Unintended Consequences of Rewards for Student Attendance: Results from a Field Experiment in Indian Classrooms*

Sujata Visaria†
Rajeev Dehejia‡
Melody M. Chao†
Anirban Mukhopadhyay†

December 12, 2015

Abstract

In an experiment in non-formal schools in Indian slums, a reward scheme for attending a target number of school days increased average attendance when the scheme was in place, but had heterogeneous effects after it was removed. Among students with high baseline attendance, the post-incentive attendance returned to previous levels and test scores were unaffected. Among students with low baseline attendance, post-incentive attendance dropped even below previous levels, and test scores decreased. These students also reported lower interest in school material and lower expectations of themselves. Thus incentives might have unintended consequences in the long term for the very students they are most expected to help.

Keywords: educational economics, incentives, attendance, motivation, experiment

---

*We are indebted to Dr. Pankaj Jain, Hiral Adhyaru, Sonal Mody and numerous class teachers and supervisors at Gyan Shala for their support and cooperation on this project. We cannot thank Putul Gupta enough for her terrific management of the project in the field. We have received very helpful comments from Mark Rosenzweig, Yasutora Watanabe, and participants at the IEMS 8th Asian Conference on Applied Microeconomics/Econometrics. Funding for the field implementation of this project was received through the Research Project Competition at The Hong Kong University of Science and Technology (Grant RPC10BM11). All errors are our own.

†Hong Kong University of Science and Technology

‡New York University

1 Introduction

A growing literature has examined whether incentives can lead students from underprivileged backgrounds to increase effort and improve performance (Angrist and Lavy 2009; Kremer, Miguel, and Thornton 2009; Fryer 2011; Bettinger 2012; Levitt et al. 2012). The underlying assumption behind incentive interventions is that the target students have sub-optimally low motivation to exert effort at school. This may because they are unaware of the benefits of schooling, are too impatient to work for benefits that will accrue far in the future, or lack the self-control to trade off current costs against future benefits. A nearer-term incentive that rewards them for say, reading a book or attending school, can provide the "carrot" that will change their behavior.

Problems of impatience and self-control notwithstanding, some students do exert effort and achieve high test scores. Those examining incentive interventions do not expect the largest gains among these students: since they already exert high effort, any gains at the margin will be small. Instead, incentives are expected to have large treatment effects on children who have lower baseline academic outcomes due to the lack of motivation. For such students, the promise of a reward creates the motivation to do the task. This increased effort can lead to improved academic performance. This could happen if the incentivized effort is large enough, or if the increased effort becomes a habit (Charness and Gneezy 2009).[1]

However it is also possible for incentives to backfire: in particular, the extrinsic motivation provided by an incentive could crowd out students intrinsic motivation to study and learn (Gneezy, Meier, and Rey-Biel 2011).[2] Attaching a price to a task that was initially enjoyable can make it less enjoyable (Deci and Ryan 1985). As Gneezy, Meier, and Rey-Biel (2011) argue, if the incentive is large enough it might create the extrinsic motivation to increase effort in order to earn the reward. However once it is removed, the lack of extrinsic motivation

[1] Charness and Gneezy (2009) find that when university students were given high-powered incentives to attend a gym, they became more likely to exercise even after the incentives were discontinued.

[2] A large literature in psychology also discusses the crowd-out of intrinsic motivation (see for example, Deci, Koestner, and Ryan (1999)).

coupled with the lower intrinsic motivation could lower student effort to a level below what it would have been if no reward had been offered.

Two points emerge from this discussion. One, if incentives increase extrinsic motivation and do not change intrinsic motivation, then they should have the largest (positive) effects on students with low baseline motivation. If they do lower intrinsic motivation, then it is unclear how this affects observed behavior. Presumably students with high motivation might have more of it to lose, but it is possible that the decrease is still not enough to change effort or performance. In contrast, those with low motivation may be relatively disengaged to start with and so the crowd-out might worsen their effort and performance. Most studies have focused on the average effects across these two subgroups, which makes it difficult to identify the channels at work.

Second, crowding-out is best detected by studying students behavior after the reward has been discontinued. Although researchers have examined long-term effects of incentives to exercise, stop smoking, and engage in pro-social behavior (Gneezy, Meier, and Rey-Biel 2011), few papers in education have examined effects after the incentive period ended. A notable exception is Rodriguez-Planas (2012), who examines the effect of the high-school Quantum Opportunity Program on youths two years and five years after the program ended. Although she cannot identify specific mechanisms that caused the positive effects of the program to fade over time, she also finds an interesting difference in the fade-out by subgroups: educational and employment outcomes in the long-term were better for treated females, but not for treated males.[3]

In this paper we report on a field experiment where the attendance of students of non-formal schools in Indian slums was monitored and an incentive was offered for meeting an attendance target. To evaluate whether the effect of the incentive varies by students' baseline motivation levels, we examine separately students with low and with high prior attendance rates, both during and

[3]We do not find evidence for such a gender difference. This is not surprising, since the children in our study are only in primary school, and both the intervention and the social context are completely different. Unlike in Rodriguez-Planas (2012), this intervention does not provide mentoring or protection against sanctions, and in any case boys and girls are not generally engaging in risky behavior where mentoring or sanctions might have differential impacts by gender.

after the 39-day reward period. We find that both in the pooled sample as well as within the two subgroups, the incentive increased student attendance while it was in place.[4] However, the two subgroups were affected very differently after the incentive period ended. Students in the incentive group who had high baseline attendance attended school at the same rate as their counterparts in the control group. However, those with low baseline attendance were even less likely to attend school than they would have been if the incentive had not been offered.

Our results show that it is instructive to examine the effects of incentives for students with low and high initial motivation separately. However, the effects of incentives are not in line with the ideas that incentives primarily help students with low motivation, or that they hurt students with high motivation. The incentive appears to have had no long-term effects on students who had high motivation to begin with. Instead, it had negative long-term impacts on students with low motivation to begin with, a group that arguably has the most to gain from improving attendance.

Scores on a test administered three months after the incentive scheme were also affected in the same manner: the test scores of students with high prior attendance were unaffected by the incentive scheme, but those of students with low prior attendance became lower than if there had been no incentive at all. The reward also lowered these students' liking for school subjects, and lowered their expectations of themselves. Thus, in contrast to the existing literature, we find that although the incentive motivated students while it was in place, it had unintended negative consequences in the longer term for students with low baseline motivation, whom it was intended to help.

