

---

# Educator Incentives and Educational Triage in Rural Primary Schools

---

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

## ABSTRACT

*In low-income countries, primary school student achievement is often far below grade level, and dropout rates remain high. Further, some educators actively encourage weaker students to drop out before reaching the end of primary school to avoid the negative attention that a school receives when its students perform poorly on their national primary leaving exams. We report the results of an experiment in rural Uganda that sought to both promote learning and reduce dropout rates. We offered bonus payments to Grade 6 (P6) teachers that rewarded each teacher for the math performance of each of their students relative to comparable students in other schools. This pay for percentile (PFP) incentive scheme did not improve overall P6 math performance, but it did reduce dropout rates. PFP treatment raised attendance rates a full year after treatment ended, from 0.56 to 0.60. In schools with math books, treatment increased attendance rates from 0.57 to 0.64, and PFP also improved performance on test items covered by P6 books. PFP did not improve any measure of attendance, achievement, or attainment in schools without books.*

---


*Daniel O. Gilligan and Naureen Karachiwalla are at International Food Policy Research Institute (IFPRI). Ibrahim Kasirye is at Economic Policy Research Centre (EPRC). Adrienne M. Lucas is at University of Delaware, J-PAL, CGD, and NBER. Derek Neal is at University of Chicago and NBER. The authors thank seminar participants at American University, Carnegie Mellon, the University of Chicago, the University of Wisconsin, Notre Dame, IFPRI, RISE 2018, and SREE for useful comments and suggestions. The authors thank Azeem Shaikh for guidance on corrections for multiple testing. The authors thank Maha Ashour, Ezra*


(continued on next page)

[Submitted November 2018; accepted November 2019]; doi:10.3368/jhr.57.1.1118-9871R2

JEL Classification: I0, J3, and O1

ISSN 0022-166X E-ISSN 1548-8004 © 2022 by the Board of Regents of the University of Wisconsin System

 Supplementary materials are freely available online at: <http://uwpress.wisc.edu/journals/journals/jhr-supplementary.html>

 This open access article is distributed under the terms of the CC-BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0>) and is freely available online at: <http://jhr.uwpress.org>.

Derek Neal <https://orcid.org/0000-0002-5322-0811>

## I. 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 (2018) report, primary achievement levels remain low, and primary dropout rates remain high.

Specific low-income countries may report low achievement levels and low rates of primary completion for many reasons. In some countries, schools lack resources.<sup>1</sup> In many countries, educator accountability is weak. Both Chaudhury et al. (2006) and Bruns, Filmer, and Patrinos (2011) contend that teachers in developing countries are commonly absent from school and frequently not engaged in teaching when they are present. In recent years, several other studies have reached the same conclusion.<sup>2</sup> Finally, we contend that, in a significant number of countries, policymakers focus public attention on the results of primary leaving exams in ways that create incentives for educators to encourage weak students to drop out before they reach the end of primary school.

For example, in Uganda, almost all students who complete primary seven (P7), the final grade of primary school, take the primary leaving exam (PLE). This exam is administered by the Uganda National Examinations Board (UNEB). The UNEB not only grades these exams and informs individual students about their results but also publishes the distribution of PLE scores earned by P7 students in each school. These school-level PLE reports receive considerable attention in local news coverage, and education officials often sanction administrators and teachers who work in schools where significant numbers of P7 students fail the PLE.<sup>3</sup> This system creates clear incentives for educators to urge weak students to drop out of school before they reach P7. Further, educators know that they will never be punished for engaging in this form of educational triage. Although the government collects annual data on enrollment by grade level in each school, it is not able to track movements of individual students. So, if the education ministry observes that, in a given school, enrollment in the final year of primary school, P7, is less than the reported P6 enrollment for the previous school year, officials have no way to know whether this decline in cohort size represents students

---

*Karger, and Giang Thai for excellent research assistance and Lucy Billings and Fiona Namugenyi for skilled project management. They 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. A randomized controlled trials registry entry may be found at: <https://www.socialsciencesregistry.org/trials/1152>. Data are archived in the J-PAL Dataverse <https://doi.org/10.7910/DVN/FJOL7N>.*

---

1. See Wane and Martin (2013) and World Bank (2018).

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

3. The newspaper article "Jinja headteachers demoted over PLE," *New Vision*, from February 1, 2018 records that 11 head teachers in one district lost their positions because too many students from their schools failed the PLE.

dropping out of school, transferring to schools nearby, or moving to schools in different villages far away. Education officials in Uganda cannot punish schools that encourage their weakest student to drop out because these officials are not able to measure school-level dropout rates.<sup>4</sup>

This scenario is not unique to Uganda. More than 30 African countries use leaving exams to both certify primary completion and ration access to secondary school. Few, if any, of these countries possess the student tracking systems required to create school-level measures of dropout rates, but Uganda, Kenya, and Rwanda do report annual total enrollment by grade level, and in all three countries, enrollment drops sharply between the penultimate and final levels of primary schools. This pattern is consistent with the hypothesis that educators in these countries are actively encouraging weak students to drop out before they are eligible to take the exams that certify primary completion in their countries.<sup>5</sup>

Although leaving exam results are high-stakes outcomes for educators in a number of African countries, overall educator accountability in these countries remains weak. As in many other low-income parts of the world, teacher effort levels are frequently low, and teacher absentee rates are often high.<sup>6</sup> This combination of weak overall accountability practices and intense public scrutiny of leaving exam results provides few incentives for educators to teach well and clear incentives for educators to urge their weakest students to drop out.

Here, we describe the results of a field experiment in rural Uganda that attempts to address both of these incentive problems simultaneously. The experiment involved 302 schools and roughly 9,000 students. We randomly assigned half of the schools to an assessment-based incentive system for educators that is designed to both 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 implement PFP in P6 math classes and assess its impact on math achievement growth during P6, as well as dropout rates and several measures of primary completion. PFP targets achievement growth directly by paying bonuses to teachers based on how their students' achievement growth during P6 compares to that of comparable students in other schools. PFP could also impact dropout rates through several distinct channels. Dropout rates are high for students of all achievement levels in rural Uganda, and we contend that whether students are weak or strong, students who receive more attention from their teachers should feel more welcome in school and therefore may persist longer. Further, if students who receive more attention learn more, they may conclude that the returns from persisting in school are greater. The best students may be more willing to finish primary school, take the PLE, and pursue secondary schooling. If marginal students make significant progress, school staff may promote them to P7 and give them the opportunity to

4. In the fall of 2015, we visited a school in rural Uganda that reported P7 enrollment of 51, even though the enrollment for each level from P1 through P6 was more than 100. The head teacher, that is, the school principal, told us that this pattern reflected their efforts to make sure that students from their school do not fail the PLE.

5. See [Online Appendix Section X](#) for details.

6. For example, Bold et al. (2013) conducted unannounced school visits in seven African countries. They found that 44 percent of teachers were not present in their class, and 23 percent were absent from school. In Uganda, the corresponding rates were 57 and 28 percent. See also Patrinos (2013).

take the PLE. This channel is important. Schools are not allowed to force pupils out of school, but they can encourage them to leave by refusing to promote them. Finally, if the lowest achieving students make real progress, more of them may be willing to repeat P6 and attempt to earn promotion to P7 in the following year.

