Working paper

IGC
International Growth Centre

Educator incentives and educational triage in rural primary schools

Dan Gilligan, Naureen Karachiwalla, Ibrahim Kasirye, Adrienne Lucas, and Derek Neal

March 2018

When citing this paper, please use the title and the following reference number:
S-89237-UGA-1

DIRECTED BY
LSE
UNIVERSITY OF OXFORD

FUNDED BY
UKaid
from the British people

Educator Incentives and Educational Triage in Rural Primary Schools

Dan Gilligan
International Food Policy Research Institute (IFPRI)
Naureen Karachiwalla
International Food Policy Research Institute (IFPRI)
Ibrahim Kasirye
Economic Policy Research Institute (EPRC)
Adrienne Lucas
University of Delaware and NBER
Derek Neal
University of Chicago and NBER

March 19, 2018

Abstract

In low-income countries, primary school students often fall far below grade level and primary dropout rates remain high. Further, in some countries, educators encourage their weaker students to dropout before reaching the end of primary school. These educators hope to avoid the negative attention that authorities direct to a school when its students perform poorly on the primary leaving exams that governments use to certify primary completion and eligibility for secondary school. We report the results of an experiment in rural Uganda that sought to reduce dropout rates in grade six and seven by offering bonus payments to grade six teachers that rewarded each teacher for the performance of each of her students relative to comparable students in other schools. Teachers responded to this Pay for Percentile (PFP) incentive system in ways that raised attendance rates two school years later from .56 to .60. These attendance gains were driven primarily by outcomes in treatment schools that provide text books for grade six math students, where two-year attendance rates rose from .57 to .64. In these same schools, students whose initial skills levels prepared them to use grade six math texts enjoyed significant gains in math achievement. We find little evidence that PFP improved attendance or achievement in schools without books even though PFP had the same impact on reported teacher effort in schools with and without books. We conjecture that teacher effort and books are complements in education production and document several results that are consistent with this hypothesis.

KEYWORDS: Achievement, Dropout, Educational Triage, Incentives, Uganda

JEL classification codes:

We thank Maha Ashour, Ezra Karger, and Giang Thai for excellent research assistance and Lucy Billings and Fiona Namugenyi for skilled project management. We gratefully acknowledge funding from the International Growth Centre (IGC) (grant no. 1-VRS-VUGA-VXXXX-89237), the Post Primary Education Initiative (PPE) of the Abdul Latif Jameel Poverty Action Lab (J-PAL) (grant no. 570004L4L), the Spencer Foundation (grant no. 20160150), and the Policies, Institutions, and Markets (PIM) Research Program of the Consultative Group for International Agricultural Research (CGIAR).

Introduction

During the past three decades, low-income countries have made great strides toward providing universal access to primary education. However, in many countries, universal primary access has not produced universal primary education. According to a recent World Bank report, primary achievement levels remain low and primary dropout rates remain high.[1]

Several recent survey articles cite results that underscore the challenges facing education systems in low-income countries. Schools often lack resources and also fail to use their existing resources efficiently. Bruns et al. (2011) contend that teachers in many developing countries are commonly absent from school and frequently less than fully engaged in teaching when they are present.[2]

Nonetheless, in countries where education officials rarely hold teachers directly accountable for their performance or the performance of their students, education authorities often attach high-stakes to the results of primary leaving exams (PLE).[3] Given the importance of these exams for students, government officials scrutinize PLE results. Students who pass the PLE given in their country can demonstrate to potential employers that they completed primary school, and good scores may guarantee that they have access to secondary schooling, which is not universal in many less developed countries.

A recent newspaper article from Uganda highlights two facts. First, teacher absenteeism is a problem in Uganda.[4] Second, even in districts where education officials do not require teachers to attend school each day, they do hold educators accountable for how students from their schools score on the PLE. Education authorities in the Jinja district demoted 11 head teachers because, in each of these schools, a significant fraction of the students who took the 2017 PLE failed. District officials held a press conference to announce the demotions and to state publicly that they planned to take additional steps to make sure that PLE pass rates rise in the future.[5]

Ugandan education authorities focus on PLE results, in part, because PLE outcomes are the only measures of student performance that they collect. Each year, the Ministry of Education and Sports gathers information from each school about aggregate enrollments, the number of students repeating particular grades, and staffing, but the Ministry does not collect attendance or performance data from individual students or teachers. Further, its data do not track movements of students among schools, which means that education officials cannot calculate dropout rates or primary graduation rates for cohorts of students that enter a given grade in a given school.

Scholars throughout the social sciences have documented the harmful and often unintended consequences that arise when the output of an organization is multidimensional, but the organization builds its incentive and accountability systems around a single performance metric.[6] Holmstrom and Milgrom (1991) note that, whenever workers have the opportunity to take hidden actions that

[1] See World Bank (2018)

[2] See World Bank (2018) and Glewwe and Muralidharan (2016).

[3] Kenya, Uganda, Botswana, Tanzania, and Zambia are examples, but such systems exist in many other low-income countries in Africa and Asia.

[4] Academic research reaches the same conclusion. See Chaudhury et al. (2006) and Patrinos (2013). During our first round of data collection in rural Uganda, more than one fourth of head teachers report that the P6 math teacher was absent from school, at least once, during the previous four school days.

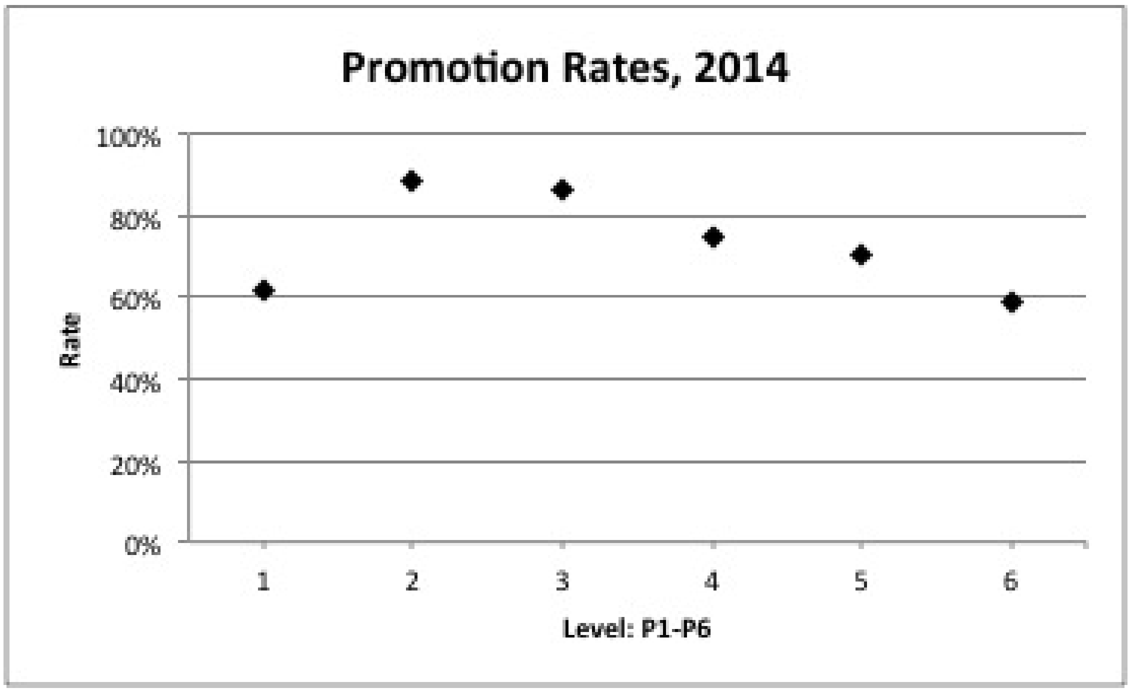
[5] See "Jinja headteachers demoted over PLE," *New Vision*, February 1, 2018. The additional steps included stricter enforcement of rules of that limit teacher absences.

[6] See Kerr (1975), Campbell (1979), and Holmstrom and Milgrom (1991).

improve a particular performance metric without improving and possibly harming their actual performance, organizations should be wary of building incentive systems around this metric. Figure 1 provides suggestive evidence that, by attaching high-stakes to PLE results, Ugandan authorities may be encouraging educators to take hidden actions that raise dropout rates.

