American Economic Review 2012, 102(4): 1241–1278
http://dx.doi.org/10.1257/aer.102.4.1241

Incentives Work: Getting Teachers to Come to School[†]

By Esther Duflo, Rema Hanna, and Stephen P. Ryan*

We use a randomized experiment and a structural model to test whether monitoring and financial incentives can reduce teacher absence and increase learning in India. In treatment schools, teachers' attendance was monitored daily using cameras, and their salaries were made a nonlinear function of attendance. Teacher absenteeism in the treatment group fell by 21 percentage points relative to the control group, and the children's test scores increased by 0.17 standard deviations. We estimate a structural dynamic labor supply model and find that teachers respond strongly to financial incentives. Our model is used to compute cost-minimizing compensation policies. (*JEL* I21, J31, J45, O15)

Many developing countries have expanded primary school access. These improvements, however, 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 (Pratham 2006). These poor learning outcomes may be due, in part, to teacher absenteeism. Using unannounced visits to measure attendance, a nationally representative survey found that 24 percent of teachers in India were absent during school hours (Kremer et al. 2005).[1] Thus, improving attendance rates is necessary to make "universal primary education" a meaningful term.

Solving the absenteeism problem poses a significant challenge (see Banerjee and Duflo 2006 for a review). In many countries, teachers are a powerful political force, able to resist attempts to enforce stricter attendance rules. As such, many governments have shifted to instead hiring "para-teachers." Para-teachers are teachers who are hired on short, flexible contracts to work in primary schools and in nonformal education centers (NFEs) that are run by nongovernmental organizations (NGOs) and local governments. Unlike government teachers, it may be feasible to implement greater oversight and incentives for para-teachers since they do not form an

* Duflo: Massachusetts Institute of Technology, 50 Memorial Drive, E52-252G, Cambridge, MA 02142, and NBER and J-PAL (e-mail: eduflo@mit.edu); Hanna: Harvard Kennedy School, Mailbox 26, 79 JFK Street, Cambridge, MA 02138, and NBER and J-PAL (e-mail: Rema_Hanna@hks.harvard.edu); Ryan: Massachusetts Institute of Technology, 50 Memorial Drive, E52-262C, Cambridge, MA 02142, and NBER (e-mail: sryan@mit.edu). 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, Konrad Menzel, Callie Scott, and Kudzaishe Takavarasha provided superb research assistance. For their helpful comments, we thank referees, Abhijit Banerjee, Rachel Glennerster, Michael Kremer, and Sendhil Mullainathan. We owe a special thanks to the referees, who made substantial suggestions that considerably improved the paper. For financial support, we thank the John D. and Catherine T. MacArthur Foundation.

† To view additional materials, visit the article page at http://dx.doi.org/10.1257/aer.102.4.1241.

[1] Teachers have some official nonteaching duties, but this absence rate is too high to be fully explained by this.

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 incentives may be an effective way to improve the quality of education, provided that they can teach effectively.

In this paper, we use both experimental 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 financial incentives based on their attendance, improves both teacher attendance and school quality.

The effect of incentives based on presence is theoretically ambiguous. While simple labor supply models predict that incentives should increase effort, there are cases where they can be ineffective. First, the incentives may not be strong enough. Second, the incentive may crowd out a 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 Goette 2007).

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.[2] 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 improving child outcomes. Thus, whether or not the incentives can improve school quality is ultimately an empirical question.

To address these questions, 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. Teacher absenteeism is high, despite the fact that Seva Mandir tries to reduce it by berating frequently absent teachers and threatening dismissal for repeated absences. In our baseline study, evaluated in August 2003, the absence rate was about 35 percent. In September of 2003, Seva Mandir gave teachers in 57 randomly selected program schools 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 nonlinear 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 attended that month. In the 56 comparison schools, teachers were paid a fixed rate for the month (Rs. 1,000)

[2] This is a legitimate concern as other incentive programs (based on test scores) have been subject to multitasking (Glewwe, Ilias, and Kremer 2010), manipulation (e.g., Figlio and Winicki 2005; Figlio and Getzler 2006) or outright cheating (Jacob and Levitt 2003). On the other hand, Lavy (2009) and Muralidharan and Sundaraman (2011) find very positive effects of similar programs.

and were reminded (as usual) that regular presence was a requirement of their job, and that they could in principle 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 at baseline and the 42 percent in the comparison schools.

While the reduced form results inform us that this program was effective in reducing absenteeism, they do not tell us what the effect of another scheme with a different payment structure would be. Moreover, they do not allow us to identify the response to the financial incentive separately from a possible independent effect of collecting daily data on absence.[3] To answer these questions, we estimate a structural dynamic model of teacher labor supply using the daily attendance data in the treatment schools. Our estimation strategy leverages the fact that the financial incentive for a teacher to attend school on a given day changes with the number of days previously worked in the month and the number of days left in the month. This is because teachers have to attend at least 10 days in the month before they begin to receive the incentive and the implied shadow value of working changes as the teacher builds up the option to work for Rs. 50 per day at the end of the month.

In order to understand the effect of the financial and monitoring incentives on teacher attendance, we estimate two complementary structural models of the teachers' labor supply functions. The two models are conceptually similar in that they both model the dynamic decision process facing teachers as they accumulate days worked towards the bonus at the end of the month. Both approaches allow for unobserved heterogeneity at the teacher level, but differ in their treatment of serial dependence in opportunity cost of working. In the first set of models, the opportunity cost is allowed to depend on whether the teacher attended work on the previous day. The second set of models posits that the opportunity cost to working is subject to an autocorrelated shock that follows an AR(1) process. The two models deliver similar results. A nice feature of the experiment is that the incentives shift discontinuously with the change in month, which is the source of the identification of the responsiveness to the bonus. As a robustness check, we combine the spirit of the regression-discontinuity approach with the structural model by estimating a model with a three-day sample window around the change in month. These results are similar to those from the earlier models.

To our knowledge, this is one of the few papers to estimate dynamic labor supply decisions with unobserved heterogeneity and a serially correlated error structure.[4] Stinebrickner (2000) discusses some of the econometric issues associated with this problem. Three related papers are Bound, Stinebrickner, and Waidmann (2010); Sullivan (2010); and Stinebrickner (2001a).[5]

[3] Another possibility is that the existence of this scheme discourages teachers in the *control* group. This would lead us to overestimate the impact of the program.

[4] See Aguirregabiria and Mira (2010); Keane, Todd, and Wolpin (2011); and Todd and Wolpin (2010) for recent surveys of the estimation of dynamic choice structural models.

[5] Stinebrickner (2001a) also estimates a dynamic model of teacher labor supply. See also Stinebrickner (2001b) and van der Klaauw (2005).

We find that teachers are responsive to the financial incentives: our estimates suggest that the elasticity of labor supply with respect to the level of the financial bonus is between 0.20 and 0.30 in our preferred specifications. In most specifications for the structural model, we do not use the control group data. Therefore, as a check on the model, we can test whether the model accurately predicts teacher presence in the control group. Models that include both serial correlation and teacher heterogeneity do well in this out-of-sample test: when we set the incentive to zero, it closely predicts the difference in attendance in the control group. In addition, the model accurately predicts number of days worked under a new incentive system initiated by Seva Mandir after the experiment.

The idea of holdout samples for validation has been used in several papers, starting with at least McFadden and Talvitie (1977) (see Keane, Todd, and Wolpin 2011). A smaller number of papers use randomized control experiments to validate a structural model. Wise (1985) estimates a model of housing demand on control group data, and validates the model using the forecast of the effect of a housing subsidy. More recently, Todd and Wolpin (2006) used data from the PROGRESA program, a conditional cash transfer program in Mexico. Using only the control villages, they estimated a structural model of fertility, school participation, and child labor. The model was validated by comparing the predicted effect of PROGRESA to the experimental estimates of program effects. Lise, Seitz, and Smith (2005) use data from the Self-Sufficiency Program in Canada to validate a search model of the labor market. As in Keane and Moffitt (1998), we estimate the model using the treatment sample because the incentive schedule provides useful variation for model identification, and use the control sample for out-of-sample model validation. Other papers that combine structural methods and experimental data (without using the control group for out of sample validation) include Attanasio, Meghir, and Santiago (forthcoming) and Ferrall (2010).

An advantage of the structural model is that the parameters can be used to estimate the effects of other possible rules (see Todd and Wolpin 2010 for different applications of this method to development policy). We use the parameters of the model to compute the optimal incentive scheme for a given number of days worked on average in a month. We calculate that Seva Mandir could achieve the same number of days worked (17) by increasing both the bonus cutoff to 21 days and the bonus to 75 rupees per day while saving 193 rupees per teacher per month, an average cost savings of 22 percent.

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; thus, the students in the 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 paper is organized as follows. Section I describes the program and evaluation strategy. The results on teacher attendance are presented in Section II, while the estimates from the dynamic labor supply model are presented in Section III. Section IV presents the results on other dimensions of teacher effort, as well as student outcomes. Section V concludes.

I. Experimental Design and Data Collection

A. *Nonformal Education Centers*

Since the enactment of the National Policy on Education in 1986, nonformal education centers (NFEs) have played an important role in India's drive toward universal primary education. They have been the main instrument for expanding school access to children in remote and rural areas. They have also been used to transition children who may otherwise not attend school into a government school. Several million children are enrolled in NFEs across India. Similar informal schools operate throughout most of the developing world (Bangladesh, Kenya, etc.).

Children of all ages may attend the NFE, though, in our sample, most are between seven and ten years of age. Nearly all of the children are illiterate when they enroll. In the setting of our study, the NFEs are open 6 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 has, on average, a tenth-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.

