NBER WORKING PAPER SERIES

EDUCATOR INCENTIVES AND EDUCATIONAL TRIAGE IN RURAL PRIMARY SCHOOLS

Daniel O. Gilligan
Naureen Karachiwalla
Ibrahim Kasirye
Adrienne M. Lucas
Derek Neal

Working Paper 24911
http://www.nber.org/papers/w24911

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2018

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), the Policies, Institutions, and Markets (PIM) Research Program of the Consultative Group for International Agricultural Research (CGIAR), and Lindy and Michael Keiser for research support through a gift to the University of Chicago's Committee on Education. We thank seminar participants at American University, Carnegie Mellon, the University of Chicago, the University of Wisconsin, Notre Dame, IFPRI, and RISE 2018. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Educator Incentives and Educational Triage in Rural Primary Schools
Daniel O. Gilligan, Naureen Karachiwalla, Ibrahim Kasirye, Adrienne M. Lucas, and Derek Neal
NBER Working Paper No. 24911
August 2018
JEL No. I0,J3,O1

ABSTRACT

In low-income countries, educators often encourage weak primary students to drop out before reaching the end of primary school in order to avoid the negative attention they receive when their students perform poorly on primary leaving exams. We conducted an experiment in rural Uganda that sought to reduce dropout rates in grade six and seven by rewarding teachers for the performance of each of their students. 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 textbooks 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 document several results that are consistent with the hypothesis that teacher effort and books are complements in education production.

Daniel O. Gilligan
International Food Policy Research Institute
1201 Eye St.. NW
Washington, DC 20005-3915
D.GIlligan@cgiar.org

Naureen Karachiwalla
International Food Policy Research Institute
2033 K Street NW
Washington, DC 20006
N.Karachiwalla@cgiar.org

Ibrahim Kasirye
Economic Policy Research Centre
P.O. Box 7841
Kampala, Uganda
ikasirye@eprcug.org

Adrienne M. Lucas
Lerner College of Business and Economics
University of Delaware
419 Purnell Hall
Newark, DE 19716
and NBER
alucas@udel.edu

Derek Neal
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and The Committee on Education
and also NBER
d-neal@uchicago.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/1152

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 not 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). Education authorities administer these exams to students who are completing the last year of primary school, and student results on these exams often affect student access to secondary education.

A recent newspaper article from Jinja district in Uganda highlights both that overall educator accountability is weak and that schools are held accountable for student performance on the PLE. The article reports that teachers are often absent and rarely punished for failing to teach their classes.[3] However, it also reports that district authorities recently demoted 11 school leaders because, in each of their schools, a significant fraction of the students who took the 2017 PLE failed. Further, 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.[4]

In many African countries, PLE outcomes necessarily receive considerable attention because they are among the few measures of primary student performance that education authorities possess. Each year in Uganda, 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 achievement data for 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 who attend a given grade in a given school. Thus, although the Ugandan Education Act of 2008 states that all children have a right to basic education, no public records document the attendance, attainment, or achievement of students who leave school without ever registering for the PLE.

This absence of public information about the academic outcomes of students who never take the PLE creates an incentive for educators to encourage weak students to drop out of school. Primary school in Uganda involves seven grades, P1 through P7. If weak students never enter P7, they cannot register for the PLE and therefore cannot fail the PLE, which means that their teachers

[1] See World Bank (2018).

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

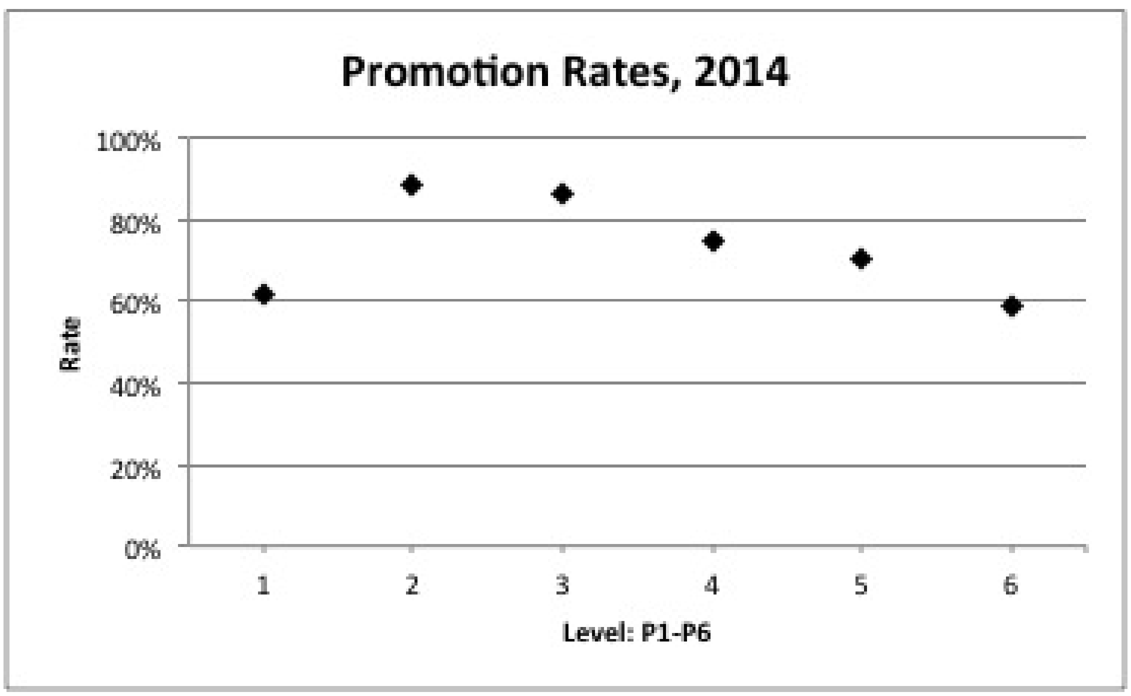
[3] See "Jinja headteachers demoted over PLE," *New Vision*, February 1, 2018. Chaudhury et al. (2006) and Patrinos (2013) also report that teacher absences are a problem in Ugandan schools. Further, during our first round of data collection in rural Uganda, more than one fourth of Head Teachers report that the P6 math teacher was absent from school, at least once, during the previous four school days.

[4] These Head Teachers, who are equivalent to school principals in the US, were demoted to regular teaching positions. These demotions likely involved substantial reductions in pay. The additional steps included stricter enforcement of rules that limit teacher absences. However, as written, these rules represent rather weak requirements.

cannot be held accountable for their low achievement levels. Even though Ugandan educators cannot legally force children to drop out of primary school, they can encourage weak students to drop out by directing all or most of their attention to students at the top of their classes, and they can counsel weak students to seek employment rather than return to school. In addition, any school can pressure a student to leave school by simply refusing to promote the student beyond a certain grade level.

Figure 1 presents aggregate promotion rates by primary grade level in Uganda. 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. Existing data do not allow us to determine what fraction of those who are not promoted at a given level in a given year choose to drop out of school, but dropout rates are substantial in Uganda, and it is reasonable to conjecture that this sharp drop in promotion rates over the upper primary grades reflects, in part, deliberate efforts by educators to encourage weak students to drop out before they reach P7 and register to take the PLE.[5] Further, during our field work for this project, we interviewed educators in Uganda who admitted that they engage in this form of educational triage.

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 $n = 1, 2, 3, 4, 5, 6$. Data are from the 2014 and 2015 Educational Statistical Abstracts published by the Ministry of Education, Technology, Science, and Sports.

[5] In its 2014 National Education Profile for Uganda, the Education Policy and Data Center reports that, in 2011, primary completion rates in Uganda were just over fifty percent for both boys and girls.

Table 1 presents the enrollment records for one school that we visited in the fall of 2015. 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, i.e. the school principal, 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.