The rest of this paper is organized as follows. Section 2 describes the empirical context. Section 3 describes the experimental intervention and data. Section 4 presents the empirical results. Section 5 discusses the implications of the study and concludes.

[4]The effect on the low baseline attendance group is large in magnitude but imprecisely estimated.

2 The Empirical Context

Our experiment was conducted in collaboration with Gyan Shala, a non-government organization that runs non-formal education centers (henceforth referred to as GS classes) in the slums of Ahmedabad in the state of Gujarat in western India. In 2010, Gyan Shala had 343 classes operating across 5 areas in the city (CfBT Education Services 2010). Each Gyan Shala education center is a single-grade class housed in a single room in a slum, usually rented from a local resident. Students pay no fees. The median class in our sample has 22 students, all of whom are from the same or from neighboring slums.[5] Each classroom has basic school supplies. While teaching is mainly lecture-based, each student has a workbook with exercises that he/she must work on during class. Three subjects are taught: language (Gujarati), mathematics and science.

Gyan Shala's mission is to provide children with low socioeconomic status a high quality education at a low cost. Operational costs are kept low by hiring teachers who do not have a teaching qualification, and therefore would not be hired by formal schools. Most teachers have only a Class 12 school certificate. To ensure quality, Gyan Shala trains these teachers intensively: the typical school year includes 30 training days. The teachers closely follow day-wise lesson plans that they receive from a "design team" made up of subject specialists who hold bachelor's or master's degrees. A supervisor visits each class once a week to observe and provides inputs as needed. When students in particular classes find individual topics difficult to understand, design team members visit the classroom to gauge the problem and to help the teacher. The information gathered is fed back into future lesson plans.

The parents of Gyan Shala students are for the most part self-employed or casual workers in the unorganized sector. They have low education levels and therefore limited ability to support their children's learning at home. Gyan Shala hopes to provide these parents with an attractive alternative to the local municipal school, while also demonstrating that a good education need not be expensive. An independent evaluation conducted by Educational Initiatives

[5] An important consideration for Gyan Shala is that children be able to walk to school unescorted, since this lowers the time and transport costs of attending school and helps to lower absenteeism.

(EI) in 2010 found that Gyan Shala students outperformed their peers in municipal schools on language, mathematics and science by wide margins (Educational Initiatives Private Limited 2010). On average Gyan Shala students were also better able to answer the more difficult, “non-straightforward” questions on EI’s tests. A short-lived experimental intervention where Gyan Shala’s teaching techniques were adopted in municipal schools also generated significant impact, with treatment municipal schools outperforming control municipal schools (Educational Initiatives Private Limited 2010).

Gyan Shala’s main effort has been to run classes for grades 1, 2 and 3. Our experiment was conducted in grade 2 and grade 3 classes, but we report here only the results for grade 3 classes because those are the only students who took a test administered by Educational Initiatives (EI), that provides us with an independent assessment of their achievement.[6] The EI examination only tested mathematics and science.

The goal of this study was to examine the effect of an incentive for student effort, on student performance.[7] The administrators at Gyan Shala identified attendance as the appropriate task to target. Research in higher-income countries has shown that student attendance is correlated with performance (Roby 2004; Paredes and Ugarte 2011). It is likely that this relationship is even stronger in our context, where parents can provide limited support at home. At an unannounced visit made by our investigators two months into the 2011-12 school year, 75 percent of students in sample classes were present. This number matches the 75 percent average attendance rate for Gujarat state reported by previous research (Educational Consultants India Limited 2007). While considerably lower than the standards set by school boards in some developed countries, this number is also not so low that it might be mainly caused by structural factors outside students control.[8] Gyan Shala administrators believed that the

[6]Our results are qualitatively unchanged when we include Grade 2 students in the analysis on attendance.

[7]This is part of a larger project aimed at understanding the impact of economic and psychological interventions on student achievement. For more detail, see Chao et al. (2015). In this paper we focus on the intervention that rewarded students for good attendance. The psychological intervention was implemented orthogonally to the reward intervention and we do not examine its effect in this paper.

[8] For example, the California legislature defines as a chronic truant a student who is absent from school without a valid excuse for ten or more percent of school days in one school

bulk of their students' absence was truancy: missing school for reasons such as wanting to play instead, it being a festive season, or because their siblings had a day off at their school.

3 The Data and the Experimental Intervention

Our study took place during the academic year 2011-12. The academic year begins in June and ends in April. Our sample consists of roughly 12 students randomly sampled from each of 68 grade 3 classes, that are spread evenly across all 5 city zones where Gyan Shala operates. Figure 1 summarizes the sequence of events in our study. Investigators made six unannounced visits to the classrooms; we label these visits Time 0 through Time 5. At all six visits, they took roll-call of the sample students to check if they were present.[9] At three of these visits (Time 1, Time 3 and Time 5) they also conducted 10-minute surveys with the sample students. Survey questions were about the students like and dislike for particular subjects, and their expectations and attitudes about learning and exerting effort on difficult tasks. At Time 1 students were also asked to provide demographic information about themselves and their family members.

Table 1 presents summary statistics from this Time 1 interview, and checks whether there were significant differences between the control and treatment classes. About half of the 799 sample students were female. They were on average 9 years old. Since we did not interview their parents, we had to rely on the childrens reports of household assets to infer socioeconomic status. We also measured their height and weight, on the grounds that their body mass index may be correlated with their socioeconomic status. Note however that all children are residents of low-income neighborhoods and so variation in SES is likely to be small. The average child had a body mass index of 13.8, which places them between the 3rd and 5th percentiles of a normal international population (World Health Organization 2007).

year (California Department of Education 2015).

[9] All visits were scheduled to begin at least an hour after the school day began, so as to not miss late-comers. However, since the Gyan Shala classes are located within the students' own neighborhoods, a teacher could have sent word to summon absent students to class when the investigator arrived. To prevent this from contaminating our attendance measure, we instructed the investigator to assign a separate code (E for "entered during visit") to any child who entered the classroom after she had entered it. In our analysis such students are considered absent.

Ninety-three percent of children reported that at least one person in their household owned a mobile phone. A quarter reported that their parents had a motorized vehicle. Three quarters had a toilet in the house, and a little over a third had a VCR or DVD player. Computers were almost non-existent. There were no significant differences between the control and treatment groups on these dimensions.