We introduced PFP for one year among P6 math classes in rural Uganda. Although this treatment lasted for only one school 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 0.56 to 0.60. However, the overall achievement gains associated with PFP during P6 are small and not statistically significant, and PFP treatment in P6 did not increase the number of students who eventually passed the PLE at the end of P7.

Mbiti et al. (2019a) report the results of an experiment in Tanzania that involved random assignment of schools to three treatments. The first treatment provided cash grants to schools. These schools spent almost two-thirds of these grants on books and other instructional materials. The second treatment allowed teachers and head teachers to earn bonuses for each student who passed an exam based on the national curriculum. The third treatment enrolled schools in both the cash grant program and the incentive program. The grant program alone had no impact on student test scores. The teacher incentive program produced some improvements in student test scores. However, the combination treatment produced large significant gains in student achievement, and the authors establish that the gains associated with the combination treatment are statistically larger than the sum of the estimated treatment effects from the cash grant and teacher incentive treatments. They interpret their results as evidence that educator effort and instructional resources are complements in education production.

Our experiment did not randomize access to additional instructional materials. Thus, we cannot directly examine complementarities between any additional teacher effort that PFP may induce and the instructional resources available to teachers. However, we document several results that are consistent with the conclusions that Mbiti et al. (2019a) reach. Roughly half of the schools in our study provide P6 math books for their students. Among these schools, PFP increased the probability that P6 students would remain in their current school through the following school year by seven percentage points, from 0.57 to 0.64. Although PFP produced no significant gains in average math achievement for students with math books, more able students in PFP schools with books appear to perform better on exam items that were closely related to the content of P6 math books. Yet, in schools without books, we found no evidence that PFP improved achievement or attendance for any group of students. Further, in the absence of PFP incentives, books are not correlated with any measure of achievement, attendance, or attainment.

Since we did not randomly assign books to schools, this pattern of results provides suggestive evidence that teacher effort and instructional resources like books are complements in the production of achievement. However, in Section V, we document several additional patterns that are consistent with the hypothesis that, when students have access to books, they gain more from any improvement in teacher effort than incentives systems like PFP may produce. Further, we conduct several auxiliary analyses that produce no evidence that the presence of books in a school serves as a proxy for unmeasured aspects of the school that independently impact achievement growth or the school's capacity to improve achievement growth.

Our work adds to the growing literature on teacher incentive programs in low-income countries. Glewwe, Ilias, and Kremer (2010) report results from a teacher incentive experiment in rural Kenya that involved students in upper primary school. Muralidharan and Sundararaman (2011) report results from an experiment in rural India that involved elementary school students. Both studies found that incentives improved test scores. However, Glewwe, Ilias, and Kremer (2010) provide considerable evidence that the test score improvements they document were generated by test preparation activities that improved student familiarity with a well-established national exam but did not produce real improvements in subject mastery. In contrast, Muralidharan and Sundararaman (2011) provide evidence that teacher incentives linked to a new set of exams produced real student learning gains.

Loyalka et al. (2018) describe a teacher incentive pay experiment that assigned sixth grade classrooms in rural China to a control group and several different teacher incentive schemes, one of which was a variant of PFP. The authors found no significant 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 0.15 standard deviations.<sup>7</sup> Mbiti, Romero, and Schipper (2019b) report results from an experiment in Tanzania that involved two different performance pay schemes for teachers in Grades 1–3. In this case, both PFP and a scheme built around score thresholds improved student achievement. However, the threshold scheme produced larger gains.<sup>8</sup>

In the next section, we describe the details of the PFP incentive system and the protocol for our experiment. We then present achievement, attendance, and attainment results for the full sample. Next, we document several positive impacts of PFP treatment among student with books, and we also show that PFP treatment has no impacts on student outcomes in schools without books. We present a number of results that are consistent with the hypothesis that the presence of P6 math books in schools enhanced the effectiveness of PFP treatment, though we recognize that our design did not involve random assignment of books to schools. We conclude by discussing directions for future research.

## II. Experimental Design

The results in Barlevy and Neal (2012) imply that PFP has more desirable properties when education officials can both construct fair contests among educators and make credible commitments to measure the academic progress of all students. Below, we explain how PFP works and how we attempt to satisfy these two conditions in our implementation.

---

7. This effect is statistically different from zero but not statistically different from the smaller, insignificant treatment effects associated with the gain and level score incentive schemes.

8. Barrera-Osorio and Raju (2017) also 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 several of 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 from those used in previous experiments.

### A. How PFP Works

Assume there are  $J$  teachers in a school system, indexed by  $j = 1, 2, \dots, J$ . Each of these teachers teaches one class of  $N$  students. Let  $n = 1, 2, \dots, N$  index distinct levels of initial achievement, and assume that all classes contain exactly one student who begins the year at each of these levels.

Next, consider the following contest scheme: Collect each of the  $J$  students who share a particular level of initial achievement, for example, all with achievement rank  $n = 1$  in their class. Place all such students in a contest group or league, and for each student, calculate their within-league percentile rank in the end-of-year achievement distribution. Pay each teacher,  $j = 1, 2, \dots, J$  a bonus proportional to the within-league percentile rank of their student. Repeat this process for groups of students defined by the other  $N - 1$  baseline achievement levels. Barlevy and Neal (2012) call this scheme “pay for percentile” because the bonus for each student is proportional to the student’s end-of-year percentile rank within their league.<sup>9</sup>

Our first visits to schools occurred at the beginning of the school year. During this first round of visits, we tested all students. We told control teachers that we were conducting research on learning outcomes for students in Uganda but told them nothing about our plans for subsequent rounds of data collection. In treatment schools, we ended the Round 1 visits by informing P6 math teachers that they were going to participate in a performance pay contest. We then described how PFP works and told them that after we graded the Round 1 tests, we would form contest groups so that each student would compete against students in other schools who received the same Round 1 score. We also told treatment teachers that we were going to return at the end of the school year and conduct a second round of testing. We added that the performance of each student on this Round 2 test would determine the student’s final percentile rank in their contest group, which would then determine the bonus their teacher would receive for their performance. We stressed that these Round 2 visits would not be announced.<sup>10</sup>

Fair contests are the key feature of PFP. Teachers would have little incentive to devote effort to a particular student if they believed that this student would compete against other students who were either clearly superior or quite inferior. In the latter case, the teacher would expect that, even if they gave the student extra attention, the student would win few contests. In the former case, the teacher would expect that, regardless of their effort choice, the student would win most contests. However, if all contests are fair, teachers expect to benefit significantly from devoting extra effort to each of their students. Therefore, we repeatedly stressed to treatment teachers 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 a similar score on our Round 1 assessment.