Primary school in Uganda involves seven grades, P1 through P7. Figure 1 shows that, at the end of P1 through P6, promotion rates are always significantly less than one, and these rates drop more than 20 percentage points between P3 and P6.

Figure 1



Notes: The promotion rate for level P(n) in 2014 is the enrollment in level P(n+1) in 2015 minus the number of P(n+1) repeaters in 2015 divided by 2014 enrollment in P(n), where $n = 1, 2, 3, 4, 5, 6$.

The sharp drop in promotion rates over the upper primary grades may, in part, reflect deliberate efforts by educators to shape the composition of students who reach P7 and register to take the PLE. If weak students never enter P7 and register for the PLE, they cannot fail the PLE, and their teachers cannot be punished for failure to learn.

Although educators in Uganda cannot legally force children to drop out of primary school, they can take steps that encourage weak students to drop out. For example, they can direct all or most of their attention to students at the top of their classes. They can also refuse to promote weak students, and they can encourage weak students to seek employment instead of returning to school. Since such actions are hidden from education authorities, who possess no data on school-level dropout rates, education officials cannot sanction teachers who employ these tactics.

During our field work for this project, we discovered that educators in rural Uganda fully understand that they benefit when they encourage weak students to dropout. Table 1 presents the enrollment records for a school that we visited in the fall of 2015. Notice how grade-level enrollment declines after grade four and declines sharply between grades six and seven. During our visit to this school, we asked the Head Teacher if this pattern had anything to do with the public attention given to PLE results. She confirmed that this was the case. She told us that she would receive negative attention and possibly sanctions from district education officials if her students took the PLE and failed. So, she felt her only option was to "narrow the bridge" between sixth and seventh grade.

Table 1

Enrollment By Grade in A Rural Ugandan Primary School

		Girls	Boys	Total
	P1	69	76	145
	P2	54	62	116
	P3	59	60	119
Grade	P4	74	54	128
	P5	80	40	120
	P6	65	39	104
	P7	37	14	51
	Total	438	345	783

Notes: We transcribed these data from official enrollment data on a chart outside the Head Teacher's office.

Better Incentives

In public schools in rural Uganda, teachers are often absent, and when they are present in school, the existing public focus on PLE outcomes creates incentives for them to devote their attention to students who have a reasonable chance of passing the PLE. Further, the government's practice of punishing schools when their students fail the PLE causes some educators to actively encourage their weaker students to leave school.

Below, we describe the results of a field experiment in rural Uganda that examines whether or not a specific assessment-based incentive system for educators can reduce dropouts rates and promote learning among students at all achievement levels. We find that by implementing the Pay For Percentile (PFP) incentive scheme developed in Barlevy and Neal (2012) for one year, we increase the probability that students who begin P6 in a given school will still be attending that school almost two years later. This result is mostly driven by outcomes among students who attend treatment schools that possess textbooks. PFP generated no significant attainment gains among students in schools without books, but in schools that provide math books for P6 students, the introduction of PFP raises the likelihood that entering P6 students remain in their current school during the next two school years by seven percentage points, from .57 to .64. Further, among boys in schools with books, PFP increases the probability that a student who begins P6 will complete P7 on time and take the PLE from .41 to .48.

In keeping with our attendance results, PFP produced no achievement gains during P6 among students without access to textbooks. However, in schools with books, PFP improved P6 math achievement among students in the top half of the distribution of initial P6 math skills. Among this population, students in PFP schools with books scored as much as .19 standard deviations higher than comparable students in control schools with books. However, PFP has no impact on measured achievement among students in the lower half of the round one achievement distribution, whether or not these schools possess math books.

Our measures of teacher effort indicate that both teachers in schools with books and teachers in schools without books responded to PFP by supplying more teaching effort. However, all the measurable gains associated with PFP treatment occur in schools with books. We conjecture that teacher effort and textbooks are complements in the production of student learning, and in section 5, we present several results that are consistent with this hypothesis.

We proceed as follows. In the next section, we briefly describe the PFP incentive scheme. Then, we describe our experiment. We first present results that describe how introducing PFP during P6 impacted student attendance, attainment, PLE participation, and PLE results over the subsequent two school years. We then examine how PFP affected achievement growth during P6. Next, we present suggestive evidence that our results support the view that teacher effort and textbooks are complements in the production of math skill. We also discuss how our results fit into the existing and growing literature on the returns from investments in school resources and how targeting strategies impact these returns. We conclude by describing directions for future research.

1 How Pay for Percentile Works

Imagine an environment where J teachers each teach one class with N students who begin the year in the same grade. Let $j = 1, 2, ..., J$ index teachers, and let $n = 1, 2, ...N$ index students within a classroom. Assume that the distribution of initial achievement levels in each class is identical, and that within each class, no two students have the same initial achievement. Without loss of generality, let $n = 1, 2, ..., N$ rank students within each class according to their initial achievement level. Finally, assume that a measurement technology allows the education authority that supervises these J educators to rank all students based on their end of year achievement levels.

In this setting, consider the following contest scheme: Collect each of the J students who occupy rank $n = 1$ in the initial achievement distribution for her class. Place all such students in a league, and for each student, calculate her within-league percentile rank in the end of year achievement distribution. Pay each teacher, $j = 1, 2, ..., J$ a bonus proportional to the within-league percentile rank of her student. Repeat this process for baseline ranks $n = 2, 3, ...N$.

Barlevy et al. (2012) show that there exists a scaling factor for these bonus payments such that all J teachers choose efficient levels of effort for all tasks that influence the achievement growth of all N students in each classroom. The scaling factor in question is the Lazear et al. (1981) prize for a contest between two, $J = 2$, educators who each devote effort to a single task that promotes learning for one, $N = 1$, student.

PFP rewards teachers for every contest that each of her students wins. Thus, PFP should induce each teacher to take on additional tasks that generate learning benefits for each of her students. Further, because PFP contests are properly seeded, no teacher believes that her students have little chance to be competitive or believes that her students have a clear advantage over their competitors. In sum, PFP provides clear incentives for educators to devote more effort to their students without inducing the educational triage responses that plague both the PLE system in Uganda and accountability schemes in the US that are built around proficiency levels or other target scores.[7]

Since the efficiency properties of PFP rest on educators believing that they are competing in properly seeded contests, we placed restrictions on the contests in our experiment. We chose only rural, government schools with one P6 stream and an expected class size within a predetermined range. Further, we told PFP teachers that they would be competing against teachers in other rural, government schools with only one P6 stream and comparable P6 enrollment.

2 Sample Design

Given the reliance of PFP on fair contests, we sought to create a sample of homogeneous schools. We used the Ugandan Education Management Information System (EMIS) to identify government-operated schools in rural areas of the 13 Luganda speaking districts within the Buganda sub-region of Uganda. We dropped all schools that reported 2014 EMIS enrollment for P6 of either less than 40 or more than 70 students. Then, given information gleaned from school vistis, we kept all schools with exactly one P6 stream and one P6 math teacher.

We identified 324 parishes that contained at least one school that satisfies our selection criteria. If a parish contained more than one eligible school, we randomly chose one eligible school for that parish. In the resulting sample of 324 schools, some schools located near parish boundaries were within 2km of another school. We wanted to minimize the likelihood that teachers in the experiment would know each other personally. So, we evaluated the location of the 324 schools in a random order. We kept the first school for our final sample, and as we evaluated the remaining schools, we kept each school that was not within 2km of any school already selected for our final sample. This process eliminated 22 schools, leaving a sample of 302 schools in 302 parishes.

Within this 302 school study sample, we formed six strata. We first divided the sample into schools that did or did not report having P6 math books during our validation visits.[8] Within these subsamples, we defined three predicted P6 enrollment cells (large, medium, or small). Within each of these six strata, we ranked schools by their past PLE performance. In three strata, we assigned treatment to schools with odd ranks. In the remaining three strata, we assigned treatment to

[7] See Neal et al. (2010) and Neal (2011).

[8] We discovered, during our round one data collection, that these reports were not accurate, presumably because the validation reports concerning books rarely came from the P6 math teacher.

schools with even ranks. In total, we gathered data from 151 control schools and 151 treatment schools.

However, we only employ data from 299 schools, 149 treatment and 150 control. One treatment teacher informed us during his round one interview that he was in the process of leaving the school to take a new job. Since his replacement was not yet present, we were not able to treat this school. In two other schools, the data gathered during round two did not allow us to definitively determine whether or not the round one P6 Math teacher was still the P6 Math teacher at the end of the school year.