B. *The Incentive Program*

Seva Mandir runs about 150 NFEs in the tribal villages of Udaipur, Rajasthan. Udaipur is a sparsely populated, hard-to-access region. Thus, it is difficult to regularly monitor the NFEs, and absenteeism is high. A 1995 study (Banerjee et al. 2005) found that the absence rate was 40 percent, while our first observation in the schools included in our study (in August 2003, before the program was announced) found that the rate was about 35 percent.

Before 2003, Seva Mandir relied on occasional visits to the schools, as well as reports by the local village workers, to monitor teacher attendence. They then use bimonthly teacher meetings to talk to delinquent teachers. Given the high absence rate, they were aware that the level of supervision was insufficient.

Therefore, starting in September 2003, Seva Mandir implemented an external monitoring and incentive program on an experimental basis. They chose 120 schools to participate, with 60 randomly selected schools serving as the treatment group and the remaining 60 as the comparison group.[6] In the 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

[6] After randomization but prior to the announcement of the program, seven of these schools closed. The closures were equally distributed among the treatment and controls schools, and were not due to the program. We thus have 57 treatment schools and 56 comparison schools.

to precisely track each school's openings and closings.[7] Rolls were collected every two months at regularly scheduled teacher meetings, and payments were distributed every two months. 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.

At the start of the program, Seva Mandir's monthly base salary for teachers was Rs. 1,000 ($23 at the real exchange rate, or about $160 at purchasing power parity) for at least 20 days of work per month. In the treatment schools, teachers received a Rs. 50 bonus ($1.15) for each additional day they attended in excess of the 20 days (where holidays and training days, or about 3 days per month on average, are automatically credited as working 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. 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 reminded that regular attendance was required and that they could, in principle, be dismissed for poor attendance. No teacher was fired during the span of the evaluation, however.[8]

C. *Data Collection*

Vidhya Bhawan (a consortium of schools and teacher training institutes) and the Abdul Latif Jameel Poverty Action Lab (J-PAL) collected the data. 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.[9] Second, Seva Mandir provided us with access to the camera and payment data for the treatment schools.

We collected data on teacher and student activity during the random check. For schools that were open during the visit, the enumerator noted the school activities: how many children were sitting in the classroom, whether anything was 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 teachers could adjust their behavior. In addition, the enumerator also conducted a roll call and 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

[7] The time and date 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).

[8] 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.

[9] Teachers understood that the random checks were not linked with an incentive. We cannot rule out the fact that the random check could have increased attendance in comparison schools. We have no reason to believe, however, that this would differentially affect the attendance of comparison and treatment teachers.

competency exams to all children enrolled in the NFEs in August 2003: a pretest in August 2003, a mid-test in April 2004, and a post-test in September 2004. The pretest 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.

D. *Baseline and Experiment Integrity*

Preprogram school quality was similar across the treatment and control groups prior to the program onset. Before the program was announced in August 2003, the evaluators visited 41 schools in the treatment group and 39 in the comparison.[10] 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 very small and never significant.

Baseline academic achievement, as measured by the pretest, was the same for students across the two types of schools (Table 1, panel E). On average, students in both groups appeared to be at the same level of preparedness before the program. There is no significant difference in either probability to take the written test or scores on the written tests.

II. Results: Teacher Attendance

A. *Reduced Form Results: Teacher Behavior*

The effect on teacher absence was both immediate and long-lasting. Figure 1 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.

[10] 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.

Table 1—Baseline Data

	Treatment (1)	Control (2)	Difference (3)
Panel A. Teacher attendance			
School open	0.66	0.64	0.02
			(0.11)
	41	39	80
Panel B. Student participation (random check)			
Number of students present	17.71	15.92	1.78
			(2.31)
	27	25	52
Panel C. Teacher qualifications			
Teacher test scores	34.99	33.54	1.44
			(2.02)
	53	54	107
Panel 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
F-stat (1,110)			1.21
p-value			(0.27)
Panel E. Baseline test scores			
Took written exam	0.17	0.19	−0.02
			(0.04)
	1,136	1,094	2,230
Total score on oral exam	−0.08	0.00	−0.08
			(0.07)
	940	888	1,828
Total score on written exam	0.16	0.00	0.16
			(0.19)
	196	206	402

Notes: Teacher Performance Measures from Random Checks include only schools that were open during the random check. Children who could write were given a written exam. Children who could not write were given an oral exam. Standard errors are clustered by school.

As Figure 1 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. At the end of the study, however, they only had enough resources to keep the program operating in the treatment schools. 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 2 presents a detailed breakdown of the program effect on absence rates for different time periods. On average, the teacher absence rate was 21 percentage points lower (or about half) in the treatment than in the comparison schools

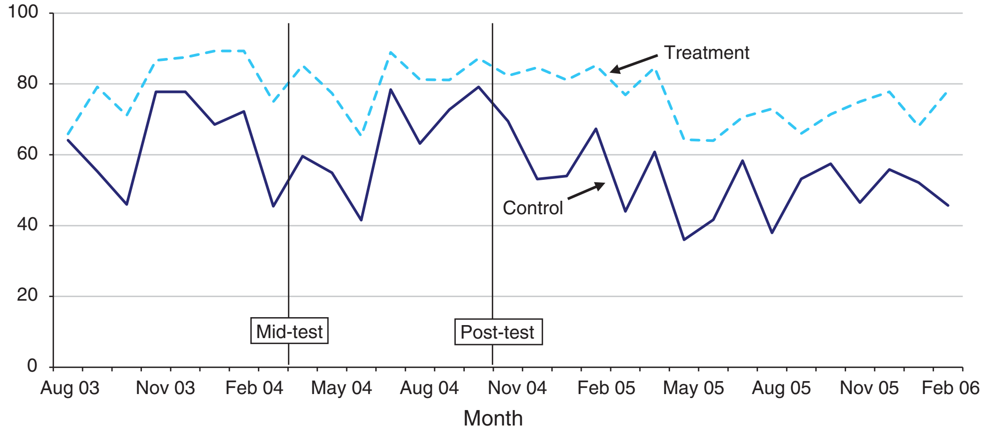

FIGURE 1. PERCENTAGE OF SCHOOLS OPEN DURING RANDOM CHECKS

Notes: 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. 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 2—TEACHER ATTENDANCE

September 2003–February 2006			Difference between treatment and control schools		
Treatment (1)	Control (2)	Diff (3)	Until mid-test (4)	Mid- to post-test (5)	After post-test (6)
Panel A. All teachers					
0.79	0.58	0.21	0.20	0.17	0.23
		(0.03)	(0.04)	(0.04)	(0.04)
1,575	1,496	3,071	882	660	1,529
Panel 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	1,545	423	327	795
Panel 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	1,382	412	300	670

Notes: 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. Standard errors are clustered by school. Panels B and C only include the 109 schools where teacher tests were available.

(panel A).[11] 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

[11] 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 et al. 2005), both because individual teacher absenteeism remained high and because teachers did not coordinate to come on different days.

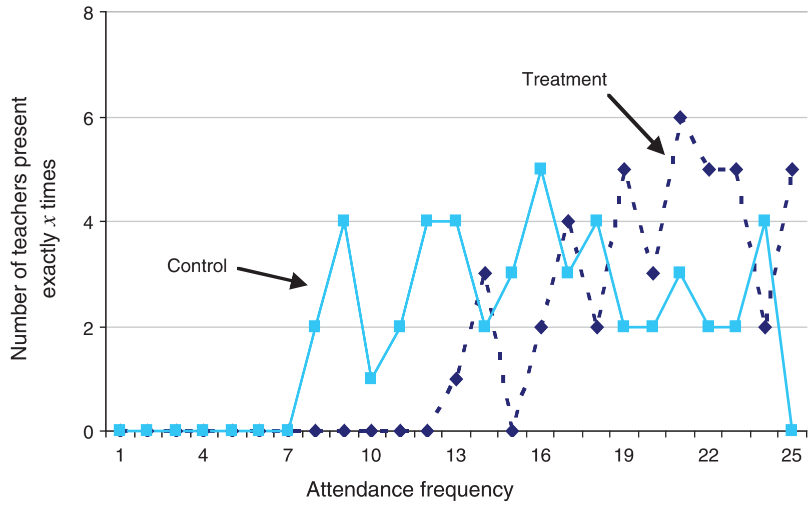

Figure 2. Impact of the Cameras
(*Out of at least 25 visits*)

to the program, while panel C shows the impact for teachers with below-median scores. The program impact on attendance was larger for below-median teachers (a 24 percentage point increase versus a 15 percentage point increase). This was due to the fact, however, 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 2 plots the observed density of absence rates in the treatment and comparison schools for 25 random checks. The figure clearly shows that the program shifted the entire distribution of absence for treatment teachers. 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. Thus, the program was effective on 2 margins: it eliminated very delinquent behavior (less than 50 percent attendance) and increased the number of teachers with 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. Out of the 1,337 cases where we have both camera data and a random check for a day, 80 percent matched. 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," often because the photos were not separated by 5 hours. 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.

One interesting question is whether the effect of the program would be very different in the long run, because the program would induce different teachers to join schools with cameras. As of October 2009, the program was still in place in the same schools (Seva Mandir has recently extended it to all schools). We monitored the schools for a year, from September 2006 to September 2007. After 4 years, teacher attendance was still significantly higher in the camera schools (72 percent versus 61 percent). Thus, this program seems to have a very long-lasting effect on teacher attendance.