9. Barlevy and Neal (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.

10. Our treatment involves two components: the announcement of end-of-year tests and the promise that the results of these tests will determine bonus payments. Given this design, the control group outcomes represent outcomes given existing Ugandan accountability practices, and the treatment group outcomes represent outcomes when PFP provides additional performance incentives.

To credibly promise educators that we would seed contests correctly, we created a Round 1 assessment that contains items drawn from the P1, P2, P3, P4, and P5 curricula. If instead this Round 1 test had been a standard assessment that included mainly P5 and P6 questions, more than half of the students in our sample would have likely ended up in one large contest group for students who missed every question on the Round 1 assessment. Thus, many of the implicit contests within this group would not have been fair. Some of these students would not have yet mastered P1 material, while others would have been closer to a P3 achievement level.

PFP rewards educators when their students perform better on end-of-year exams than students in other schools who are in the same contest group. This means that PFP not only requires start-of-year assessments that facilitate the creation of fair contest groups but also end-of-year assessments that produce reliable measures of the final levels of achievement that determine bonus payments. Thus, we also stressed that the Round 2 assessment would not be a standard P6 test with mostly P6 level questions. We told treatment teachers that the Round 2 assessment at the end of the school year would include items from each of the P1 through P5 curriculum guides and also some items from the P6 guide.<sup>11</sup> Without this assurance, some teachers may have rationally chosen to ignore their weakest students. Many of the students in these schools began P6 so far below grade level that heroic teacher efforts could not have prepared them to answer standard P6 questions by the end of the year.

### ***B. Sample Design***

Since the efficiency properties of PFP hinge on contestants believing that they are competing in properly seeded contests, we began by creating a sample of rural, government schools with only one P6 stream<sup>12</sup> and an expected class size within a predetermined range. In early 2016, 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 subregion 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 satisfied 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 two kilometers 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 two kilometers of any school already selected for our final sample. This process eliminated 22 schools, leaving a sample of 302 schools in 302 parishes.

---

11. We also told treatment teachers that the Round 1 assessment contained questions from the P1–P5 curricula. Yet, in an effort to avoid coaching, we did not allow P6 teachers to see either assessment, and we did not provide practice sheets or model questions.

12. “Stream” is the Ugandan term for a section.

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.<sup>13</sup> 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. Then we randomly selected three strata and 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 (Gilligan et al. 2021) 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 1 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 2 did not allow us to definitively determine whether or not the Round 1 P6 math teacher was still the P6 math teacher at the end of the school year.

### *C. Round 1*

Figure 1 is a timeline that presents the sequence of events in our study. Round 1 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.<sup>14</sup> Given these advance notices, the P6 math teacher for each school was present for our Round 1 interviews.

During these visits, we interviewed each P6 student in attendance, the P6 math teacher, and the head teacher. While one enumerator interviewed the P6 teacher and the head teacher, the other supervised the administration of our Round 1 math assessment to all P6 students who were present in each school. After the students finished their exams, we told treatment teachers that they would be participating in a performance pay contest during the coming school year. They learned that, for each student, they would receive a bonus payment of 20,000 Shillings times the student's percentile rank in their contest group, for example, the teacher of the median performer in a contest group would receive 10,000 Shillings for that student's performance.<sup>15</sup>

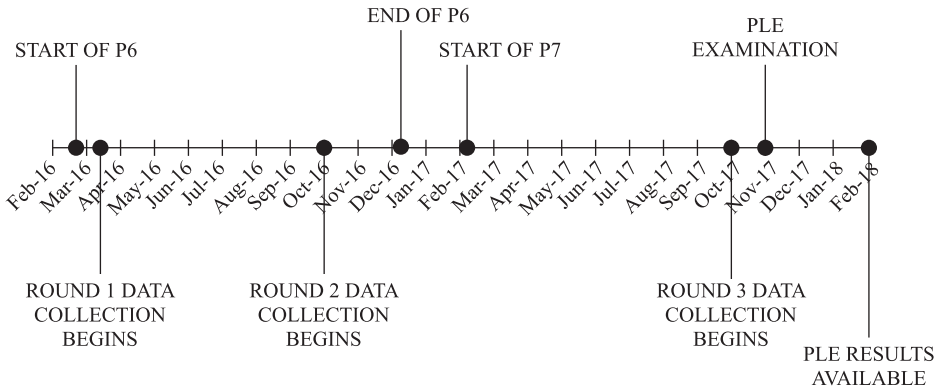
To make sure that treatment teachers understood PFP, we had each treatment teacher fill out a worksheet that asked them to calculate the bonus payments that a teacher would earn given a scenario involving the assessment outcomes of five students in a hypothetical

13. These validation visits, which were effectively "Round 0," took place about one month before Round 1 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 1 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.

14. We did not provide advance notice that we would be testing the students.

15. In March 2016, 20,000 Shillings were worth about six US dollars. We told treatment teachers that they would only earn bonus payments for the performance of students who were present and tested during these Round 2 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 2 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.





**Figure 1**  
*Timeline*

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.

On average, 30 students were present in each treatment school during Round 1. Further, we tested just over three students per school in Round 2 who claimed to be students who were absent in Round 1 but were listed on the Round 1 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 roughly 330,000 Shillings per teacher, which is about six weeks pay for a new teacher in Uganda and between two and three weeks pay for more experienced teachers.<sup>16</sup>

#### ***D. Subsequent Rounds***

In October of 2016, we returned to our 302 schools for Round 2 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. Teachers in treatment schools faced no incentives linked to any outcomes other than the Round 2 test scores. So, PFP treatment ended when these Round 2 assessments were complete.<sup>17</sup>

In October, 2017, roughly one year after PFP treatment ended, 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.

16. See <http://www.publicservice.go.ug> (accessed May 25, 2021) for salary information. A small number of teachers received no payment because they did not finish the school year.

17. We graded the Round 2 tests, and we paid PFP bonuses to treatment teachers in spring 2017, shortly after the next school year began.

**Table 1**  
*Balance Tests*

	Treatment	Control	Difference	<i>p</i> -Value	<i>N</i>
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	8,864
Enjoys school	0.876	0.857	0.019	0.33	8,864
Age	13.1	13.2	-0.086	0.14	8,864
Girl	0.551	0.534	0.017	0.22	8,864
R1 achievement	-0.048	0.048	-0.096	0.14	8,786
Cement floor	0.474	0.478	-0.004	0.86	8,864
Electricity	0.427	0.409	0.018	0.28	8,864
Radio	0.835	0.837	-0.002	0.86	8,864

Notes: 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 primary leaving exam 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. R1 denotes Round 1. 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 (2009) recommend estimating regressions with strata dummies to conduct balance tests on school-level characteristics, and Bugni, Canay, and Shaikh (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 0.01.

Students took the PLE in early November 2017. In February 2018, we obtained individual PLE results from the Uganda National Examinations Board (UNEBC) 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 UNEBC data. The PLE data [Online Appendix](#) provides more details about the matching procedure.