The enrollment patterns in Table 1 are not anomalous. Nationwide data for Uganda from 2015 show that total P7 enrollment was less than two thirds of P6 enrollment. Further, in the sample of rural Ugandan schools that we analyze below, less than half of the students who begin P6 go on to complete P7 within two school years.[6]

We see similar patterns in Kenya, where students take the Kenyan Certificate of Primary Education exam at the end of eighth grade. Glewwe et al. (2009) report that because the government

[6]We cannot know exactly how many students complete P7 because we have no way to track students who allegedly transfered. However, in our control schools, only 42 percent of the students in our round one sample of P6 students go on to complete P7 and take the PLE in their round one school. If the dropout rate among transfers is at all comparable to the dropout rate in the rest of the sample, we know that less than half of our round one control sample completed P7 and took the PLE on time.

holds primary schools accountable for student scores on this leaving exam, some schools encourage weak students to drop out of school at the end of seventh grade, and national data show that recent P8 enrollment was less than 75 percent of P7 enrollment. In Rwanda, students take their PLE at the end of P6. In 2016, P6 enrollment in Rwanda was less than 60 percent of P5 enrollment.[7]

The appendix table in section 10 shows that more than thirty African countries use leaving exams to both certify primary completion and ration access to secondary school. PLE results are high-stakes outcomes for students, and education authorities in these countries rarely produce other information about primary student performance. Given this context, it is not surprising that PLE results often receive great public scrutiny. However, scholars throughout the social sciences have documented the harmful and possibly unintended consequences that frequently arise when organizations build incentive and accountability systems around a single performance metric.[8] Teachers in Uganda and many other African countries know that, if they can convince a weak student to drop out before she is eligible to register for her leaving exam, no record of her failure to learn will ever exist. Thus, when education officials in these countries pursue educator accountability by attaching high-stakes to PLE results, they likely induce educators to interact with weaker students in ways that encourage them to leave school.

Better Incentives

Below, we describe the results of a field experiment in rural Uganda that examined whether or not a specific assessment-based incentive system for educators can reduce dropout rates and promote learning among students at all achievement levels. The Pay For Percentile (PFP) incentive scheme developed in Barlevy and Neal (2012) rewards educators for the academic performance of each of their students. We introduced PFP for one year among P6 math classes in rural Uganda. Although this treatment lasted for only one year, it increased the probability that students who began P6 in a given school would still be attending this school at the end of the next school year from .56 to .60. This result was mostly driven by outcomes among students who attended 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 raised the likelihood that entering P6 students would 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 increased the probability that a student who began P6 would 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. Given the wide range of initial achievement levels in our schools, our assessments included questions from the P1 through P6 curricula. In treatment schools with books, PFP treatment improved performance on P4, P5, and P6 questions, i.e. the items most closely related to the exercises in P6 math texts, by .118 standard deviations. Further, among students in the top half of the round one achievement distribution, PFP in schools with books improved achievement on P4-P6 items by .186 standard deviations. In sum, PFP produced achievement gains in schools with books on items related to the exercises in these books, and these gains are largest among the students who are better prepared to use the books.

[7] See *Basic Education Statistical Booklet* (2014) Ministry of Education, Science, and Technology, Kenya, and Education Statistical Yearbook (2016), Ministry of Education, Republic of Rwanda.

[8] See Kerr (1975), Campbell (1979), Holmstrom and Milgrom (1991). Neal (2018) summarizes this literature.

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 academic 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. Several of our results support the view that teacher effort and textbooks are complements in the production of student achievement. We also discuss how our results fit into a growing literature that documents how the returns from investments in particular types of school resources depend on context. Our results support the view that, when new programs provide additional resources for students, the programs must ensure that the resources match current student achievement levels and that educators are motivated to use these new resources effectively. We conclude by discussing the potential policy implications of our results and 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 occupies 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 each teacher for the performance of each of her students. 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.

Our working hypothesis is that teachers in our control group are providing less than efficient levels of instructional effort for all of their students, especially their weaker students. Further, we conjecture that many are actually pressuring their weaker students to drop out before P7. Thus, we expect PFP treatment in P6 to induce more teacher effort for each student in each treatment class.

Our experiment is designed to determine whether or not PFP incentives elicit additional teacher effort that offsets or mitigates the educational triage incentives that plague the PLE system in Uganda. Since PLE systems are ubiquitous in Africa, our results speak to policy issues that likely impact primary completion throughout the continent.

2 Sample Design

The efficiency properties of PFP hinge on contestants believing that they are competing in properly seeded contests. Therefore, we chose only rural, government schools with one P6 stream[9] and an expected class size within a predetermined range. Further, we made sure that PFP teachers understood that they would be competing solely against teachers in other rural, government schools with only one P6 stream and comparable P6 enrollment.

To create a sample of eligible 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, 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.[10] 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 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

[9] Stream is the Ugandan term for a section.

[10] These validation visits, which were effectively round zero, took place about one month before round one began. We used a short survey to gather information about our sample schools from the Head Teachers. We used this information to make sure schools were eligible and to define our strata. We discovered, during our round one data collection, that the validation data concerning the presence of P6 math books were not accurate, presumably because these reports typically came from the Head Teacher and not the P6 math teacher.

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 interview the Head Teacher and the P6 math teacher.[11] 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 P1 through P6 curricula. After the students finished their exams, we then told treatment teachers that they would be participating in a performance pay contest during the coming school year.

PFP is designed to reward educators for how well each of their students performs at the end of the school year relative to students who began the year with comparable skills. Thus, we could not have implemented PFP faithfully using standard P5 or P6 assessments. If our round one test had included mainly P5 and P6 questions, more than half of the sample would have likely been placed in one large contest group for students who missed every question on the round one assessment. Further, teachers would have found it almost impossible to bring many of these students up to the level of P6 questions by the end of year, given that a significant fraction of them had not yet mastered P1 and P2 material. To credibly promise educators that we would seed contests correctly and measure differences in academic progress among students in all contest groups, we needed to create assessments for our beginning and end of P6 visits, i.e. rounds one and two, that cover a wide range of skill levels.

Our round one assessment contains 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.[12]

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.

[11] We did not provide advance notice that we were going to be testing the students.

[12] We also had each treatment teacher fill out a worksheet that illustrates how to map student performance into bonus payments under PFP. This exercise demonstrates that the performance of each student matters.

2.2 Subsequent Rounds of Data Collection

In October of 2016, we returned to our 302 schools for round two data collection. We administered a second math assessment, and we conducted a second round of interviews with the pupils, 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. Section 9, 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 -.096 standard deviations is not quite statistically significant, but it is academically noteworthy. Therefore, in all regression analyses of student outcomes, we include round one math achievement as a control.

2.4 PFP Implementation

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, but we noted that these return visits would not be announced.

We repeatedly stressed that each treatment 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 round one 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, e.g. the teacher of the median performer in a contest group would receive 10,000 Shillings for that student's performance.[13]

[13] In March 2016, 20,000 Shillings were worth about six US dollars.