B. *The Impact of Financial Incentives: Preliminary Evidence*

The program had two components: the daily monitoring of teacher attendance (which complements Seva Mandir's usual random checks and reports from village or zonal workers) and an incentive that was linked to attendance. In addition, a monthly random check was performed in both treatment and control schools by the research team. We believe, however, that teachers did not perceive that this last check was part of Seva Mandir's program: it was performed by a separate team, associated with a different organization, and the informed consent signed by all teachers disclosed that these data would not be shared with Seva Mandir.

Aside from the financial incentive, teachers may respond to the fact that Seva Mandir now obtains daily attendance data (what we refer to as "the monitoring effect"), either because of a fear of being fired if the data reveals that they are absent most of the time, or because Seva Mandir may punish them for absence. At the bimonthly teacher trainings, teacher absence is discussed and Seva Mandir workers berate teachers whom they believe to be frequently absent. On the other hand, since Seva Mandir continues to inspect treatment and control schools on a random basis, it is also possible that teachers believe that daily data on attendance does not increase the chance that they are punished for absence in expectation: they may believe that one absence found out during a surprise visit would be as costly as several absences in the detailed data. Whether there is a direct effect of obtaining daily attendance data (an "increased monitoring effect") is thus an open question.

Ideally, to disentangle the effect of the financial incentive from a direct increased monitoring effect, we would have provided different types of monitoring and incentive systems in different, randomly selected schools. Some teachers could have been monitored daily, but without receiving incentives. Some could have received a small incentive, while others could have received a larger one. This was not feasible. The nonlinear nature of the incentive scheme, however, provides us with an opportunity to try to isolate the effect of the financial incentives, assuming that the effect of the threat of monitoring does not follow exactly the same time pattern as the incentive. 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, the same amount he would earn if he did not work any other days that month. Although he is still monitored (and may worry that if he does not attend at all in a month he may be punished), his monetary incentive to work in these last few days is zero. At the start of the next month, the clock is reset. He now has incentive to start 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

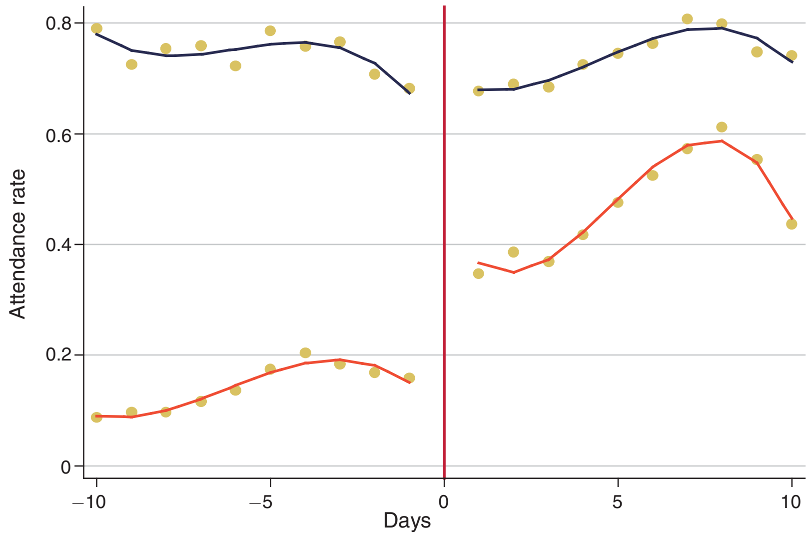

FIGURE 3. RDD REPRESENTATION OF TEACHER ATTENDANCE AT THE START AND END OF THE MONTH

Notes: The top lines represent the months in which the teacher is in the money, while the bottom lines represent the months in which the teacher is not in the money. The estimation includes a third-order polynomial of days on the left and right side of the change of month.

another teacher who has worked 10 days by the 21st day of the month. For every day he works in the 5 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 ten days in that month.

This leads to a simple regression discontinuity design test for whether financial incentives matter, under the assumption that the teacher's outside option if he does not go to school does not also jump discontinuously when the month changes. Figure 3 gives a graphical representation of the approach. It shows a regression of the probability that a teacher works if she is in the money by day 21 of the month (with four 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 for teachers 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 since, in the data, we see that 65 percent of the teachers who are out of the money in a month will be in the money the following month. 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 toward the ten day threshold. Correspondingly, we do not see a sharp drop in presence for the teachers who had been in the money the previous month.

Table 3 presents these results in regression form. Specifically, for teachers in the treatment group, we created a dataset that contains their attendance records for the first and the last 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,

TABLE 3—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	2,813	2,813	27,501	27,501
R^2	0.06	0.22	0.08	0.16
Sample	First and last day of month	First and last day of month	First ten and last ten days of month	First ten and last ten 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: 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.

where $Work_{itm}$ is a dummy variable equal to 1 if teacher i works in day t in the pair of days m (t is either 1 or 2):

$$(1) \qquad Work_{itm} = \alpha + \beta 1_{im}(d > 10) + \gamma Firstday_t + \lambda 1_{im}(d > 10) \times Firstday_t + \upsilon_i + \mu_m + \epsilon_{itm},$$

where $1_{im}(d > 10)$ is a dummy equal to 1 for both days in the pair m if the teacher had worked more than ten days in the month of the first day of the pair, and 0 otherwise. $Firstday_t$ is a dummy that indicates that this is the first day of the month (i.e., the second day of the pair). 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 ten days before).[12]

The results indicate that teachers are more likely to attend school at the beginning of a month if they were not in the money in the previous month, which we do not see for teachers who were in the money. This holds after controlling for teacher fixed

[12] Note that even with teacher fixed effects, β does not have a causal interpretation, because the shocks may be auto-correlated. For example, a teacher who has been sick the entire month, and thus has worked fewer than 10 days, may also be less likely to work the first day of the next month. Because when a month starts and finishes is arbitrary and should not be related to the underlying structure of shocks, however, a positive γ 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, since it would suggest that only teachers who are "out of the money" experience a "first day of the month" effect.

effects (column 4), and even if we restrict the sample to the first and last day of the month (columns 1 and 2).

These results imply that teachers are responsive to the financial incentives, unless there are other factors affecting teachers that happen to have exactly the same structure. Shocks to teachers' outside options are unlikely to change discontinuously when the calendar month changes, as no other teacher activity is linked to the calendar month. It is possible that the daily attendance monitoring could be month-specific if teachers were afraid that Seva Mandir would use the sum of monthly absence to punish or berate the teacher (and possibly to fire them). In this case, they would also worry about total absence in a month and each month would be a new beginning. If teachers thought that Seva Mandir's probability to fire them changed discontinuously every month that their attendance was less than ten days, it would be impossible to separate that from the incentives, since it would have exactly the same time structure.

We cannot directly test the above case since we do not observe any dismissals in the data (let alone teachers' belief about the probability of dismissal) and we do not have data on nonpecuniary punishment. The very fact that no teacher was fired, even though some teachers were absent almost the entire month, however, suggests that Seva Mandir takes a much longer perspective when they consider teacher performance. Indeed, according to Seva Mandir's head of the education unit, it is a teacher's record over an entire year or more that determines their assessment, not how it is distributed across a month. In fact, in the control group, we find no relationship between the calendar day in the month and the chance that we see a teacher at work.[13]

Furthermore, even if Seva Mandir paid attention to monthly totals, we may expect them to matter in the opposite direction. Seva Mandir's official policy is that teachers should attend at least 20 days per month. Therefore, a teacher who has attended 20 days would have had no reason to attend any more days from an "official policy" perspective. We expect the opposite from a financial perspective, and this is also what we observe. Conversely, it is reasonable to think that Seva Mandir would be particularly likely to punish teachers who have attended very few days, so that the incentive to attend to avoid displeasing Seva Mandir should be strong for teachers who have attended fewer than ten days in a month, precisely when the financial incentive is the weakest.

Thus, these findings suggest that teachers respond to the incentives. Without more structure, however, it is not possible to conclude what part of the effect of the program was due to financial incentives per se. 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.

III. A Dynamic Model of Labor Supply

We propose and estimate a simple partial equilibrium model of dynamic labor supply, which incorporates the teacher response to the varying incentives over a month.

[13] This result is available from the authors upon request.

Let m signify the month and t the day within the month, where $t = \{1, \ldots, T_m\}$.[14] The teacher's utility function over consumption, C_{tm}, and leisure, L_{tm}, each day in the month is as follows:

$$U_{tm} = U(C_{tm}, L_{tm}) = \beta C_{tm}(\pi_m) + (\mu_{tm} - P)L_{tm}, \tag{2}$$

where P is the nonpecuniary cost of missing work.[15] We have assumed that utility is linear in consumption and that consumption and leisure are additively separable. This formulation implies that there will not be a dependency in behavior between months. For example, a teacher would not decide to work more in one month because she worked little in previous months.[16]

Consumption is a function of earned income, π_m. Since we assume that there is no discounting within months and utility is linear in consumption, we can assume that the teacher consumes all her income on the last day of the month, when she is paid.[17] The parameter β converts consumption, measured in rupees, into utility terms. We let L_{tm} equal one if the teacher does not attend work on that day and zero otherwise.

The coefficient on the value of leisure, μ_{tm}, has a deterministic and stochastic component:

$$\mu_{tm} = \mu + \epsilon_{tm}. \tag{3}$$

The deterministic component, μ, is the difference between the value of leisure and the intrinsic value of being in school, including any innate motivation. To the extent that teachers value teaching, or do not want to disappoint students and parents, μ will be less positive. The stochastic shock, ϵ_{tm}, captures variation in the opportunity cost of attending work on a given day; we assume that it has a normal distribution.

