option from a distribution with much higher outside options, characterized by $\mu_2 = 3.616$ and variance $\sigma_2^2 = 0.260$. The elasticity with respect to the bonus declines to 0.306 (though it remains highly significant), and the semi-elasticity with respect to the bonus cutoff falls to -1.29%.

Model V thus shows that taking teacher heterogeneity into account is important. To get a feel for the importance of the serially-correlated errors, we estimated Model V with the AR(1) process

turned off. The results are given by Model VI (Column 6). Without serial correlation in the error process, we estimate that the marginal utility of income is higher, $\beta = 0.019$, than in Model V. The estimate of β moves closer to what was estimated in Model II, which is intuitive: we are now attributing some of the correlation in the data, which was due to the auto-correlation, to the response to the incentive. The two peaks of the unobserved heterogeneity move closer to each other, and variance of both types remains about the same. The model estimates that the proportion of the population drawing from the slacker type is now 13.1 percent.

In summary, the estimation shows that taking into account both auto-correlation in the structure of shocks *and* heterogeneity in teacher types is important when estimating the model. Our preferred specification (Model V) leads to estimates of labor supply elasticity that are reasonable (of the order of 0.3), and highly significant.

3.3.5 Policy Rules Simulations and Out of Sample Tests

To provide a sense of the fit of each model, we report the predicted number of days worked under each specification. This is not a good test for Models I and II, which use all the days worked to compute the parameters of the model, and should therefore do a good job of matching the average number of days worked. However, note that this is not a parameter that our method of simulated moments estimation tried to match (since we matched only the first 5 days of teacher behavior), so it provides a good test of goodness of fit for these models. The actual average days worked was 20.16. Our preferred specification (Model V) matches this figure very well: it predicts 20.23 days worked (even closer than the Model II, which predicted 19 days). We can see that the other specifications fit the data much less well, again underscoring the importance of accounting for both auto-correlation and heterogeneity.

Figure 6A plots the density of days worked predicted by model V, and its 95% confidence interval, and compares it to the actual density observed in the data.[24] Since the estimation is not calibrated to match this shape (we are only using the history of the first few days in our estimation), the fit is surprisingly good. The model reproduces the general shape of the distribution, although the mode of the distribution in our predicted fit is to the left of the mode in the data. The truth is always within the 95% confidence interval of the prediction.

A unique feature of this experiment is that we have two real out-of-sample tests for the fit of the model. A first test is to compute would happen if the incentive was brought to zero. The estimates vary significantly from one model to the next, so we focus on our preferred specification

[24] The 2.5% line is not particularly informative since it is 0 all the time.

(Model V). This model suggests that the teachers would work on average about half the time if there were no incentives (13.5 days, or 52 percent). Of course, we lack data on daily attendance for the comparison group. However, using the random check data, we estimate that the control group teachers are present 58 percent of the times (Table 3). The random check data estimates that the treatment group teachers are present 79 percent of the time (similar to the 78 percent predicted by our model), or 21 percent more than the comparison group. Therefore, the predicted number of days for both the treatment and the control group and their difference (26 percent) are quite close to the actual numbers we observe, and are statistically indistinguishable. This suggest that the effect of the program on the difference between treatment and control is entirely explained by the response to the financial incentives. In contrast, the other models severely under predict the number of days worked with no incentives, perhaps due to the fact that they over-estimate the elasticity.

A change in the incentive system at Seva Mandir, after the first version of this paper was written and our model was estimated, provides us with another very nice counterfactual experiment. In December 2006, Seva Mandir increased the minimum monthly payment to Rs 700, which teachers receive if they work 12 days or less (rather than 10 days). For each additional day they work, teachers earn an additional Rs 70 per day. This policy change pulls the teachers in two directions: on the one hand, the zone during which there are no incentives increases, which should reduce work effort. On the other hand, the incentive per day worked also increases, which should increase work effort.[25] Seva Mandir provided us with the camera data in the summer 2007, a few months after the change in policy. The average number of days worked since January 2007 increased very slightly, from 20.17 to 20.4 days. The predicted number of days worked for each model is reported in the last row in Table 6. Here again, our preferred specification (Model V) performs very well: it predicts 20.86 days worked under the new incentive scheme, a small increase from the predicted number of days under the main scheme (as in the actual data). Figure 6B shows the actual distribution of days worked and the predicted one. If anything, the model does an even better job of predicting the distribution of days worked in this out of sample test than in the original data. In contrast, the other models continue to perform rather poorly. In particular, the iid models without fixed effects over-predicts this counterfactual (which is a result of its high estimated elasticity), in fact predicting more days worked than are available in most months (including holidays etc).

[25] We suggested that this change was not necessarily the optimal one, but Seva Mandir, who increased the salary of its para-worker across the board to Rs 1,400 were keen to increase the number of day necessary to obtain a "full salary" (which someone now gets if they work 22 days).

3.3.6 Summary

This section has set up and estimated a simple structural model, which posits utility maximization under a non-linear incentive schemes, ignoring discounting, income effect, loss of motivation due to the monitoring scheme, or "focal" number of days. Estimating this model taking into account both heterogeneity and auto-correlation proved to be a complex exercise in practice, but the results show that the model fits both the actual number of days worked under with the incentive scheme and the number of days worked under a new incentive schemes later imposed by Seva Mandir. Interestingly, our preferred specification suggests that the monitoring system *per se* did not affect behavior, either positively or negatively (the behavior of the control group is similar to what would have been predicted by setting the incentives to zero in our model). Teachers attended school more often because the incentives made a difference.

4 Was Learning Affected?

The program was effective in increasing the teachers' attendance, mainly because teachers seem to have responded to the financial incentives. In this section, we examine its impact on actual learning. First, we explore whether teachers decreased their effort while in school to compensate for the fact that they now have to attend more regularly (section 4.1). Second, we look at student presence (section 4.2) and child learning (section 4.3). Finally, in section 4.4, we provide a rough cost-benefit analysis of the program in terms of the cost per additional kid-day in school and per additional standard deviation of test score.

4.1 Teacher Behavior

Though the program increased teacher attendance and the length of the school day, it could still be considered ineffective if the teachers compensated for increased attendance by teaching less. We used the activity data that was collected at the time of the random check to determine what the teachers were doing once they were in the classroom. Since we can only measure the impact of the program on teacher performance for schools that were open, the fact that treatment schools were open more may introduce selection bias. That is, if teachers with high outside options (who are thus more likely to be absent) also tended to teach less when present, the treatment effect may be biased downward since more observations would be drawn from among low-effort teachers in the treatment group than in the comparison group. Nevertheless, Table 7 shows that there was no significant difference in teacher activities: across both types of schools, teachers were as likely to be

in the classroom, to have used the blackboard, and to be addressing students when the enumerator arrived. This does not appear to have changed during the duration of the program.

The fact that teachers did not reduce their effort in school suggests that the fears of multitasking and loss of intrinsic motivation were perhaps unfounded. Instead, our findings suggest that once teachers were forced to attend (and therefore to forgo the additional earnings from working elsewhere or their leisure time), the marginal cost of teaching must have been small. This belief was supported during in-depth conversations with 15 randomly selected NFE teachers regarding their teaching habits in November and December 2005. We found that teachers spent little time preparing for class. Teaching in the NFE follows an established routine, with the teacher conducting the same type of literacy and numeracy activities every day. One teacher stated that he decides on the activities of the day as he is walking to school in the morning. Other teachers stated that, once they left the center, they were occupied with household and field duties, and, thus, had little time to prepare for class outside of mandatory Seva Mandir training meetings. Furthermore, despite the poor attendance rates, many teachers displayed a motivation to teach. Teachers stated that they felt good when the students learned, and liked the fact they were helping disadvantaged students get an education. Plus, most stated that they actually liked teaching, once they were in the classroom: "The day is all teaching, so I just try to enjoy the teaching."

The teachers' general acceptance of the incentive system may be an additional reason why multitasking was not a problem. Several months into the program, teachers filled out feedback forms. Seva Mandir also conducted a feedback session at their bi-annual sessions, which were attended by members of the research team. Overall, teachers did not complain about the principle of the program, although many teachers had some specific complaints about the inflexibility of the rules. For example, many did not like the fact that a day was not valid even if a teacher was present 4 hours and 55 minutes (the normal school day is six hours, but an hour's slack was given). Others stated that assembling eight children on time at the beginning of the day is difficult, or that they disliked the fact that the program did not plan for sick leave or leave for extenuating circumstances, such as a funeral. On the other hand, many felt empowered by the fact that the onus of performing better was actually in their hands: "Our payments have increased, so my interest in running the center has gone up." Others described how the payment system had made other community members less likely to burden the teacher with other responsibilities once they knew that a teacher would be penalized if he did not attend school. This suggests that the program may actually have stronger effects in the long run, as it signals a change in the norms of what teachers are expected to do.