At the Time 0 visit conducted about 6 weeks after the school year had begun, investigators found 75 percent of the sample students present in class. This is in line with the administrative attendance records, according to which these students were present for 78 percent of days during the first two months of the school year. We do not find significant differences across treatment (mean = 0.02) and control (mean = -0.03) classes in the z-score of the students scores on the previous year's final exam (conducted by Gyan Shala). Students provided a rating for their liking for each of the three subjects they were taught, on a 7-point scale. To assess how much a student liked a school subject, we first asked them to choose either a smiling, neutral or sad face to indicate how they felt about the subject. If they chose the smiling face, they were then asked to choose one of three happy faces where the faces and smiles were small, medium or large, to indicate how intensely they liked it. If instead they chose the sad face, they were shown three unhappy faces to choose from, where the faces, frowns and tears became incrementally larger.

As can be seen, mathematics was very popular among students, with an average rating of 2.5 on a scale ranging from -3 to +3. The difference between control and treatment schools was not significant. Science was relatively less popular, with an average rating of 2. To elicit students' opinions about their ability to pick up new skills, we asked them if they thought they could learn solve a crossword puzzle. (They knew what crossword puzzles were because they had been introduced to them shortly before the Time 1 interview.) Ninety-six percent of students answered in the affirmative. We also tried to elicit students' optimism about their ability to perform at a challenging task. To do this, we told them about a hypothetical child attempting a difficult sum, and asked them to predict the child's performance, on a scale of 1 to 5.[10] The average prediction was

[10]Although we could have framed this question in terms of the students themselves, i.e. ask-

2.2. The difference between treatment and reward schools was not statistically significant. We therefore conclude that the control and treatment groups were balanced on observables.

After the Time 0 (August), Time 1 (September-October) and Time 2 (November) visits had taken place, in December the supervisors introduced the incentive scheme in randomly selected classes. In each city zone, classes were first stratified by neighborhood and then randomized so that classes with and without the incentive scheme were in different neighborhoods. This helped to address the concern that students in control classes might hear about the incentive scheme. The scheme promised a reward to all students in the class who attended more than 85% of school days during the 39-day period between 14th December and 31st January. To inform students about the scheme, the supervisors put up on the wall a chart with each student's name and each school date during the incentive period. Next, following a script that the research team had prepared, they told the students that when they skipped school, it became harder for them to understand the material that was taught, and this also affected their ability to learn subsequent material. The school had decided that any student who attended school regularly would receive a reward. Their attendance would be marked on the chart every day during the specified period. At the end of this period, all students who had attended more than 33 days would be eligible for a reward. The students were then shown samples of the reward (each reward was two pencils and a brightly colored eraser shaped like an animal), and were told that the supervisor would give them one of these as a reward.[11] On each day during the reward period, the teacher was asked to fill in the chart, but not to mention it directly to any student. In the classes that were assigned to the control group, the supervisors gave each teacher a similar chart to fill in every day. The chart was not made public, and the supervisor did not make any announcements in class.[12]

ing them to imagine that they were attempting a difficult sum, we chose to frame it in terms of a hypothetical child (whose gender was kept ambiguous) in the hope of avoiding implicit biases about gender.

[11] Although these rewards had small monetary value, we had found in a pilot the previous year that they were appealing to the students.

[12] Thus our incentive scheme involved both the promise of a reward to students who met the attendance threshold, and a public monitoring of each student's attendance. We are thus unable to disentangle the pure effect of a reward scheme absent any public monitoring. It can be argued

The Time 3 visits took place during the incentive period, thus allowing us to examine how students responded to the scheme while it was in place. At the end of the incentive period, our project coordinator collected all the charts and identified the students who had earned the reward, all of whom received their rewards from the supervisors at small ceremonies in the classroom. All rewards were distributed within two weeks of the end of the incentive period. No further announcements about attendance were made.

Two further visits took place at Time 4 and Time 5, roughly one month and two months after the incentive period ended. Finally, in March, all grade 3 students took a test in mathematics and science, administered by Educational Initiatives (EI).[13] Their tests were aimed at uncovering student ability, and so did not directly test the material covered in the classroom. Questions were designed to test a variety of types of knowledge, ranging from fact and concept recognition to complex problem-solving and analysis skills. Thus rote learning was unlikely to guarantee a high test score. Note also that since Gyan Shala teachers strictly follow a daily lesson plan, they were unlikely to be able to teach to the test.

All test questions were multiple-choice. Students were given question papers, the exam administrator read aloud an exam question, asked students to circle the correct alternative, and then moved on to the next question. Test administrators unaffiliated with Gyan Shala then took these question papers and filled in an optimal mark recognition (OMR) sheet for the student. To minimize costs, Gyan Shala opted to have a random subsample of exam scripts graded. These were then processed, and the test scores were delivered both to EI and to Gyan Shala. EI then prepared a summary report of the students performance in each class. This report also classifies each question in the test according to the type of knowledge it was testing. Using this information, we classify the questions as “simple”, “intermediate” and “complex” and analyze not just the total

however that any reward scheme in such a setting would have to perforce involve a public monitoring component, so as to implement it transparently and ensure that there is common knowledge between the student and the teacher/incentivizer about the student’s attendance and eligibility for the reward.

[13]Educational Initatives provides an independent testing service. The scores on tests administered by EI have been used to evaluate student performance in previous research on education in India (Muralidharan and Sundararaman 2011).

scores in the math and science tests, but also the scores in each category. We have test score data for 584 students.[14]

4 Empirical Specification and Results

4.1 Attendance

In this section we examine the effect of the incentive scheme on attendance. We examine separately the effect when the incentive was in place and after it had been removed.

We start by depicting the key patterns as seen in the raw means from the data. Next we run regressions with additional controls and student fixed effects, to provide rigorous estimates of the effects of the incentive scheme. As noted above, we have 799 students in the sample. For each of these students we have data on whether they were present in school at six different points in time (Time 0 - Time 5).

As column 2 in Table 2 shows, average attendance rates vary from a low of 72 percent to a high of 86 percent over the 6 visits. Columns 5-7 and columns 8-10 show how the attendance rates varied between the control classes and the incentive classes, and how the subsequent attendance rates differed between baseline (Time 0) attenders and non-attenders. In each of these two subgroups of students, in the control classes, attendance dipped from Time 1 to Time 2 and then increased in Time 3. Recall that the intervention was the promise of a reward for attending 85 percent or more of school days during a 39-day period in December-January. Since the Time 3 visit took place during this 39-day period, the difference in attendance between the incentive and reward classes at Time 3 reflects the effect of the incentive on attendance. At Time 3, 90 percent of incentive class students were present, compared to 81.5 percent of control class students.