Teachers who do not go to work face two types of penalties. First, an agent who does not attend school on a particular day is assumed to pay a nonpecuniary cost, P. This term captures the idea that teachers are verbally rebuked by their supervisors if they have been found to be absent during the month: each day of absence makes it more likely that the teacher is found to be shirking.[18] Second, we introduce the possibility that an agent is fired for poor attendance; we denote the probability of being fired in a given period by $p_m(t, d)$, where this probability depends on the

[14] T is 25 in most months. Note that out of the 25 days, there are also several days of training and holidays, which are automatically credited as days worked for the teachers by the payment algorithm. Our estimation procedure follows the same rule, but when we report the number of days worked, we report it out of the days where teachers actually had to make a decision, which is on average 22 days per month.

[15] Although it is clear that μ and P are not separately identified, we have written the costs and benefits of missing work in this way to make explicit the difference between these two countervailing forces on the teacher's labor decision. Moreover, this makes explicit what we mean by "monitoring effect": the experiment may affect P, but we assume it does not affect μ.

[16] A more general utility function might generate such behavior. For example, if utility were logarithmic, and teachers could borrow and save between months, they could decide to work little in a particular month, where the opportunity cost of working is high, and borrow against work in future months. This would make teacher behavior dependent on the entire history of work so far.

[17] Alternatively, we can assume that the consumption of π_m is spread over the next month. This would not change the estimation under strong separability between consumption and leisure.

[18] This way of introducing the nonpecuniary cost is a restrictive functional form assumption, but is necessary to identify the incentive effects. As we explain above, we think that this is a reasonable functional form.

number of days previously worked, d, by time t in month m.[19] Teachers can be fired in the morning before attending work; they will receive a one-time payment of F, the outside option to being a teacher; F may potentially be related to μ, the opportunity cost of working a day (e.g., as a day laborer or tending their field).[20]

Letting d_{m-1} denote the number of days worked in a month, the agent's income earned in the last period in the treatment group is given by the following function:

$$\pi_m = 500 + 50 \max\{0, d_{m-1} - 10\}. \tag{4}$$

In the control group, the agent's income in the last period is Rs. 1,000, irrespective of attendance. The payoff function is such that teachers who work every day in a month will receive more in the treatment group than in the control group.

The teacher is assumed to maximize the present value of lifetime utility.[21] Thus, control group teachers face a simple repeated binary choice problem. The Bellman equation for them every day of the month except the last day is

$$\begin{aligned} V_m(t,d;\epsilon_{tm}) = {} & p_m(t,d) \cdot F + (1 - p_m(t,d)) \\ & \times \max\{\mu - P + \epsilon_{tm} + EV_m(t+1,d;\epsilon_{t,m+1}), \\ & EV_m(t+1,d+1;\epsilon_{t,m+1})\}, \end{aligned} \tag{5}$$

where, without loss of generality, we have set the current-period utility of attending work to zero. The expectation over future value functions is taken with respect to the distribution of next period's shock, $\epsilon_{t,m+1}$. Agents weigh the marginal change in the possibility of being fired in future periods against the immediate benefits of skipping work. From equation (5), it is clear that μ and P are not separately identified. Therefore, without loss of generality, we redefine the outside option of not working for the control group as $\tilde{\mu} = \mu - P$; $\tilde{\mu}$ could easily be negative if P is large enough. At the end of each day, for $t < T_m$, t increases by one and d increases by one if the teacher worked that day. After time T_m, the stated variables of time and days worked reset to zero. On the last day of the month, the value function is almost identical, with $\beta \times 1{,}000$ added to the utility of not being fired.

Teachers in the treatment group face a very different decision problem. First, the structure of the financial incentives induces an additional dynamic concern, as teachers trade off immediate gratification against the possibility of increased wages at the end of the month. Second, the cameras provide Seva Mandir with better information on absences, which can lead to changes in both P, the nonpecuniary cost paid for each absence, and the probability of being fired $p_m(t,d)$. How P should change with

[19] In principle, the probability of being fired can be a function of the teacher's complete past work history, but for expositional clarity we consider a specification that only depends on days worked in the current month. If the probability of being fired depended on the entire work history, the complete life cycle dynamic-programming problem would need to be solved, which would greatly complicate our estimation.

[20] To slightly anticipate our results below, we are not able to identify F in the data, since we do not observe any firing in our study period; it will therefore be impossible to estimate it in more detail.

[21] We assume that there is no discounting within or across months. With idiosyncratic shocks to the outside option, our model is equivalent to one where there is discounting across months but not within months, as idiosyncratic errors imply that the relevant decision horizon is only the current month.

the financial incentive is an open question: on the one hand, Seva Mandir now has perfect information on presence, whereas in the control group, they visit the school infrequently, so most absences go undetected. On the other hand, one can imagine that Seva Mandir puts more weight on an absence they find during one of their inspection visits, and thus that the expected cost of a missed day is similar in both groups. Moreover, if Seva Mandir feels that teachers are sufficiently punished for not attending school by the financial penalty, they may lower the nonpecuniary punishment relative to a situation without incentive, which would lower P. To emphasize that it may differ from the control group, we denote the punishment in the treatment group by $\overline{P}$.

Given this payoff structure, for $t < T_m$, the value function for each teacher is as follows:

$$(6)\quad V_m(t,d;\epsilon_{tm}) = p_m(t,d)\cdot F + (1 - p_m(t,d)) \times \max\{\mu - \overline{P} + \epsilon_{tm} + EV_m(t+1,d;\epsilon_{t,m+1}),\ EV_m(t+1,d+1;\epsilon_{t,m+1})\}.$$

At time T_m:

$$(7)\quad V_m(T_m,d;\epsilon_{T_m,m}) = p_m(T_m,d)\cdot F + \big(1 - p_m(T_m,d)\big) \times \max\{\mu - \overline{P} + \epsilon_{T_m,m} + \beta\pi(d) + EV_{m+1}(1,0;\epsilon_{t,m+1}), \beta\pi(d+1) + EV_{m+1}(1,0;\epsilon_{t,m+1})\}.$$

Note that the term $EV_{m+1}(1,0;\epsilon_{t,m+1})$ enters into both arguments of the maximum operator in equation (7). Since the expectation of this term is independent of any action taken today, in the context of the present model 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_m and working backward, which breaks an infinite-horizon dynamic program into a repeated series of independent finite-time horizon dynamic programs.

Equation (7) also motivates several of our normalizing assumptions. First, the mean of the shock and the mean level of utility of not working are not separately identified; as a result, we set the mean of the shock to be equal to zero. Second, equation (7) is only identified up to scale, as multiplying both sides by a positive constant does not change the work decision. Therefore, we follow a common standard in the discrete choice literature and normalize the variance of the error term to one. Third, as in equation (5), μ and $\overline{P}$ are not separately identified, and without loss of generality we let $\overline{P} = 0$. We note, however, that we can calculate the difference between $\tilde{\mu}$ and $\mu - \overline{P}$ by comparing the predicted attendance rates across the treatment and control groups when we set the financial incentives to zero. This difference identifies the effect of the cameras on teacher attendance absent the financial incentives.

A. *Estimators*

We estimate several specifications of the general dynamic program described by equations (6) and (7).[22] The models vary according to what we assume about μ and the distribution of ϵ. We start with the simplest i.i.d. model, and progressively add observed and unobserved heterogeneity and allow for autocorrelation in the shock to the outside option, μ. We first estimate these models using only the treatment group data and, following Todd and Wolpin (2006), use the means from the control group as an out-of-sample check under the hypothesis that the nonpecuniary cost of not working (P) is the same in the treatment and the control group. We also estimate a second set of models, where we allow the outside option to vary with teacher test scores and the average attendance of the control group in the local block area. The inclusion of block-level control group attendance scores controls for spatially correlated shocks to working at the month level, for example, particularly hot or cold weather makes attending school unattractive to both groups.[23] We also allow the outside option to vary with teacher scores to control for the fact that teachers with higher scores may be more diligent and thus may work more often.

Before describing the empirical specifications, note that we never observe any teachers being fired in the data. Therefore, a consistent estimator for $p_m(t,d)$ in the model above is $\hat{p}_m(t,d) = 0$. We therefore proceed as if the teachers perceive the probability of being fired as being identically equal to zero. This may not be completely correct (e.g., teachers may believe that if they do not come at all in the year they will be fired). No teacher was ever fired, however, despite the fact that 2.3 percent of the teacher-months in the treatment data have recorded zero attendance, with the worst teacher missing 50 percent of the days in 2005. In the control group, of teachers with 20 or more random checks, 34 percent of teachers were present 50 percent of the time or less, and 8 percent present less than 35 percent of the time. For these reasons, we are fairly comfortable positing that the probability of being fired is equal to zero regardless of work history, at least in the range of work history we observe.

Models with i.i.d. Errors.—The simplest model that we estimate is one where all agents share the same marginal utility of income, β, and average outside option of not working, μ, and the shocks to the utility of not working are i.i.d. We use all of the days in the month in the estimation by utilizing the empirical counterpart of equation (6) for $t < T$:

$$\text{(8)} \quad \Pr(work; t, d, \theta) = \Pr\big(\mu + \epsilon_{tm} + EV(t+1,d) < EV(t+1,d+1)\big)$$
$$= \Pr\big(\epsilon_{tm} < EV(t+1,d+1) - EV(t+1,d) - \mu\big)$$
$$= \Phi\big(EV(t+1,d+1) - EV(t+1,d) - \mu\big),$$

[22] We briefly discuss the identification of these models below.