Table 2: Balance Tests

	Treatment	Control	Difference	p-Value	N
School Variables					
Class Size	30.0	29.3	0.680	0.66	299
Low Teacher Education	0.698	0.673	0.025	0.65	299
Teacher Age	35.8	36.6	-0.815	0.42	299
Female Teacher	0.208	0.160	0.048	0.29	299
English Instruction	0.745	0.673	0.072	0.17	299
PLE Practice	0.557	0.520	0.037	0.52	299
Books	0.463	0.507	-0.044	0.45	299
Student Variables					
Does Homework	0.944	0.930	0.014	0.37	8864
Enjoys School	0.876	0.857	0.019	0.33	8864
Age	13.1	13.2	-0.086	0.14	8864
Girl	0.551	0.534	0.017	0.22	8864
R1 Achievement	-0.048	0.048	-0.096	0.14	8786
Cement Floor	0.474	0.478	-0.004	0.86	8864
Electricity	0.427	0.409	0.018	0.28	8864
Radio	0.835	0.837	-0.002	0.86	8864

Notes: R1 denotes Round One. Low Teacher Education equals one if the P6 Math teacher does not have a teaching diploma of any kind. The variable PLE Practice is an indicator that equals one if the school gives PLE practice exams to their students. Books captures the provision of P6 math texts. This indicator equals one even for schools that require some students to share a book. For school-level variables, we report results from standard t-tests. For student-level variables, we use a HAC estimator and treat schools as clusters to estimate standard errors. Bruhn and McKenzie (2008) recommend estimating regressions with strata dummies to conduct balance tests on school-level characteristics, and Bugni et al. (2018) derive conditions that justify this approach. Our sampling scheme does not satisfy these conditions because we do not assign treatment randomly within strata but instead seek to achieve rough balance within strata on past PLE performance. Nonetheless, to satisfy interested readers, we performed both the school-level and student-level tests using regressions that control for strata fixed effects. The results are quite similar. In seven cases, the resulting p-values are equal to or slightly greater than those reported above, and no p-value falls by more than .01.

To make sure that treatment teachers understood PFP, we had each treatment teacher fill out a worksheet that asked them to calculate the bonus payments that a teacher would earn given a scenario involving the assessment outcomes of five students in a hypothetical class. More than 75 percent filled out the entire worksheet correctly on their first attempt. Further, only four treatment teachers needed more than two tries to get a perfect score. Thus, our treatment teachers were not only literate and numerate but also understood how PFP works.

We also told treatment teachers that students who were enrolled at round one but not present for testing would compete against other absent students from other schools. However, we do not include students who were absent in round one but tested in round two in our analyses, in part, because we have no way to verify that they were actually enrolled at round one.

We further told treatment teachers that they would only earn bonus payments for the performance of students who were present and tested during these round two visits, but ex post, we used a slightly more generous payment rule. For the purpose of calculating bonus payments, we treated absent students as students who took the round two assessment but got every question wrong. We then gave these students a percentile rank equal to the fraction of students in their league who were either absent or took the assessment and got no questions correct.[14]

Table 2 shows that, on average, 30 students were present in each treatment school during round one. Further, we tested just over three students per school in round two who claimed to be students who were absent in round one but listed on the round one student registers. The maximum bonus that a teacher can win for the performance of a given student is 20,000 Shillings, and each contest among students must have one winner and one loser. So, overall we paid out just over 330,000 Shillings per teacher, which is 33 students times one half times 20,000 Shillings. Class-size is not correlated with round two performance, so the average of our 149 payouts to treatment teachers is just over 329,000 Shillings, which is about six weeks pay for a new teacher in Uganda and between two and three weeks pay for more experienced teachers.[15]

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 registers 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.[16] 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

[14] We exclude students who were not tested in round two from all of our achievement analyses. We include the two students who took the round one assessment, returned to take the round two assessment, and actually got every question wrong.

[15] See http://www.public service.go.ug for salary information.

[16] All of the results we present here are estimated impacts of the intention to treat (ITT). In both treatment and control schools, roughly 13 percent of R1 teachers are no longer teaching their R1 class of P6 students at R2. We contend that the ITT impacts are policy relevant because officials cannot mandate that teachers remain on their jobs.

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.

Students who scored higher on our round one math assessment are more likely to remain in school, enter P7, and take the PLE at the end of P7. However, more than one fourth of the best P6 students do not complete P7 on time, and many weak P6 students do.[17] Therefore, if PFP induces educators to devote more attention to all of their students, we expect students throughout the round one achievement distribution to feel more welcome in school. Whether or not these students experience learning gains, this effect could reduce dropout rates.

We also expect some students to 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, among some students, even small improvements in P6 achievement may make educators less eager to pressure them to leave school.[18]

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_{nj} 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, and ε_{nj} captures unobserved factors that influence measured round two achievement for student n in school j. When calculating standard errors, we assume that our individual error terms, ε_{nj}, are independent across schools, but we allow an unrestricted pattern of correlation among the error terms associated with students who attend the same school.

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, which occurred at the end of the first school year in our experiment. The second column demonstrates how treatment changes the probability that students are still attending their round one school in round three, which occurred at the end of the second school year. 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 in round 3. These students moved directly from P6 to P7 during our study. This indicator equals zero for those who are not attenders and for attenders who are still in P6.

[17] Among control students who score two standard deviations above the mean in round one, the predicted probability of completing P7 at the end of the following school year is less than seventy percent. Among those who score a full standard deviation below the mean, the corresponding rate is more than one fourth.

[18] Students who passed the PLE but earned Division 4 marks, i.e. the weakest performers among those who passed, answered about forty percent of the P4 questions correctly in the round 2 assessment at the end of P6. Students who took the PLE and failed answered one third of these questions correctly. Both groups missed roughly ninety percent of the P5 questions, although the former performed marginally better.

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

	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
Both	0.018	0.042	0.034	0.023	0.010
	(0.017)	(0.018)	(0.019)	(0.019)	(0.018)
	p = 0.28	p = 0.02	p = 0.08	p = 0.23	p = 0.59
$\bar{Y}_c$	0.71	0.56	0.43	0.42	0.34
N	8788	8788	8788	8770	8770
Male	0.017	0.041	0.042	0.036	0.007
	(0.020)	(0.022)	(0.023)	(0.023)	(0.021)
	p = 0.41	p = 0.06	p = 0.06	p = 0.12	p = 0.73
$\bar{Y}_c$	0.69	0.55	0.41	0.40	0.34
N	4012	4012	4012	3998	3998
Female	0.018	0.042	0.025	0.011	0.010
	(0.020)	(0.021)	(0.022)	(0.022)	(0.021)
	p = 0.36	p = 0.05	p = 0.26	p = 0.61	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. Present 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

Schools Without Books

	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
	0.022	0.013	0.028	0.019	0.009
Both	(0.025)	(0.026)	(0.026)	(0.026)	(0.024)
	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
	0.009	-0.005	0.011	0.005	-0.014
Male	(0.029)	(0.030)	(0.030)	(0.029)	(0.026)
	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
	0.031	0.028	0.040	0.029	0.026
Female	(0.029)	(0.030)	(0.031)	(0.030)	(0.029)
	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

Schools With Books

	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
	0.014	0.072	0.038	0.026	0.008
Both	(0.023)	(0.025)	(0.028)	(0.029)	(0.026)
	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
	0.025	0.089	0.074	0.067	0.029
Male	(0.028)	(0.029)	(0.033)	(0.034)	(0.031)
	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	0.057	0.006	-0.010	-0.011
Female	(0.027)	(0.030)	(0.033)	(0.033)	(0.030)
	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.

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

We define all five outcomes based on a student's relationship to her baseline school. When schools reported, in round three, that a student had not attended her baseline school at all during year two of our study, we asked why. In a substantial number of cases, schools reported that these students were attending other schools. Yet, we have no way to verify these reports. Some of these students may have told their baseline school that they were going to attend another school and never did, and others may have transferred to a different school but stopped attending school before the date of our round three data collection.

Still, we do know that these students are not attending their baseline schools. So, we code them as students who are not attending in round three and not participating in the PLE. We do not expect this choice to have a significant impact on our estimates of the impact of treatment on attendance and attainment outcomes. In both our treatment and control samples, schools report in round three that roughly fourteen percent of the students we tested at baseline have transferred to another school, and in both samples, about half of these students were also not present for testing during round two. Thus, neither the prevalence or timing of these reported transfers are correlated with treatment status.