[14]In linear probability regressions, neither assignment to treatment nor baseline attendance at Time 0 predict the probability that we observe a test score for a student. Note that not all 584 of these students are in our sample of 799 students above. Restricting the sample to the intersection of the two sets leaves us with 308 students. When we evaluate the effect of the intervention on only these 308 students, our results are qualitatively unchanged.

Time 4 and Time 5 visits occurred after the reward period had ended, and therefore allow us to see if the incentive had a persistent effect even after it had been discontinued. As we can see, both at Time 4 and Time 5, attendance was lower than at Time 3 for all students. However at Time 4, incentive students still remained more likely to attend than control students. This effect wore off with time, so that in both subgroups (high and low baseline attenders), incentive students were less likely to be present at Time 5 than control students.

In Table 3 we run linear probability regressions according to the specification below, in order to provide rigorous estimates of the effect of the intervention.

$$\begin{aligned} y_{ict} = \quad & \alpha_i + \beta_1 \text{Time } 3_t + \beta_2 \text{Time } 4_t + \beta_3 \text{Time } 5_t + \beta_4 (\text{Reward}_c \times \text{Time } 3_t) \\ + \quad & \beta_5 (\text{Reward}_c \times \text{Time } 4_t) + \beta_6 (\text{Reward}_c \times \text{Time } 5_t) + \epsilon_{ict} \end{aligned}$$

Here y_{ict} is a binary variable that takes value 1 if student i in class c was present at the investigator visit at time t, and is zero otherwise. The α_i represent student fixed effects that capture all time-invariant personal and location-specific characteristics that may influence attendance. Thus if there there are fixed personal characteristics correlated with low socio-economic status preventing a student from attending school regularly, these do not affect our results.[15] The Time 0 observations are removed from the sample because as we shall see below they are used to classify students by baseline attendance levels. Standard errors are clustered at the class level to control for intra-class correlation in attendance.[16]

In line with our discussion above, we interpret the coefficient β_1 as the time effect on attendance rate at Time 3 among control students. The coefficient β_4 tells us how the incentive students' attendance differed from this effect, and thus picks up the effect of the incentive at the time when the incentive was in effect. Column 1 in Table 3 shows that the incentive increased the likelihood that the average student attended school. At the Time 3 visit, the probability that the investigators found a sample student present in the control classes was

[15] Student fixed effects also absorb the dummy for Reward_c.

[16] These results are robust to fixed-effects logit specifications instead of linear probability models.

the same as before ($\hat{\beta}_1$ = 0.035, s.e. = 0.026, no statistically significant), but in the reward classes, the likelihood was 10.9 percentage points higher ($\hat{\beta}_1 + \hat{\beta}_4$ = 0.109, s.e. = 0.022, p-value = 0.000). The coefficient β_4 can thus be interpreted to imply that at Time 3, the incentive increased the average student's attendance by a statistically significant 7.4 percentage points. Given that the reward period was 39 days, this implies that the incentive caused the average child to attend school for an additional 2.9 days.

Thus we find that while the incentive was in place, it caused attendance to increase. This is line with expectations: if the incentive is attractive, it can increase student effort (Gneezy, Meier, and Rey-Biel 2011). However if incentives reduce intrinsic motivation, then after the incentive is discontinued, student motivation should be even lower: not only would they no longer have the extrinsic motivation to attend, they would also have lower intrinsic motivation. This could make them even less likely to attend than the control students. Accordingly, we examine the effect of the reward 1 month after (Time 4) and 2 months after (Time 5) the reward period ended. Again, at Time 4 and at Time 5, control students were no more likely to attend school than before ($\hat{\beta}_2$ = -0.018, s.e. = -0.024, p-value = 0.464, and $\hat{\beta}_3$ = -0.028, s.e. = -0.025, p-value = 0.279). The incentive did not change this non-effect either ($\hat{\beta}_5$ = 0.014, s.e. = 0.039, p-value = 0.720; $\hat{\beta}_5$ = -0.034, s.e. = 0.039, p-value = 0.389), suggesting that the positive effect of the incentive scheme did not persist after the incentive was removed.

However, as discussed earlier, there is reason to believe that the incentive has different long-term effects on students with low and high baseline motivation to attend school. Accordingly, in columns 2 and 3 we divide the sample into two subgroups using as a proxy for baseline motivation their attendance during the Time 0 unannounced visit. In column 2, we focus on baseline attenders, and find that the incentive increased their likelihood of attending school by a statistically significant 6.7 percentage points. After the incentive was removed, their attendance rate was no different from the control group (at either Time 4 or Time 5). This is in line with the idea that there was either no reduction in their intrinsic motivation, or that the reduction was too small to change their attendance.

In column 3 we focus on baseline non-attenders (absent at Time 0). Although the magnitude of the incentive effect is large at 9.3 percentage points, it is imprecisely estimated. As a result we do not have rigorous evidence that the reward increased these students' attendance. Strikingly however, at Time 5, these students were 13.9 percentage points less likely to attend school that similar baseline non-attenders in control classes. Thus, in contrast to the previous literature, we do find a negative long-term effect of the incentive, but only among students who had low baseline attendance. The incentive lowered their attendance rate in the post-incentive period below what it would have been if no incentive had been offered.

4.2 Test scores

We see the same pattern in student performance. In Table A10, we run regressions with the specification

$$y_{ic} = \beta_0 + \beta_1 \text{Reward}_c + \beta_2 X_{ic} + \epsilon_{ic}$$

where the dependent variable is student i's standardized score on the Educational Initiatives test administered at the end of the school year.[17] Controls include the student's z-score on the final exam (administered by Gyan Shala) in the previous year, the student's gender, the city zone where the class is located, and a dummy variable for the psychological intervention that was conducted in an orthogonal design to the reward intervention. Standard errors are clustered at the class-level.

As column 1 in Table A10 shows, although the average treatment effect on the aggregate test score is negative, it is not statistically different from zero. This is also true when we analyze the mathematics (column 4) and science (column 7) scores separately.

However, as we see in columns 2, 5 and 8, this null effect was driven by baseline attenders (present at Time 0). As we noted above, the incentive increased these students' attendance during the incentive period, but had no effect on it afterwards. It is perhaps not surprising that the very small increase in days attended had no direct mechanical effect on their test scores. However the in-

[17]The score is standardized with respect to the mean score across all students in the 68 classes in the sample.

centive also does not appear to have increased test scores through other means, such as for example, by making them more interested in school.