[23] Formally, this specification embeds a block-level shock, ξ_{tmb}, in individual i's utility of not working, $\mu_{itmb} = \mu_i + \epsilon_{itm} + \xi_{tmb}$. Aside from sampling error, the only reason that control group average attendance rates would vary spatially is due to ξ_{tmb}.

where $\Phi(\cdot)$ is the standard normal distribution. Each of the value functions in equation (8) is computed using backward recursion from period T_m. Let w_{imt} be an indicator function equal to one if teacher i worked on day t in month m, and zero otherwise. The log-likelihood function for the model without serial correlation in the error terms is then

$$(9) \qquad LLH(\theta) = \sum_{i=1}^{N} \sum_{m=1}^{M_i} \sum_{t=1}^{T_m} \big[w_{imt} \Pr(work; t, d, \theta) + (1 - w_{imt})(1 - \Pr(work; t, d, \theta)\big],$$

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.

Models with Serial Correlation.—It is reasonable to think that the shock to a teacher's outside option may be correlated over periods. For example, when a teacher is sick, she may be sick for a few days. Indeed, serial correlation is prevalent in the data (see online Appendix Table 1). The table shows empirical sequences of days worked in the last five days of a month by teachers who were already "in" and "out" of the money. These teachers did not face any dynamic incentives within the month. Intuitively, a lack of autocorrelated shocks should imply that the distribution of sequences is uniform. While the table with teachers who are definitely out of the money is inconclusive, due to a small sample size, the table for teachers in the money is more clear. The two most frequent sequences are the two most positively autocorrelated, 00111 and 11100. At the other end of the spectrum, the most negatively autocorrelated sequence, 10101, is the second least frequent, appearing ten times less frequently than 00111. This suggestive evidence motivates several specifications that can handle serial correlation in shocks to the outside option.

Our first approach allows for a simple form of serial correlation in the preference shock by allowing the value of leisure to depend on the observed lagged absence:[24]

$$(10) \qquad \mu_{mt} = \mu + w_{m,t-1} \cdot \gamma,$$

where, as above, $w_{m,t-1}$ is an indicator function for whether the agent worked in the previous period. If $\gamma > 0$, then we expect that working today increases the probability of working tomorrow, and therefore days worked or missed will be clustered together in a month. The likelihood of this model is given by

$$(11) \qquad LLH(\theta) = \sum_{i=1}^{N} \sum_{m=1}^{M_i} \sum_{t=1}^{T_m} \big[w_{imt} \Pr(work; t, d, \theta, w_{m,t-1}) + (1 - w_{imt})[1 - \Pr(work; t, d, \theta, w_{m,t-1})\big].$$

[24] We thank an anonymous referee for suggesting this specification.

The probability of working today expressly depends on whether the agent worked yesterday through equation (10).[25] This specification has the advantage of using all of the data within a month and is simple to implement. An important advantage is that we can also estimate this specification using only a narrow window around the change in month. The main source of identifying variation in this case is that, when the month changes, the financial incentives to work increase for teachers who were "out of the money" at the end of the previous month, while they decrease for those who were previously "in the money." This specification of the structural model incorporates the spirit of the regression discontinuity approach.

The second approach is to model the shock process as following an AR(1) process:

$$\epsilon_{mt} = \rho \epsilon_{m,t-1} + \nu_{mt}, \tag{12}$$

where ρ is the persistence parameter and ν_{mt} is a draw from the standard normal distribution. Autocorrelation could be either positive (illness) or negative (teacher has a task to accomplish). Irrespective of whether ρ is positive or negative, we can no longer directly apply the estimator used in the i.i.d. case. This is because ϵ_{mt} will be correlated with d, as teachers with very high draws on ϵ_{mt} 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 is that $\epsilon = 0$ is invalid, and will bias our estimates of the other parameters.

Our solution is to consider only the sequence of days worked at the beginning of the month.[26] Heuristically, we match the empirical frequencies of sequences of N days worked at the beginning of each month as closely as possible to the frequencies predicted by our model. This results in $2^N - 1$ linearly independent moments, where we have subtracted one to correct for the fact that the probabilities must sum to one. In our estimation, we match sequences of length $N = 5$, which generates 31 moments. In this approach, we treat the draw of the error term at the beginning of each month as coming from the unconditional distribution of ϵ. This is justified by the observation that true distribution of ϵ, conditioning on the work history in the previous month's first 5 days, is essentially identical to the unconditional distribution after 25 days, even for high ρ values.[27] The MSM estimator does not directly exploit the variation from the discontinuous change in incentives at the end of the month, as this would require an enormous number of moments. For example, modeling the sequences across a month with 25 days would require at least sequences of 26 days, generating $2^{26} - 1 = 67{,}108{,}863$ moments. The discontinuity and the nonlinear payment rule is still the source of identification, however, as expressed through changes in the probability of working throughout the month as a function of days previously worked.

Observed and Unobserved Heterogeneity.—We consider two extensions to the specifications above. First, we incorporate observed characteristics into the

[25] We drop the first month of observations for each teacher since we lack data on whether the teacher worked the last day in the previous month.

[26] This model is estimated using the method of simulated moments (see the online Appendix).

[27] Simulation results are available upon request from the authors.

specification for μ. We introduce the attendance in the control group in the same geographic block as a shifter for the outside option, to exploit the informational content of absence in the control group as the monthly fluctuations in the attendance of control and treatment teachers in Figure 2 suggest that there the behavior of these teachers may be correlated across months.[28] We also allow the teacher's score on Seva Mandir's admission exam to shift μ. In our current model, this will reflect teacher heterogeneity. In a richer model, it will also capture dependency in behavior between months.

Second, we relax the assumption that the outside option is equal across all agents after conditioning for observed heterogeneity. We estimate specifications with either fixed effects or random coefficients. In the fixed effects model, teacher types are fixed across time, and we allow μ_i to be estimated separately for each teacher.[29] In the random coefficient models, we estimate two specifications which differ through the distribution of outside options. In the first specification, μ_{im} is drawn anew from a normal distribution each month. In the second specification, we allow for a mixture of two types, where each type is distributed normally with proportion p and $(1-p)$ in the population.

B. *Parameter Estimates*

We present the results of these various specifications in Table 4. We present the main parameters of the model, as well as the implied labor-supply elasticity, the percentage increase in the average number of days worked caused by a one percent increase in the value of the bonus and the semielasticity with respect to the bonus cutoff, and the percentage increase in the average number of days worked in response to an increase in one day in the minimum number of days necessary for a bonus.

The first two columns present the results from specifications without any controls for autocorrelated shocks. Model I estimates a common β and μ for all teachers. The estimate for β indicates that teachers respond positively to the financial incentives, and will work more often the closer they are to being in the money. The predicted number of days worked in the treatment group, 17.23, tracks very closely to the empirical number (17.16). Because the estimated opportunity cost of working, $\mu = 1.564$, is greater than zero, however, this model vastly underpredicts the number of days that teachers work in the control group. Teachers in the control group attended, on average, 12.9 days of work per month; Model I predicts that they would work 1.31 days. A potential explanation for this result is that teachers vary in their outside options. Therefore, Model II relaxes the assumption of a common μ and allows for teacher fixed effects. The estimated β is lower but still positive; the model still underpredicts attendance in the control group.

It is possible that these models are picking up the confounding effects of serial correlation in the errors and the financial incentives. Therefore, we next estimate

[28] This suggests, in the framework of our model, that the leisure shock has a component that varies at the level of the geographic block and the month. The only reason that control group average attendance rates would differ across blocks in different months (aside from sampling variation) is because of this component.

[29] The model with fixed effects has the usual panel model bias, although we expect it will be attenuated here given the relatively long length of the panel.

Table 4—Results from the Structural Model

Parameter	Model I (1)	Model II (2)	Model III (3)	Model IV (4)	Model V (5)	Model VI (6)	Model VII (7)	Model VIII (8)
β	0.049	0.027	0.055	0.057	0.013	0.017	0.017	0.016
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
μ_1	1.564		1.777	1.778	−0.428	−0.304	−0.160	−0.108
	(0.013)		(0.013)	(0.021)	(0.045)	(0.042)	(0.092)	(0.057)
ρ			0.422	0.412	0.449			
			(0.030)	(0.021)	(0.043)			
σ_1^2				0.043	0.007	0.252	0.418	0.235
				(0.012)	(0.019)	(0.015)	(0.052)	(0.028)
μ_2					1.781			
					(0.345)			
σ_2^2					0.050			
					(0.545)			
ρ					0.024			
					(0.007)			
Yesterday shifter						0.094	0.024	0.095
						(0.010)	(0.009)	(0.014)
Attendance								−0.132
								(0.095)
Test score								−0.005
								(0.002)
Heterogeneity	None	FE	None	RC	RC	RC	RC	RC
Three-day window	No	No	No	No	No	No	Yes	No
LLH	10,269.13	9,932.71				9,286.03	3,320.70	9,287.33
ϵ_{Bonus}	1.09	0.592	1.299	1.82	0.196	0.298	0.279	0.283
	(0.147)	(0.062)	(0.123)	(0.136)	(0.053)	(0.026)	(0.038)	(0.064)
ϵ_{bonus_cutoff}	−18.26	−1.90	−16.94	−14.07	−0.14	−0.074	−0.454	−0.100
	(2.023)	(0.564)	(0.889)	(1.609)	(0.144)	(0.050)	(0.252)	(0.137)
Predicted days worked	17.23	17.30	16.87	16.28	16.75	18.381	17.596	18.213
	(0.361)	(0.153)	(0.260)	(0.566)	(0.391)	(0.391)	(0.809)	(0.974)
Days worked $BONUS = 0$	1.31	6.96	1.35	1.174	12.90	9.774	11.314	10.605
	(0.041)	(0.101)	(0.049)	(0.072)	(0.281)	(0.605)	(0.916)	(1.454)
Out-of-sample prediction	21.47	19.975	21.48	21.550	17.77	20.157	19.281	19.948
	(0.046)	(0.164)	(0.030)	(0.060)	(0.479)	(0.287)	(0.753)	(0.678)

Notes: Models I, II, VI, VII, and VIII are estimated using maximum likelihood. Models III, IV, and V 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 semielasticity 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.

two sets of models, one set (III, IV, and V) controlling for serial correlation using an AR(1) specification for the per-period error term, and one set (VI, VII, and VIII) using a shifter for the outside option that depends on whether the teacher worked the previous day.