2.1 Protocol

Round one data collection began in March of 2016, less than one month into the 2016 academic year. During this round, a team of enumerators visited each of our 302 schools. The night before each school visit, enumerators informed the school staff that a survey team would be arriving the next day with written approval from the district education office to conduct interviews and test P6 math students. Given these advance notices, the P6 math teacher for each school was present for our round one interviews.

During these visits, we interviewed each P6 student in attendance, the P6 math teacher, and the Head Teacher. We also asked the P6 math teacher to assess each of her P6 students. While one enumerator interviewed the P6 teacher, the other supervised the administration of our round one math assessment. We tested all students who were present using an exam that we created based on government publications that describe the primary school curricula for grades P1 through P5 in Uganda. We did not include P6 items because these students had only been attending P6 for roughly one month. We did include P6 items in the round two assessment.

PFP is designed to reward educators for the progress that each of their students makes during a school year, and achievement levels vary widely among P6 students in these schools. In order to credibly promise educators that we would reward them for the learning gains of each of their students, we needed to create assessments that could detect the learning gains of each student.

Our round one assessment contains many items based on the P1, P2, P3, P4, and P5 curricula. We did not repeat any items on our round two assessments. However, we chose similar items from the P1 through P5 curriculum guides and also added completely new items from the P6 guide. We made sure that educators in the PFP treatment group were aware that our assessments would contain items from all levels of the P1-P6 curricula. We did so to assure them that our assessments would measure the academic progress of all of their students, and not just those who were able to master P6 material.

For both assessments, we used a two-parameter IRT model to create an estimate of latent math skill for each student. We then created standardized versions of these scores that have mean zero and standard deviation one.

2.2 Subsequent Rounds of Data Collection

In October of 2016, we returned to our 302 schools. We not only gave students a second math assessment, we conducted a second round of interviews with the P6 math teacher and the head teacher.

In October, 2017 we returned for a third round of data collection. We did not test students, but we did gather information about their attendance during the current term, their attendance during the past week, and whether or not each enrolled student was still in P6 or had been promoted to P7. We also gathered data about PLE registrations.

Students took the PLE in early November of 2017. In February of 2018, we obtained individual PLE results from the Uganda National Examinations Board (UNEB) for all students in the 13 districts that constitute our sampling frame. We used names and PLE testing center numbers to match students in our sample to the individual records in the UNEB data. The PLE data Appendix provides more details about the matching procedure.

2.3 Balance

Table 2 presents key descriptive statistics from round one for both our treatment and control samples. There is no evidence that the students in our treatment and control schools differ in terms of educational resources. None of these group differences in school level resources are statistically significant. Further, the differences that exist do not fit a pattern. Treatment schools are more likely to have a teacher with a low education level and are less likely to have books for students, but these same schools are more likely to use PLE practice exams and teach students in English. Students are demographically quite similar in treatment and control schools, and the differences that exist are not statistically significant. Students in treatment schools do score lower on the round one math assessment. This difference of -.09 standard deviations is not quite statistically significant, but it is academically noteworthy. Therefore, in all regression analyses of subsequent student outcomes, we include round one math achievement as a control.

2.4 PFP Script

During round one, we told control teachers that we were conducting research on learning outcomes for students in Uganda. We did not tell them about the existence of the PFP treatment group or our plans to return for a second round of testing at the end of the school year. In treatment schools, we ended round one visits by informing P6 math teachers that we were going to return at the end of the school year and test their students again. We told them that the return visits would not be announced, and we told them that they would only earn bonus payments for the performance of students who were present and tested during these end of year visits.

We repeatedly stressed that each student would only be competing against students in other rural, government schools with comparable P6 enrollment. We also stressed that each student would compete only against other students who received similar scores on our baseline assessment. Treatment teachers learned that, for each of their students tested at the end of the school year, they would receive a bonus payment of 20,000 Shillings times the student's percentile rank in her contest group.[9]

[9]In March 2016, 20,000 Shilings were worth six US dollars.

Table 2: Balance Tests

	Treatment	Control	Difference	p-Value	N
School Variables					
Class Size	30	30	0.68	0.66	299
Low Teacher Education	0.70	0.67	0.02	0.65	299
Teacher Age	36	36	-0.81	0.42	299
Female Teacher	0.21	0.16	0.05	0.29	299
English Instruction	0.74	0.67	0.07	0.17	299
PLE Practice	0.56	0.52	0.04	0.52	299
No Books	0.54	0.49	0.04	0.45	299
Student Variables					
Does Homework	0.94	0.93	0.01	0.37	8864
Enjoys School	0.88	0.86	0.02	0.33	8864
Age	13	13	-0.09	0.14	8864
Girl	0.55	0.53	0.02	0.22	8864
Baseline Achievement	-0.04	0.04	-0.09	0.14	8786
Cement Floor	0.47	0.48	-0.004	0.86	8864
Electricity	0.43	0.41	0.02	0.28	8864
Radio	0.84	0.84	-0.002	0.86	8864

Notes: For student level variables, we use a standard HAC estimator and treat schools as clusters to estimate standard errors. The variable PLE Practice is an indicator that equals one if the school gives PLE practice exams to their students. Low Teacher Education equals one if the P6 Math teacher does not have a teaching diploma of any kind. No Books captures the absence of P6 math texts.

Table 2 shows that, on average, roughly 30 students were present in each school during round one. Just over two thirds of these students were present for testing in round two, and among those tested, the expected bonus payment for each student was 10,000 Shillings.[10] The expected total bonus payment per teacher is just over 400,000 shilings, which represents roughly two months salary for new teachers and roughly two weeks salary for the most experienced head teachers.

3 Attainment

In all of our analyses of student outcomes, we restrict attention to the sample of students tested during round one. We impose this restriction for several reasons. To begin, we were not able to accurately identify the sample of students who were actively attending P6 in a given school at the time of our round one visits. School registries contain many students who do not attend the school and some who attend quite infrequently, and we are not confident that the schools possess accurate attendance records for these students. Further, we use the round one math score as a control in all of our empirical models, and these scores are not available for students who were not present during round one.

Our experiment is motivated by evidence that Ugandan educators behave in ways that encourage weaker students to drop out of school before P7. As we note above, these behaviors take several different forms. A teacher may devote little attention to a weak student. A head teacher can tell a student that she must repeat P6, and a head teacher may also tell the student that she is not likely to ever move up to P7.

If PFP does induce educators to devote more attention to all of their students, we expect students of all ability levels to feel more welcome in school. We also expect that some students will make additional academic progress that will cause their teachers and head teachers to believe that they have less incentive to encourage these students to drop out. Our data suggest that many students who are still clearly below P6 achievement levels at the end of P6 have a reasonable chance of passing the PLE given a full year of P7 to prepare or the opportunity to prepare over two years by repeating P6 and then proceeding to P7. Thus, even small additional gains in achievement may make educators less eager to pressure particular students to leave school.

Tables 3a and 3b present results from linear probability models that take the following form

$$y_{nj} = c + treat_j \alpha + score_{nj} \beta + \varepsilon_{nj}$$

Here, y_{ij} is an attendance or attainment outcome for student $n = 1, 2, ..., N_j$ who was tested during round one in school $j = 1, 2, ..., J$. The indicator variable $treat_j$ equals one if school j is a treatment school and zero if it is a control school. The conditioning variable $score_{nj}$ is the score that student n in school j earned on the round one assessment. Tables 3a and 3b present the OLS estimates of, $\hat{\alpha}$, from these regressions for different estimation samples.

The five columns in these tables present results for five different indicator variables. The first column presents the effects of treatment on the probability that a student is present for testing in round two. The second column demonstrates how treatment changes the probability that students

[10] PFP involves seeded contests. Each contest has one winner and one loser. Thus, the expected bonus payment per student is one half of 20,000 Shilings

are still attending their round one school in round three. Here, we count students as attenders if they are present or have been present on any of the previous four school days. The third column presents results for an indicator that equals one for attenders who are enrolled in P7. It equals zero for those who are not attenders and for attenders who are still in P6.

The final two columns deal with PLE outcomes. Column four reports the effects of treatment on the probability of taking the PLE in Novermber, 2017. The final column reports the effects of treatment on the probability of passing the PLE.