### *E. Balance*

Table 1 presents key descriptive statistics from Round 1 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 1 math assessment. This difference of  $-0.096$  standard deviations is not quite statistically significant, but it is academically noteworthy. Therefore, in all regression analyses of student outcomes, we include Round 1 math achievement as a control.<sup>18</sup>

### III. Academic Outcomes

We designed our experiment to examine whether or not PFP treatment in the penultimate year of primary school could simultaneously improve student achievement and reduce dropout rates. After our study began, we raised more money that allowed us to examine PLE outcomes. In Uganda, participation in the PLE marks the completion of primary school for a student. The student's PLE performance determines which, if any, secondary schools will accept them.

For both the Round 1 and Round 2 math assessments, we used a two-parameter item response theory (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.

In all of our analyses of student outcomes, we restrict attention to the sample of students tested during Round 1. 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 1 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.<sup>19</sup> Further, we use the Round 1 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 1.

Our experiment is motivated, in part, by evidence that Ugandan educators behave in ways that encourage weaker students to drop out of school before P7. These behaviors may take several different forms. A teacher may devote little attention to a weak student and encourage the student to leave school and seek a job or vocational training. A head teacher can tell a student that they must repeat P6, and a head teacher may also tell the student that they are not likely to ever move up to P7.<sup>20</sup>

18. [Online Appendix Tables 1a and 1b](#) present our key attainment results based on models that do not control for Round 1 achievement. [Online Appendix Table 2](#) presents achievement results produced by models that regress gain scores on an indicator for PFP treatment.

19. 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 Round 1 teachers were no longer teaching their Round 1 class of P6 students at Round 2. We contend that the ITT impacts are policy relevant because officials cannot mandate that teachers remain on their jobs.

20. We learned about these approaches in conversations with both head teachers and regular teachers during the field visits that we conducted while designing our study.

At all Round 1 achievement levels, significant numbers of students do not complete primary school. Although students who scored higher on our Round 1 math assessment were more likely to remain in school, enter P7, and take the PLE at the end of P7, more than one-fourth of the best P6 students do not complete P7, and many weak P6 students do.<sup>21</sup> Therefore, if PFP induces educators to devote more attention to all of their students, we expect students throughout the Round 1 achievement distribution to feel more welcome in school. Whether or not these students experience learning gains, this effect could reduce dropout rates. In addition, if strong students experience significant learning gains, they may become more interested in finishing primary school and progressing to secondary school.

We also expect some weak 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.<sup>22</sup>

Table 2 presents results from regression models that take the following form

$$y_{nj} = c + \text{treat}_j \alpha + \text{score}_{nj} \beta + \varepsilon_{nj}$$

Here,  $y_{nj}$  is an achievement, attendance or attainment outcome for student  $n = 1, 2, \dots, N_j$  who was tested during Round 1 in school  $j = 1, 2, \dots, J$ . The indicator variable  $\text{treat}_j$  equals one if school  $j$  is a treatment school and zero if it is a control school. The conditioning variable  $\text{score}_{nj}$  is the score that student  $n$  in school  $j$  earned on the Round 1 assessment, and  $\varepsilon_{nj}$  captures unobserved factors that influence measured Round 2 achievement for student  $n$  in school  $j$ . When calculating standard errors, we assume that our individual error terms,  $\varepsilon_{nj}$ , are independent across schools, but we allow an unrestricted pattern of correlation among the error terms associated with students who attend the same school.

Table 2 presents the OLS estimates,  $\hat{\alpha}$ , from six regressions. The first column presents the effect of PFP treatment on math achievement at the end of P6. The next five columns present the effects of PFP on five indicator variables that capture different aspects of student attendance or attainment.

The first column shows that PFP did not impact P6 math achievement. The estimated treatment impact is positive, but it represents a small and statistically insignificant improvement. We have no theoretical reason to believe that PFP should create different academic gains for girls versus boys, but given the significant literature on gender differences in academic outcomes among students in Africa, we also include separate

21. Among control students who score two standard deviations above the mean in Round 1, the predicted probability of completing P7 at the end of the following school year is less than 70 percent. Among those who score a full standard deviation below the mean, the corresponding rate is more than one-fourth.

22. Students who passed the PLE but earned Division 4 marks, that is, the weakest performers among those who passed, answered about 40 percent of the P4 questions correctly in the Round 2 assessment at the end of P6. Students who failed the PLE, answered one-third of these questions correctly. Both groups missed roughly 90 percent of our P5 questions, although the former performed marginally better.

**Table 2**  
*Effects of PFP on Academic Outcomes*

	Round 2 Achievement Full Test	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
Both	0.018 (0.030) $p=0.54$	0.018 (0.017) $p=0.28$	0.042 (0.018) $p=0.02$	0.034 (0.019) $p=0.08$	0.023 (0.019) $p=0.23$	0.010 (0.018) $p=0.59$
$\bar{Y}_c$	0.07	0.71	0.56	0.43	0.42	0.34
$N$	6,183	8,788	8,788	8,788	8,770	8,770
Male	0.005 (0.036) $p=0.88$	0.017 (0.020) $p=0.41$	0.041 (0.022) $p=0.06$	0.042 (0.023) $p=0.06$	0.036 (0.023) $p=0.12$	0.007 (0.021) $p=0.73$
$\bar{Y}_c$	0.15	0.69	0.55	0.41	0.40	0.34
$N$	2,731	4,012	4,012	4,012	3,998	3,998
Female	0.029 (0.034) $p=0.39$	0.018 (0.020) $p=0.36$	0.042 (0.021) $p=0.05$	0.025 (0.022) $p=0.26$	0.011 (0.022) $p=0.61$	0.010 (0.021) $p=0.64$
$\bar{Y}_c$	-0.01	0.73	0.57	0.44	0.44	0.35
$N$	3,452	4,776	4,776	4,776	4,772	4,772

Notes: The first column presents the effect of PFP treatment on end of P6 math achievement. Our achievement measure is the IRT ability parameter implied by a student's full set of answers. We have normalized this score to have mean zero and a standard deviation of one. The remaining five outcomes are indicator variables. Present Round 2 equals one if students were present on the day of our Round 2 visits. Attending Round 3 captures attendance during the day of our Round 3 visits or the four prior school days.  $\bar{Y}_c$  is the control sample mean. To estimate standard errors, we use a standard HAC estimator and treat schools as clusters. As in Table 1, to satisfy interested readers, we ran versions of these regressions that control for strata fixed effects. Since these strata are approximately orthogonal to treatment by construction, it is no surprise that we see no noteworthy changes in the estimated coefficients, and the standard errors are quite similar, although the  $p$ -value on the Round 3 attendance effect among males does fall to 0.05

estimates of PFP treatment effects for boys and girls in Table 2 and subsequent tables. In a few cases, our estimates of PFP treatment effects differ notably by gender, but none of these differences are statistically significant.<sup>23</sup>

The remaining columns document the effects of treatment on various attendance and attainment indicators. The second column records the impact on PFP treatment on the probability that a student was present on the day we returned for Round 2 testing and data collection, which occurred at the end of the first school year in our experiment. The third column reports how treatment changes the probability that students are still attending their Round 1 school in Round 3, which occurred at the end of the second school