Model III adds the AR(1) error process to Model I. The model estimates imply that agents respond strongly to the financial incentives and finds strong evidence of positive serial correlation ($\rho = 0.422$), but the control group prediction is still too low. Model IV adds in one degree of unobserved heterogeneity by allowing μ to be drawn anew from a normal distribution at the beginning of each month. These estimates are largely the same as in Model III, with similar poor results for the control group prediction. Finally, Model V adds in a second type of unobserved heterogeneity. The outside option is drawn from one of two normal distributions with a probability p. The estimates from Model V suggest that there are two types of workers in the data: a majority with a μ less than zero, and a small proportion ($p = 0.024$) who have a μ drawn from a much higher distribution. In contrast to the previous models, this model predicts that most teachers have a negative μ, which implies that teachers in the control group will work most days. Model V predicts control group attendance of 12.9 days per month, which is the same as the empirical attendance rate.[30]

While Model V does an excellent job of predicting attendance in the control group, one drawback is that it uses only the first five days of data from each month. We also estimate several models that incorporate serial correlation through a shifter on the outside option, which depends on whether the teacher worked in the previous period. This is a simple method of introducing serial correlation into the model while retaining the ability to use the maximum likelihood approach that can be estimated on the entire dataset.

In Model VI, we estimate a random coefficient specification with μ drawn from one normal distribution and a simple shifter, *yesterday*, which is added to μ if the teacher did not work in the previous period. The results of this model are similar to Model V, as β and the mean of the outside option are estimated to be similar. This model estimates a higher level of variance in the outside option, and finds that not working yesterday makes not working again today more likely. If a teacher did not work on the previous day, μ is shifted up by 0.094. Holding the financial incentives fixed, for the average teacher with $\mu = -0.304$, this implies a decrease in the probability of working in the current period by 4.5 percent. Model VII is the same specification as Model VI, but uses a restricted sample of three days on either side of the change in the month to approximate a structural version of a regression discontinuity model. The counterfactual predictions are similar to Model VI. This is very reassuring, as this identifying variation in this specification is the sharp change in incentives around the change in month.

Model VIII is the same specification of Model VI, estimated on the full sample, with the inclusion of μ shifters for control group attendance in the same geographic block and teacher test scores. Both variables enter μ positively, so the estimates of -0.132 and -0.005 for attendance and test scores, respectively, imply that teachers work more when teachers in the control group geographically proximate to them work more and when they have higher scores, as the negative coefficients on these coefficients imply that the μ decreases with these two variables. The coefficient on the behavior of the control group teachers is significant, which is consistent with the parallel seasonal pattern we observed in the reduced form. The addition of these two controls improves the efficiency of the estimator, but does not significantly change the other parameters or the predicted number of days worked.

[30] A natural question arises: why stop at two types of heterogeneity? Bajari et al. (forthcoming) use the same data as an application of an estimator that is nonparametric with respect to the distribution of μ. Their estimator allows for up to 80 discrete types and 40 continuous types for μ, holding the other parameters at their values from Model V. The results suggest that the mixture of two normals captures the unobserved heterogeneity extremely well. The Bajari et al. result suggests that most of the weight is put on a few points centered around -0.5 and a small diffuse mass of weight around points at higher levels of μ near 2.0, just as the two-type Model V does.

The similarity of results from different estimation methods for the model is encouraging. Particularly reassuring is the fact that the estimate of β, or all the implied elasticities for the model, is very similar (ranging between 0.2 and 0.3) when using all the data (Model VI and VIII), the first five days (Model V), or the three-day window (Model VII). This suggests that the identification based on the shift in incentive at the end of the month drives the results in all the models. In these three models, the mean outside option (which includes the punishment for not working) is negative for the majority of the teachers.[31] This suggest that, taking into account the nonpecuniary cost of absence, teachers are willing to work more than half the days, even without financial incentives. This is consistent with the fact that the teachers work a little over half the time in the control group.

Note that all these models predict about the same rate of absence in the control group as we predict when setting the incentive to zero for our treatment group teacher. It suggests that $\overline{P}$ is close to P, or that there is no direct impact of the daily monitoring.

C. *Goodness-of-Fit 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 the models estimated using maximum likelihood (Models I–III and VI–VIII), 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. Note, however, that this is not a parameter that our method of simulated moments estimation tried to match (since we matched only the first five days of teacher behavior), so it provides a partial goodness of fit metric for these models. Moreover, the model using the three-day window does not match this moment mechanically either.

Figure 4A plots the density of days worked predicted by Model V, and its 95 percent confidence interval, and compares it to the actual density observed in the data. Since the estimation is not calibrated to match this shape, as we only used the history of the first five days in the 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 by one day. The model tends to slightly overpredict the frequency of 17 to 21 days worked and underpredict the frequency between 3 and 10. With the exception of a small proportion of teachers who work few days in a month, the true distribution lies comfortably within the 95 percent confidence interval of the prediction.[32]

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 a 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

[31] In the model with two types, the average outside option is estimated to be -0.375.

[32] As an extra test of goodness of fit, Table 2 in the online Appendix shows the empirical moments and the predicted moments from Model V. Our model generally does well in predicting the patterns of days worked, underpredicting some of the extreme not-work/work sequences (00011, 00111, and 01111) and overpredicting others (11100 and 11110).

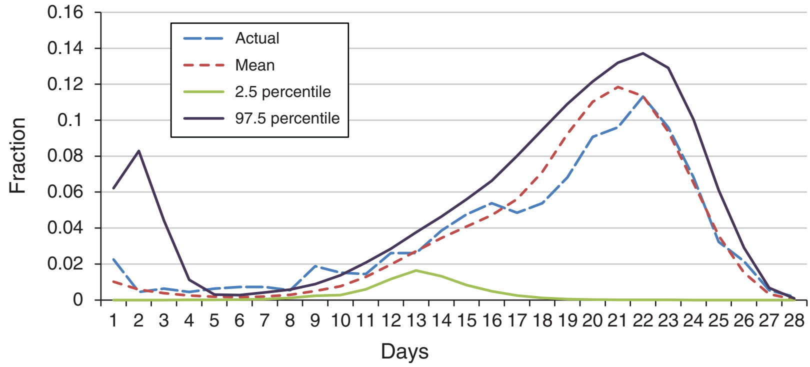

Figure 4A. Predicted Fit from Model V

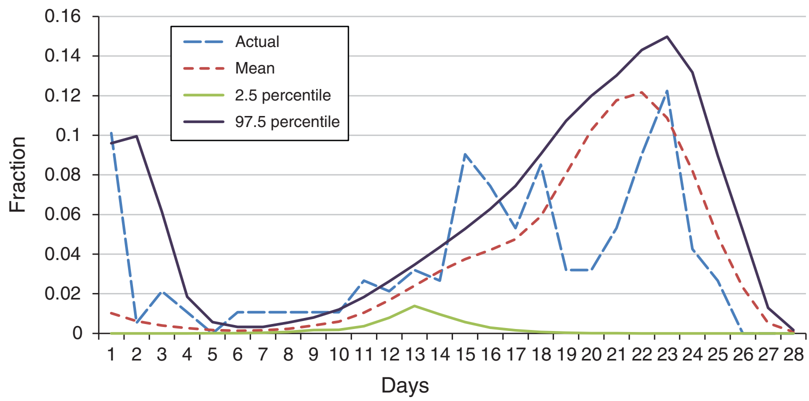

Figure 4B. Counterfactual Fit from Model V

fewer (rather than 10 days). For each additional day they work, teachers earn an additional Rs. 70. Seva Mandir provided us with the camera data in the summer of 2007, a few months after the change in policy. The average number of days worked since January 2007 increased very slightly, from 17.16 to 17.39 days. The predicted number of days worked for each model is reported in the last row in Table 4. Here again, our preferred specifications (Models V–VIII) performs well: they predict between 17.8 and 20.2 days worked under the new incentive scheme, an increase from the predicted number of days under the main scheme (as in the actual data). Figure 4B shows the actual distribution of days worked and the predicted one for Model V. The model does a good job of predicting the distribution of days worked in the out-of-sample test, although the empirical distribution has more variance than in the original experiment.

D. *Counterfactual Optimal Policies*

A primary benefit of estimating a structural model of behavior is the ability to calculate outcomes under economic environments not observed in the data. In our

Table 5—Counterfactual Cost-Minimizing Policies

Expected days worked (1)	Bonus cutoff (2)	Bonus (3)	Expected cost (4)	Test score gain over control group (13 days) (5)
14	0	0	500	0.04
15	21	25	521	0.07
16	22	75	664	0.11
17	21	75	672	0.15
18	20	75	755	0.18
19	20	100	921	0.22
20	20	125	1,112	0.26
21	16	225	2,642	0.29
22	11	275	4,604	0.33

case, we are interested in finding the cost-minimizing combination of the two policy instruments, the size of the bonus and the threshold to get into the bonus, that lead to a minimum number of days worked in a month. Using Model V as our foundation, we calculated the expected number of days worked and expected size of the financial payout for a wide range of potential policies under our preferred model with autocorrelation and two types of heterogeneity.[33] We let the minimum number of days to obtain a bonus range from 0 to 23, which is the upper limit of days that a teacher could work in any month. At the same time, we varied the bonus paid for each day over the cutoff from 0 to 300 Rs./day in increments of 25 Rs./day. Table 5 shows the lowest-cost combinations of those two policy variables that achieved a minimum expected number of days worked under Model V.[34] The table also shows the gain in test scores for each of these combinations (calculated using the estimate of the effect of each extra day on presence and test score, which we estimate below). As with any simulations, it is worth pointing out that, as we move further from the range of parameters under which we have estimated the models, the validity may decrease.