Table 3a presents results for the full sample. The estimated impacts in the first column imply that, overall and among boys and girls separately, PFP treatment raises attendance rates in round two by roughly two percentage points. However, none of these impacts are statistically significant. In both treatment and control schools, roughly seventy percent of students tested in round one are present for testing in round two.

However, during year two, we see attendance rates in treatment schools diverge more significantly from those in control schools. Column two shows that, in control schools, 56 percent of the students we interviewed in round one were still attending their original school when we returned to collect round three data. PFP treatment is associated with a four percentage point increase in this attendance rate. When we examine boys and girls separately, we see the same 4 percentage point increase in round three attendance. The p-values associated with the estimated impacts for the full sample, the boys sample, and the girls sample are .02, .06, and .05 respectively.

Only 43 percent of our round one students in control schools are both present at round three and enrolled in P7. Our results indicate that PFP treatment raises the probability of this outcome by three percentage points, but here the p value is .08, and the estimated impact among girls is slightly smaller and less statistically significant. We see no significant impacts on overall PLE outcomes.

Our results indicate that PFP treatment created changes in school environments that raised attendance rates more than 18 months later. Table 3b demonstrates that student outcomes in treatment schools that possess books drive these results. In control schools with books, the attendance rate in round three was .57, and among schools with books, our results imply that PFP raised this rate by 7 percentage points overall, 9 percentage points among boys, and 6 percentage points among girls.

Further, among schools with books, we see some evidence that PFP impacts promotion and PLE participation, at least among boys. Boys in treatment schools with books were seven percentage points more likely to attend P7 during round three and also seven percentage points more likely to take the PLE. Note that only about forty percent of boys in our control schools took the PLE.

Table 3a: The Effects of PFP on Attendance and Attainment

	Interviewed Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
Both	0.02 (0.02) p = 0.28	0.04 (0.02) p = 0.02	0.03 (0.02) p = 0.08	0.02 (0.02) p = 0.23	0.01 (0.02) p = 0.59
$\bar{Y}_c$	0.71	0.56	0.43	0.42	0.34
N	8788	8788	8788	8770	8770
Male	0.02 (0.02) p = 0.41	0.04 (0.02) p = 0.06	0.04 (0.02) p = 0.06	0.04 (0.02) p = 0.12	0.01 (0.02) p = 0.73
$\bar{Y}_c$	0.69	0.55	0.41	0.40	0.34
N	4012	4012	4012	3998	3998
Female	0.02 (0.02) p = 0.36	0.04 (0.02) p = 0.05	0.03 (0.02) p = 0.26	0.01 (0.02) p = 0.61	0.01 (0.02) p = 0.64
$\bar{Y}_c$	0.73	0.57	0.44	0.44	0.35
N	4776	4776	4776	4772	4772

Notes: The five outcomes are indicator variables. Interviewed Round 2 equals one if students were present on the day of our round two visits. Attending Round 3 captures attendance during the day of our round three visits or the four prior schools days. $\bar{Y}_c$ is the control sample mean. To estimate standard errors, we use a standard HAC estimator and treat schools as clusters.

Table 3b: The Effects of PFP on Attendance and Attainment

Panel A: Schools Without Books

	Interviewed Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
	0.02	0.01	0.03	0.02	0.01
Both	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
	p = 0.39	p = 0.60	p = 0.29	p = 0.46	p = 0.70
$\bar{Y}_c$	0.70	0.56	0.41	0.41	0.33
N	4703	4703	4703	4695	4695
	0.01	-0.01	0.01	0.01	-0.01
Male	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	p = 0.75	p = 0.86	p = 0.71	p = 0.85	p = 0.60
$\bar{Y}_c$	0.69	0.55	0.40	0.40	0.33
N	2125	2125	2125	2118	2118
	0.03	0.03	0.04	0.03	0.03
Female	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	p = 0.30	p = 0.36	p = 0.20	p = 0.34	p = 0.37
$\bar{Y}_c$	0.72	0.56	0.42	0.42	0.33
N	2578	2578	2578	2577	2577

Panel B: Schools With Books

	Interviewed Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
	0.01	0.07	0.04	0.03	0.01
Both	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
	p = 0.54	p < 0.01	p = 0.17	p = 0.36	p = 0.75
$\bar{Y}_c$	0.72	0.57	0.45	0.44	0.36
N	4085	4085	4085	4075	4075
	0.02	0.09	0.07	0.07	0.03
Male	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	p = 0.37	p < 0.01	p = 0.03	p = 0.05	p = 0.35
$\bar{Y}_c$	0.69	0.55	0.42	0.41	0.34
N	1887	1887	1887	1880	1880
	0.003	0.06	0.01	-0.01	-0.01
Female	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	p = 0.91	p = 0.06	p = 0.85	p = 0.75	p = 0.71
$\bar{Y}_c$	0.75	0.58	0.47	0.47	0.38
N	2198	2198	2198	2195	2195

Notes: See Table 3a.

Among schools with books, PFP had little impact on promotion rates and PLE participation among girls. Almost all of the increase in round three attendance among girls in treatment schools was driven by a large increase in the fraction of girls who repeat P6[11] Finally, PFP did not raise PLE pass rates significantly for boys or girls. This is true regardless of whether or not PFP treatment schools possess books.[12]

Among schools without math books, we find no statistically significant impacts of PFP treatment on any measure of attendance or attainment for boys or girls. The differences between our esimtates of PFP impacts for specific subsamples of students in schools with books as opposed to those in schools without books are often not statistically significant. However, the estimated differences in the effects of PFP on round three attendance among boys are clearly significant and the corresponding difference for the full sample is marginally significant.[13]

In the next section, we examine how PFP impacts achievement. Here, we find even more striking differences between the implied effects of PFP among schools with books versus those without books. Although we find significant positive achievement effects of PFP among schools with books, we find no significant impacts of PFP on achievement among schools without books. Further, our estimates of the impacts on PFP on achievement among schools without books are negative in every sample given every measure of round two math achievement. This pattern provides additional evidence that our PFP treatment produced real benefits for students in schools that possessed P6 math books but did little to benefit students in schools without these books.

4 P6 Achievement

The PFP design seeks to direct educator attention to each student. In rural Uganda, this goal raises concerns about assessment design. Existing research and the results from our round one assessment show that many pupils in rural Uganda begin P6 far below grade level. On average, students in the bottom fourth of our round one achievement distribution got less than half of the questions from the P1 and P2 curricula correct. Further, the vast majority of these students answered none of the questions from the P4 and P5 curricula correctly.

If the teachers in our treatment sample believed that our round two assessment would consist primarily of questions drawn from the P6 curriculum with some easier questions from P5 and possibly P4, our PFP treatment would have provided little incentive for them to direct effort to the students in the bottom fourth or more of our round one achievement distribution. These students did not yet possess clear command of P2 material. There is no reason to believe that their teachers could have taught them in ways that would have allowed them to move up three or four grade levels in one year. Therefore, the best efforts of these teachers would not have changed the expected scores of their weakest students on a standard P6 assessment.

[11] Among schools with P6 math books, less than 11 percent of round one girls from control schools repeat P6. The corresponding fraction among treatment schools is .164, which represents an increase of more than 50 percent in the P6 repetition rate.

[12] Appendix Tables 1a and 1b present parallel results from regression models that do not contain a control for round one math performance. The implied effecs of PFP on attendance and attainment are only slightly smaller because the -.09 standard deviation gap in round one math score between treatment and control schools is associated with an expected difference in attendance rates of far less than one percent. Appendix Tables 2a and 2b present round two achievement regressions that do not contain a control for round one achievement.

[13] The p=values on two-sided tests of equality are .024 and .099, respectively.

As we note above, to address this concern, we made sure that both our round one and round two assessments contained many items from the P1, P2, and P3 curricula. Further, we informed PFP teachers that our round two assessment would include items from each of the P1 through P6 curricula in order to enhance our capacity to measure the progress made by students at all initial P6 achievement levels.

Given this design feature, we present achievement results for three different measures of round two achievement. The first is a pupil-specific IRT ability parameter based on the full round two assessment. The second and third are ability parameters derived from subtests of the round two assessment that contain items from the P1 through P3 curricula and P4 through P6 curricula respectively. Given the wide range of round one achievement levels, we also present results separately for students who scored in the bottom or top half of the round one achievement distribution.

Tables 4a and 4b follow the format of Tables 3a and 3b. Table 4a presents results for all schools. Table 4b presents separate results for schools without math books and with math books, respectively.