23. See Evans and Yuan (2019) for a recent meta-analysis on gender differences in the impacts of various educational interventions.

year. Here, we count students as attenders if they are present or have been present on any of the previous four school days. The fourth 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. The final two columns deal with PLE outcomes. Column 5 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 attendance and attainment outcomes based on a student's relationship to their baseline school. When schools reported in Round 3 that a student had not attended their baseline school at all during the second year 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 3 data collection. We code them as students who are not attending in Round 3 and not participating in the PLE.<sup>24</sup>

Column 2 shows that, in both treatment and control schools, roughly 70 percent of students tested in Round 1 are present for testing in Round 2, which took place six to seven months later. Although the estimated treatment impacts in Column 2 show that attendance rates in Round 2 are roughly two percentage points higher in PFP schools, this difference is not statistically significant.<sup>25</sup>

However, we see attendance rates in treatment schools diverge more significantly from those in control schools in Round 3. In control schools, 56 percent of the students we interviewed in Round 1 were still attending their original school when we returned 18 to 19 months later to collect Round 3 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 four percentage point increase in Round 3 attendance. The *p*-values associated with the estimated impacts for the full sample, the boys sample, and the girls sample are 0.02, 0.06, and 0.05, respectively.

Only 43 percent of our Round 1 students in control schools are both present at Round 3 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 0.08. We see no significant impacts on overall PLE outcomes.

24. 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 3 that roughly 14 percent of the students we tested at baseline have transferred to another school, and in both samples, about one-half of these students were also not present for testing during Round 2. Thus, neither the prevalence nor timing of these reported transfers are correlated with treatment status.

25. This small difference in test-taking rates should have a negligible impact on our estimate of the effect of PFP on Round 2 math achievement. To confirm this conjecture, we calculate propensity scores based on the relationship between Round 1 scores and Round 2 attendance within the control sample, and then use inverse probability weighting to estimate the impact of PFP on Round 2 math achievement. Rounded to three decimal places, the estimated effect remains 0.018 standard deviations, as in Table 2. We also ran the regressions described in Table 4 below using inverse probability weighting, and we again found that this adjustment produced trivial changes in our results.

The four percentage point increase in Round 3 is significant statistically and academically, but some may worry that the single hypothesis  $p$ -value of 0.0184, which we round up to 0.02 in Table 2, overstates its statistical significance. Since we report six treatment impacts for the full sample, the Bonferroni-corrected  $p$ -value on this effect is 0.11.

Bonferroni's procedure is a conservative correction for multiple hypothesis testing. Anderson (2008) recommends a different approach. He suggests creating index values for groups of related outcomes and then estimating the impact of treatment on these index values. Our Round 2 and Round 3 attendance indicators are pure attendance measures and, therefore, form a natural group. The P7 attendance indicator is not a pure attendance measure because it captures both attendance and promotion to P7. The other three measures are primarily measures of achievement, attainment, or both.<sup>26</sup>

We formed the first principal component of our Round 2 and Round 3 attendance indicators. We then regressed this index on our treatment indicator. PFP treatment raises this index by 0.08 standard deviations, and the  $p$ -value on this effect is 0.03. If we form the first principal component of these two indicators and the indicator for attending P7 in Round 3, our estimated PFP treatment effect is again 0.08 standard deviations, with a  $p$ -value of 0.04. Table 2 provides no evidence that PFP improved overall math achievement or final educational attainment, but our PFP treatment appears to have improved attendance a full year after treatment ended.

#### IV. Heterogeneous Impacts

In Section I.B above, we note that our design assigns treatment and control status to schools within sampling blocks that we defined using data on expected class size and the expected availability of P6 math books in the school. We gathered these data during sample validation visits that we conducted roughly one month before we began Round 1 data collection.

We adopted this approach because we wanted our treatment and control samples to be balanced on features of the classroom environment that may influence the expected gains from changing various instructional practices in response to PFP. The total number of students in a classroom affects the costs and benefits of employing lectures versus group work versus one-on-one tutoring. Further, when students have P6 math books, teachers not only have more chances to let students work through problems that build their understanding, but they likely also find it easier to allow some students to work at their own pace while they give special attention to others who need it.

Ex post, we never found a way to accurately measure effective class size. Many school rosters contained the names of numerous students who were not present during our Round 1 or Round 2 visits. The official P6 rosters for these schools were often much

---

26. PLE participation marks primary completion in Uganda. Several hundred pupils took the PLE who were not attending school at the end of P7, and more than 100 students who were attending P7 in Round 3 did not take the PLE. The Round 1 P6 math scores we have for the former group are more than 0.3 standard deviations higher than those for the latter group. Students who attend P7 but know they cannot pass the PLE have little incentive to pay the costs of taking it.

larger than P6 attendance at Round 1 or Round 2, and we never found a way to accurately identify the set of students who were on the P6 roster for a given school but not actually attending the school.<sup>27</sup>

Further, we found that the relationships between PFP treatment and the outcomes examined in Table 2 did not differ by the school-level attendance counts in Round 1. In the median school in our sample, 45 pupils were present for Round 1 testing. We created results that parallel those presented in Table 2 for schools below and above this median size. In the full sample, the sample of girls, and the sample of boys, the impacts of PFP treatment on the six outcomes we examine in Table 2 are similar in large versus small schools. None of the differences are statistically significant, and these differences follow no pattern. For some outcomes, we see larger point estimates of PFP treatment effects in large classes. For other outcomes, we see the opposite.

On the other hand, the impacts of PFP treatment on Round 3 attendance do differ between the sample of schools that provide math books and the sample of schools without books. Table 3 demonstrates that, among schools without math books, PFP treatment has no discernible impact on Round 3 attendance. The overall four percentage point increase in attendance that we attribute to PFP treatment in Table 2 is driven almost entirely by outcomes in the sample of schools with books. In control schools with books, the attendance rate in Round 3 was 0.57, and Table 3 shows that the expected attendance rate among PFP schools with books is 0.072 higher overall, 0.089 greater among boys, and 0.057 greater among girls. The corresponding results for the sample of schools without books are 0.013,  $-0.005$ , and 0.028. These latter impacts are not statistically different from zero. Further, the full sample treatment effects on Round 3 attendance for schools with and without books are statistically different, given a 10 percent significance level.<sup>28</sup>

The  $p$ -value on our estimate of the impact of PFP on Round 3 attendance among students in schools with books is less than 0.01. However, concerns about multiple testing remain since we are now estimating separate treatment impacts for schools with and without books. As before, let us ignore the gender-specific results and focus on outcomes in samples that contain boys and girls. Table 3 contains 12 estimated PFP treatment effects. The Bonferroni-corrected  $p$ -value for the impact of PFP on Round 3 attendance in schools with books is 0.048.