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, which took place six to seven months later.

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 18 to 19 months later 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. The estimated impact among boys is slightly larger with an associated p-value of .06, but 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 this result. 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 .072 overall, .089 among boys, and .057 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 more than seven percentage points more likely to attend P7 during round three and almost seven percentage points more likely to take the PLE. Note that only about forty percent of boys in our control schools took the PLE.

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.[19] 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.[20]

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 estimates 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 difference between the estimated effects of PFP on round three attendance among boys is clearly significant and the corresponding difference for the full sample is marginally significant.[21]

4 P6 Achievement

In this section, we examine how PFP impacts achievement. Our achievement results provide strong additional support for the conclusion that our PFP treatment produced real benefits for students in schools that possessed P6 math books but did little to benefit students in treatment schools without these books. In the next section, we present evidence that the value of PFP treatment hinges on the presence of books because books complement teacher effort and not because books are simply a proxy for some unmeasured dimension of school quality.

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.[22] 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. Many of these students did not yet possess clear command of P1 material. There is no reason to believe that their teachers could have taught them in ways that would have allowed them to move up four or five grade levels in one year. Therefore, the best efforts of these teachers would have had little impact on the expected scores of their weakest students on a standard P6 assessment.

[19] 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.

[20] Appendix Tables 1a and 1b present parallel results from regression models that do not contain a control for round one math performance. The raw difference between round two attendance rates in treatment and control schools is .035. The p-value associated with this difference is .06.

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

[22] See World Bank (2018) pages 3-8.

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
	0.018	0.003	0.042	-0.001	-0.002	-0.002	0.038	0.011	0.067
Both	(0.030)	(0.028)	(0.035)	(0.031)	(0.030)	(0.034)	(0.040)	(0.038)	(0.048)
	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
	0.005	-0.012	0.036	0.003	-0.0001	0.019	0.012	-0.014	0.039
Male	(0.036)	(0.035)	(0.042)	(0.044)	(0.043)	(0.049)	(0.047)	(0.047)	(0.058)
	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
	0.029	0.014	0.049	-0.006	-0.003	-0.020	0.063	0.033	0.097
Female	(0.034)	(0.033)	(0.040)	(0.037)	(0.037)	(0.040)	(0.045)	(0.044)	(0.055)
	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

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
	-0.031	-0.023	-0.029	-0.036	-0.026	-0.050	-0.026	-0.017	-0.033
Both	(0.036)	(0.036)	(0.042)	(0.038)	(0.038)	(0.040)	(0.051)	(0.050)	(0.061)
	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
	-0.041	-0.029	-0.038	-0.044	-0.037	-0.041	-0.036	-0.016	-0.052
Male	(0.047)	(0.045)	(0.056)	(0.060)	(0.058)	(0.069)	(0.064)	(0.061)	(0.081)
	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
	-0.022	-0.019	-0.020	-0.031	-0.019	-0.056	-0.016	-0.019	-0.012
Female	(0.041)	(0.042)	(0.046)	(0.043)	(0.048)	(0.046)	(0.057)	(0.058)	(0.066)
	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

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
	0.072	0.032	0.118	0.033	0.025	0.042	0.113	0.044	0.186
Both	(0.047)	(0.044)	(0.056)	(0.049)	(0.046)	(0.056)	(0.061)	(0.059)	(0.073)
	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
	0.054	0.005	0.113	0.047	0.037	0.071	0.067	-0.012	0.144
Male	(0.054)	(0.054)	(0.061)	(0.064)	(0.063)	(0.070)	(0.070)	(0.073)	(0.079)
	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
	0.087	0.053	0.125	0.020	0.014	0.017	0.157	0.095	0.226
Female	(0.054)	(0.050)	(0.066)	(0.061)	(0.057)	(0.067)	(0.067)	(0.064)	(0.088)
	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 4a.

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-P6 subtest by about .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 are statistically significant. In contrast, among schools with books, we see important gains on the P4-P6 subtest. Consider the third column of Panel B in Table 4b. Among schools with books, PFP raised achievement on the P4-P6 subtest by .118 standard deviations. We see similar results when we examine boys and girls separately, although these estimated impacts are slightly less significant statistically.[23]

The final column shows that the performance of students in the top half of the round one achievement distribution accounts 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-P6 subtest that are, on average, .186 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. Further, among students with higher round one achievement, the treatment gains on the P4-P6 subtest drive a parallel gain of .113 standard deviations on the full test, with a p-value of .07. These estimated impacts of PFP on round two achievement are not only statistically significant from zero but also statistically different from the corresponding estimates of PFP impacts among treatment schools without books.[24]

Yet, even in treatment schools with books, we see little evidence that PFP improves performance on the P1-P3 subtest. This may reflect that P6 math books are not the best tools for raising achievement on P1-P3 items. The most common P6 math text in Uganda is *Primary Mathematics:*

[23]Tables 4a and 4b present results from regression of round two achievement on round one achievement and an indicator for treatment. One can also assess the impact of treatment on gain scores. We defined three different gain scores by taking the differences between our three round two achievement measures and our round one measure. Appendix Tables 2a and 2b contain treatment vs control differences in mean gain scores. These results are quite similar to those presented in Tables 4a and 4b.

[24]In the full sample and the above median R1 achievement sample, the p-values for tests of equal treatment impacts on the P4-P6 subtest are .035 and .022 respectively. Among higher round one achievers, the p-value on the test for equal treatment impact on the full test is .079.

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-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.[25] Students without books did not benefit from PFP. Further, among students with access to math books, the gains associated with treatment primarily reflect improved performance on the items that are closely related to the content of these books, among the students who were more prepared to use these books. Weaker students, who began the year far below grade level, may not have been prepared to use P6 math texts effectively. Books allow teachers to assign and grade more problems. Further, books allow teachers to let one group of students work on problems while they give special attention to other students. However, students who are working independently in class or doing homework exercises may benefit little from their books if their true current achievement levels are far below P6.

4.1 Comparisons with Other PFP Studies

Three existing papers describe results from previous PFP experiments. Loyalka et al. (2018) ran a teacher incentive experiment among sixth 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. Mbiti, Romero and Schipper (2018) report results from an experiment in Tanzania that involved two different performance pay schemes for teachers in grades one through three. In this case, both PFP and a scheme built around several score thresholds that mark different levels of achievement improved student learning. However, PFP did not produce systematically larger learning gains. Further, in several grades and subjects, the threshold scheme produced better, although not statistically different, student performance. Fryer et al. (2013) report results from a PFP experiment 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.[26]

Our study most closely parallels Loyalka et al. (2018). Both studies involve sixth grade classrooms, and in both cases, teachers earned bonuses tied to the performance of their students in math. When comparing our results, three environmental differences are noteworthy. First, all students in the China study have books. Second, students in China began the year much closer to their current grade level than our students in Uganda. Finally, the assessments used in the China study are standard collections of grade-specific items. Taken together, these facts imply that our estimated treatment effect on the P4-P6 subtest among students who scored above the median on our round one assessment and enjoyed access to math books is the most natural point of comparison between the two studies. Here, we estimate that PFP raises round two math achievement by .186 standard deviations, which is quite close to the corresponding result in China.

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

[26] 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.

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 are not surprised that PFP created academic gains in schools with books. However, it is not clear why PFP had no positive impacts on attainment or achievement among schools without books. In this section, we present evidence that PFP did induce teachers in all schools to provide more effort. Thus, we discount the possibility that treatment teachers in schools without books did not respond to the PFP incentive scheme. Rather, we conjecture that the stark contrasts between our results for schools with and without books reveals 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.

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 our effort index was almost one fourth of a standard deviation higher among teachers in treatment schools.[27]

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 that PFP teachers responded in schools with books.[28] 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.