The results for the full sample contain little evidence that PFP had significant impacts on student achievement. Among girls in the top half of the baseline achievement distribution, PFP raised round two achievement on the P4 through P6 subtest by .1 standard deviations, but none of the other estimated impacts even border on statistical significance.

However, Table 4b documents a striking contrast between our estimates of PFP impacts among schools with books versus our estimates among schools without books. Among schools without books, all of our estimated impacts of PFP on achievement are negative, although none is statistically significant. In contrast, among schools with books, we see important gains on the P4 through P6 subtest that appear to be driven by the performance of students in the top half of the baseline achievement distribution.

Consider the third column of Panel B in Table 4b. Among schools with books, PFP raised achievement on the P4 through P6 subtest by .12 standard deviations. We see similar results when we examine boys and girls separately, although these estimated impacts are slightly less significant statistically.

The final column shows that the performance of students in the top half of the round one achievement distribution account for most of the overall achievement gains attributable to PFP. In PFP schools with books, students who scored above the round one median score earned round two scores on the P4 through P6 subtest that are, on average, .19 standard deviations higher than the scores of their peers in control schools. This gap is larger and more statistically significant among girls than boys, but both enjoyed significant learning gains.

Table 4a: The Effects of PFP on Achievement

	All R1 Achievement Levels			Below R1 Median			Above R1 Median		
	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest
	0.02	0.003	0.04	-0.001	-0.002	-0.002	0.04	0.01	0.07
Both	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)
	p = 0.54	p = 0.91	p = 0.24	p = 0.97	p = 0.96	p = 0.95	p = 0.34	p = 0.78	p = 0.16
$\bar{Y}_c$	0.07	0.08	0.03	-0.54	-0.52	-0.50	0.61	0.62	0.51
N	6183	6183	6183	2995	2995	2995	3188	3188	3188
	0.01	-0.01	0.04	0.003	-0.0001	0.02	0.01	-0.01	0.04
Male	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)
	p = 0.88	p = 0.73	p = 0.40	p = 0.94	p = 1.00	p = 0.71	p = 0.79	p = 0.77	p = 0.50
$\bar{Y}_c$	0.15	0.14	0.13	-0.48	-0.49	-0.43	0.67	0.65	0.58
N	2731	2731	2731	1283	1283	1283	1448	1448	1448
	0.03	0.01	0.05	-0.01	-0.00	-0.02	0.06	0.03	0.10
Female	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)
	p = 0.39	p = 0.66	p = 0.23	p = 0.88	p = 0.93	p = 0.62	p = 0.16	p = 0.45	p = 0.08
$\bar{Y}_c$	-0.01	0.02	-0.05	-0.58	-0.55	-0.55	0.56	0.59	0.44
N	3452	3452	3452	1712	1712	1712	1740	1740	1740

Notes: We define the Below and Above R1 Median subsamples relative to the median of the entire R1 score distribution. $\bar{Y}_c$ is the control sample mean of the round two achievement measure for a given column in the sample defined by a given row. These subsample means are defined relative to the mean scores among all round two test takers, including those who were not present for our round one visit and therefore not included in our analysis samples. To estimate standard errors, we use a standard HAC estimator and treat schools as clusters.

Table 4b: The Effects of PFP on Achievement: Without Books and With Books

Panel A: Schools Without Books

	All R1 Achievement Levels			Below R1 Median			Above R1 Median		
	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest
	-0.03	-0.02	-0.03	-0.04	-0.03	-0.05	-0.03	-0.02	-0.03
Both	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)
	p = 0.39	p = 0.52	p = 0.49	p = 0.34	p = 0.49	p = 0.22	p = 0.61	p = 0.73	p = 0.58
$\bar{Y}_c$	0.10	0.11	0.06	-0.49	-0.46	-0.47	0.61	0.60	0.51
N	3275	3275	3275	1536	1536	1536	1739	1739	1739
	-0.04	-0.03	-0.04	-0.04	-0.04	-0.04	-0.04	-0.02	-0.05
Male	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.08)
	p = 0.38	p = 0.53	p = 0.51	p = 0.47	p = 0.52	p = 0.55	p = 0.58	p = 0.79	p = 0.52
$\bar{Y}_c$	0.18	0.17	0.16	-0.45	-0.45	-0.41	0.67	0.64	0.60
N	1436	1436	1436	648	648	648	788	788	788
	-0.02	-0.02	-0.02	-0.03	-0.02	-0.06	-0.02	-0.02	-0.01
Female	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)
	p = 0.59	p = 0.65	p = 0.67	p = 0.48	p = 0.69	p = 0.23	p = 0.78	p = 0.74	p = 0.86
$\bar{Y}_c$	0.03	0.06	-0.03	-0.52	-0.48	-0.52	0.55	0.57	0.43
N	1839	1839	1839	888	888	888	951	951	951

Panel B: Schools With Books

	All R1 Achievement Levels			Below R1 Median			Above R1 Median		
	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest
	0.07	0.03	0.12	0.03	0.02	0.04	0.11	0.04	0.19
Both	(0.05)	(0.04)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
	p = 0.13	p = 0.46	p = 0.04	p = 0.50	p = 0.59	p = 0.45	p = 0.07	p = 0.46	p = 0.01
$\bar{Y}_c$	0.03	0.04	0.00	-0.60	-0.59	-0.53	0.62	0.64	0.50
N	2908	2908	2908	1459	1459	1459	1449	1449	1449
	0.05	0.01	0.11	0.05	0.04	0.07	0.07	-0.01	0.14
Male	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)
	p = 0.32	p = 0.92	p = 0.07	p = 0.46	p = 0.56	p = 0.31	p = 0.34	p = 0.87	p = 0.07
$\bar{Y}_c$	0.12	0.11	0.09	-0.52	-0.54	-0.45	0.66	0.67	0.56
N	1295	1295	1295	635	635	635	660	660	660
	0.09	0.05	0.12	0.02	0.01	0.02	0.16	0.09	0.23
Female	(0.05)	(0.05)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.06)	(0.09)
	p = 0.11	p = 0.29	p = 0.06	p = 0.74	p = 0.81	p = 0.80	p = 0.02	p = 0.14	p = 0.01
$\bar{Y}_c$	-0.05	-0.02	-0.07	-0.65	-0.63	-0.58	0.57	0.61	0.45
N	1613	1613	1613	824	824	824	789	789	789

Notes: See Table 4b.

Two existing papers describe results from previous PFP experiments. Loyalka et al. (2018) ran a teacher incentive experiment among six grade math classes in rural China. They found no effects of incentive schemes based on simple formulas that map student gain scores or level scores into bonus payments for educators, but they found that PFP raised math achievement by .15 standard deviations. Fryer et al. (2013) report results from a PFP experiments in Chicago Heights, IL, which is an economically disadvantaged suburb of Chicago. This experiment involved students in all elementary grades, K-8. Here, the introduction of PFP raised math scores by .185 standard deviations.[14]

When comparing our results to the results from these previous studies, several features of the study environments merit attention. Students in China and Chicago Heights began the year much closer to their current grade level than our students in Uganda, and the assessments used in these two previous studies contained standard collections of grade-specific items. Further, all students in China and Chicago Heights have books. Given these considerations, our P4-P6 subtest outcome among students who attended schools with books and scored above the median on our round one assessment, seems like the most natural point of comparison with these two previous studies. Here, our estimate of the impact of PFP on round two math achievement is almost identical to the corresponding result from Chicago Heights and slightly greater than the result from China.

Given these gains on the P4 through P6 items, more able students in PFP schools with books scored .11 standard deviations higher on the full P1 though P6 round two assessment than their peers in control schools, even though there is little evidence that PFP raised achievement on the P1 through P3 subtest for students at any initial achievement level.

The most common P6 math text in Uganda is *Primary Mathematics: Pupil's Book 6* by MK publishers. We have compared the exercises in this text to the items on our round two assessment. Almost all of our P5 and P6 items are variations on exercises in this text, while a few of our P4 items are related but easier versions of these exercises. On the other hand, none of the items that we chose to represent the P1 through P3 curricula resemble these exercises. All of these items are much less challenging than the exercises in any standard P6 math text.