This simulation shows two general trends: the cost-minimizing cutoff generally decreases and the bonus increases in the expected number of days worked that the policymaker wants to achieve. Both of these trends lead to drastically increasing costs as the target increases. This result follows directly from the model: as we increase the target, the marginal teacher has increasingly higher opportunity costs of working. This becomes quite expensive, as soon it is necessary to incentivize the "slacker" teacher types in our sample. It is interesting to note that for about the same amount of money spent on both the treatment and control groups in the actual experiment (roughly 1,000 Rs./month), teachers under the optimal counterfactual policy would have worked approximately 20 days, an improvement of roughly 16 percent and 56 percent over the treatment group and control group, respectively. The counterfactual calculations show that while the actual intervention successfully increased teacher attendance, the NGO could have induced higher work effort with

[33] We also calculated the optimal policies under Model VI to test the robustness of our results. Those policies are reported in online Appendix Table 4. The results are roughly comparable to those under Model V, with an increasing per-day bonus and an inverted-U shaped cutoff function.

[34] The expected outcomes were subject to a very small amount of variance as we drew model primitives from their estimated distributions 50 times for each combination of policy instruments.

Table 6—Teacher Performance

	September 2003–February 2006			Difference between treatment and control schools		
	Treatment (1)	Control (2)	Diff. (3)	Until mid-test (4)	Mid- to post-test (5)	After post-test (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)
	1,239	867	2,106	643	408	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)
	1,239	867	2,106	643	480	983
Blackboards utilized	0.92	0.93	−0.01	−0.03	0.01	−0.01
			(0.01)	(0.02)	(0.02)	(0.02)
	990	708	1,698	613	472	613

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

approximately the same expenditure by doubling the bonus threshold and nearly tripling the per-day bonus. This is due to the fact that teachers in our sample appear to be more likely than not to attend school even without incentives and be forward-looking. A higher threshold avoids rewarding inframarginal days, and provides incentives to teachers to work to accumulate the number of days necessary to get the larger prize.

IV. Was Learning Affected?

A. *Teacher Behavior*

Though the program increased teacher attendance and the length of the school day, it would be 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 6 shows that there were no significant differences 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, the marginal cost of teaching must have been small. This belief was supported by in-depth conversations with 15 randomly selected NFE teachers regarding their teaching habits in November and December of 2005. We found that teachers spent little time preparing for class as teaching in the NFE follows an established routine. 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 NFE, they were occupied with household and field duties, and thus had little time to prepare for class outside of mandatory trainings. Furthermore, despite the poor attendance rates, many teachers displayed a motivation to teach. They stated that they felt good when the students learned and liked the fact that they were helping to educate disadvantaged students.

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 biannual 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 6 hours, but slack of 1 hour was given). On the other hand, many felt empowered as 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 them 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.

B. *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 estimated directly 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 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 7, 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

TABLE 7—CHILD ATTENDANCE

	September 2003–February 2006			Difference between treatment and control schools		
	Treatment (1)	Control (2)	Diff (3)	Until mid-test (4)	Mid- to post-test (5)	After post-test (6)
Panel A. Attendance conditional on school open						
Attendance of students present at pretest exam	0.46	0.46	0.01	0.02	0.03	0.00
			(0.03)	(0.03)	(0.04)	(0.03)
	23,495	16,280	39,775			
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)
	12,956	10,737	23,693			
Panel B. Total instruction time (presence)						
Presence for students present at pretest exam	0.37	0.28	0.09	0.10	0.10	0.08
			(0.03)	(0.03)	(0.04)	(0.03)
	29,489	26,695	56,184			
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)
	16,274	17,247	33,521			
Panel C. Presence, by student learning level at program start (for those who did not leave)						
Took oral pretest	0.50	0.36	0.14	0.11	0.14	0.15
			(0.03)	(0.03)	(0.05)	(0.04)
	14,778	14,335	29,113			
Took written pretest	0.48	0.39	0.10	0.07	0.07	0.11
			(0.06)	(0.07)	(0.06)	(0.07)
	1,496	2,912	4,408			

Notes: Standard errors are clustered at the level of the school. Child attendance data were collected during random checks. The attendance at the pretest exam determined the child enrollment at the start of the program.

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.

Treatment schools, however, 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's impact on child instruction time is reported in panel B of Table 7. 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 (Glewwe and Kremer 2006; Banerjee et al. 2005). The effect on presence does not appear to be affected by student ability (proxied by whether or not the child could take a written test in the pretest). 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 when 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.

C. *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, they are ineffective. Thus, the fact that it is possible to induce them to attend school more often is not particularly policy-relevant. Understanding the effect of the program on learning is therefore critical.

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.[35] 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 pretest, 1,893 also took the mid-test, and 1,760 also took the post-test. Table 8 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 their test scores to children who dropped out of the comparison schools.

Table 8 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 pretest 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.

Finally, Table 8 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

[35] As mentioned earlier, seven centers closed down prior to the start of the program. We made no attempt to test the children from these centers in the pretest.

TABLE 8—DESCRIPTIVE STATISTICS FOR MID- AND POST-TEST

	Mid-test			Post-test		
	Treatment	Control	Difference	Treatment	Control	Difference
Panel A. Attrition process						
Percent attrition	0.11	0.22	−0.10 (0.05)	0.24	0.21	0.03 (0.04)
Difference in percent written of pretest attriters-stayers	0.01	0.03	0.02 (0.06)	0.06	−0.03	0.10 (0.06)
Difference in verbal test of pretest attriters-stayers	0.05	0.08	−0.03 (0.14)	0.02	0.12	−0.10 (0.14)
Difference in written test of pretest attriters-stayers	−0.41	−0.23	−0.18 (0.34)	−0.19	−0.13	−0.06 (0.29)
Panel B. Exam score means						
Took written	0.36	0.33	0.03 (0.04)	0.61	0.57	0.04 (0.05)
Math	0.14	0.00	0.14 (0.10)	−0.08	−0.24	0.16 (0.15)
Language	0.14	0.00	0.14 (0.10)	1.71	1.60	0.11 (0.11)
Total	0.14	0.00	0.14 (0.10)	0.35	0.24	0.12 (0.11)

Notes: Test scores in panel B are normalized by the mean of the mid-test control. Standard errors are clustered by school.

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. Since child test scores are strongly autocorrelated, we obtain greater precision by controlling for the child's pretest score level.

Test Results.—In Table 9, we report the program's impact on test scores. We compare the average test scores of students in the treatment and comparison schools, conditional on a child's preprogram competency 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 score ($Score_{ijk}$) on test k (where k denotes either the mid- or post-test exam):

$$(13) \qquad Score_{ijk} = \beta_1 + \beta_2 Treat_j + \beta_3 Pre\_Writ_{ij} + \beta_4 Oral\_Score_{ij} + \beta_5 Written\_Score_{ij} + \varepsilon_{ijk}.$$

Since test scores are highly autocorrelated, controlling for a child's test scores before the program increases the precision of our estimate. The specific structure of the pretest (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 pretest), however, does not allow for a traditional difference-in-differences (DD) or "value added" (child fixed effect) strategy. Instead, we include a variable containing the child's pretest score for the oral test if he took the oral pretest and 0 otherwise ($Oral\_Score_{ij}$), the child's pretest score on the written test if he took the written test and 0 otherwise ($Written\_Score_{ij}$), and an indicator variable for whether he took the written test at the

Table 9—Estimation of Treatment Effects for the Mid- and Post-Test

	Mid-test				Post-test			
	Took written (1)	Math (2)	Lang. (3)	Total (4)	Took written (5)	Math (6)	Lang. (7)	Total (8)
Panel 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)
	1,893	1,893	1,893	1,893	1,760	1,760	1,760	1,760
Panel B. With controls								
	0.04	0.13	0.14	0.14	0.06	0.18	0.14	0.15
	(0.03)	(0.07)	(0.06)	(0.06)	(0.04)	0.13	0.08	0.09
	1,752	1,752	1,752	1,752	1,760	1,760	1,760	1,760
Panel C. Took pretest 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)
		1,550	1,550	1,550		1,454	1,454	1,454
Panel D. Took pretest 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
Panel E. Girls								
	0.07	0.18	0.18	0.19	0.07	0.22	0.13	0.17
	(0.03)	(0.07)	(0.07)	(0.07)	(0.05)	(0.12)	0.08	(0.09)
	876	876	876	876	811	811	811	811
Panel F. Boys								
	0.02	0.11	0.13	0.13	0.05	0.20	0.15	0.16
	(0.04)	(0.09)	(0.07)	(0.07)	(0.04)	(0.15)	(0.10)	(0.10)
	983	983	983	983	926	926	926	926

Notes: 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. The mid- and post-test scores are normalized by mid-test control group. Controls in panel B include Block, Teacher Test Scores, and Infrastructure Index. Standard errors are clustered by school.

pretest (Pre_Writ_{ij}).[36] This fully controls for the child's pretest achievement, and is thus similar in spirit to a DD strategy. Standard errors are clustered by school. Each cell in Table 9 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.[37]

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,

[36] At the pretest, 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).

[37] Scores are normalized such that the mean and standard deviation of the comparison group at the time of the mid-test exam are zero and one, respectively. (Specifically, we subtract the mean of the comparison group in the pretest, 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 pretest score distribution since not every child took the same test at the pretest.