Table 3 does not report any treatment impacts that are statistically different by gender. Yet, among boys who attend schools with math books, we do see noteworthy impacts of PFP treatment on not only Round 3 attendance but also P7 promotions rates and PLE participation rates. In schools with books, boys in treatment schools are almost seven percentage points more likely to take the PLE than their counterparts in control schools. These results for boys are interesting because, in the control sample, boys are less likely to take the PLE than girls who began P6 with comparable levels of math achievement.<sup>29</sup>

Since PFP did not generate statistically significant improvements in final P6 math achievement or PLE pass rates, even in schools with books, we must consider the

27. Ex post, we learned that our validation data were quite noisy. Expected P6 class sizes were not good predictors of actual P6 attendance on the day of our Round 1 visits, and reports that books would be available were not good predictors of books actually being present during our Round 1 visits.

28. If we restrict attention to the boys sample, we can reject equal treatment impacts in the books versus no books samples at a significance level of 0.025.

29. We provide more details about this gender difference in attainment below. See Section VI.



**Table 3**  
*Effects of PFP: Without Books and with Books*

	Round 2 Achievement Full Test	Present for Round 2	Attending Round 3	Attending P7 Round 3	Took PLE	Passed PLE
<b>Panel A: Schools without Books</b>						
Both	-0.031 (0.036) <i>p</i> =0.39	0.022 (0.025) <i>p</i> =0.39	0.013 (0.026) <i>p</i> =0.61	0.028 (0.026) <i>p</i> =0.29	0.019 (0.026) <i>p</i> =0.46	0.009 (0.024) <i>p</i> =0.70
$\bar{Y}_c$	0.10	0.70	0.56	0.41	0.41	0.33
<i>N</i>	3,275	4,703	4,703	4,703	4,695	4,695
Male	-0.041 (0.047) <i>p</i> =0.38	0.009 (0.029) <i>p</i> =0.75	-0.005 (0.030) <i>p</i> =0.86	0.011 (0.030) <i>p</i> =0.71	0.005 (0.029) <i>p</i> =0.86	-0.014 (0.026) <i>p</i> =0.60
$\bar{Y}_c$	0.18	0.69	0.55	0.40	0.40	0.33
<i>N</i>	1,436	2,125	2,125	2,125	2,118	2,118
Female	-0.022 (0.041) <i>p</i> =0.59	0.031 (0.029) <i>p</i> =0.30	0.028 (0.030) <i>p</i> =0.36	0.040 (0.031) <i>p</i> =0.20	0.029 (0.030) <i>p</i> =0.34	0.026 (0.029) <i>p</i> =0.37
$\bar{Y}_c$	0.03	0.72	0.56	0.42	0.42	0.33
<i>N</i>	1,839	2,578	2,578	2,578	2,577	2,577
<b>Panel B: Schools with Books</b>						
Both	0.072 (0.047) <i>p</i> =0.13	0.014 (0.023) <i>p</i> =0.54	0.072 (0.025) <i>p</i> <0.01	0.038 (0.028) <i>p</i> =0.17	0.026 (0.029) <i>p</i> =0.37	0.008 (0.026) <i>p</i> =0.75
$\bar{Y}_c$	0.03	0.72	0.57	0.45	0.44	0.36
<i>N</i>	2,908	4,085	4,085	4,085	4,075	4,075
Male	0.054 (0.054) <i>p</i> =0.32	0.025 (0.028) <i>p</i> =0.37	0.089 (0.029) <i>p</i> <0.01	0.074 (0.033) <i>p</i> =0.03	0.067 (0.034) <i>p</i> =0.05	0.029 (0.031) <i>p</i> =0.35
$\bar{Y}_c$	0.12	0.69	0.55	0.42	0.41	0.34
<i>N</i>	1,295	1,887	1,887	1,887	1,880	1,880
Female	0.087 (0.054) <i>p</i> =0.11	0.003 (0.027) <i>p</i> =0.91	0.057 (0.030) <i>p</i> =0.06	0.006 (0.033) <i>p</i> =0.85	-0.010 (0.033) <i>p</i> =0.75	-0.011 (0.030) <i>p</i> =0.72
$\bar{Y}_c$	-0.05	0.75	0.58	0.47	0.47	0.38
<i>N</i>	1,613	2,198	2,198	2,198	2,195	2,195

Notes: See Table 2.

possibility that the large impact of PFP treatment on Round 2 attendance rates in treatment schools with books has nothing to do with how books per se interact with PFP treatment. The presence of books in a school may proxy for some unobserved school characteristic that shapes how teachers and students respond to PFP. Further, this factor could improve attendance without producing noteworthy achievement gains. In the next section, we examine how PFP affected performance on different types of math questions among students with different baseline achievement levels. We show that while PFP treatment in schools with books did not improve overall math performance, it did improve performance on the material covered by P6 math books, especially among the students who were better prepared to use P6 books. These patterns are consistent with the hypothesis that, at least among students who are closer to grade level, PFP treatment produced learning gains when combined with books.

## V. Baseline Skills, Item Difficulty, and Books

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 1 assessment show that many pupils in rural Uganda begin P6 far below grade level.<sup>30</sup> On average, students in the bottom fourth of our Round 1 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 2 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 1 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.

For this reason, we told teachers that our assessments were designed to measure the progress made by all P6 students. We stressed that our Round 2 assessment would include items from each of the P1 through P6 curricula. The Round 1 test asked 30 questions. The Round 2 test asked 37 questions. We used IRT methods and results from pilot studies to select questions that showed significant discrimination.

We must include items that cover the entire P1–P6 curricula to implement PFP correctly, but this design feature also allows us to learn more about ways that PFP treatment may have interacted with the presence of books in P6 classrooms. The most common P6 math text in Uganda is *Primary Mathematics: Pupil's Book 6* by MK Publishers. We have compared the exercises in this text to the items on our Round 2 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

30. See World Bank (2018, p. 3–8).

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. Thus, even if one assumes that PFP improved teacher effort levels, there is little reason to believe that correlations between PFP treatment and performance on P1–P3 items should differ depending on whether or not schools provide P6 math books for students.<sup>31</sup> Yet, if the presence of books is complementary with additional teacher effort induced by PFP, we expect to see the most compelling evidence of this interaction effect when we examine performance on the P4–P6 items, which closely resemble the exercises in P6 math books.

Further, if one assumes that students must master the P1–P3 material before approaching the P4–P6 curricula, it seems reasonable to conjecture that, weaker students may have struggled to answer the P4–P6 items on our Round 2 assessment whether or not they were in PFP treatment schools and whether or not their schools provided books. Students who began the year at P1 or P2 levels of mastery needed to make tremendous progress before they could even approach the material in a standard P6 math book. Few students in any setting are able to progress three or four grade levels in one year.<sup>32</sup>

In sum, even if we suppose that PFP induces greater effort from teachers, and we also assume that teacher effort and instructional resources like books are complements in education production, there are good reasons to conjecture that the strength of this complementarity depends on the match between baseline student achievement levels and the content of the books in question. Access to P6 books should not have directly affected how PFP influenced the performance of any students on P1–P3 items since these books did not cover material from the P1–P3 curricula. Further, because weaker students needed to make substantial progress before they could approach the material covered in P6 books, it may be more than optimistic to hope that the combination of greater teacher effort and access to books could allow these weaker students to command the P4–P6 curricula. Finally, we do expect that access to books should improve the performance of more able students on test items that resemble those in P6 math books.