The growing literature on the value of targeting instructional resources to individual achievement levels may help us understand the patterns of results in Table 4b.[15] Students without books did not benefit from PFP. Further, among students with access to math books, only the more able students made gains, and these gains were driven by improved performance on the items that are most closely related to the content of these books. It may be that, weaker students, who began the year far below grade level, were not prepared to use P6 math texts effectively.

5 Books and Teacher Effort

PFP is designed to provide incentives for teachers to make better use of all the educational resources at their disposal. Thus, we should not be surprised that PFP created the largest achievement gains in schools with books among students who were academically prepared to use these books. However, this line of reasoning does not explain why PFP had no positive impacts on attainment or

[14]This effect is the average effect over two arms, one that provided individual incentives and another that provided team incentives. Yet another treatment arm combined PFP with a loss aversion treatment, i.e. teachers competed to keep bonus payments they received at the beginning of the school year, and this treatment produced even larger achievement gains.

[15]See Banerjee et al. (2017) for a recent summary.

achievement among schools without books. In this section, we present evidence that PFP did induce teachers in all schools to provide more effort. We conjecture that the stark contrasts between our resuls for schools with and without books reveal that books and educator effort are complements in the production of student learning. Below, we present several results that support this conjecture.

Table 5 describes the impacts of PFP on several measures of teacher effort. Each measure is derived from data collected in round two. The variable "Days Present" is the number of days during the past five schools days that the P6 math teacher has been present at school. We gathered this information from the Head Teacher. Our two "Hours" measures record the hours per week that the P6 math teacher spends preparing lessons and grading assignments. These measures are self-reports from the P6 teacher. Our effort index is the first principle component of the other three measures.

Table 5: The Effects of PFP on Teacher Effort

	0.18 (0.12) p = 0.13	0.10 (0.09) p = 0.30	0.26 (0.10) $p < 0.01$	0.24 (0.12) p = 0.04
$\bar{Y}_c$	4.29	2.13	2.22	-0.12
N	299	299	299	299

Notes: Days Present is the number of days in the past five schools days, including the round two visit day, that the P6 Math teacher was present at school, according to the Head Teacher. The Hours Preparing and Hours Grading variables are constructed from self-reports by the P6 math teachers. These survey items asked teachers to choose from a menu of thirty minute time intervals to describe their time allocations. To turn these responses into hours of work, we assigned time allocations that equal the midpoints of the chosen intervals. The Effort Index is the first principal component of the other three variables.

Our results indicate that PFP teachers supply more effort. All of the estimated effects of treatment on effort are positive. The increase in hours spent grading and the overall improvement in our effort index are statistically significant and represent noteworthy changes in behavior. Treatment teachers increased the time they spent grading assignments by more than ten percent, and the average value of this effort index was almost one fourth of a standard deviation higher among teachers in treatment schools.[16]

We do not report separate results for schools with and without books because the results in both subsamples are quite similar to those for the full sample. With respect to our measures of behavior changes, treatment teachers in schools without books responded to PFP the same way

[16] Our effort index is scaled to have a mean of zero and a standard deviation of one.

that PFP teachers responded in schools with books.[17] These results suggest that the gains from PFP are not concentrated in schools with books because the teachers in these treatment schools responded more to the PFP incentive scheme than PFP teachers in schools without books.

Given our conjecture about the complementarity between teacher effort and books, we examine correlations between the presence of books and rates of achievement growth. Tables 6a and 6b report results from regressions of round two achievement on round one achievement and an indicator for whether or not a student's school provided P6 math books for students. Table 6a presents results for control schools. Table 6b presents results for treatment schools.

Table 6a shows that, among students with similar round one scores, round two achievement in control schools is not significantly correlated with the presence of books. This is true for all three measures of round two achievement in the full sample and all the subsamples that we analyze. Further, most of the estimated correlations between the presence of books and achievement growth are negative.

Table 6a: Correlations Between Books and Round Two Achievement: Controls Schools

	All R1 Achievement Levels			Below R1 Median			Above R1 Median		
	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest
	-0.03	-0.02	-0.02	-0.04	-0.05	-0.01	-0.02	0.01	-0.05
Both	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)
	p=0.47	p=0.58	p=0.65	p=0.38	p=0.21	p=0.88	p=0.71	p=0.91	p=0.46
N	3044	3044	3044	1442	1442	1442	1602	1602	1602
	-0.01	0.01	-0.02	-0.03	-0.04	-0.01	0.00	0.05	-0.03
Male	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.08)
	p=0.84	p=0.89	p=0.71	p=0.64	p=0.49	p=0.88	p=0.94	p=0.49	p=0.65
N	1370	1370	1370	612	612	612	758	758	758
	-0.05	-0.05	-0.02	-0.05	-0.06	-0.01	-0.05	-0.03	-0.06
Female	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.07)	(0.07)
	p=0.32	p=0.31	p=0.68	p=0.37	p=0.23	p=0.89	p=0.48	p=0.63	p=0.43
N	1674	1674	1674	830	830	830	844	844	844

Notes: This table contains OLS regression coefficients from student-level regressions of round two achievement measures on round one achievement and an indicator for the presence of P6 math books in the student's school. The entries are estimated coefficients on the indicator for math book availability. To estimate standard errors, we use a standard HAC estimator and treat schools as clusters.

Table 6b shows that, in treatment schools, round two achievement is correlated with the presence of books. Treatment students who scored above the median in round one scored higher in round

[17] The esimtated PFP impacts on hours grading and the overall effort index among schools with books are quite close to those in the full sample and those based on the sample of schools without books. The hours grading effect is .29, and the PFP impact on overall effort is .22.

two if they enjoyed access to books. Further, their performance on the P4 through P6 items on the round two assessment drives most of this result. In treatment schools, but not in control schools, access to textbooks is a strong predictor of performance on items related to the material in P6 math books.

We have been careful to describe the results in Tables 6a and 6b as estimates of correlations. Our experiment did not involve random assignment of textbooks to schools. So, we cannot interpret these correlations as casual impacts. Further, because we do not know what forces caused some schools to acquire books and others not to acquire books, the patterns in Tables 6a and 6b do not provide direct support for our conjecture concerning the complementarity of teacher effort and books.

Table 6b: Correlations Between Books and Round Two Achievement: Treatment Schools

	All R1 Achievement Levels			Below R1 Median			Above R1 Median		
	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest
Both	0.07 (0.04) p=0.09	0.03 (0.04) p=0.44	0.13 (0.05) p=0.02	0.03 (0.04) p=0.56	-0.01 (0.04) p=0.88	0.08 (0.05) p=0.10	0.12 (0.05) p=0.03	0.07 (0.05) p=0.18	0.17 (0.07) p=0.02
N	3139	3139	3139	1553	1553	1553	1586	1586	1586
Male	0.08 (0.05) p=0.11	0.04 (0.05) p=0.44	0.13 (0.06) p=0.04	0.06 (0.06) p=0.35	0.03 (0.06) p=0.61	0.09 (0.07) p=0.19	0.11 (0.07) p=0.10	0.05 (0.07) p=0.44	0.16 (0.08) p=0.05
N	1361	1361	1361	671	671	671	690	690	690
Female	0.06 (0.05) p=0.21	0.02 (0.04) p=0.62	0.12 (0.06) p=0.05	-0.003 (0.05) p=0.96	-0.04 (0.05) p=0.51	0.06 (0.06) p=0.30	0.12 (0.06) p=0.05	0.08 (0.05) p=0.16	0.17 (0.08) p=0.04
N	1778	1778	1778	882	882	882	896	896	896

Notes: See Table 6a.

Nonetheless, round one achievement, in our full sample, is also uncorrelated with the presence of P6 math books.[18] This result and the results in Table 6a show that, absent additional incentives for educators, students with access to books do not perform better than students without access to books. Further, the results in Table 4b show that, among students with books, PFP incentives raised achievement among more able students on the items covered by their books. Finally, Table 6b shows that, among students whose teachers provide more effort in response to PFP, those with access to books performed better on material covered by their books.

[18] Round one scores are, on average, .066 standard deviations lower among students in schools with books, but this difference is far from statistically significant.

Taken together, these results are what we expect if baseline levels of effective teacher effort were quite low in our schools, PFP induced more effective effort from all treatment teachers, and teacher effort and books act as strong complements in the production of math skill.

5.1 Broader Gains in Attendance and Attainment

Before concluding, we must note an important difference between our achievement results and our attainment and attendance results. We did not present separate estimates of the impacts of PFP on attendance and attainment for students who are above versus below the median round one math score because we found that these differences are typically not large. Further, and as a rule, PFP generated larger attendance and attainment gains among students who performed below the median score in round one.