4.2 Child Presence

On the feedback forms, many teachers claimed that the program increased child attendance: "This program has instilled a sense of discipline among us as well as the students. Since we come on time, the students have to come on time as well." Unfortunately, conditional on whether a school was open, the effect of the program on child attendance cannot be directly estimated without bias, because we can only measure child attendance when the school is open. For example, if schools that were typically open also attracted more children, and the program induced the "worst" school (with fewer children attending regularly) to be open more often in the treatment schools than in the comparison schools, then this selection bias will tend to bias the effect of the program on child attendance downwards. The selection bias could also be positive, for example if the good schools generally attract students with better earning opportunities, who are more likely to be absent, and the "marginal" day is due to a weak schools catering to students with little outside opportunities. Selection bias is a realistic concern (and likely to be negative) since, for the comparison schools, there is a positive correlation between the number of times a school is found open and the number of children found in school. Moreover, we found that the effect of the program was higher for schools with originally weak teachers, which may attract fewer children.

Keeping this caveat in mind, child attendance was not significantly different in treatment and comparison schools. In Table 8, we present the child attendance rates in an open school, by treatment status (Panel A). An average child's attendance rate was the same in treatment and comparison schools (46 percent). Excluding children who left the NFE, child attendance is higher overall (62 percent for treatment and 58 percent for comparison schools), but the difference is not significant.

However, treatment schools had more teaching days. Even if the program did not increase child attendance on a particular day, the increase in the number of days that the school was open should result in more days of instruction per child. The program impact on child instruction time is reported in Panel B of Table 8. Taking into account days in which the schools were closed, a child in a treatment school received 9 percentage points (or 30 percent) more days of instruction than a child in a comparison school. This corresponds to 2.7 more days of instruction time a month at treatment schools. Since there are roughly 20 children per classroom, this figure translates into 54 more child-days of instruction per month in the treatment schools than in comparison schools. This effect is larger than that of successful interventions that have been shown to increase child attendance, such as the PROGRESA program of conditional cash transfers, which increased

enrollment by 3.4 percent in primary schools and had no impact on attendance (Schultz, 2004); de-worming, which increased attendance by 7.5 percentage points (Miguel and Kremer, 2004); a child incentive program (Kremer, Miguel, and Thornton, 2004), which increased attendance by 5 percentage points; and a child scholarship program, which increased attendance by 8 percentage points (Kremer et al., 2004). The effect is comparable to that of adding a second teacher in Seva Mandir NFEs (Banerjee, Jacob and Kremer, 2005), which increased the number of days of instruction per month by 3.1. The effect on presence does not appear to be affected by student ability (proxied by the whether or not the child could take a written test in the pre-test). While presence increased slightly more for those who could not write prior to the program (14 versus 10 percentage points), this difference is not significant.

In summary, since children were as likely to attend class on a given day in the treatment schools as in the comparison schools, and because the school was open more often, children received significantly more days of instruction in the treatment schools. This finding suggests that the high teacher absence rate we observed is not likely to be the efficient response to a lack of interest by the children: if it were the case that children came to school 55 percent of the time because they could not afford to attend more than a certain number of days, then we would see a sharp reduction in child attendance in treatment schools on days where the school was open. On the other hand, we do not see a sharp increase in the attendance of children in the treatment schools. This suggests that either the teacher absence rate is not the main cause of the children's irregular attendance, or that the children have not yet had time to adjust. The latter explanation is not entirely plausible, however, since the program has now been in place for over two years, and we do not see a larger increase in the attendance of children in the later periods than in the earlier period.

4.3 Child Learning

Children in the treatment schools, on average, received about 30 percent more instruction time than children in the comparison schools, with no apparent decline in teacher effort. Some, however, argue that because para-teachers are less qualified than other teachers, it is not clear that they are effective despite the support and in-service training they get from NGOs like Seva Mandir. If para-teachers are indeed ineffective, the fact that it is possible to induce them to attend school more is not particularly relevant for policy. Understanding the effect of the program on learning is, therefore, critical.

4.3.1 Attrition and Means of Mid- and Post-Test

Before comparing test scores in the treatment and comparison schools, we must first ensure that selective attrition does not invalidate the comparison. There are two possible sources of attrition.[26] First, some children leave the NFEs, either because they drop out of school altogether or because they start attending regular primary schools. Second, some children were absent on testing days. To minimize the impact of attrition on the study, we made considerable attempts to track down the children (even if they had left the NFE to attend a formal school or had been absent on the testing day) and administered the post-test to them. Consequently, attrition was fairly limited. Of the 2,230 students who took the pre-test, 1,893 also took the mid-test, and 1,760 also took the post-test. Table 9 shows the attrition rate in both types of schools, as well as the characteristics of the attriters. At the time of the mid-test, attrition was higher in the comparison group than in the treatment group. At the time of the post-test, attrition was similar across both groups, and children who dropped out of the treatment schools were similar in terms of test scores to those that dropped out of the comparison schools.

Table 9 also provides some simple descriptive statistics, comparing the test scores of treatment and comparison children. The first row presents the percentage of children who were able to take the written exam, while subsequent rows provide the mean exam score (normalized by the mid-test comparison group). Relative to the pre-test and mid-test, many more children, in both the treatment and comparison schools, were able to write by the post-test. On the post-test, students did slightly worse in math relative to the mid-test comparison, but they performed much better in language.

Table 9 also shows the simple differences in the mid- and the post-test scores for students in the treatment and comparison schools. On both tests, in both language and math, the treatment students did better than the comparison students (a 0.16 standard deviation increase and 0.11 standard deviations in language at the post-test score), even though the differences are not significant at a 95 percent confidence level. Since child test scores are strongly auto-correlated, we obtain greater precision by controlling for the child's pre-test score level.

[26] As mentioned earlier, 7 centers closed down or failed to open prior to the start of the program. These closures were unrelated to the program, and equally distributed among treatment and comparison schools. We made no attempt to test the children from these centers in the pre-test.

4.3.2 Test Results

In Table 10, we report the program impact on the mid-test and the post-test scores. We compare the average test scores of students in the treatment and comparison schools, conditional on a child's pre-program competency and preparedness level. In a regression framework, we model the effect of being in a school j that is being treated (Treat_j) on child i's test score (Score_{ijk}) on test k (where k denotes either the mid- or post-test exam):

$$\text{Score}_{ijk} = \beta_1 + \beta_2 \text{Treat}_j + \beta_3 \text{Pre\_Writ}_{ij} + \beta_4 \text{Oral\_Score}_{ij} + \beta_5 \text{Written\_Score}_{ij} + \varepsilon_{ijk}. \quad (18)$$

Because test scores are highly autocorrelated, controlling for a child's test scores before the program increases the precision of our estimate. However, the specific structure of the pre-test (i.e. there is not one "score" on a comparable scale for each child because the children either took the written or the oral test in the pre-test) does not allow for a traditional difference-in-difference (DD) or "value added" (child fixed effect) strategy. Instead, we include a variable containing the child's pre-test score for the oral test if he took the oral pre-test and 0 otherwise (Oral_Score_{ij}), the child's pre-test score on the written test if he took the written test and 0 otherwise ($\text{Written\_Score}_{ij}$), and an indicator variable for whether he took the written test at the pre-test (Pre_Writ_{ij}).[27] This fully controls for the child pre-test achievement, and is thus similar in spirit to a DD strategy. Standard errors are clustered by school. Each cell in Table 10 represents the treatment effect (β_2) obtained in a separate regression. For ease of interpretation, the mid-test results (Columns 1 to 4) and post-test results (Columns 5 to 8) are expressed in the standard deviation of the distribution of the mid-test score in the comparison schools.[28]

The tables reveal that the program had a significant impact on learning, even as early as the mid-test. Children in treatment schools gained 0.16 standard deviations of the test score distribution in language, 0.15 standard deviations in math, and 0.17 overall (Panel A). Including controls for school characteristics—location, teacher test scores, and the infrastructure index of school—does not significantly change our findings (Panel B). Children who could write at the time of the pre-test gained the most from the program. For example, they had mid-test test scores 0.25

[27] At the pre-test, children were given either the oral or the written score. At the mid- and post-test, every child took the oral part, and every child who could write took the written exam (all children were given a chance to try the written exam; if they could not read, they were given a zero for the written test).