Table 4 contains two panels. The top panel contains P6 achievement results for the sample of schools without P6 math books. The bottom panel presents results for schools with books. These panels document the impacts of PFP treatment and three different measures of achievement. The first is the pupil-specific IRT ability parameter derived from the full Round 2 assessment. This is the achievement measure used to define the PFP achievement impacts reported in Column 1 of the two panels of Table 3. The other two achievement measures are ability parameters derived from subtests of the Round 2 assessment that contain items from the P1–P3 curricula and P4–P6 curricula, respectively. In all cases, we report impacts of PFP treatment on these three achievement measures for three different samples of the students: the full sample, students in the bottom half of the baseline achievement distribution, and students in the top half of the baseline distribution. As before, we also present gender-specific results.

---

31. One can never rule out the possibility that having books in a P6 classroom indirectly improves performance on P1–P3 items because, given books, motivated teachers are able to give more attention to students who are still attempting to master the P1–P3 curricula while more advanced students work independently on the P4–P6 material in their books.

32. [Online Appendix Table 3](#) shows that, even in PFP schools with books, lower-achieving students missed roughly five of six questions from the P4–P6 curricula.

**Table 4**  
*Effects of PFP on Achievement: Without Books and with Books*

	All Round 1 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
<b>Panel A: Schools without Books</b>									
Both	-0.031 (0.036) $p=0.39$	-0.023 (0.036) $p=0.52$	-0.029 (0.042) $p=0.49$	-0.036 (0.038) $p=0.34$	-0.026 (0.038) $p=0.49$	-0.050 (0.040) $p=0.22$	-0.026 (0.051) $p=0.61$	-0.017 (0.050) $p=0.73$	-0.033 (0.061) $p=0.59$
$\bar{Y}_c$	0.10	0.11	0.06	-0.49	-0.46	-0.47	0.61	0.60	0.51
$N$	3,275	3,275	3,275	1,536	1,536	1,536	1,739	1,739	1,739
Male	-0.041 (0.047) $p=0.38$	-0.029 (0.045) $p=0.53$	-0.038 (0.056) $p=0.51$	-0.044 (0.060) $p=0.47$	-0.037 (0.058) $p=0.52$	-0.041 (0.069) $p=0.55$	-0.036 (0.064) $p=0.58$	-0.016 (0.061) $p=0.79$	-0.052 (0.081) $p=0.52$
$\bar{Y}_c$	0.18	0.17	0.16	-0.45	-0.45	-0.41	0.67	0.64	0.60
$N$	1,436	1,436	1,436	648	648	648	788	788	788
Female	-0.022 (0.041) $p=0.59$	-0.019 (0.042) $p=0.65$	-0.020 (0.046) $p=0.67$	-0.031 (0.043) $p=0.48$	-0.019 (0.048) $p=0.69$	-0.056 (0.046) $p=0.23$	-0.016 (0.057) $p=0.78$	-0.019 (0.058) $p=0.74$	-0.012 (0.066) $p=0.86$
$\bar{Y}_c$	0.03	0.06	-0.03	-0.52	-0.48	-0.52	0.55	0.57	0.43
$N$	1,839	1,839	1,839	888	888	888	951	951	951

(continued)

Table 4 (continued)

	All Round 1 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	Full Test	P1–P3 Subtest	P4–P6 Subtest
<b>Panel B: Schools with Books</b>												
Both	0.072 (0.047)	0.032 (0.044)	0.118 (0.056)	0.033 (0.049)	0.025 (0.046)	0.042 (0.056)	0.113 (0.061)	0.044 (0.059)	0.186 (0.073)			
$\bar{Y}_c$	$p=0.13$	$p=0.47$	$p=0.04$	$p=0.50$	$p=0.59$	$p=0.45$	$p=0.07$	$p=0.47$	$p=0.01$			
$N$	2,908	2,908	2,908	1,459	1,459	1,459	1,449	1,449	1,449			
Male	0.054 (0.054)	0.005 (0.054)	0.113 (0.061)	0.047 (0.064)	0.037 (0.063)	0.071 (0.070)	0.067 (0.070)	–0.012 (0.073)	0.144 (0.079)			
$\bar{Y}_c$	$p=0.32$	$p=0.92$	$p=0.07$	$p=0.46$	$p=0.56$	$p=0.31$	$p=0.34$	$p=0.87$	$p=0.07$			
$N$	1,295	1,295	1,295	635	635	635	660	660	660			
Female	0.087 (0.054)	0.053 (0.050)	0.125 (0.066)	0.020 (0.061)	0.014 (0.057)	0.017 (0.067)	0.157 (0.067)	0.095 (0.064)	0.226 (0.088)			
$\bar{Y}_c$	$p=0.11$	$p=0.29$	$p=0.06$	$p=0.74$	$p=0.81$	$p=0.80$	$p=0.02$	$p=0.14$	$p=0.01$			
$N$	1,613	1,613	1,613	824	824	824	789	789	789			

Notes: We define the below and above R1 median subsamples relative to the median of the entire Round 1 score distribution.  $\bar{Y}_c$  is the control sample mean of the Round 2 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 2 test takers, including those who were not present for our Round 1 visit and therefore not included in our analysis samples. To estimate standard errors, we use a standard HAC estimator and treat schools as clusters.

The patterns we observe are consistent with our conjectures concerning how the presence of books could complement any improvements in teacher effort induced by PFP. The performance of more able treatment students in schools with books on test items from the P4–P6 curricula stands out, but treatment produces few if any gains elsewhere.

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 find several significant impacts of PFP treatment on measured achievement.<sup>33</sup> Students who scored above the Round 1 median score and attend a PFP school that has P6 math books earned Round 2 scores on the P4–P6 subtest that are, on average, 0.186 standard deviations higher than the scores of comparable students in control schools with books.<sup>34</sup> Even though there is little evidence that these same students performed better on the P1–P3 subtest, their performance on the P4–P6 subtest is so strong that, among students with books and higher baseline achievement, PFP treatment in schools improves scores on the full test by 0.113 standard deviations ( $p=0.07$ ). Finally, even though weaker students in PFP schools with books did not perform significantly better on the P4–P6 subtest, the performance of their more able peers on this subtest was so strong that overall scores on the P4–P6 subtest are 0.118 standard deviations higher for PFP students with books than for control students with books. These treatment effects are both noteworthy and statistically different from the parallel results among schools without books.<sup>35</sup>

The presence of achievement gains among treatment students who both have P6 books and initial achievement levels that allow them to use these books is noteworthy because the results in Table 3 indicate that students in schools with books enjoyed all of the attendance gains generated by PFP.<sup>36</sup> Nonetheless, these results come from nine different regressions that involve three different achievement measures and three different estimation samples. Thus, it is natural to ask whether or not our key finding, the 0.186 impact on the P4–P6 test among more able students, is significant given potential corrections for multiple testing. The  $p$ -value on this effect is 0.013. When we consider this result as one of nine estimated impacts for students with books, the Bonferroni correction yields a  $p$ -value of 0.117.