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

[28] The estimated 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.

Table 5: The Effects of PFP on Teacher Effort

	Days Present	Hours Preparing	Hours Grading	Effort Index
	0.18	0.10	0.26	0.24
	(0.12)	(0.09)	(0.10)	(0.12)
	p = 0.13	p = 0.30	p < 0.01	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.

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 6b shows that, in treatment schools, round two achievement is correlated with the presence of books. Further, the pattern of achievement differences between treatment students with and without books mirrors the contrasts between treatment and control students with books presented in Table 4b. Among treatment students, access to books is positively correlated with round two achievement on the P4-P6 subtest. Further, this correlation is driven largely by outcomes among treatment students in the top half of the round one achievement distribution. Among these students, scores on the P4-P6 subtest are .171 standard deviations higher among those with access to books. Further, among these same students, those with books score .118 standard deviations higher on the full test.

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 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
Both	-0.031 (0.043) p=0.47	-0.024 (0.043) p=0.58	-0.021 (0.045) p=0.65	-0.039 (0.044) p=0.38	-0.054 (0.042) p=0.21	-0.007 (0.048) p=0.88	-0.022 (0.057) p=0.71	0.007 (0.059) p=0.91	-0.047 (0.063) p=0.46
N	3044	3044	3044	1442	1442	1442	1602	1602	1602
Male	-0.010 (0.051) p=0.84	0.007 (0.052) p=0.89	-0.021 (0.056) p=0.71	-0.028 (0.061) p=0.64	-0.043 (0.062) p=0.49	-0.009 (0.065) p=0.88	0.005 (0.068) p=0.94	0.048 (0.069) p=0.49	-0.035 (0.077) p=0.65
N	1370	1370	1370	612	612	612	758	758	758
Female	-0.048 (0.048) p=0.32	-0.050 (0.049) p=0.31	-0.021 (0.052) p=0.68	-0.047 (0.052) p=0.37	-0.062 (0.052) p=0.23	-0.007 (0.054) p=0.89	-0.046 (0.064) p=0.48	-0.032 (0.068) p=0.63	-0.057 (0.072) 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: 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
	0.071	0.029	0.126	0.025	-0.006	0.080	0.118	0.068	0.171
Both	(0.041)	(0.037)	(0.054)	(0.043)	(0.043)	(0.049)	(0.054)	(0.050)	(0.071)
	p=0.09	p=0.44	p=0.02	p=0.56	p=0.88	p=0.10	p=0.03	p=0.18	p=0.02
N	3139	3139	3139	1553	1553	1553	1586	1586	1586
	0.081	0.038	0.127	0.058	0.031	0.094	0.108	0.051	0.163
Male	(0.050)	(0.048)	(0.061)	(0.062)	(0.059)	(0.072)	(0.066)	(0.065)	(0.082)
	p=0.11	p=0.44	p=0.04	p=0.35	p=0.61	p=0.19	p=0.10	p=0.44	p=0.05
N	1361	1361	1361	671	671	671	690	690	690
	0.061	0.022	0.124	-0.003	-0.036	0.063	0.121	0.076	0.175
Female	(0.048)	(0.044)	(0.062)	(0.053)	(0.054)	(0.061)	(0.060)	(0.054)	(0.083)
	p=0.21	p=0.62	p=0.05	p=0.96	p=0.51	p=0.30	p=0.05	p=0.16	p=0.04
N	1778	1778	1778	882	882	882	896	896	896

Notes: See Table 6a.

Nonetheless, we must note that round one achievement is also uncorrelated with the presence of P6 math books in both treatment and control schools. In fact, in both treatment and control schools, round one achievement is lower in schools with P6 math books, although neither of these deficits is statistically significant. These patterns lead us to discount the possibility that the presence of books serves a proxy for school quality or community support for education.

We see several patterns in our two rounds of achievement results that fit the same story. Absent PFP treatment, student achievement was not correlated with student access to books in either round one or round two. Further, absent books, student achievement in round two was not correlated with PFP treatment. However, the combination of PFP treatment and access to books was associated with better round two performance on the material covered in P6 math books, especially among students who were better prepared to use these books. All of these results should be expected 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.

6 Comparisons of Achievement and Attainment Gains

Before concluding, we must note a contrast between our achievement results and our results for attendance and attainment. 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, when differences exist, the attendance and attainment gains enjoyed by PFP students in schools with books are usually slightly larger among those with below median achievement in round one.

Yet, Tables 4a and 4b indicate that PFP did not produce statistically significant achievement gains among lower achieving students, whether or not they enjoyed access to books. In treatment schools with books, boys in the bottom half of the round one achievement distribution did score .071 standard deviations higher on the P4-P6 subtest than their peers in control schools with books, and among those who scored between the 25th percentile and the median in round one, this gap is .124 standard deviations. However, neither of these score improvements is statistically significant.

Even though these achievement gains are not statistically significant, they are of interest since the attainment gains associated with PFP are large and statistically significant among boys in the lower half of the round one achievement distribution. For example, among boys in treatment schools with books who scored below the round one median, PFP raised P7 attendance rates by almost 11 percentage points and raised PLE participation rates by 8 percentage points.[29] These attainment gains may be driven by the attainment outcomes of a portion of low-achieving boys in treatment schools with books who enjoyed substantial achievement gains, or it may be that promotion decisions are quite elastic to small improvements in command over the P4-P6 material.

Among girls in the lower-half of the round one achievement distribution, we see little evidence that treatment in schools with books produces either achievement growth or attainment gains. However, treatment did boost round three attendance among these girls. It is possible that PFP improved the interactions between low-achieving girls and their teachers in ways that made them more willing to return to school and repeat P6.

Nonetheless, we do not understand why the girls in treatment schools with books who scored in the top-half of the round one achievement distribution enjoyed substantial achievement gains on the P4-P6 subtest but did not enjoy any attainment gains in terms on promotion rates, PLE participation, or PLE performance. We hope to explore this question in future research, but for now, we note that, although boys enjoy higher average round one scores than girls in both our treatment and control samples, girls in our control sample were more likely than boys to finish P7, take the PLE, and pass the PLE. These gender gaps are apparent in the patterns of control mean outcomes presented in Tables 3a-4b.

Further, we discovered using regressions that, among control school students with books, girls are seven percentage points more likely to take the PLE than boys with similar round one scores, but among treatment students with books, the corresponding gender gap in PLE participation probabilities is essentially zero at -.006. These estimated gender gaps are statistically different, and we find results that follow a similar pattern when we examine gender gaps in PLE pass rates.[30] Students must complete P7 to take the PLE. So, these results suggest that, in schools with books, PFP treatment eliminates a significant female attainment advantage, even though Table 4b shows that, in the same schools, PFP generates substantial achievement gains among high-achieving girls.

7 Related Research

Our results contribute to several literatures. Other studies have shown that measured student achievement in rural schools in low-income countries increases in experiments where teachers receive

[29]The p-values associated with these treatment impacts are .01 and .06 respectively.