[28] Scores are normalized such that the mean and standard deviation of the comparison group at the time of the mid-test exam is zero and one, respectively. (Specifically, we subtract the mean of the comparison group in the pre-test, and divide by the standard deviation.) This allows for comparison across samples, as well as with the results from other studies. We could not normalize with respect to the pre-test score distribution since not every child took the same test at the pre-test.

standard deviations higher in treatment schools than in comparison schools (Panel D). Interestingly, the children who could write at the time of the pre-test do not increase their attendance rate in response to the greater teacher attendance rate relatively more than those who could not write at the time of the pre-test. Therefore, it is not that they have relatively more days of schooling than the students who could not write as a result of the program, but rather that they seem more equipped to make the most out of the additional days of schooling that they receive.

We compare the impact of the program on girls versus boys in Panels E and F. Girls gained as much, if not more, from the program as boys. On the mid-test, 7 percentage-points more of girls in the treatment schools were able to write relative to the comparison schools, compared to only 2 percentage-points of boys (this five percentage point difference is significant).

The differences between students in the treatment and comparison schools persisted in the post-test (Columns 5 to 8). Children in treatment schools gained 0.21 standard deviations in language, 0.16 in math, and 0.17 overall (Panel A). Similar to the mid-test, much of the gains came from children who could write at the time of the pre-test. The post-test also suggests that girls gained slightly more from the program than the boys, but these differences are not significant. The treatment effect of 0.17 standard deviations is similar to other successful educational interventions, such as the Tennessee Star experiment in the United States (Kruger and Whitmore, 2001), the Balsakhi Remedial Education Program in India during its first year (Banerjee, et al., 2005), and an incentive program for girls in Kenya (Kremer, Miguel and Thornton, 2004).

Finally, we examined other sources of heterogeneity in program effect (school infrastructure, teacher level of education, teacher proficiency measured by the test scores). The results (omitted for brevity, but available from the authors), suggest little or no heterogeneity in program effect along these dimensions.

4.3.3 Leaving the NFE

NFEs prepare children, who might not otherwise attend school, to enter government schools at the age-appropriate grade level. To enter the government schools, children must demonstrate proficiency for a grade, either by passing an exam or through vetting by a government teacher. The ability of his students to join government schools is, therefore, a strong signal of success for a NFE teacher. The program increased the number of children graduating to the government schools. As shown in Table 11, 26 percent of students in the treatment schools graduated to the government schools, compared to only 16 percent in the comparison schools (by February 2006). This 10 percentage point difference implies a 62 percent increase in the graduation rate and is

significant.

In the final row of Table 11, we present the dropout rates for children who left school entirely (i.e. left the NFE and did not join a government school). The dropout rate is slightly lower for the treatment schools, but this difference is insignificant.

4.3.4 Estimating the Effect of Teacher Presence on Learning

The previous sections presented the reduced form analysis of the effect of the incentive program on child learning. Table 12 interprets what these estimates can tell us about the impact of teacher attendance.[29] Columns 1 to 3 report simple correlations between the teacher attendance rate and the child test scores. Specifically, we report the coefficient estimate of the number of times a school was found open (Open_j) on a regression of either the mid-test or post-test scores:

$$\text{Score}_{ijk} = \beta_1 + \beta_2 \text{Open}_j + \beta_3 \text{Pre\_Writ}_{ij} + \beta_4 \text{Oral\_Score}_{ij} + \beta_5 \text{Written\_Score}_{ij} + \varepsilon_{ijk}. \tag{19}$$

We continue to control for the child's pre-test score and to cluster standard errors by school.

Column 1 reports OLS estimation of Equation 2 for the comparison schools. In this case, the random check data are used to estimate the number of times a school is found open. The coefficient is 0.20, indicating that the test scores of children in centers open 100 percent of the time would be 0.10 standard deviations higher than those of children in a center open 50 percent of the time. Note that this coefficient is insignificant.

This point estimate is similar to those reported in other studies (Chaudhury, et al., 2005a) and, taken at face value, would imply that the effect of teacher attendance on learning is not that large. Chaudhury et al. (2005a) conjectures that the measurement of absence rates based on a few random visits per school have considerable error, and may thus bias the results downwards. Consistent with this theory, the effect on the post-test scores, where having more months of random check data allows us to better estimate the absence rate per school, becomes larger (0.58 standard deviations). Our study provides a much more direct test of this hypothesis, since, for treatment teachers, the photograph data gives us the actual attendance. We present the OLS estimate of the effect of attendance for treatment teachers using the random check data (Column 2) and camera data (Column 3). Overall, the effect of teacher attendance is larger in the treatment schools than the comparison schools (0.39 in Column 2 to 0.20 in Column 1, both obtained with random check data). More interestingly, consistent with the measurement error hypothesis, the effect of teacher

[29] This estimate are the effect of being present at a random check, which combines the effect of having come at all, and having come for a longer time.

attendance is larger and much more significant when using the more accurate measure of attendance from the camera data, especially for the mid-test scores (the estimate is 0.87 standard deviations in the Column 3 as compared to 0.39 in Column 2). For the post-test, where we have a much more accurate measure of attendance from the random check data, the results from the two methods are similar (0.98 in Column 3 versus 1.17 in Column 2).

Finally, in Column 4, we pool both samples and instrument Open_j (as measured by the random check) with the treatment status of the school to obtain exogenous variation in the percentage of time the school was found open. Since we have shown that the program had a direct effect on the length of the school day, as well as whether or not the school opened at all, the 2SLS estimate captures the joint effect of outright absence and of a longer school day. The 2SLS estimates are higher than the OLS results found in Column 1, and they are indistinguishable from the OLS results in Column 3, obtained with the precisely measured absence rate. This suggests that the relatively low correlation between teacher absence and test scores that was observed in previous studies is indeed likely to be due to measurement error in the teacher absence data. The more precise IV estimates suggest that even a 10 percentage point reduction in the absence rate would result in a 0.10 standard deviation increase in child test scores.

Extrapolating these estimates (which must be done with caution, since the local effect may be different from the overall effect), we can conclude that the effect of being enrolled in an NFE for a year with a teacher present every day is about one standard deviation. This point estimate is similar to the effect of attending remedial education classes with a para-teacher for one year in urban India for children who are enrolled in regular primary school, but have not yet achieved basic numeracy or literacy (Banerjee et al., 2005). In Banerjee et. al (2005), the point estimate was 1.12 standard deviation. Both of these studies therefore suggest that, para-teachers can be extremely effective teachers, at least when an NGO provides them with proper training.

4.4 Cost-Benefit Analysis

This paper shows that a straightforward monitoring and incentive program can effectively reduce teacher truancy. The benefits of this type of program, relative to its costs, are high, and comparable to other successful education programs in developing countries that have been evaluated with randomized evaluations.

Table 13 presents an estimate of the program's administrative costs for one year. The average teacher salary was nearly Rs 1,000 in the treatment schools. Since the flat salary paid to comparison teachers was also Rs 1,000, the program did not increase the expenditures on teacher salaries. Other

program costs (administration, developing the photographs, and buying the cameras) amounted to Rs 5,379 per center per year. This cost corresponds to 40 percent of a teacher's yearly salary, but only Rs 268 ($6) per child per year (assuming about 20 children per teacher).[30] Expressed in terms of cost per outcome, this program cost approximately 11 cents for each additional instruction day per child, $60 per additional school year, and $3.58 per 0.10 standard deviations of increased test scores.

The cost per standard deviation increase in test scores compares favorably to most education programs evaluated through randomized evaluation (see Kremer, Miguel and Thornton (2004) for cost-effectiveness comparison of programs in Africa). It is more expensive than a remedial education program evaluated in India (Banerjee et al., 2005), but that program was conducted in urban India, where the monitoring costs are substantially cheaper.

The cost per additional year of schooling of the camera program is much higher than that of the de-worming program in Africa (evaluated to be only $3.53 per additional year of schooling), but lower than that of any other programs evaluated in Africa, such as a child incentive program ($90 per extra year) or a child sponsorship program that delivered uniforms to children ($99 per extra year).[31] It is also just over half of the cost of the two-teacher program, previously implemented in Seva Mandir in India, which, evaluated at the current teacher's salary, cost $115 per extra year of schooling.

Thus, the camera program is a relatively cost-effective program, both in terms of increasing instruction time and in terms of increasing learning (even in its pilot form, which used an expensive way to develop the photographs).

Finally, these estimates suggest that the long-run returns to the program are quite high. Duraisamy (2000) estimates an 8 percent return from an additional year of primary school in India. The camera program increased the school year by 0.09, and therefore, we expect a rate of return to wages of 0.72 percent from the program. In 2000, GDP per capita in India was $2683 (Penn World Tables). To calculate the effect on the net present value of discounted wages, we assume that

[30] This estimate does take into account the opportunity cost for teachers and children. Note, however, that the effects are larger than they could be if the program was implemented on a large scale, and more cost-effective technology (such as digital cameras) could be used.