Tables 4a and 4b indicate that PFP did not produce measurable achievement gains among these lower achieving students. However, in treatment schools with books, something about the interactions between these students and their teachers changed that improved attendance rates at the end of the following school year. This pattern suggests that there are factors that we do not measure that impact attendance and attainment outcomes.[19]

6 Conclusion

We began by presenting evidence that educators in Uganda pressure weak students to drop out of primary school before they reach P7. We argue that teachers and head teachers engage in these educational triage practices to avoid negative attention and the sanctions that some schools suffer when their students take the PLE and fail. At one level, Uganda education authorities simply do not possess the information they need to hold educators accountable for these triage behaviors and many other dimensions of their performance. The government does not collect student-level achievement data for students in P1 through P6, and among P7 students, PLE scores are the only measures of individual achievement. Further, information systems do not exist that allow education authorities to accurately track the attendance of students or teachers. Finally, the government does not possess any reliable measures of dropout rates or rates of transfers among schools.

Given the government's limited information about student outcomes and teacher behavior, policy choices that attach high-stakes to PLE outcomes produce educational triage. Our results suggest that, if education authorities collect more information about student achievement, they may be able to reduce dropout rates and promote learning by employing well designed incentive schemes like PFP.

However, this conclusion comes with two important caveats. Our results suggest that even motivated teachers need learning materials, e.g. books, in order to help their students learn. Further, our results are consistent with the growing evidence that these learning materials are more valuable when they are targeted to the baseline achievement levels of students.

Finally, although our results suggest that effective teacher incentive provision combined with policies that increase student access to vital instructional resources may mitigate educational triage

[19] Along these lines, note the gender differences in control sample means in Tables 3a and 4a. Boys enjoy higher levels of math achievement, but girls have better attendance and higher attainment outcomes. The results in Table 3b suggest that, at least in schools with math books, PFP treatment narrows this female attainment advantage.

and promote primary completion, education authorities in Uganda cannot achieve universal primary completion simply by implementing these policies in upper primary grades, e.g. P5, P6, or P7. They must also implement policies that improve school performance in the early primary grades. Figure 1 implies that many students drop out of school before reaching P5, and many of the students in our sample began P6 with a tenuous command of the material in the P1 and P2 curriculum.[20] While a growing body of evidence suggests that these students can make progress, given instruction and materials that are intentionally remedial, Uganda will not approach universal primary completion as long as significant fractions of students enter P6 at P1 or P2 achievement levels.

[20] See the Overview chapter in World Bank (2018) for more evidence that achievement levels for many students in Uganda and other countries in Sub-Saharan Africa fall well below grade level.

7 PLE Data Appendix

Here, we describe how we matched students in our schools to their PLE records. The match requires records from two data sets:

- The round three data on PLE registration gathered by our round three enumerators. Our enumerators collected names, candidate numbers, and PLE testing centers for all students who were registered for the PLE.
- We obtained data on individual PLE outcomes for all students tested in the districts that contain our sample schools. The data contain the name, candidate number, PLE testing center, and PLE outcomes for each registered student.

Our merge process involved several steps:

1. We cleaned the student names in our round three data. These cleaning procedures involved correcting problems with spelling and spacing of characters for less than 100 records.
2. We cleaned the PLE data as well. We removed duplicate observations. We removed five records that marked a student as not showing up for the exam even though another record in the data provided exam results for the student in question. We dropped 16 records that contain results for eight students. In each case, there were two records for each of these eight students, and the PLE results conflicted within each record pair.
3. We matched these two data sets on name and PLE testing center. We required exact matches on both. We found that the candidate numbers were not reliable keys for matching.

8 Appendix Tables

Appendix Table 1a
The Effects of PFP on Attendance and Attainment:
No Control for Round One Achievement

	Interviewed Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
	0.02	0.04	0.02	0.01	-0.01
Both	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	p = 0.34	p = 0.06	p = 0.31	p = 0.63	p = 0.75
$\bar{Y}_c$	0.71	0.56	0.43	0.42	0.34
N	8788	8788	8788	8770	8770
	0.01	0.03	0.02	0.02	-0.02
Male	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	p = 0.50	p = 0.16	p = 0.32	p = 0.49	p = 0.50
$\bar{Y}_c$	0.69	0.55	0.41	0.40	0.34
N	4012	4012	4012	3998	3998
	0.02	0.04	0.02	0.003	-0.0002
Female	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
	p = 0.40	p = 0.08	p = 0.46	p = 0.89	p = 0.99
$\bar{Y}_c$	0.73	0.57	0.44	0.44	0.35
N	4776	4776	4776	4772	4772

Notes: See Table 3a.

Appendix Table 1b
The Effects of PFP on Attendance and Attainment: No Control for Round One Achievement Without Books and With Books

Panel A: Schools Without Books

	Interviewed Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
	0.02	0.01	0.02	0.01	-0.002
Both	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	p = 0.41	p = 0.75	p = 0.50	p = 0.70	p = 0.95
$\bar{Y_c}$	0.70	0.56	0.41	0.41	0.33
N	4703	4703	4703	4695	4695
	0.01	-0.01	-0.003	-0.01	-0.03
Male	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	p = 0.79	p = 0.67	p = 0.93	p = 0.79	p = 0.32
$\bar{Y_c}$	0.69	0.55	0.40	0.40	0.33
N	2125	2125	2125	2118	2118
	0.03	0.03	0.04	0.03	0.02
Female	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
	p = 0.30	p = 0.41	p = 0.28	p = 0.44	p = 0.54
$\bar{Y_c}$	0.72	0.56	0.42	0.42	0.33
N	2578	2578	2578	2577	2577

Panel B: Schools With Books

	Interviewed Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
	0.01	0.06	0.02	0.01	-0.01
Both	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
	p = 0.65	p = 0.01	p = 0.49	p = 0.81	p = 0.66
$\bar{Y_c}$	0.72	0.57	0.45	0.44	0.36
N	4085	4085	4085	4075	4075
	0.02	0.08	0.05	0.04	-0.00005
Male	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
	p = 0.48	p = 0.01	p = 0.14	p = 0.23	p = 0.99
$\bar{Y_c}$	0.69	0.55	0.42	0.41	0.34
N	1887	1887	1887	1880	1880
	0.001	0.05	-0.01	-0.02	-0.03
Female	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
	p = 0.98	p = 0.10	p = 0.87	p = 0.51	p = 0.47
$\bar{Y_c}$	0.75	0.58	0.47	0.47	0.38
N	2198	2198	2198	2195	2195

Notes: See Table 3b

Appendix Table 2a
The Effects of PFP on Achievement:
No Control for Round One Achievement

	All R1 Achievement Levels			Below R1 Median			Above R1 Median		
	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest
	-0.04	-0.05	-0.005	-0.08	-0.08	-0.05	0.06	0.03	0.09
Both	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.06)
	p = 0.60	p = 0.46	p = 0.94	p = 0.08	p = 0.08	p = 0.19	p = 0.29	p = 0.58	p = 0.15
$\bar{Y}_c$	0.07	0.08	0.03	-0.54	-0.52	-0.50	0.61	0.62	0.51
N	6183	6183	6183	2995	2995	2995	3188	3188	3188
	-0.07	-0.09	-0.03	-0.08	-0.08	-0.03	0.04	0.01	0.06
Male	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)
	p = 0.30	p = 0.20	p = 0.64	p = 0.22	p = 0.20	p = 0.55	p = 0.52	p = 0.89	p = 0.33
$\bar{Y}_c$	0.15	0.14	0.13	-0.48	-0.49	-0.43	0.67	0.65	0.58
N	2731	2731	2731	1283	1283	1283	1448	1448	1448
	-0.003	-0.02	0.02	-0.09	-0.09	-0.07	0.08	0.05	0.11
Female	(0.08)	(0.08)	(0.08)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)
	p = 0.97	p = 0.82	p = 0.78	p = 0.11	p = 0.12	p = 0.14	p = 0.22	p = 0.44	p = 0.11
$\bar{Y}_c$	-0.01	0.02	-0.05	-0.58	-0.55	-0.55	0.56	0.59	0.44
N	3452	3452	3452	1712	1712	1712	1740	1740	1740