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 pretest gained the most from the program. For example, they had midline test scores 0.25 standard deviations higher in treatment schools than in comparison schools (panel D). Interestingly, the children who could write at the time of the pretest 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 pretest. 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 better equipped to make the most out of the additional days of schooling that they receive.

We compare the program's impact 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, seven percentage points more of girls in the treatment schools were able to write relative to the comparison schools, compared to only two 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 pretest. 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 Balsakhi Remedial Education Program in India during its first year (Banerjee et al. 2005).

Leaving the NFE.—Nonformal Education Centers prepare children, who might not otherwise attend school, to enter government schools at an age-appropriate grade. To do so, children must demonstrate proficiency 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 graduation rate to the government schools. As shown in Table 10, 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 10, 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.

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 11 interprets what these estimates can tell us about the impact of teacher attendance.[38] Columns 1 to 3 report simple correlations between the teacher attendance rate and the child test scores. Specifically, we report the coefficient

[38] This estimate is 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.

Table 10—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	1,136	1,061	2,197

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

Table 11—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)
Panel A. Mid-test (September 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	1,015	1,015	1,893
Panel B. Post-test (September 03–October 04)				
Took written	0.31	0.51	0.59	0.33
	(0.15)	(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	1,760

Notes: 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. The mid- and post-test scores are normalized by the mid-test control group. Standard errors are clustered by school.

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:

$$Score_{ijk} = \beta_1 + \beta_2 Open_j + \beta_3 Pre\_Writ_{ij} + \beta_4 Oral\_Score_{ij} + \beta_5 Written\_Score_{ij} + \varepsilon_{ijk}. \tag{14}$$

We continue to control for the child's pretest 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 estimate is similar to those reported in other studies (Kremer et al. 2005) and, taken at face value, would imply that the effect of teacher attendance on learning is not that large. Kremer et al. (2005) conjectures that the measurement of absence rates based on a few random visits per school has 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 more direct test of this hypothesis since the photograph data gives us the actual attendance of treatment teachers. 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 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 column 3 as compared to 0.39 in column 2). For the post-test, where we have a 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 ten 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 (1.12 standard deviations in Banerjee et al. 2005). Both of these studies therefore suggest that para-teachers can be effective teachers, at least when an NGO provides them with proper training.

V. Conclusion

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. Absenteeism fell from an average of 42 percent in the comparison schools to 21 percent in the treatment schools, without affecting the teachers' effort while in school. As a result, the students in treatment schools benefited from about 30 percent more instruction time. The program had an 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.

This paper contributes to a small but growing literature that exploits both structural modeling and carefully controlled randomized experiments to answer an economic question. On the substantive front, our results suggest that providing incentives for attendance in nonformal schools can increase learning levels. The question arises, however, as to whether incentive programs can be instituted for government teachers, who tend to be politically powerful. It may prove difficult to institute a system in which they would be monitored daily using a camera or similar device. Our findings suggest, however, that the barriers currently preventing teachers from attending school regularly (e.g., distance, other activities) are not insurmountable. Given political will, it is possible that solutions to the absence problem could be found in government schools as well.[39]

A recent experiment demonstrates the external validity of these results outside the NGO context (Banerjee, Duflo, and Glennerster 2008). Following the results of the cameras program, the government of Rajasthan created a similar system for government nurses, whose absence rate was about 44 percent. The nurses were monitored using time and date stamps. 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 carried out, the program led to about a 50 percent reduction in absenteeism. After a few months, however, the government started granting a large number of "exemptions" (although the monitoring did continue). The absence rate in the treatment group quickly converged to that of the control group. This further confirms that 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 incentive systems for government employees. Our findings also imply, however, that para-teachers can be effective. If implementing monitoring within the government system turns out to be impossible, our results provide support for the policy of increasing teaching staff through the hiring of para-teachers.

REFERENCES

Aguirregabiria, Victor, and Pedro Mira. 2010. "Dynamic Discrete Choice Structural Models: A Survey." *Journal of Econometrics* 156 (1): 38–67.

Attanasio, Orazio, Costas Meghir, and Ana Santiago. Forthcoming. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresa." *Review of Economic Studies.*

Bajari, Patrick, Jeremy T. Fox, Kyoo il Kim, and Stephen P. Ryan. Forthcoming. "A Simple Nonparametric Estimator for the Distribution of Random Coefficients." *Quantitative Economics*.

Banerjee, Abhijit, and Esther Duflo. 2006. "Addressing Absence." *Journal of Economic Perspectives* 20 (1): 117–32.

[39] 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.

Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster. 2008. "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System." *Journal of the European Economic Association* 6 (2–3): 487–500.

Banerjee, Abhijit, Suraj Jacob, Michael Kremer, Jenny Lanjouw, and Peter Lanjouw. 2005. "Moving to Universal Education: Costs and Tradeoffs." Unpublished.

Benabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78.

Bound, John, Todd Stinebrickner, and Timothy Waidmann. 2010. "Health, Economic Resources and the Work Decisions of Older Men." *Journal of Econometrics* 156 (1): 106–29.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.102.4.1241.

Fehr, Ernst, and Lorenz Goette. 2007. "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment." *American Economic Review* 97 (1): 298–317.

Fehr, Ernst, and Klaus M. Schmidt. 2004. "Fairness and Incentives in a Multi-task Principal-Agent Model." *Scandinavian Journal of Economics* 106 (3): 453–74.

Ferrall, Christopher. 2010. "Explaining and Forecasting Results of the Self-Sufficiency Project." Queen's University Department of Economics Working Paper 1165.

Figlio, David N., and Lawrence S. Getzler. 2006. "Accountability, Ability and Disability: Gaming the System?" In *Advances in Applied Econometrics Vol. 14,* edited by Michael Baye and John Maxwell, 35–49. Bingley, UK: Emerald Group Publishing Limited.

Figlio, David N., and Joshua Winicki. 2005. "Food for Thought: The Effects of School Accountability Plans on School Nutrition." *Journal of Public Economics* 89 (2–3): 381–94.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3): 205–27.

Glewwe, Paul, and Michael Kremer. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." In *Handbook of the Economics of Education Volume 2,* edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 945–1017. Amsterdam: Elsevier.

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

Jacob, Brian A., and Steven D. 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 Robert Moffitt. 1998. "A Structural Model of Multiple Welfare Program Participation and Labor Supply." *International Economic Review* 39 (3): 553–89.

Keane, Michael, Petra Todd, and Kenneth I. Wolpin. 2011. "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications." In *Handbook of Labor Economics Volume 4*, edited by O. Ashenfelter and D. Card, 331–461. Amsterdam: Elsevier.

Kremer, Michael, Nazmul Chaudhury, F. Halsey Rogers, Karthik Muralidharan, and Jeffrey Hammer. 2005. "Teacher Absence in India: A Snapshot." *Journal of the European Economic Association* 3 (2–3): 658–67.

Kreps, David M. 1997. "Intrinsic Motivation and Extrinsic Incentives." *American Economic Review* 87 (2): 359–64.

Lavy, Victor. 2009. "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics." *American Economic Review* 99 (5): 1979–2011.

Lise, Jeremy, Shannon Seitz, and Jeffrey Smith. 2005. "Equilibrium Policy Experiments and the Evaluation of Social Programs." National Bureau of Economic Research Working Paper 10283.

McFadden, Daniel, and A. P. Talvitie and Associates. 1977. "Validation of Disaggregate Travel Demand Models: Some Tests." In *Demand Model Estimation and Validation*. Berkeley, CA: Institute of Transportation Studies, University of California, Berkeley.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77.

Pratham. 2006. *Annual Status of Education Report*.

Stinebrickner, Todd R. 2000. "Serial Correlated Variables in Dynamic, Discrete Choice Models." *Journal of Applied Econometrics* 15 (6): 595–624.

Stinebrickner, Todd R. 2001a. "A Dynamic Model of Teacher Labor Supply." *Journal of Labor Economics* 19 (1): 196–230.

Stinebrickner, Todd R. 2001b. "Compensation Policies and Teacher Decisions." *International Economic Review* 42 (3): 751–79.

Sullivan, Paul. 2010. "A Dynamic Analysis of Educational Attainment, Occupational Choices, and Job Search." *International Economic Review* 51 (1): 289–317.

Todd, Petra E., and Kenneth I. Wolpin. 2010. "Structural Estimation and Policy Evaluation in Developing Countries." *Annual Review of Economics* 2 (1): 21–50.

Todd, Petra E., and Kenneth I. Wolpin. 2006. "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility." *American Economic Review* 96 (5): 1384–1417.

van der Klaauw, Wilbert. 2005. "The Supply and Early Careers of Teachers." Unpublished.

Wise, David. 1985. "A Behavioral Model vs. Experimentation: The Effects of Housing Subsidies on Rent." In *Methods of Operations Research 50*, edited by P. Brucker and R. Pauly, 441–89. Frankfurt am Main, Germany: Verlag Anton Hain.

This article has been cited by:

1. Benjamin A. Olken, Junko Onishi, Susan Wong. 2014. Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia. *American Economic Journal: Applied Economics* **6**:4, 1-34. [Abstract] [View PDF article] [PDF with links]
2. Dana Burde,, Leigh L. Linden. 2013. Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools. *American Economic Journal: Applied Economics* **5**:3, 27-40. [Abstract] [View PDF article] [PDF with links]
3. Sylvain Chassang,, Gerard Padró i Miquel,, Erik Snowberg. 2012. Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments. *American Economic Review* **102**:4, 1279-1309. [Abstract] [View PDF article] [PDF with links]