[31] In making this comparison, it is worth noting that Kremer and Miguel (2004) use the cost of the de-worming program if implemented on a large scale, whereas we use the cost of the program as implemented in this small scale pilot. However, the cost of the program they actually evaluated was only about three times larger than what they used for the cost-benefit evaluation, which still makes the de-worming program a more cost-effective way to improve instruction time.

The cost per year of the PROGRESA program in primary schools is substantially larger ($5,902.45). However, the PROGRESA program is primarily a transfer program to families, and its cost effectiveness should probably not be based on its effect on school outcomes alone.

sixty percent of the output per worker in India is wages and that wage gains from higher school participation are earned over forty years and discounted at ten percent per year. This results in a long-run wage increase of $125. From this, we subtract both the opportunity costs of the children and of the teachers. Assuming that children are half as productive as adults, children would incur an opportunity cost of $104. Teachers would incur an opportunity cost of approximately $10 per child ($209/20 children). Under these assumptions, the program increases the net present value of wages by about $11. Given the program costs of $6 per child, we find a benefit-cost ratio of 1.83.

5 Conclusion

Addressing the startlingly high rates of teacher absenteeism is crucial to improving the quality of education in developing countries. School systems have often failed to carry out their own rules regarding teacher attendance and, in practice, teachers are rarely penalized for unexcused absences. Therefore, teachers face few incentives to attend class.

In this paper, we show that direct monitoring, combined with simple and credible financial incentives based on teacher attendance, leads to large increases in attendance among para-teachers in informal schools. The program studied in this paper reduced the teacher absence rate from an average of 42 percent in the comparison schools to 21 percent in the treatment schools, without affecting teacher effort while in school. As a result, students in program schools benefited from about 30 percent more instruction time. The program had a statistically and economically significant impact on test scores: after one year, child test scores in program schools were 0.17 standard deviations higher than in comparison schools. Children were also much more likely to be admitted to government schools. Despite being implemented on a small scale, the program was very cost-effective.

The paper offers a methodological contribution as well as a substantive one. On the methodological front, the paper demonstrate the power of combining structural estimation with carefully controlled randomized experiments. The incentive scheme studied in this paper gave rise to an intuitive dynamic programming problem. A regression discontinuity design analysis provides the first indication that these dynamic incentives are taken seriously by the teachers. The model was estimated using the rich daily attendance data generated by the camera program. While the framework is simple, the estimation of a model allowing both heterogeneity and auto-correlation was a challenging empirical exercise, requiring the use of the methods of simulated moments. A very nice feature of our setup is that we could perform two out of sample tests of the model: a comparison

with the predicted impact of the program with the experimental results (something that is only possible when structural models are used in tandem with randomized evaluations) and a comparison of the predicted model effect with a changes in rules that was implemented by Seva Mandir, and for which we obtained the data *after* fitting the model on the original experimental data. This provides strong validation of our results, and suggests that it is very important to account for both auto-correlation and heterogeneity. Only the model that incorporates these two features does well in the out of sample tests.

On the substantive front, these results suggest that extending Seva Mandir's incentive program to the many non-formal schools who are serving an increasingly large number of children in India has the potential to increase learning levels. The question arises as to whether the program can be instituted for regular teachers in government schools. Teachers in government schools are often more politically powerful than para-teachers in informal or private schools. Thus, it may prove difficult to institute a system in which government teachers would be monitored daily using a camera or similar device (such as a date-time stamps), and other methods may prove necessary. However, our findings suggest that the barriers currently preventing teachers from attending school regularly (e.g. distance, other activities, lack of interest by children) are not insurmountable. Given political will, it is possible that solutions to the absence problem could be found in government schools as well. A sign that these results did indeed generate interest in the government is that Seva Mandir was awarded the annual Government of India "Digital Learning Award" of 2007 for this project.

A recent experiment, the results of which are reported in Banerjee, Duflo and Glennerster (2007) demonstrates the external validity of these results outside the NGO context, as well as the difficulty of extending such programs in government settings. Following a discussion of the results of the cameras program, the government of Rajasthan put in place a similar system for government nurses, whose absence rate is as bad as that of NFE teachers (the absence rate in a baseline survey was 44 percent (Banerjee, Deaton and Duflo, 2004)). The nurses were monitored using time and date stamps, one to three days per week depending on the type of nurse. The monitoring system was administered by independent monitors, but the data was provided to the government, which was responsible for administering punishments and bonuses. The announced incentive system was severe: it called for a 50 percent reduction in the pay of nurses who were absent 50 percent of the time, and termination of persistently absent nurses. In the first few months, when these punishments were indeed carried out, the program was very effective, leading to a reduction of about 50 percent in absence rate on monitored days (similar to the camera program for teachers).

However, the local administration stopped enforcing the incentives after a few months (although the monitoring did continue), in particular by granting a large number of ex-post "exemption." This led the absence rate in the treatment group to quickly converge to that of the control group. This experiment confirms that regular monitoring is effective, but only when coupled with real incentives, as is suggested by the results of our structural model.

The program for nurses suggests that barriers exist to the implementation of real incentive systems for government teachers (Kremer and Chen (2001) find similar results for teachers in Kenya). However, the findings in this paper also imply that para-teachers can be effective teachers. If implementing strict monitoring within the government system turns out to be impossible, our results provide support for the policy of many developing countries to increasing teaching staff through the hiring of para-teachers.

6 Bibliography

Attanasio, Orazio, Costas Meghir and Santiago (2006),"Education Choices in Mexico. Using a Structural Model and a Randomised Experiment to Evaluate Progresa," MIMEO, UCL.

Banerjee, Abhijit, Rukmini Barnerji, Esther Duflo, Rachel Glennerster, Stuti Khemani (2007), "The Impact of Information, Awareness and Participation on Learning Outcomes: Evidence from a Randomized Evaluation" MIMEO, MIT.

Banerjee, Abhijit, Esther Duflo and Rachel Glennerster (2007) "Putting Band-Aid on a Corpse. Incentives for nurses in the Indian Public Health Care System" forthcoming,*Journal of the European Economics Association.*

Banerjee, Abhijit, and Esther Duflo (2006), "Addressing Absence," *Journal of Economic Perspectives* 20(1): 117-132

Banerjee, Abhijit, Angus Deaton and Esther Duflo (2004), "Wealth, Health and Health Services in Rural Rajasthan," *American Economic Review Papers and Proceedings*, 94(2), 326-330.

Banerjee, Abhijit, Suraj Jacob and Michael Kremer, with Jenny Lanjouw and Peter Lanjouw (2005), "Moving to Universal Education: Costs and Tradeoffs" MIMEO, MIT.

Banerjee, Abhijit, Shawn Cole, Esther Duflo and Leigh Linden (2007), "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics*, 122(3):1235-1264, August.

Benabou, Roland, and Jean Tirole (2006) "Incentives and pro-social Behavior" *American Economic Review*, 95(6):1652-1678

Bound, John, Todd Stinebrickner, and Timothy A. Waidman (2005), "Using a Structural Retirement Model to Simulate the Effect of Changes to the OASDI and Medicare Programs," ERIU Working Paper, University of Michigan.

Card, David and Dean Hyslop (2005) "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare Leavers." *Econometrica*, 73(6):1723-1770

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, F. Halsey Rogers (2005a), "Missing in Action: Teacher and Health Worker Absences in Developing Countries," *Journal of Economic Perspectives*20(1):91-116

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, F. Halsey Rogers (2005b), "Teacher Absence in India: A Snapshot," *Journal of the European Economic Association.*3(2) 658-667.

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

Duraisamy, P. (2000), P. Changes in Returns to Education in India, 1983-94: By Gender, Age-Cohort and Location," Yale Center Discussion Paper # 815.

Duthilleul, Yael (2004) "International Perspectives on Contract teachers and their Impact on Meeting Education for All. The Case of Cambodia, Nicaragua, India" MIMEO UNESCO International Institute for Education Planning.

Education for All Forum (2000), EFA Country Assessment Country Reports.

Fehr, Ernst, and Schmidt (2004), "Fairness and Incentives in a Multi-task Principal–Agent Model" *Scandinavian Journal of Economics*, 106(3), 453–474.

Fehr, Ernst and Lorenz Gotte (2002), "Do Workers Work More if Wages are High? Evidence from a Randomized Field Experiment," University of Zurich Working Paper 125, forthcoming, American Economic Review

Figlio, David and Lawrence S. Getzler (2002) "Accountability, Ability and Disability: Gaming The System," NBER Working Paper 9307.

Figlio, David and Josh Winicki (2002), "Food for Thought? The Effects of School Accountability Plans on School Nutrition," NBER Working Paper 9319.