In column 3, we see the opposite result: the reward lowered test performance of baseline non-attenders. Their average score was 0.59 standard deviations lower than their counterparts in the control classes. We see a similar effect on the mathematics score (-0.48σ, column 6) and the science score (-0.59σ, column 9). Thus after the incentive was removed, these students both attended school less, and performed worse than if they had not faced the incentive.

4.3 Possible Mechanisms

4.3.1 Lower Scores on Difficult Questions

In order to unpack the lower test performance, we examine separately the students' scores on questions of different difficulty levels.[18] As Table 5 shows, baseline non-attenders' scores on simple questions were unaffected by the incentive (columns 4 and 10). Scores were lower on the more difficult questions: intermediate and complex questions in mathematics (column 5, $\hat{\beta}_1$ = -0.475, s.e. = 0.210, p = 0.028) and column 6, $\hat{\beta}_1$ = -0.567, s.e. = 0.231, p = 0.017) and intermediate questions in science (column 11, $\hat{\beta}_1$ = -0.715, s.e. = 0.276, p = 0.012). (The coefficient for complex mathematics questions in column 3 is negative, but not significantly different from zero.) Thus the incentive appears to have lowered these students' ability or willingness to answer difficult test questions. The incentive did not have a significant effect on any of the test scores for baseline attenders.

4.3.2 Lower Liking for School Subjects

After the incentive was removed, baseline attenders in the incentive classes gave lower ratings to their enjoyment of school subjects than their counterparts in control classes. In Table 6, we use data from the student interviews at Times 3 and 5 to run student fixed-effects regressions according to the specification:

[18]For a list of the knowledge categories that were tested and our classification of test questions into the "simple", "intermediate" and "complex" categories, see the Appendix.

$$y_{ict} = \alpha_i + \beta_1 \text{Time } 5_t + \beta_2(\text{Reward}_c \times \text{Time } 5_t) + \epsilon_{ict}$$

The dependent variable is student i's rating at time t of mathematics, or of science (on a 7-point scale). Student fixed effects control for time-invariant observable and unobservable characteristics of the students. The coefficient β_2 captures whether after the scheme ended, students in the incentive classes rated mathematics or science differently than students in the control classes. As we see in column 1, the average student in control classes rated mathematics 0.15 points higher (on a mean of 2.46) at Time 5 than at Time 3. The reward had no differential effect ($\hat{\beta}_2$ = 0.007, s.e. = 0.110, p = 0.949). However, when we split the sample by students' baseline attendance, we see in column 3 that among baseline non-attenders, the coefficient $\hat{\beta}_2$ is negative, although imprecisely estimated.

A potential concern with column 3 is that since the investigators conducted the interviews when they visited the classrooms, students who were absent at the time of the visit were less likely to be interviewed. If, as we have shown above, the incentive lowered attendance at the Time 5 visit, then in column 3 we might be disproportionately estimating the effect of the incentive not on representative baseline non-attenders, but on those who chose to attend at Time 5, perhaps because they enjoyed school. To avoid this sort of sample selection bias, at each interview visit (Times 1, 3 and 5), our investigators were required to make up to three efforts to find these students and interview them. This involved asking around to find out where and when the student would be available, and making follow-up visits accordingly. Note that since the Gyan Shala classes are in the same neighborhoods as the students' homes, it is relatively easy to locate homes and interview the students there if they are available. As a result, 79% of students who were absent on the day of the Time 5 visit, were nevertheless interviewed. Although this is significantly lower than the 95% interview rate of those who were present in school during the Time 5 visit, it gives us a sample size large enough to measure these children's liking for school subjects. Therefore, in column 4 we restrict the sub-sample to baseline non-attenders students who were also absent at the Time 5 visit. If repeated absence is indicative of disinterest, then both the incentive and the control students in this sub-sample should have low ratings for school subjects. Within this sample we find that the incentive caused students to give a significantly lower rating ($\hat{\beta}_2$ = -0.923, s.e. = 0.372, p = 0.018).

When we consider the effect on students' rating for science on this subsample in column 8, the sign on β_2 is negative but not statistically different from zero. We conclude that the incentive reduced baseline non-attenders' enjoyment of mathematics. This is consistent with the psychological insight that intrinsic motivation is a key determinant of liking: as the intrinsic motivation to study a particular subject dwindles, liking for that subject correspondingly decreases.

4.3.3 Lower Optimism about Ability to Perform and Learn

Finally, in Table 7 we analyse two other interview questions that measure students' opinion about their performance at challenges, and their ability to learn. Students were told about a hypothetical student attempting a challenging sum and asked to predict how he or she would perform on a scale of 1 to 5. As we have seen in Table **??**, at baseline, the average student predicted the child would receive 2.2 stars from the teacher, and there was no significant difference between control and incentive classes. However as column 3 shows, the incentive caused baseline non-attenders to predict a statistically significant 0.3 stars lower. This difference becomes even larger when we restrict the sample to students who were absent at both Time 0 and Time 5 (column 4, $\hat{\beta}_1$ = -0.469, s.e. = 0.174, p = 0.011).

We also tried to elicit students' confidence in their ability to learn something new. Since teachers had newly introduced students to crossword puzzles, we asked them if they thought they could learn to solve one.[19] Once again, among baseline non-attenders, the incentive lowered the belief they could learn this new skill (columns 7 and 8, although the coefficient in column 8 is imprecisely estimated). Thus, the reduction in attendance and test scores caused by the intervention appears to be correlated with lower self-reported enjoyment of school subjects, optimism about their ability to perform at a challenging task, and learn a new skill.

[19]Crossword puzzles were part of a worksheet exercise that students saw a few weeks before the Time 1 interviews. We asked students this question at all three interviews.

5 Conclusion

We have identified two issues that have received relatively little attention in the experimental incentive literature in education. One, even if incentives have positive effects on motivation while they are in place, they might have negative effects after they are removed. This makes it important to examine not just their impacts not just in the immediate term but also in the longer term. Two, if incentives lower intrinsic motivation, this might have more substantial behavioral impacts on students who had low motivation to start with. This could happen if a decrease in motivation has a larger effect on student effort and outcomes when the motivation level is low, than if it is high.