[30]The conditional gender gap in PLE pass rates favors girls by almost five percentage points in control schools with books, and the p-value associated with this gap is .04. In treatment schools with books, the parallel gap is just over one percentage point and not statistically significant. However, we cannot reject the null that these gaps are the same.

bonus payments linked to measures of educator performance derived from student scores. In Glewwe et al. (2010), the authors found evidence that coaching students for a specific exam rather than better teaching may have created these improvements. However, Muralidharan and Sundararaman (2011) and Loyalka et al. (2018) report similar results in settings where coaching should not be a concern because the experiments employed new exams that teachers had not seen before. Further, Mbiti, Romero and Schipper (2018) found that performance pay raised student achievement on low-stakes audit exams that did not impact teacher pay.[31] In the Glewwe et al. (2010), Muralidharan and Sundararaman (2011) and Loyalka et al. (2018) samples, all or most of the students enjoyed access to books. However, in Mbiti, Romero and Schipper (2018), which dealt with Tanzanian students in grades one through three, roughly two-thirds of students did not have books.[32]

The results from a different experiment involving first, second, and third graders in Tanzania provide support for our claim that teacher effort and books are complements in the production of student skill. In Mbiti, Muralidharan, Romero and Schipper (2018), reseachers assigned 140 schools to a control group. Then, seventy schools received resource grants that they could use for classroom investments. Another seventy schools participated in a performance pay program for educators that linked bonus payments to whether or not students scored above fixed proficiency scores on end-of-year assessments. Finally, another seventy schools both received resource grants and participated in the performance pay program.

The grants did generate improvement in classroom resources, and on average, schools spent over thirty-five percent of their grant money on textbooks. However, these grants alone did not produce improved learning outcomes for students. Further, in schools that participated in the incentives-only treatment, students enjoyed, at best, small achievement gains. Nonetheless, students in schools that both received grants and participated in the performance pay program enjoyed large achievement gains. At the end of year two, these students typically scored higher in all grades and subjects than comparable students in control schools. Further, in all cases, these score gains were roughly .2 standard deviations or more.

Mbiti, Muralidharan, Romero and Schipper (2018) find that providing grants alone does not improve achievement, even though these grants were used to purchase books, practice exams, and other instructional materials. Several previous studies found that providing books directly to all children in rural schools may do little to improve student outcomes.[33] Our results suggest that providing better incentives for educators may accomplish little if their students do not have books or other materials that facilitate learning. Yet, our results and the Mbiti, Muralidharan, Romero and Schipper (2018) results indicate that by combining performance incentives with programs that improve student access to books and other learning materials, policy makers can improvement student achievement.

Our finding that PFP treatment produced the most noteworthy achievement gains among stu-

[31] Barrera-Osorio and Raju (2017) describe a performance pay experiment in Pakistan that produced few learning gains. The authors conjecture that the program was less effective than those evaluated in the papers cited above because government officials rather than researchers ran the program. However, the program also involved a complicated school-level incentive scheme that was quite different than those used in previous experiments.

[32] Glewwe et al. (2010) report that International Child Support provided textbooks or school resource grants to 75 of the 100 schools in their incentive experiment either before or during the experiment. Das et al. (2013) report that the government provides free textbooks to students in Andhra Pradesh, the province where Muralidharan and Sundararaman (2011) ran their study. In personal correspondence, Prashant Loyaloka confirmed that the Chinese schools in Loyalka et al. (2018) provide books for students. In response to our inquiry, Isaac Mbiti reported that 67 percent of the Tanzanian students in Mbiti, Romero and Schipper (2018) had no books.

[33] See Glewwe et al. (2009), Glewwe and Muralidharan (2016).

dents who were more prepared to use the books present in their schools echoes several recent studies that highlight the importance of matching instructional resources to baseline achievement levels. Glewwe et al. (2009) found that providing textbooks to schools in Kenya produced no gains for students in the bottom four quintiles of the baseline achievement distribution. However, students in the top quintile did enjoy achievement gains when they received books. Glewwe et al. (2009) argue that Kenya education policies are designed to serve the most advantaged students and that the textbooks schools use match the needs of these students.

Banerjee et al. (2017) reviews a number of experiments that sought to implement "teaching at the right level" (TaRL) in Indian schools. In a series of experiments, researchers learned that students who are far below grade level can make substantial progress given instruction tailored to their baseline achievement levels. Along the way, researchers learned that success required getting instructors, either teachers or volunteers, to embrace the mission of targeting the learning needs of students who have fallen behind.

Muralidharan et al. (2017) report results from another set of experiments in India. They note that students enjoyed large gains in math and Hindi when they participated in an after-school program that used Mindspark software to tailor instruction to the learning needs of individual students. They also note that Berry and Mukherjee (2016) evaluate a similar after-school program in India that did not employ software to deliver targeted instruction, and this program produced no learning gains.[34]

Our PFP treatment is designed to focus educators on the learning needs of all of their students, and our attendance results suggest that even the weaker students in our treatment schools were more likely to stay in school the following academic year. However, our treatment did not provide materials or training that would help teachers provide differentiated instruction for students at different round one achievement levels, and this may explain why PFP produced few measurable achievement gains among students who were ill-prepared to use P6 math books.

8 Conclusion

We began by presenting evidence that educators in Uganda pressure weak students to drop out of primary school before they reach P7. 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 education authorities in Uganda may be able to reduce dropout rates and promote learning by employing well designed incentive schemes like PFP.

We only spent about three dollars per student on bonus payments. Further, the government can implement PFP at scale without tracking student movements among all schools or identifying all dropouts. Authorities just need a technology that would allow them to verify the population of enrolled students in each school at the beginning of each school year and then determine which of these students were present for testing at the end of each year. Education officials in rural Uganda do not now have this capacity, but they may soon have it. During the summer of 2017, Uganda began creating a national registry of school-aged children.[35]

[34] Kerwin and Thornton (2018) report that implementation details surrounding training and resources interact in complicated ways that significantly impact the benefits that students derive from the Mango Tree literacy program.

[35] See https://www.nira.go.ug/wp-content/uploads/Publish/Handbook.pdf and https://allafrica.com/stories/201705010523.html.

Our results likely have policy implications for many other low-income countries. The appendix table in section 10 lists more than thirty African countries that use PLE systems to ration access to all levels of secondary schooling. In addition, other countries in Africa and Asia use similar systems to ration access to some level of upper secondary schooling.[36] In many of these countries, PLE results are high-stakes outcomes for students, and governments collect little data on other student outcomes. Thus, it is likely the educational triage practices we highlight exist beyond the borders of Uganda.

Nonetheless, our results suggest that countries like Uganda cannot produce the learning outcomes they desire simply by providing better incentives for educators. 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 match 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 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. In Uganda, 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 curricula.[37] A growing body of evidence suggests that these students can make progress, given instruction and materials that are intentionally remedial. However, Uganda will likely not approach universal primary completion without adopting reforms that promote much more learning and much lower dropout rates in grades P1 through P5.

[36] Examples include China, Ghana and India.

[37] 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.

9 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 fewer 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.

10 Primary Leaving Exams and Access to Secondary Schooling in African Countries

Country	Years of Schooling Prior to Primary School Exam(s)	Requirements for Secondary School Admission	Useful Websites
Angola	6	Includes passing school-level evaluations in grades 2, 4, and 6 and taking national exams at the end of grade 6.	http://www.ibe.unesco.org
Benin	6	Includes passing the Certificat d'Études (CEP).	http://www.ibe.unesco.org
Burkina Faso	6	Includes passing the Certificat d'Études Primaires (CEP).	http://www.ibe.unesco.org
Burundi	6	Includes obtaining the Certificat de Réussite through the Concours National.	http://www.iipe-poledakar.org http://www.uis.unesco.org
Cameroon	6	Includes passing the Certificat d'Études Primaires Élémentaires (CEPE) in the Francophone system or the First School Leaving Certificate (FCLC) in the Anglophone system.	http://www.ibe.unesco.org http://www.uis.unesco.org
Central African Republic	6	Includes passing the Certificat d'Études Fondamentales 1 (CEF1).	http://www.ibe.unesco.org
Chad	6	Includes passing the Certificat de Fin d'Études Primaires Elémentaires Tchadien (CEPET).	http://www.uis.unesco.org
Comoros	6	Includes passing the Certificat d'Études Primaires Elémentaires (CEPE).	http://www.uis.unesco.org
Democratic Republic of the Congo	6	Includes passing the Test de Fin d'Études Primaires (TENAFEP).	http://www.uis.unesco.org http://www.ibe.unesco.org
Republic of the Congo	6	Includes passing the Certificat d'Études Primaires Élémentaires (CEPE) and entrance exam.	http://www.ibe.unesco.org
Ethiopia	4, 8	Includes passing national exams at the end of grade 4 to continue to grade 5 and the Primary School Certificate Exam at the end of grade 8 for admission to secondary schools.	http://www.ibe.unesco.org
Gabon	5	Includes passing the Certificat d'Études Primaires Élémentaires (CEPE) and entrance exams.	http://www.ibe.unesco.org
Guinea	6	Includes passing the Certificat d'Études Primaires Élémentaires (CEPE).	https://www.afdb.org http://www.ibe.unesco.org
Ivory Coast	6	Includes passing the Certificat d'Études Primaires Élémentaires (CEPE).	http://www.uis.unesco.org
Kenya	8	Includes passing the Kenya Certificate of Primary Education exam.	http://www.ibe.unesco.org