Glewwe, Paul, Nauman Ilias and Michael Kremer (2003), "Teacher Incentives," MIMEO, Harvard.

Hanushek, Eric, John Kain and Steven Rivkin (2005),"Teachers, Schools, and Academic Achievement" *Econometrica*, 73(2), 417-458.

Holmstrom, Bengt and P. Milgrom (1991), "Multi-Task Principal-Agent Problems: Incentive Contracts, Asset Ownership and Job Design," *Journal of Law, Economics and Organization*, VII, 24–52.

Jacob, Brian and Steve Levitt (2003), "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics,* 118(3), 843-77.

Keane, Michael and Ken Wolpin (1994), "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76, 648-672.

Kremer, Michael and Daniel Chen (2001), "An Interim Report on a Teacher Attendance Incentive Program in Kenya," MIMEO, Harvard University.

Kremer, Michael, Edward Miguel and Rebecca Thorntorn (2004) "Incentives to Learn" NBER Working Paper #10971.

Kreps, David (1997), "Intrinsic Motivation and Extrinsic Incentives," *American Economic Review*, 87(2), 359-364.

Krueger, Alan and Diane M. Whitmore (2001), "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR," *Economic Journal*, 111, 1-28.

Lavy, Victor (2002), "Evaluating the Effect of Teachers' Performance Incentives on Pupils' Achievements" *Journal of Political Economy* December 1286-1317

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Multinomial Discrete Response Models Without Numerical Integration," *Econometrica*, 57:995-1026.

Miguel, Edward and Michael Kremer (2004), "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72 (1): 159-217.

Muralidharan , Kharthik and Venkat Sundaraman (2007) "Teacher Incentives in Developing Countries: Experimental Evidence from India" MIMEO, Harvard.

Pakes, A. and D. Pollard, (1989), "Simulation and the Asymptotics of Optimization Estimators," Econometrica, 57, 1027-1057.

Pratham (2006) *Annual Status of Education Report, 2005*

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

Stinebrickner, Todd (2000), "Serially Correlated State Variables in Dynamic, Discrete Choice Models," *Journal of Applied Econometrics*, 15, 595-624.

Todd, Petra and Kenneth Wolpin (2007) "Assessing the Impact of a School Subsidy Program in Mexico: Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," forthcoming, *American Economic Review*.

World Bank (2004), *Making Service Work for Poor People," World Development Report, Washington and Oxford*: World Bank and Oxford University Press.

Figure 1: Photographs from Program

Table 1: Is School Quality Similar in Treatment and Control Groups Prior to Program?

	Treatment (1)	Control (2)	Difference (3)
A. Teacher Attendance			
School Open	0.66	0.64	0.02
			(0.11)
	41	39	80
B. Student Participation (Random Check)			
Number of Students Present	17.71	15.92	1.78
			(2.31)
	27	25	52
C. Teacher Qualifications			
Teacher Test Scores	34.99	33.62	1.37
			(2.01)
	53	56	109
Teacher Highest Grade Completed	10.21	9.80	0.41
			(0.46)
	57	54	111
D. Teacher Performance Measures (Random Check)			
Percentage of Children Sitting Within Classroom	0.83	0.84	0.00
			(0.09)
	27	25	52
Percent of Teachers Interacting with Students	0.78	0.72	0.06
			(0.12)
	27	25	52
Blackboards Utilized	0.85	0.89	-0.04
			(0.11)
	20	19	39
E. School Infrastructure			
Infrastructure Index	3.39	3.20	0.19
			(0.30)
	57	55	112
Fstat(1,110)			1.21
p-value			(0.27)

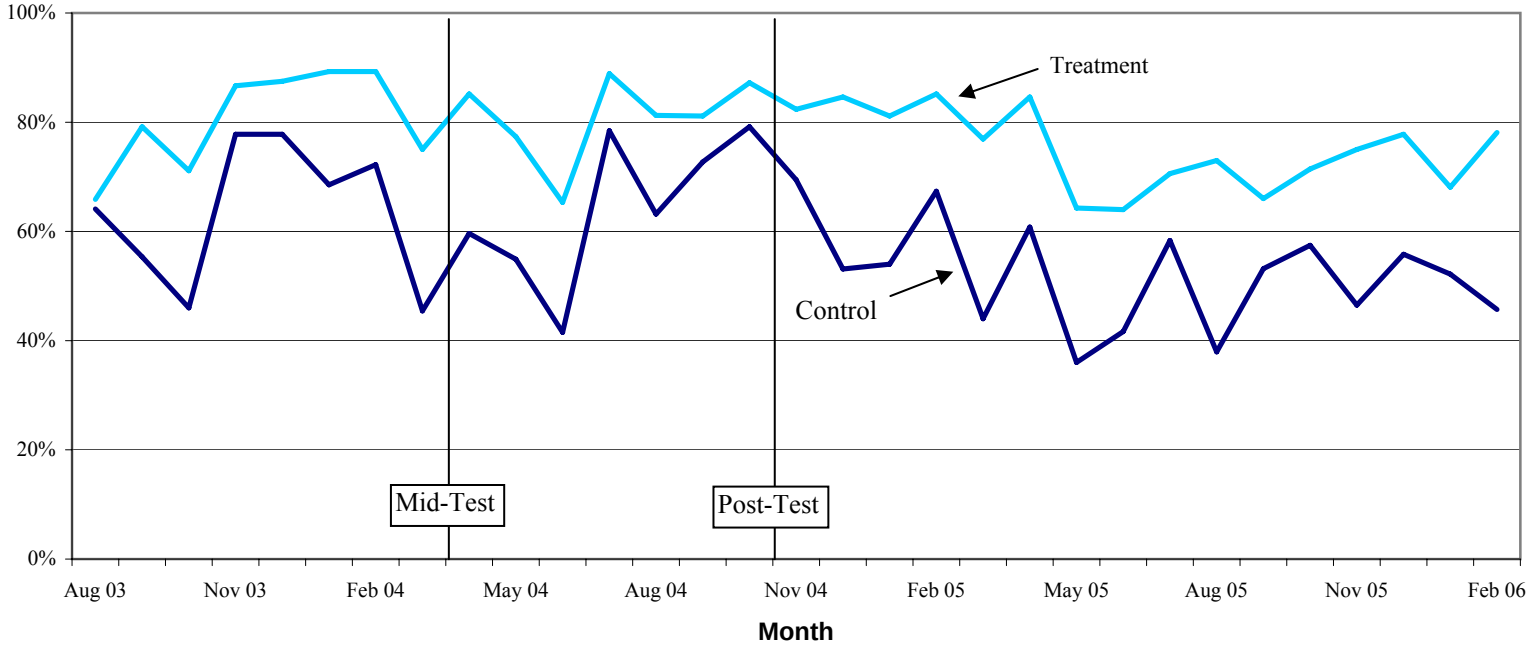
Notes: (1) Teacher Performance Measures from Random Checks only includes schools that were open during the random check. (2) Infrastructure Index: 1-5 points, with one point given if the following school attribute is sufficient: Space for Children to Play, Physical Space for Children in Room, Lighting, Library, Floor Mats

Table 2: Are Students Similar Prior To Program?

	Levels			Normalized by Control		
	Treatment	Control	Difference	Treatment	Control	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Can the Child Write?					
Took Written Exam	0.17	0.19	-0.02			
			(0.04)			
	1136	1094	2230			
	B. Oral Exam					
Math Score on Oral Exam	7.82	8.12	-0.30	-0.10	0.00	-0.10
			(0.27)			(0.09)
	940	888	1828	940	888	1828
Language Score on Oral Exam	3.63	3.74	-0.10	-0.03	0.00	-0.03
			(0.30)			(0.08)
	940	888	1828	940	888	1828
Total Score on Oral Exam	11.44	11.95	-0.51	-0.08	0.00	-0.08
			(0.48)			(0.07)
	940	888	1828	940	888	1828
	C. Written Exam					
Math Score on Written Exam	8.62	7.98	0.64	0.23	0.00	0.23
			(0.51)			(0.18)
	196	206	402	196	206	402
Language Score on Written Exam	3.62	3.44	0.18	0.08	0.00	0.08
			(0.46)			(0.20)
	196	206	402	196	206	402
Total Score on Written Exam	12.17	11.41	0.76	0.16	0.00	0.16
			(0.90)			(0.19)
	196	206	402	196	206	402

Notes: (1) Children who could write were given a written exam. Children who could not write were given an oral exam. (2) Standard errors are clustered by school.

Figure 2: Percentage of Schools Open during Random Checks



Note: (1) The program began in September 2003. August only includes the 80 schools checked before announcement of program. September includes all random checks between August 25 through the end of September. (2) Child learning levels were assessed in a mid-test (April 2004) and a post-test (November 2004). After the post-test, the "official" evaluation period ended. Random checks continued in both the treatment and control schools.