We find empirical evidence that incentives might have different longer-term effects in groups with low versus high baseline motivation. Specifically, students with high baseline attendance (and presumably high baseline motivation) were influenced positively by the incentive while it was in effect, but were unaffected by it after it had been discontinued. This could be interpreted to mean that the incentive did not create a "habit" for these students to attend school more than their non-incentivized peers. However students with low baseline attendance were negatively affected. Not only did the incentive lower their attendance in the post-incentive period, it also lowered their test performance three months after the incentive scheme ended. We also find that the incentive lowered their enjoyment of the material taught in school, and their optimism and confidence about their ability to perform and learn.

In any incentive scheme, it is likely that some students will fail to earn the incentive because they do not meet the target. When an attendance target is absolute (as it was in our case), students with high attendance levels earn the incentive more easily, and the losers are disproportionately those with low attendance levels to start with. This paper shows that the incentive scheme can have unintended negative consequences for this very set of students, which is the group that the incentive scheme intended to help.

A few caveats are in order. First, it could be argued that if students were unable to attend school due to circumstances beyond their control, then the reward scheme might have imposed an extremely challenging standard that only served to make their constraints more salient and discouraged them further. We took care to choose a reasonable attendance target. as Table 2 shows, the average control student attended on 78 percent of school days during the incentive period, so that 85 percent represented only a 9% increase. According to school administrators, much of the absence could be explained by students' choices not to attend school rather than systemic problems at home or elsewhere.[20] Having said this, it is possible that some students in the sample were discouraged by failing to meet the attendance target, and that since a baseline non-attender was more likely to miss the target, this discouragement effect was disproportionately strong among that group of students. Since we have daily data from the incentive period for all classes in both the treatment and control groups, in unreported results we examine separately baseline non-attenders who met the incentive target of 85% of school days during the incentive period, and who did not. Among those who met the target, longer-term attendance (as measured at the Time 5 visit) did not decline significantly. Among those who failed to meet the target, the incentive lowered the attendance rate by 16.8% points (report stat. sig.). It is possible that the incentive scheme made these students' poor attendance salient to them and thereby de-motivated them even further. This underscores a central message of this paper, that rewards can have negative consequences on the students that educators intend to help the most.

Second, the attendance target could have been relative, so that students were rewarded for improving their attendance by a certain percentage of their own baseline. Then students with low baseline attendance could have earned rewards with relatively small (absolute) increases in attendance and would have been less likely to be discouraged. This would have required catering the target to each student individually, and given that there is a distribution of student attendance within each classroom, would have required within-class variation in attendance targets.[21] Not only would this have been difficult to administer, it would have been difficult to ensure that each student understood what their

[20]Note also that the reward period was deliberately chosen during a period when there are no long-drawn festivals that often cause students to miss school.

[21]However this might have discouraged students with high baseline attendance, since some of them might have missed their own target even if they increased absolute attendance by more than their low baseline peers did.

own target was.[22] Although pedagogical best practices prescribe that each student be set an achievement target that is appropriate for them individually, it is rare, especially in developing country contexts where teaching resources are scarce, that different standards of achievement are applied to different students. Thus our experiment tests an incentive scheme that closely simulates one that might be implemented in such a setting. It cautions educators and policymakers that such a scheme could end up hurting students whose effort and motivation need the greatest boost, while also not generating significant benefits for those who are already performing at a high level.

References

Angrist, Joshua, and Victor Lavy. 2009. "The Effects of High Stake High School Achievement Awards: Evidence from a Randomized Trial." *American Economic Review* 99 (4): 1384–1414.

Berry, James. 2014. "Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India." Mimeograph.

Bettinger, Eric P. 2012. "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." *The Review of Economics and Statistics* 94 (3): 686–698.

CfBT Education Services. 2010. "The Gyan Shala Programme: An Assessment." Technical Report. Technical Report.

Chao, Melody M., Rajeev Dehejia, Anirban Mukhopadhyay, and Sujata Visaria. 2015. "Effects of Lay Theories and Incentive Mechanisms on Human Capital Formation: Evidence from a Field Experiment in Low-Income Indian Schools." Mimeograph.

Charness, Gary, and Uri Gneezy. 2009. "Incentives to Exercise." *Econometrica* 77 (3): 909–931.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin* 25 (6): 627–668.

Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-determination in Human Behavior*. New York, NY: Plenum Press.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–1278.

[22] In Bettinger (2012)'s study, for eighth and ninth graders the eligibility to receive cash rewards was randomized at the student level. However, once they were selected into the incentive group, all students were assigned the same target. In Berry (2014)'s experiment, all students were offered rewards of the same value for meeting the same targets, but the type of reward was randomized at the student level.

Educational Consultants India Limited. 2007. "Study of Students Attendance in Primary & Upper Primary Schools: Abridged Report." Technical Report. Technical Report.

Educational Initiatives Private Limited. 2010. "Test of Student Learning for Gyanshala: Assessment Report." Technical Report. Technical Report.

Fryer, Roland. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *Quarterly Journal of Economics* 126:1755–1798.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25 (4): 191–210.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *The Review of Economics and Statistics* 91 (3): 437456.

Levitt, Steven D., John A. List, Susanne Neckerman, and Sally Sadoff. 2012. "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance." Technical Report. NBER Working Paper 18165.

California Department of Education. 2015. "Truancy." Technical Report, Department of Education. Accessed on 19th November 2015.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 3977.

Paredes, Ricardo D., and Gabriel A. Ugarte. 2011. "Should Students Be Allowed to Miss?" *The Journal of Educational Research* 104:194201.

Roby, Douglas E. 2004. "Research on School Attendance and Student Achievement: A Study of Ohio Schools." *Educational Research Quarterly* 28 (1): 314.

Rodriguez-Planas, Nùria. 2012. "Longer-term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the United States." *American Economic Journal: Applied Economics* 4 (4): 121–139.

World Health Organization. 2007. "Growth Reference data for 5-19 years." Technical Report. Accessed on 21st March 2015.