Notes: See Table 4a

Appendix Table 2b
The Effects of PFP on Achievement:
No Control for Round One Achievement
Without Books and With Books

Panel A: Schools Without Books

	All R1 Achievement Levels			Below R1 Median			Above R1 Median		
	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest
	-0.09	-0.08	-0.08	-0.15	-0.14	-0.12	-0.01	-0.004	-0.02
Both	(0.09)	(0.09)	(0.08)	(0.06)	(0.06)	(0.05)	(0.07)	(0.07)	(0.07)
	p = 0.31	p = 0.36	p = 0.33	p = 0.01	p = 0.02	p = 0.02	p = 0.85	p = 0.95	p = 0.79
$\bar{Y}_c$	0.10	0.11	0.06	-0.49	-0.46	-0.47	0.61	0.60	0.51
N	3275	3275	3275	1536	1536	1536	1739	1739	1739
	-0.13	-0.11	-0.11	-0.15	-0.15	-0.11	-0.03	-0.01	-0.04
Male	(0.10)	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)	(0.07)	(0.07)	(0.09)
	p = 0.20	p = 0.24	p = 0.24	p = 0.07	p = 0.08	p = 0.16	p = 0.71	p = 0.90	p = 0.61
$\bar{Y}_c$	0.18	0.17	0.16	-0.45	-0.45	-0.41	0.67	0.64	0.60
N	1436	1436	1436	648	648	648	788	788	788
	-0.06	-0.05	-0.05	-0.14	-0.13	-0.13	0.01	0.004	0.01
Female	(0.10)	(0.10)	(0.09)	(0.07)	(0.07)	(0.06)	(0.08)	(0.08)	(0.08)
	p = 0.57	p = 0.59	p = 0.58	p = 0.04	p = 0.06	p = 0.03	p = 0.93	p = 0.96	p = 0.90
$\bar{Y}_c$	0.03	0.06	-0.03	-0.52	-0.48	-0.52	0.55	0.57	0.43
N	1839	1839	1839	888	888	888	951	951	951

Panel B: Schools With Books

	All R1 Achievement Levels			Below R1 Median			Above R1 Median		
	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest	Full Test	P1-P3 Subtest	P4-P6 Subtest
	0.02	-0.01	0.08	-0.01	-0.02	0.01	0.14	0.07	0.21
Both	(0.11)	(0.10)	(0.10)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.09)
	p = 0.82	p = 0.89	p = 0.45	p = 0.90	p = 0.81	p = 0.83	p = 0.10	p = 0.41	p = 0.02
$\bar{Y}_c$	0.03	0.04	0.00	-0.60	-0.59	-0.53	0.62	0.64	0.50
N	2908	2908	2908	1459	1459	1459	1449	1449	1449
	-0.01	-0.06	0.05	0.01	-0.002	0.04	0.11	0.02	0.19
Male	(0.10)	(0.10)	(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
	p = 0.90	p = 0.56	p = 0.58	p = 0.95	p = 0.98	p = 0.61	p = 0.20	p = 0.77	p = 0.05
$\bar{Y}_c$	0.12	0.11	0.09	-0.52	-0.54	-0.45	0.66	0.67	0.56
N	1295	1295	1295	635	635	635	660	660	660
	0.06	0.02	0.10	-0.02	-0.03	-0.01	0.16	0.10	0.23
Female	(0.13)	(0.13)	(0.12)	(0.09)	(0.09)	(0.08)	(0.10)	(0.10)	(0.11)
	p = 0.66	p = 0.85	p = 0.42	p = 0.79	p = 0.73	p = 0.90	p = 0.11	p = 0.30	p = 0.04
$\bar{Y}_c$	-0.05	-0.02	-0.07	-0.65	-0.63	-0.58	0.57	0.61	0.45
N	1613	1613	1613	824	824	824	789	789	789

Notes: See Table 4b

Appendix Table 3a: Correlations Between Books and Attendance/Attainment Measures: Control Schools

	Interviewed Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
Both	0.02 (0.02) p=0.45	0.01 (0.03) p=0.57	0.04 (0.03) p=0.10	0.03 (0.03) p=0.20	0.03 (0.02) p=0.19
N	4373	4373	4373	4373	4373
Male	0.01 (0.03) p=0.85	0.004 (0.03) p=0.89	0.03 (0.03) p=0.35	0.02 (0.03) p=0.57	0.01 (0.03) p=0.59
N	2038	2038	2038	2038	2038
Female	0.03 (0.03) p= 0.27	0.02 (0.03) p=0.43	0.06 (0.03) p=0.07	0.05 (0.03) p=0.11	0.05 (0.03) p=0.11
N	2335	2335	2335	2335	2335

Notes: See Table 6a. Here the outcomes are not achievement measures but the attendance and attainment outcomes in Table 3a.

Appendix Table 3b: Correlations Between Books and Attendance/Attainment Measures: Treatment Schools

	Interviewed Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
Both	0.01 (0.02) p=0.69	0.07 (0.03) p<0.01	0.05 (0.03) p=0.07	0.04 (0.03) p=0.16	0.03 (0.03) p=0.25
N	4415	4415	4415	4415	4415
Male	0.02 (0.03) p=0.47	0.10 (0.03) p<0.01	0.09 (0.03) p<0.01	0.08 (0.03) p=0.02	0.06 (0.03) p=0.06
N	1974	1974	1974	1974	1974
Female	0.002 (0.03) p=0.95	0.05 (0.03) p=0.07	0.02 (0.03) p=0.49	0.01 (0.03) p=0.76	0.01 (0.03) p=0.75
N	2441	2441	2441	2441	2441

Notes: Notes: See Table 6a. Here the outcomes are not achievement measures but the attendance and attainment outcomes in Table 3a.

References

Banerjee, Abhijit, Banerji, Rukmini, Berry, James, Duflo, Esther, Kannan, Harini, Mukerji, Shobhini, Shotland, Marc and Walton, Michael. (2017). 'From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application', *Journal of Economic Perspectives* 31(4), 73–102.

Bank, World. (2018), World Bank Development Report 2018: Learning to Realize Education's Promise, Technical report, World Bank.

Barlevy, Gadi and Neal, Derek. (2012). 'Pay for Percentile', *American Economic Review* 102(5), 1805–1831.

Bruns, Barbara, Filmer, Deon and Patrinos, Harry A. (2011), *Making Schools Work through Accountability Reforms*, The World Bank, chapter 5, pp. 211–251.

Campbell, Donald T. (1979). 'Assessing the Impact of Planned Social Change', *Evaluation and Program Planning* 2(1), 67–90.

Chaudhury, N., Hammer, J., Kremer, M. Gutmann, Muralidrharan, K. and Rogers, F.H. (2006). 'Missing in Action: Teacher and Health Worker Absence in Developing Countries', *Journal of Economics Perspectives* 20(1), 91–116.

Fryer, R., Levitt, S., List, J. and Sadoff, S. (2013). 'Enhancing the Efficacy of Teacher Incentives through Loss Aversioin: A Field Experiment', *working paper* .

Glewwe, Paul and Muralidharan, Karthik. (2016), Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications, *in* **Eric A. Hanushek, Stephen Machin and Ludger Woessmann**., eds, 'Handbook of the Economics of Education', Elsevier, chapter 10, pp. 653–744.

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

Kerr, Steven. (1975). 'On the Folly of Rewarding A, While Hoping for B.', *Academy of Management Journal* 18(4), 769–783.

Loyalka, Prashant, Sylvia, Sean, Liu, Chengfang, Chu, James and Shi, Yaojiang. (2018), Pay by Design: Teacher Performance Pay Design and the Distribution of Student Achievement.

Patrinos, Harry A. (2013), The Hidden Cost of Corruption: Teacher Absenteeism and Loss in Schools, *in* **Gareth Sweeney, Krina Despota and Samira Lindner**., eds, 'Global Corruption Report: Education', Routledge, chapter 2.8, pp. 70–73.

The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about our work on our website
www.theigc.org

For media or communications enquiries, please contact
mail@theigc.org

Subscribe to our newsletter and topic updates
www.theigc.org/newsletter

Follow us on Twitter
@the_igc

Contact us
International Growth Centre,
London School of Economic and Political Science,
Houghton Street,
London WC2A 2AE

IGC
International Growth Centre

DIRECTED BY
LSE
UNIVERSITY OF OXFORD

FUNDED BY
UKaid from the British people

Designed by soapbox.co.uk