Table 3: Teacher Attendance

Sept 2003-Feb 2006			Difference Between Treatment and Control Schools		
Treatment	Control	Diff	Until Mid-Test	Mid to Post Test	After Post Test
(1)	(2)	(3)	(4)	(5)	(6)
			A. All Teachers		
0.79	0.58	0.21	0.20	0.20	0.23
		(0.03)	(0.04)	(0.04)	(0.04)
1575	1496	3071	882	660	1529
			B. Teachers with Above Median Test Scores		
0.78	0.63	0.15	0.15	0.15	0.14
		(0.04)	(0.05)	(0.05)	(0.06)
843	702	1545	423	327	795
			C. Teachers with Below Median Test Scores		
0.78	0.53	0.24	0.21	0.14	0.32
		(0.04)	(0.05)	(0.06)	(0.06)
625	757	1382	412	300	670

Notes: (1) Child learning levels were assessed in a mid-test (April 2004) and a post-test (November 2004). After the post-test, the "official" evaluation period was ended. Random checks continued in both the treatment and control schools. (2) Standard errors are clustered by school. (3) Panels B and C only include the 109 schools where teacher tests were available.

Figure 3: Impact of the Cameras
(out of at least 25 visits)

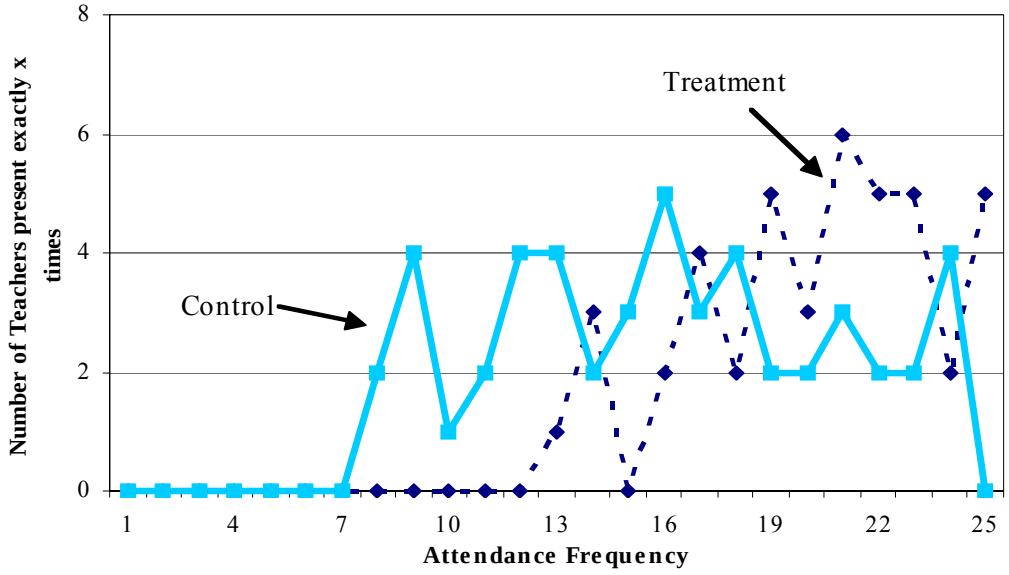


Table 4: Comparing Random Checks to Photo Data for Treatment Schools

Scenario	Number	Percent of Total
A. Possible Scenarios		
School Open and Valid Photos	879	66%
School Open and Invalid Photos	179	13%
School Closed and Valid Photos	88	7%
School Closed and Invalid Photos	191	14%
B. Out of 179 where School is Open, the photos are invalid because….		
School not open for full 5 hours	43	24%
Only one photo	90	50%
Not enough Children	36	20%
Instructor not in Photo	9	5%
Don't Know	1	1%
C. Out of 88 where School is Closed and the photos are valid…..		
Random check completed after the school closed	13	15%
Camera broke/excused meeting	21	24%
Teacher left in the middle of the day	54	61%

Figure 4: Difference in the Percent of Open Schools Between Treatment and Control, By Hour

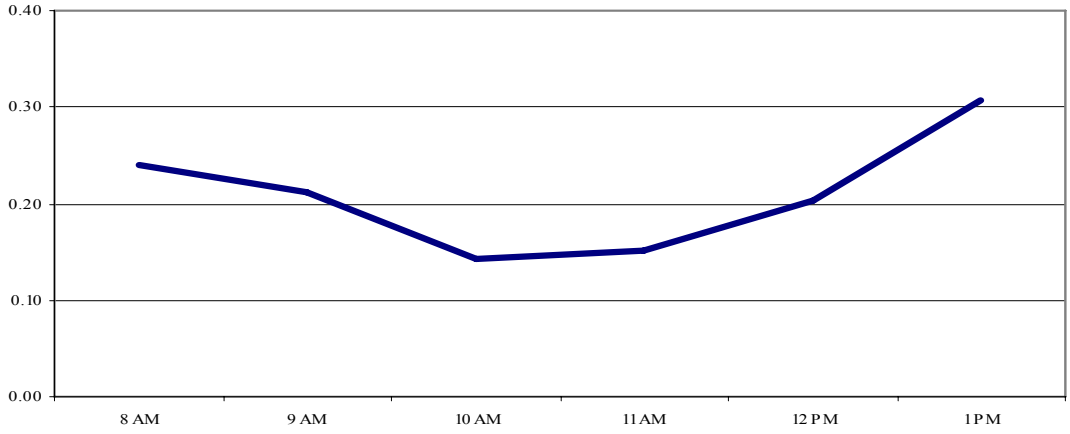
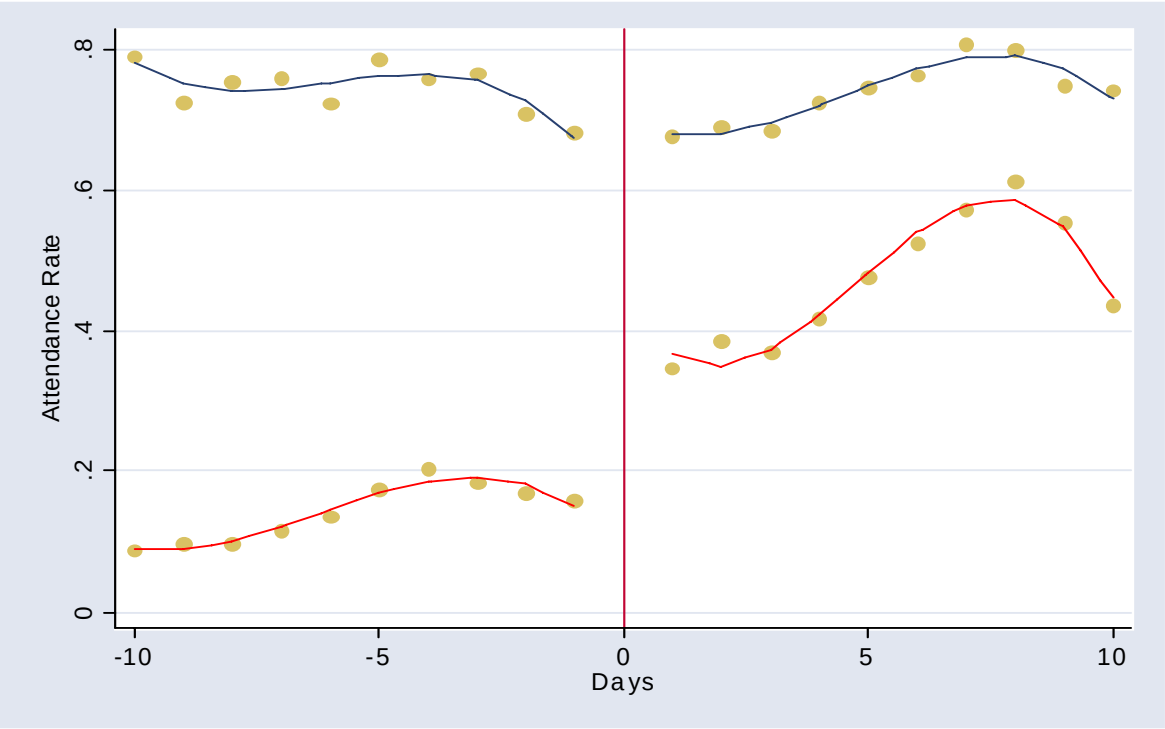


Table 5 : Do Teachers Work More When They are "In the Money"?