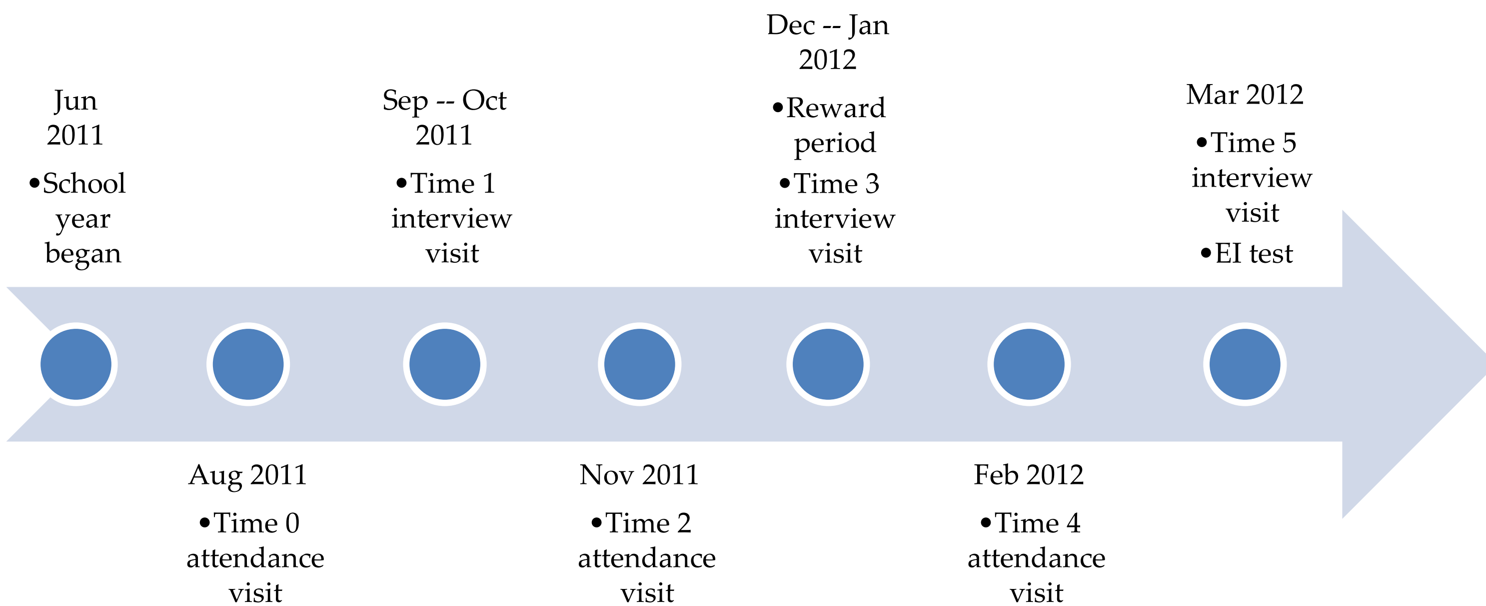

Figure 1: Sequence of Events

Table 1: Sample Characteristics

	N (1)	All (2)	No reward (3)	Reward (4)	T-test of differences (5)
Student characteristics					
Female	799	0.51	0.49	0.54	*0.257*
		(0.02)	(0.03)	(0.03)	
Year of birth	769	2002.8	2002.8	2002.8	*0.785*
		(0.06)	(0.08)	(0.08)	
Body Mass Index (kg/m2)	768	13.83	13.85	13.81	*0.842*
		(0.11)	(0.15)	(0.16)	
Household assets					
Mobile phone	768	0.93	0.92	0.93	*0.810*
		(0.01)	(0.01)	(0.01)	
VCR/DVD	768	0.36	0.37	0.35	*0.791*
		(0.03)	(0.04)	(0.04)	
Computer	768	0.01	0.01	0.01	*0.659*
		(0.00)	(0.01)	(0.01)	
Autorickshaw/motorbike/car	799	0.24	0.22	0.26	*0.268*
		(0.02)	(0.03)	(0.03)	
Toilet in the house	768	0.73	0.69	0.78	*0.148*
		(0.03)	(0.04)	(0.05)	
School-related variables					
Present at Time 0	799	0.75	0.74	0.75	*0.817*
		(0.02)	(0.03)	(0.03)	
Administrative attendance record	797	0.78	0.79	0.78	*0.585*
		(0.01)	(0.01)	(0.01)	
z-score on previous year's exam	783	0.00	0.02	-0.03	*0.687*
		(0.06)	(0.08)	(0.09)	
Likes maths (range = [-3, 3])	621	2.46	2.51	2.41	*0.367*
		(0.06)	(0.08)	(0.08)	
Likes science (range = [-3, 3])	621	1.99	2.09	1.87	*0.158*
		(0.08)	(0.10)	(0.11)	
Score on a difficult sum (range = [1, 5])	768	2.24	2.30	2.17	*0.481*
		(0.09)	(0.12)	(0.13)	
Able to solve a crossword puzzle (range = 0,1)	759	0.96	0.96	0.96	*0.723*
		(0.01)	(0.01)	(0.01)	

Means are computed from the baseline student survey data. t-tests account for correlation at the class level. Standard errors are in parentheses. Column 4 reports p-values for t-tests of differences between columns (2) and (3).

Table 2: Attendance Rates

		All students			Present at Time 0			Absent at Time 0		
	N	Both conditions	No reward	Reward	Both conditions	No reward	Reward	Both conditions	No reward	Reward
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Random visits										
Time 0	799	74.59	74.10	75.13	100.0	100.00	100.00	0.00	0.00	0.00
Time 1	799	85.36	85.90	84.82	87.58	89.64	85.37	78.82	75.00	83.16
Time 2	799	71.84	70.26	73.56	74.66	74.76	74.56	63.55	57.41	70.53
Time 3	799	85.61	81.53	90.05	88.26	86.08	90.59	77.83	68.52	88.42
Time 4	799	77.47	76.26	78.80	80.37	80.91	79.79	68.97	62.96	75.79
Time 5	799	74.22	75.30	73.04	76.68	77.67	75.61	67.00	68.52	65.26
Reward period										
Average attendance	798	80.38	78.49	82.45	83.54	82.24	84.94	71.13	67.78	74.93
Above the 85% threshold	798	52.51	48.32	57.07	57.82	54.87	60.98	36.95	29.63	45.26

Tablenotes.