Primary Leaving Exams in African Countries - continued

Lesotho	7	Includes passing the Primary School Leaving Examination (PSLE).	http://www.ibe.unesco.org
Madagascar	5	Includes passing the Certificat d'Études Primaires Élémentaires (CEPE).	http://www.ibe.unesco.org http://www.uis.unesco.org
Malawi	8	Includes passing the Primary School Leaving Certifcation Examination (PSLCE).	http://www.ibe.unesco.org
Mauritania	6	Includes passing the Concours d'Entrée en Première Année Secondaire/ Certificat d'Études Fondamentales (CEPAS/CEF).	http://www.uis.unesco.org
Mauritius	6	Includes passing the Certificate of Primary Education (CPE).	http://www.uis.unesco.org
Mozambique	5, 7	Includes passing national exams at the end of grade 5 to continue to grade 6 and national exams at the end of grade 7 for admission to secondary schools.	http://www.ibe.unesco.org
Niger	6	Includes passing the Certificat de Fin d'Études du Premier Degré (CFEFD) or Certificat d'Études primaires Élémentaires Franco-Arabe (CEPE- FA).	http://www.ibe.unesco.org
Rwanda	6	Includes obtaining a national certificate of grade 6 by passing national exams.	http://www.ibe.unesco.org
Senegal	6	Includes earning sufficient rank in the Certificat de Fin d'Études Élémentaires (CFEE).	http://www.ibe.unesco.org
Sierra Leone	6	Includes passing the National Primary School Examination (NPSE).	http://worldbank.org
Sudan	8	Includes passing the basic education certificate examination.	http://www.ibe.unesco.org
Swaziland	7	Includes passing the Swaziland Primary Certificate exam.	http://www.ibe.unesco.org
Tanzania	7	Includes passing the Primary School Leaving Examination (PSLE) in Mainland Tanzania or national exam for admission to selective secondary schools in Zanzibar.	http://www.ibe.unesco.org
Togo	6	Includes passing the Certificat d'Études du Premier Degré (CEPD).	http://www.ibe.unesco.org
Tunisia	6	Includes passing national exam for admission to pilot middle schools.	http://www.ibe.unesco.org
Uganda	7	Includes passing the Primary Leaving Examinations (PLE).	http://www.uis.unesco.org
Zambia	7	Includes passing the Grade 7 Composite Examination.	http://www.uis.unesco.org

Notes: This table describes how results on primary leaving exams impact secondary school admission in many African countries. We do not include countries that only use exit exams to ration upper secondary education. The countries listed here all consider PLE results when allocating slots in secondary schools. Column 2 describes the typical years of schooling before students take their PLE. Column 3 summarizes how PLE results impact secondary admission decisions. Our data come primarily from UNESCO Institute for Statistics Central Data Catalog and English translations of UNESCO International Bureau of Educations World Data on Education 2010-11 reports.

11 Appendix Tables

Appendix Table 1a
Mean Treatment vs Control Differences
in Attendance and Attainment

	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
Both	0.016 (0.017) p = 0.34	0.035 (0.019) p = 0.06	0.021 (0.021) p = 0.31	0.010 (0.021) p = 0.63	-0.007 (0.022) p = 0.75
$\bar{Y_c}$	0.71	0.56	0.43	0.42	0.34
N	8788	8788	8788	8770	8770
Male	0.014 (0.020) p = 0.50	0.031 (0.022) p = 0.16	0.024 (0.024) p = 0.32	0.017 (0.024) p = 0.49	-0.016 (0.024) p = 0.50
$\bar{Y_c}$	0.69	0.55	0.41	0.40	0.34
N	4012	4012	4012	3998	3998
Female	0.017 (0.020) p = 0.40	0.038 (0.022) p = 0.08	0.018 (0.024) p = 0.46	0.003 (0.024) p = 0.89	-0.0002 (0.026) p = 0.99
$\bar{Y_c}$	0.73	0.57	0.44	0.44	0.35
N	4776	4776	4776	4772	4772

Notes: The results come from regressions of individual attendance or attainment outcomes on the school treatment indicator. To estimate standard errors, we use a standard HAC estimator and treat schools as clusters.

Appendix Table 1b
Mean Treatment vs Control Differences
in Attendance and Attainment: Without Books and With Books

Schools Without Books

	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
Both	0.021	0.009	0.020	0.011	-0.002
	(0.025)	(0.027)	(0.029)	(0.028)	(0.030)
	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
Male	0.008	-0.013	-0.003	-0.008	-0.031
	(0.029)	(0.031)	(0.032)	(0.031)	(0.032)
	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
Female	0.030	0.026	0.036	0.025	0.022
	(0.029)	(0.031)	(0.033)	(0.033)	(0.035)
	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

Schools With Books

	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
Both	0.010	0.064	0.021	0.007	-0.014
	(0.023)	(0.026)	(0.030)	(0.031)	(0.032)
	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
Male	0.020	0.078	0.051	0.043	-0.00005
	(0.028)	(0.030)	(0.035)	(0.035)	(0.035)
	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
Female	0.001	0.051	-0.006	-0.024	-0.027
	(0.027)	(0.031)	(0.036)	(0.036)	(0.038)
	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 Appendix Table 1a.

Appendix Table 2a
Mean Treatment vs Control Differences
in Achievement Gain Scores

	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
	0.033	0.019	0.064	0.020	0.020	0.046	0.032	0.004	0.062
Both	(0.029)	(0.027)	(0.036)	(0.030)	(0.029)	(0.034)	(0.038)	(0.037)	(0.047)
	p = 0.24	p = 0.47	p = 0.08	p = 0.50	p = 0.49	p = 0.18	p = 0.41	p = 0.92	p = 0.20
$\bar{Y}_c$	0.002	0.01	-0.03	0.18	0.20	0.23	-0.16	-0.15	-0.26
N	6183	6183	6183	2995	2995	2995	3188	3188	3188
	0.030	0.015	0.070	0.024	0.023	0.064	0.003	-0.025	0.031
Male	(0.036)	(0.035)	(0.044)	(0.043)	(0.043)	(0.050)	(0.048)	(0.048)	(0.058)
	p = 0.41	p = 0.68	p = 0.11	p = 0.58	p = 0.60	p = 0.20	p = 0.95	p = 0.59	p = 0.60
$\bar{Y}_c$	0.03	0.02	0.01	0.22	0.22	0.28	-0.13	-0.14	-0.21
N	2731	2731	2731	1283	1283	1283	1448	1448	1448
	0.038	0.023	0.062	0.017	0.018	0.032	0.059	0.028	0.092
Female	(0.032)	(0.031)	(0.041)	(0.036)	(0.036)	(0.041)	(0.043)	(0.042)	(0.054)
	p = 0.24	p = 0.45	p = 0.13	p = 0.65	p = 0.62	p = 0.44	p = 0.17	p = 0.50	p = 0.09
$\bar{Y}_c$	-0.02	0.01	-0.06	0.15	0.19	0.19	-0.19	-0.16	-0.31
N	3452	3452	3452	1712	1712	1712	1740	1740	1740

Notes: The results come from regressions of individual gain scores, defined as the differences between our measures of round two achievement and our round one achievement measure, on the school treatment indicator. To estimate standard errors, we use a standard HAC estimator and treat schools as clusters.