	(1)	(2)	(3)	(4)
Beginning of Month	0.19	0.12	0.46	0.39
	(0.05)	(0.06)	(0.04)	(0.03)
In the Money	0.52	0.37	0.6	0.48
	(0.04)	(0.05)	(0.03)	(0.01)
Beginning of the Month * In the Money	-0.19	-0.12	-0.34	-0.3
	(0.06)	(0.06)	(0.04)	(0.02)
Observations	2813	2813	27501	27501
R-squared	0.06	0.22	0.08	0.16
Sample	1st and last day of month	1st and last day of month	1st 10 and last 10 days of month	1st 10 and last 10 days of month
Third Order Polynomial on Days on each side			X	X
Teacher Fixed Effects		X		X
Month Fixed Effects		X		X
Clustered Standard Errors	X		X	

Note: (1) The dependent variable in all models is an indicator variable for whether the teacher worked on a particular day, as measured by the photographs for the treatment schools.

Figure 5: RDD Representation of Teacher Attendance at the Start and End of the Month



Note: (1) The blue lines represent the months in which the teacher is in the money, while the red line represents the months in which the teacher not in the money. (2) The estimation includes a third order polynomial of days on the left and right side of the change of month.

Table 6: Results from the Structural Model

Parameter	Model I (1)	Model II (2)	Model III (3)	Model IV (4)	Model V (5)	Model VI (6)
β	0.049	0.024	0.059	0.051	0.014	0.019
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
μ_1	1.55		2.315	2.063	-0.107	0.012
	(0.013)		(0.013)	(0.012)	(0.040)	(0.028)
ρ			0.682	0.547	0.461	
			(0.010)	(0.023)	(0.039)	
σ_1^2				0.001	0.153	0.135
				(0.011)	(0.053)	(0.027)
μ_2					3.616	1.165
					(0.194)	(0.101)
σ_2^2					0.26	0.311
					(0.045)	(0.051)
p					0.047	0.131
					(0.007)	(0.015)
Heterogeneity	None	FE	None	RC	RC	RC
$\in_{Bonus}$	3.52	1.687	6.225	10.08	0.306	0.370
	(1.550)	(0.098)	(0.634)	(1.249)	(0.038)	(0.029)
$\in_{bonus\_cutoff}$	-75.49	-16.04	-50.22	-63.11	-1.29	-1.78
	(6.506)	(1.264)	(2.612)	(3.395)	(0.479)	(0.449)
Predicted Days Worked	20.50	19.00	15.30	12.15	20.23	21.36
	(0.031)	(0.062)	(0.058)	(0.102)	(3.512)	(0.373)
Days Worked BONUS=0	1.60	6.02	1.29	1.318	13.55	11.81
	(0.597)	(0.234)	(0.875)	(0.863)	(5.251)	(0.669)
Out of Sample Prediction	26.16	18.886	15.08	12.956	20.86	21.57
	(0.059)	(0.253)	(0.635)	(0.520)	(3.793)	(0.456)

Note: Models I and II are estimated using maximum likelihood. Models III through VI are estimated using the method of simulated moments with an optimal weighting matrix. We report the elasticity of days worked with respect to the bonus, ϵ_{Bonus}, and the semi-elasticity with respect to a bonus cutoff, ϵ_{Bonus_cutoff}. The last three rows report the expected number of days worked under the original incentives, a counterfactual where BONUS=0, and the second set of financial incentives.

Figure 6A: Predicted Fit From Model V

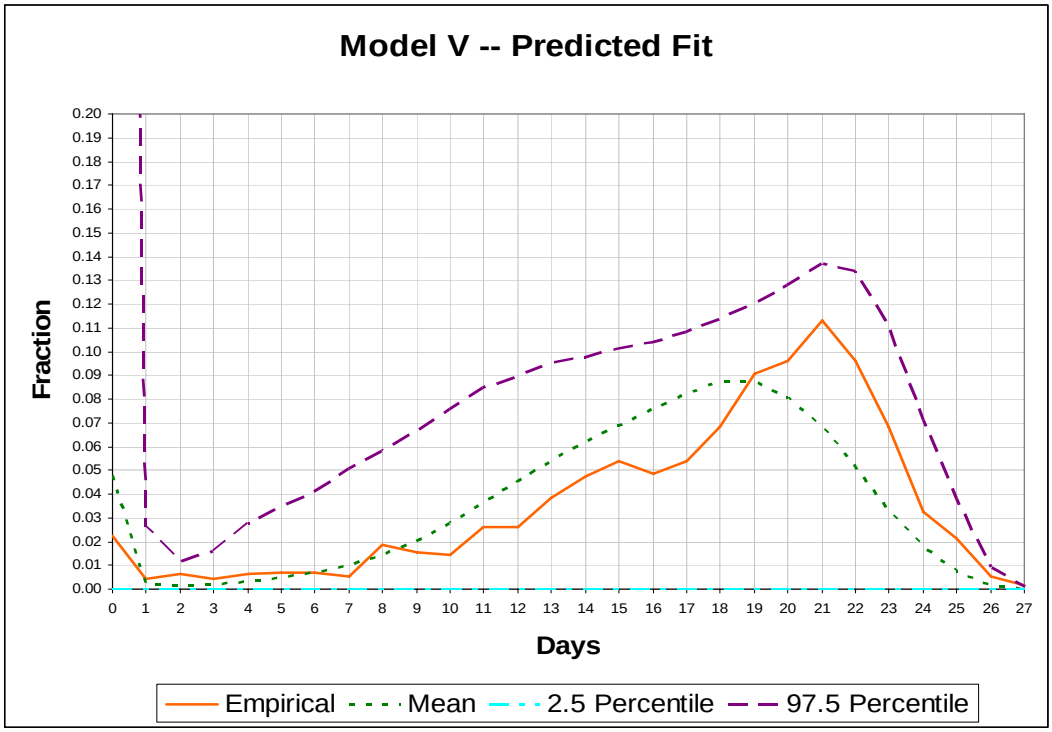


Figure 6B: CounterFactual Fit From Model V

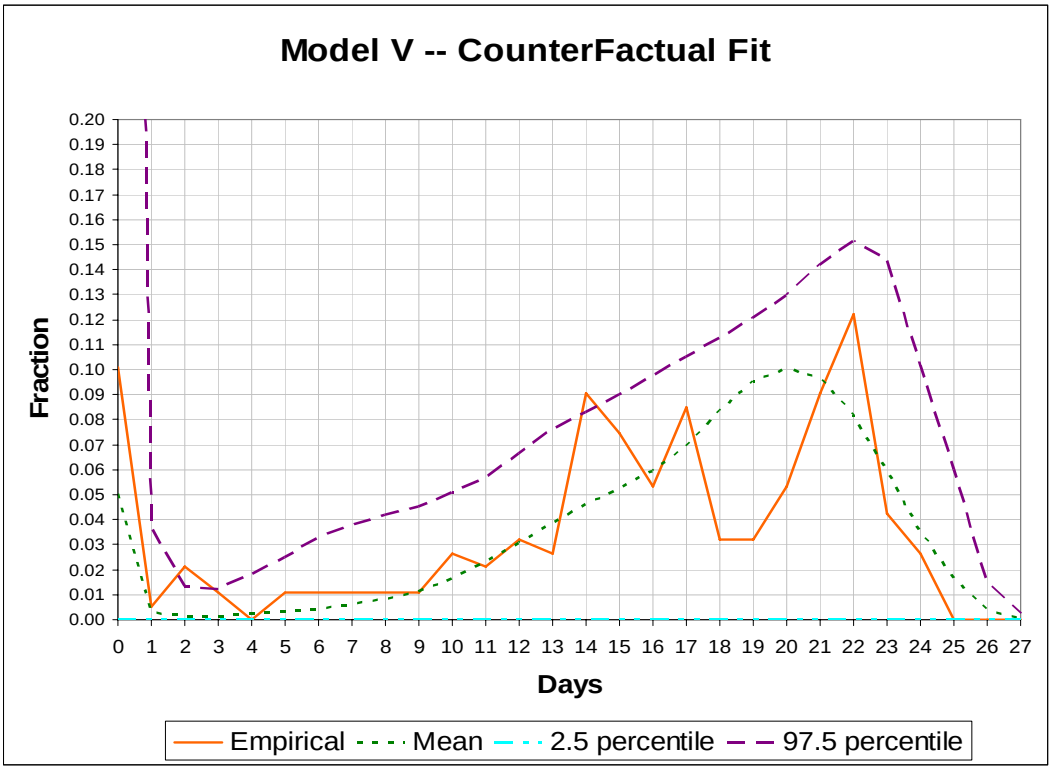


Table 7: Teacher Performance

	Sept 2003-Feb 2006			Difference Between Treatment and Control Schools		
	Treatment	Control	Diff	Until Mid-Test	Mid to Post Test	After Post Test
	(1)	(2)	(3)	(4)	(5)	(6)
Percent of Children Sitting Within Classroom	0.72	0.73	-0.01	0.01	0.04	-0.01
			(0.01)	(0.89)	(0.03)	(0.02)
	1239	867	2106	643	480	983
Percent of Teachers Interacting with Students	0.55	0.57	-0.02	-0.02	0.05	-0.04
			(0.02)	(0.04)	(0.05)	(0.03)
	1239	867	2106	643	480	983
Blackboards Utilized	0.92	0.93	-0.01	-256766.00	0.01	-0.01
			(0.01)	(0.02)	(0.02)	(0.02)
	990	708	1698	613	472	613

Notes: (1) Teacher Performance Measures from Random Checks only includes schools that were open during the random check. (2) Standard errors are clustered by school.