Table 3: Effect of reward scheme on attendance at unannounced visits

	All students (1)	Present at Time 0 (2)	Absent at Time 0 (3)
Time 3	0.035	0.039	0.023
	(0.026)	(0.027)	(0.046)
Time 4	-0.018	-0.013	-0.032
	(0.024)	(0.024)	(0.055)
Time 5	-0.028	-0.045	0.023
	(0.025)	(0.029)	(0.049)
Reward × Time 3	0.074**	0.067*	0.093
	(0.034)	(0.034)	(0.069)
Reward × Time 4	0.014	0.011	0.022
	(0.039)	(0.039)	(0.078)
Reward × Time 5	-0.034	0.002	-0.139**
	(0.039)	(0.045)	(0.068)
Constant	0.786***	0.811***	0.712***
	(0.008)	(0.009)	(0.017)
Observations	*3,995*	*2,980*	*1,015*
R-squared	*0.015*	*0.015*	*0.022*
Number of students	*799*	*596*	*203*

All columns report student fixed-effects linear probability regressions, where the dependent variable takes value 1 if the student was present at the unannounced visit, and 0 otherwise. Standard errors are clustered at the class level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Effect of reward scheme on test scores

	All students (1)	Present at Time 0 (2)	Absent at Time 0 (3)	All students (4)	Present at Time 0 (5)	Absent at Time 0 (6)	All students (7)	Present at Time 0 (8)	Absent at Time 0 (9)
		Aggregate			Mathematics			Science	
Reward	-0.062	0.055	-0.586**	-0.055	0.036	-0.483**	-0.052	0.069	-0.594**
	(0.202)	(0.216)	(0.235)	(0.207)	(0.233)	(0.202)	(0.182)	(0.179)	(0.278)
Constant	0.187	0.270	0.052	0.112	0.165	0.051	0.207	0.293	0.053
	(0.286)	(0.291)	(0.430)	(0.290)	(0.322)	(0.413)	(0.245)	(0.246)	(0.372)
Observations	*584*	*419*	*152*	*584*	*419*	*152*	*583*	*418*	*152*
R-squared	*0.076*	*0.101*	*0.151*	*0.059*	*0.070*	*0.126*	*0.077*	*0.107*	*0.141*

All columns report OLS regressions. The dependent variable is the student's z-score on Educational Initiatives's test. The student's z-score on the previous year's final exam is controlled for. A female dummy, zone dummies and a dummy for the orthogonal psychological intervention are included. Standard errors in parentheses are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Effect of reward scheme on test scores, broken by difficulty level

	Present at Time 0			Absent at Time 0			Present at Time 0			Absent at Time 0		
	Simple	Intmdt	Complex	Simple	Intmdt	Complex	Simple	Intmdt	Complex	Simple	Intmdt	Complex
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Mathematics						Science					
Reward	0.142	0.046	-0.141	-0.075	-0.475**	-0.567**	0.072	-0.033	0.186	-0.234	-0.715**	-0.308
	(0.183)	(0.225)	(0.220)	(0.197)	(0.210)	(0.231)	(0.190)	(0.168)	(0.145)	(0.201)	(0.276)	(0.269)
Constant	0.165	0.111	0.201	0.244	-0.067	0.167	0.330	0.164	0.311	0.170	-0.033	0.068
	(0.246)	(0.295)	(0.306)	(0.381)	(0.380)	(0.293)	(0.211)	(0.239)	(0.239)	(0.285)	(0.337)	(0.378)
Observations	*419*	*419*	*419*	*152*	*152*	*152*	*419*	*418*	*419*	*152*	*152*	*152*
R-squared	*0.059*	*0.058*	*0.066*	*0.089*	*0.111*	*0.155*	*0.057*	*0.111*	*0.074*	*0.147*	*0.119*	*0.108*

All columns report OLS regressions. The dependent variable is the student's z-score on Educational Initiative's test. The student's z-score on the previous year's final exam is controlled for. A female dummy, zone dummies and a dummy for the orthogonal psychological intervention are included. Standard errors in parentheses are clustered at the school level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 6: Effect of reward scheme on change in students' liking for maths and science

	Mathematics				Science			
	All students	Present at Time 0	Absent at Time 0	Absent at both Time 0 & Time 5	All students	Present at Time 0	Absent at Time 0	Absent at both Time 0 & Time 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time 5	0.149**	0.131*	0.159	0.650**	0.120	0.100	0.098	0.300*
	(0.060)	(0.071)	(0.133)	(0.292)	(0.103)	(0.109)	(0.185)	(0.152)
Reward × Time 5	0.007	0.053	-0.159	-0.923**	0.178	0.180	0.204	-0.073
	(0.110)	(0.115)	(0.209)	(0.372)	(0.144)	(0.156)	(0.237)	(0.353)
Constant	2.584***	2.625***	2.527***	2.346***	2.189***	2.194***	2.201***	2.316***
	(0.027)	(0.029)	(0.053)	(0.096)	(0.037)	(0.040)	(0.062)	(0.088)
Observations	*1,437*	*1,068*	*349*	*102*	*1,437*	*1,068*	*349*	*102*
Number of students	*785*	*581*	*194*	*60*	*785*	*581*	*194*	*60*
R-squared	*0.017*	*0.021*	*0.009*	*0.125*	*0.019*	*0.016*	*0.022*	*0.035*

All columns report student fixed effects regressions, where the dependent variable is the student's rating of her liking for the subject, at interviews in Time 3 and Time 5. A female dummy, zone dummies and an indicator for the orthogonal psychological intervention are included. Standard errors in parentheses are clustered at the class level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Effect of reward scheme on student confidence

	All students	Present at Time 0	Absent at Time 0	Absent at both Time 0 & Time 5	All students	Present at Time 0	Absent at Time 0	Absent at both Time 0 & Time 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Performance on a difficult sum				Ability to solve a crossword puzzle			
Reward	-0.081	-0.015	-0.309**	-0.469**	-0.012*	-0.007	-0.029*	-0.024
	(0.095)	(0.109)	(0.144)	(0.174)	(0.007)	(0.005)	(0.014)	(0.025)
Constant	2.022***	1.981***	2.059***	2.360***	1.008***	1.007***	1.007***	1.024***
	(0.114)	(0.127)	(0.143)	(0.215)	(0.005)	(0.005)	(0.007)	(0.022)
Observations	*777*	*576*	*191*	*55*	*776*	*575*	*191*	*55*
R-squared	*0.075*	*0.062*	*0.142*	*0.237*	*0.029*	*0.014*	*0.094*	*0.129*

All columns report OLS regressions using student interview data from Time 3. In columns 1, 2 and 3 the dependent variable is the number of stars (maximum = 5) the student expects a child will receive for a difficult maths sum. In columns 4, 5 and 6 the dependent variable indicates whether the student expects he/she can learn how to solve a crossword puzzle. A female dummy, zone dummies and a dummy for the orthogonal psychological intervention are included. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1.