Appendix Table 2b
Mean Treatment vs Control Differences
in Achievement Gain Scores: Without Books and 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 Subtes
	-0.013	-0.004	-0.004	-0.001	0.008	0.024	-0.031	-0.022	-0.038
Both	(0.037)	(0.036)	(0.047)	(0.038)	(0.039)	(0.044)	(0.049)	(0.048)	(0.060)
	p = 0.73	p = 0.91	p = 0.94	p = 0.97	p = 0.84	p = 0.59	p = 0.53	p = 0.64	p = 0.53
$\bar{Y}_c$	0.01	0.02	-0.03	0.19	0.22	0.21	-0.15	-0.15	-0.24
N	3275	3275	3275	1536	1536	1536	1739	1739	1739
	-0.013	0.001	0.002	-0.009	-0.003	0.033	-0.039	-0.020	-0.055
Male	(0.048)	(0.047)	(0.061)	(0.059)	(0.057)	(0.072)	(0.064)	(0.061)	(0.082)
	p = 0.79	p = 0.97	p = 0.97	p = 0.87	p = 0.96	p = 0.65	p = 0.55	p = 0.74	p = 0.50
$\bar{Y}_c$	0.03	0.01	0.01	0.23	0.23	0.27	-0.13	-0.16	-0.20
N	1436	1436	1436	648	648	648	788	788	788
	-0.012	-0.009	-0.005	0.004	0.015	0.017	-0.022	-0.026	-0.019
Female	(0.042)	(0.042)	(0.052)	(0.043)	(0.048)	(0.050)	(0.055)	(0.056)	(0.064)
	p = 0.77	p = 0.83	p = 0.92	p = 0.92	p = 0.75	p = 0.74	p = 0.69	p = 0.65	p = 0.77
$\bar{Y}_c$	-0.003	0.03	-0.06	0.17	0.21	0.17	-0.16	-0.14	-0.27
N	1839	1839	1839	888	888	888	951	951	951

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 Subtes
	0.084	0.045	0.137	0.043	0.035	0.066	0.106	0.034	0.179
Both	(0.044)	(0.041)	(0.054)	(0.047)	(0.044)	(0.054)	(0.059)	(0.057)	(0.072)
	p = 0.06	p = 0.27	p = 0.01	p = 0.36	p = 0.42	p = 0.22	p = 0.08	p = 0.55	0.01
$\bar{Y}_c$	-0.01	0.01	-0.03	0.17	0.18	0.25	-0.18	-0.16	-0.30
N	2908	2908	2908	1459	1459	1459	1449	1449	1449
	0.073	0.027	0.140	0.056	0.049	0.092	0.051	-0.034	0.130
Male	(0.053)	(0.053)	(0.061)	(0.063)	(0.063)	(0.070)	(0.071)	(0.075)	(0.079)
	p = 0.17	p = 0.61	p = 0.02	p = 0.37	p = 0.44	p = 0.19	p = 0.48	p = 0.65	p = 0.10
$\bar{Y}_c$	0.03	0.03	0.01	0.22	0.20	0.29	-0.12	-0.11	-0.23
N	1295	1295	1295	635	635	635	660	660	660
	0.094	0.060	0.136	0.031	0.024	0.044	0.155	0.093	0.225
Female	(0.049)	(0.046)	(0.063)	(0.058)	(0.055)	(0.067)	(0.065)	(0.062)	(0.087)
	p = 0.06	p = 0.19	p = 0.03	p = 0.60	p = 0.66	p = 0.52	p = 0.02	p = 0.14	p = 0.01
$\bar{Y}_c$	-0.04	-0.01	-0.07	0.14	0.17	0.21	-0.23	-0.20	-0.36
N	1613	1613	1613	824	824	824	789	789	789

Notes: See Appendix Table 2a.

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

	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
	0.018	0.014	0.042	0.033	0.031
Both	(0.023)	(0.025)	(0.026)	(0.026)	(0.023)
	p=0.45	p=0.57	p=0.10	p=0.21	p=0.19
N	4373	4373	4373	4366	4366
	0.005	0.004	0.028	0.017	0.015
Male	(0.027)	(0.030)	(0.030)	(0.031)	(0.027)
	p=0.85	p=0.89	p=0.35	p=0.58	p=0.60
N	2038	2038	2038	2032	2032
	0.029	0.024	0.056	0.049	0.045
Female	(0.026)	(0.030)	(0.031)	(0.031)	(0.028)
	p= 0.27	p=0.43	p=0.07	p=0.12	p=0.11
N	2335	2335	2335	2334	2334

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

	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
Both	0.010 (0.025) p=0.69	0.073 (0.025) p<0.01	0.052 (0.028) p=0.07	0.039 (0.028) p=0.17	0.030 (0.026) p=0.26
N	4415	4415	4415	4404	4404
Male	0.021 (0.029) p=0.47	0.098 (0.030) p<0.01	0.089 (0.033) p<0.01	0.076 (0.032) p=0.02	0.056 (0.030) p=0.06
N	1974	1974	1974	1966	1966
Female	0.002 (0.029) p=0.95	0.053 (0.030) p=0.07	0.022 (0.032) p=0.49	0.009 (0.032) p=0.77	0.009 (0.031) p=0.78
N	2441	2441	2441	2438	2438

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.*

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

Barrera-Osorio, Felipe and Raju, Dhushyanth. (2017). 'Teacher Performance Pay: Experimental Evidence from Pakistan', *Journal of Public Economics* 148, 75–91.

Berry, J. and Mukherjee, P. (2016), Pricing of Private Education in Urban India: Demand, Use, and Impact, Technical report, working paper.

Bruhn, Miriam and McKenzie, David. (2008), In Pursuit of Balance, Technical report, World Bank Policy Rsearch Working Paper 4752.

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

Bugni, Federico A., Canay, Ivan A. and Shaikh, Azeem M. (2018). 'Inference under Covariate-Adaptive Randomization', *Journal of the American Statistical Association* p. forthcoming.

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.

Das, Jishnu, Dercon, Stefan, Habyarimana, James, Krishnan, Pramila, Muralidharan, Karthik and Sundararaman, Venkatesh. (2013). 'School Inputs, Household Substitution, and Test Scores', *American Economic Journal: Applied Economics* 5(2), 29–57.

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, Ilias, Nauman and Kremer, Michael. (2010). 'Teacher Incentives', *American Economic Journal: Applied Economics* 2(3), 205–227.

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

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.

Kerwin, Jason T. and Thornton, Rebecca L. (2018). 'Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures', *SSRN working paper* .

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', *Journal of Labor Economics* forthcoming.

Mbiti, Isaac, Muralidharan, Karthik, Romero, Mauricio and Schipper, Youdi. (2018), Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania, Technical report, working paper.

Mbiti, Isaac, Romero, Mauricio and Schipper, Youdi. (2018), Left Behind by Optimal Design: The Challenge of Designing Effective Teacher Performance Pay Programs, Technical report, working paper.

Muralidharan, Karthik, Singh, Abhijeet and Ganimian, Alejandro J. (2017). 'Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India', *working paper* .

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

Neal, Derek. (2018), *Information, Incentives, and Education Policy*, Harvard University Press.

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.