Table 8: Child Attendance

	Sept 03-Feb 06			Difference Between Treatment and Control Schools		
	Treatment	Control	Diff	Until Mid-Test	Mid to Post Test	After Post Test
	(1)	(2)	(3)	(4)	(5)	(6)
A. Attendance Conditional on School Open						
Attendance of Students Present at Pre-Test Exam	0.46	0.46	0.01	0.02	0.03	0.00
			(0.03)	(0.03)	(0.04)	(0.03)
	23495	16280	39775			
Attendance for Children who did not leave NFE	0.62	0.58	0.04	0.02	0.04	0.05
			(0.03)	(0.03)	(0.04)	(0.03)
	12956	10737	23693			
B. Total Instruction Time (Presence)						
Presence for Students Present at Pre-Test Exam	0.37	0.28	0.09	0.10	0.10	0.08
			(0.03)	(0.03)	(0.04)	(0.03)
	29489	26695	56184			
Presence for Student who did not leave NFE	0.50	0.36	0.13	0.10	0.13	0.15
			(0.03)	(0.04)	(0.05)	(0.04)
	16274	17247	33521			
C. Presence, by Student Learning Level at Program Start (for those who did not leave)						
Took Oral Pre-Test	0.50	0.36	0.14	0.11	0.14	0.15
			(0.03)	(0.03)	(0.05)	(0.04)
	14778	14335	29113			
Took Written Pre-Test	0.48	0.39	0.10	0.07	0.07	0.11
			(0.06)	(0.07)	(0.06)	(0.07)
	1496	2912	4408			

Notes: (1) Standard errors are clustered at the level of the school. (2) Child attendance data were collected during random checks. (3) The attendance at the pre-test exam determined the child enrollment at the start of the program.

Table 9: Descriptive Statistics for Mid Test and Post Test

	Mid Test			Post Test		
	Treatment	Control	Difference	Treatment	Control	Difference
	A. Attrition Process					
Percent Attrition	0.11	0.22	-0.10	0.24	0.21	0.03
			(0.05)			(0.04)
Difference in Percent Written of Pre-Test attriters-stayers	0.01	0.03	0.02	0.06	-0.03	0.10
			(0.06)			(0.06)
Difference in Verbal Test of Pre-Test attriters-stayers	0.05	0.08	-0.03	0.02	0.12	-0.10
			(0.14)			(0.14)
Difference in Written Test of Pre-Test attriters-stayers	-0.41	-0.23	-0.18	-0.19	-0.13	-0.06
			(0.34)			(0.29)
	B. Exam Score Means					
Took Written	0.36	0.33	0.03	0.61	0.57	0.04
			(0.04)			(0.05)
Math	0.14	0.00	0.14	-0.08	-0.24	0.16
			(0.10)			(0.15)
Language	0.14	0.00	0.14	1.71	1.60	0.11
			(0.10)			(0.11)
Total	0.14	0.00	0.14	0.35	0.24	0.12
			(0.10)			(0.11)

Notes: (1) Test Scores in Panel B are normalized by the mean of the mid-test control. (2) Standard Errors are clustered by school.

Table 10: Estimation of Treatment Effects for the Mid- and Post-Test

Mid-Test				Post-Test			
Took Written	Math	Lang	Total	Took Written	Math	Lang	Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			A. All Children				
0.04	0.15	0.16	0.17	0.06	0.21	0.16	0.17
(0.03)	(0.07)	(0.06)	(0.06)	(0.04)	(0.12)	(0.08)	(0.09)
1893	1893	1893	1893	1760	1760	1760	1760
			B. With Controls				
0.02	0.13	0.13	0.14	0.05	0.17	0.13	0.15
(0.03)	(0.07)	(0.05)	(0.06)	(0.04)	(0.10)	(0.07)	(0.07)
1893	1893	1893	1893	1760	1760	1760	1760
			C. Took Pre-Test Oral				
	0.14	0.13	0.15		0.2	0.13	0.16
	(0.08)	(0.06)	(0.07)		(0.14)	(0.09)	(0.10)
	1550	1550	1550		1454	1454	1454
			D. Took Pre-Test Written				
	0.19	0.28	0.25		0.28	0.28	0.25
	(0.12)	(0.11)	(0.11)		(0.18)	(0.11)	(0.12)
	343	343	343		306	306	306
			E. Girls				
0.07	0.18	0.18	0.2	0.07	0.22	0.17	0.18
(0.03)	(0.07)	(0.07)	(0.07)	(0.05)	(0.12)	(0.09)	(0.09)
891	891	891	891	821	821	821	821
			F. Boys				
0.02	0.12	0.14	0.14	0.05	0.19	0.16	0.16
(0.04)	(0.09)	(0.07)	(0.07)	(0.04)	(0.15)	(0.10)	(0.10)
988	988	988	988	929	929	929	929

Notes: (1) The table presents the coefficient estimate of being in a treated school on the sum of a child's score on the oral and written exams. All regressions include controls for the child's learning levels prior to the program. (2) The mid and post test scores are normalized by mid test control group. (3) Controls in Row B include Block, Teacher Test Scores, and Infrastructure Index. (4) Standard errors are clustered by school.

Table 11: Dropouts and Movement into Government Schools

	Treatment (1)	Control (2)	Diff (3)
Child Left NFE	0.44	0.36	0.08
			(0.04)
Child Enrolled in Government School	0.26	0.16	0.10
			(0.03)
Child Dropped Out of School	0.18	0.20	-0.02
			(0.03)
N	1136	1061	2197

Notes: (1) Standard errors are clustered at the level of the school. (2) Dropouts are defined as those who were absent for the last five random checks in which a school was found open.

Table 12: Does the Random Check Predict Test Scores?

Method:	OLS	OLS	OLS	2SLS
Sample:	Control Schools	Treatment Schools	Treatment Schools	All Schools
Data:	Random Check	Random Check	Photographs	Random Check
	(1)	(2)	(3)	(4)
		A. Mid-test (Sept 03-April 04)		
Took Written	0.02	0.28	0.36	0.26
	(0.10)	(0.08)	(0.11)	(0.19)
Total Score	0.20	0.39	0.87	1.07
	(0.19)	(0.21)	(0.22)	(0.43)
N	878	1015	1015	1893
		B. Post-test (Sept 03 -Oct 04)		
Took Written	0.24	0.51	0.59	0.33
	(0.16)	(0.15)	(0.20)	(0.22)
Total Score	0.58	1.17	0.98	0.97
	(0.35)	(0.36)	(0.53)	(0.47)
N	883	877	877	1760

Notes: (1) The table presents the coefficient estimate of the teacher's attendance on the sum of a child's score on the oral and written exams. All regressions include controls for the child's learning levels prior to the program. (2) The mid and post test scores are normalized by the mid test control group. (3) Standard errors are clustered by school.

Table 13: Cost of Program Per Center over 12 Month Period

Item	Cost
A. Camera Cost	
Camera Cost[1]	1133
Film Cost	1392
Battery Cost	552
Photo Development and Printing:	1852
B. Salaries	
Teacher Salaries[2]	0
Labor Cost to Run Program[3]	450
Total Costs to Run Program	5379

Notes: (1) Assumes cameras last 3 years. (2) The average teacher salary was Rs1000 in program. Thus, in the absence of the program, it would be the same. (3) It takes approximately 50 man hours to process 115 schools per month. Assuming that a staff worker is paid Rs 10,000 per month and works a 40 hour week, it takes 1/2 hour of labor at Rs37.5 to complete one center per month.

Appendix Table 1: School Closures

	Total Schools	Prior to Post-Test	By Dec 06
	(1)	(2)	(3)
	A. Number of Schools that Closed		
Treatment	57	3	14
Control	56	4	12
Total	113	7	26
	B. Number of Schools Where the Teacher Changed at Least Once		
Treatment	57	5	7
Control	56	6	6
Total	113	11	13