33. Table 4 presents results from regressions of Round 2 achievement on Round 1 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 2 achievement measures and our Round 1 measure. [Online Appendix Table 2](#) contains treatment vs control differences in mean gain scores. These results are quite similar to those presented in Table 4.

34. These students began P6 closer to grade level and with access to books. Loyalka et al. (2018) and Fryer et al. (2012) present results from PFP experiments in China and Illinois that involve payments linked to student math achievement in classes where students have books and are typically not years below grade level. The estimated treatment impacts in these studies are 0.15 and 0.185 standard deviations, respectively.

35. In the full sample and the above median Round 1 achievement sample, the  $p$ -values for tests of equal treatment impacts on the P4–P6 subtest in schools with and without books are 0.035 and 0.022, respectively. Among higher Round 1 achievers, the  $p$ -value on the test for equal treatment impact on the full test is 0.079.

36. [Online Appendix Table 3](#) shows that our results for PFP students with books are robust to using the percent of correct answers as the achievement outcome, instead of the IRT ability parameter. The pattern of results is identical to the pattern we see in the bottom panel of Table 4. Measured on this scale, more able PFP students with books scored roughly three points higher on the P4–P6 subtest than their peers in control schools. Since even the better students in control schools with books got less than 30 percent of these questions correct, this three point gain represents roughly a 10 percent improvement. PFP has no impact on percent correct for any test or subtest among students without books.

Bonferroni provides a test for the null of no average treatment impacts on any outcomes, but since we are interested in the possibility that particular PPF students performed better on a subset of items, we also performed a permutation test to assess the null that PFP treatment has no impact on the distributions of our three achievement measures. Recall that we assigned treatment within sampling strata by ranking schools according to past PLE performance and then assigning either odd or even ranked schools to treatment. This procedure implicitly assigns treatment within pairs of schools in the same strata that have similar past PLE performance. So, we created 5,000 replication samples by randomly assigning treatment at the school level within these pairs. We ran all nine regressions on each replication sample and collected the  $p$ -values associated with each of the nine treatment impact estimates. In 6.6 percent of these replications, the smallest  $p$ -value is less than 0.013. In 3.1 percent, the smallest  $p$ -value is less than 0.013, and the estimated treatment effect in question is positive.<sup>37</sup>

Students without books did not benefit from PFP. Further, among students with access to math books, the achievement 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. These results are consistent with the view that instructional materials like books are complements to any improvements in teacher effort that performance incentives may generate. However, these patterns also parallel findings in the growing literature on the value of targeting instructional resources to individual achievement levels.<sup>38</sup> PFP students who were far below grade level were likely ill-prepared to use P6 books. Further, P6 books are likely not the best tool for improving performance on P1–P3 items. Even if one accepts the hypothesis that educator effort and instructional resources are complements in education production, one should still expect stronger complementarities when the resources that students receive match their current achievement levels.

## VI. Books and Teacher Effort

We did not randomly assign books to schools. Therefore, the results in Table 4 provide only suggestive evidence that PFP improved performance in schools with books because books are complementary with the additional teacher effort that PFP elicits. In fact, we have not yet provided any evidence that PFP improved teacher effort.

In this section, we demonstrate that PFP did improve self-reported effort measures both in schools with books and in schools without books. We then document several other patterns in our data that are consistent with, but do not provide experimental support for, the hypothesis that teacher effort and books are complements in education production.

Table 5 describes the impacts of PFP on several measures of teacher effort. Each measure is derived from data collected in Round 2. The variable “Days Present” is the

37. Note that we can turn any  $p$ -value into a  $z$ -score. Given a  $p$ -value of 0.013, the  $z$ -score for an estimated positive impact is 2.226. Among our replication samples, the largest such  $z$ -score in a given sample is greater than 2.226 in 3.1 percent of these samples. The largest  $z$ -score in absolute value is greater than 2.226 in 6.6 percent of these samples.

38. See Banerjee et al. (2017) for a recent summary.

**Table 5**  
*Effects of PFP on Teacher Effort*

	Days Present	Hours Preparing	Hours Grading	Effort Index
<b>Panel A: Full Sample</b>				
	0.18 (0.12) $p=0.13$	0.10 (0.09) $p=0.30$	0.26 (0.10) $p<0.01$	0.24 (0.12) $p=0.04$
$\bar{Y}_c$	4.29	2.13	2.22	-0.12
$N$	299	299	299	299
<b>Panel B: Schools without Books</b>				
	0.18 (0.17) $p=0.28$	0.15 (0.13) $p=0.25$	0.24 (0.13) $p=0.07$	0.26 (0.16) $p=0.11$
$\bar{Y}_c$	4.22	2.04	2.16	-0.21
$N$	154	154	154	154
<b>Panel C: Schools with Books</b>				
	0.18 (0.16) $p=0.26$	0.05 (0.14) $p=0.70$	0.29 (0.14) $p=0.05$	0.22 (0.17) $p=0.18$
$\bar{Y}_c$	4.36	2.21	2.27	-0.03
$N$	145	145	145	145

Notes: Days present is the number of days in the past five schools days, including the Round 2 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 30-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.

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 principal component of the other three measures. We normalized this index to have a mean of zero and a standard deviation of one.

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.



Some may worry that self-reported effort may be inflated due to experimenter demand effects. We share this concern but also note that both treatment and control samples may be affected. In both treatment and control samples, enumerators asked about teacher effort after they introduced themselves to the school as “members of a research team studying primary education in Uganda.”<sup>39</sup>

The bottom two panels present separate results for schools with and without books. The results in these subsamples are less precisely estimated but 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. 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.

Next, given our conjecture about the complementarity between teacher effort and books, we examine correlations between the presence of books and rates of achievement growth. Table 6 reports results from regressions of Round 2 achievement on Round 1 achievement and an indicator for whether or not a student’s school had P6 math books. The top panel presents results for control schools. The bottom panel presents results for treatment schools.

Among students with similar Round 1 scores, Round 2 achievement in control schools is not significantly correlated with the presence of books. This is true for all three measures of Round 2 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.

In treatment schools, Round 2 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 4. Among treatment students, Column 3 shows that access to books is positively correlated with Round 2 achievement on the P4–P6 subtest. Further, this correlation is driven largely by outcomes among treatment students in the top half of the Round 1 achievement distribution. Among these students, scores on the P4–P6 subtest are 0.171 standard deviations higher among those with access to books. Further, among these same students, those with books score 0.118 standard deviations higher on the full test. Performance on the P1–P3 subtest is not correlated with the presence of books in treatment or control schools.

We gave the Round 1 tests at the beginning of our Round 1 visits. Thus, no student or teachers received any treatment before students took their Round 1 exams, and Round 1 achievement is uncorrelated with the presence of P6 math books in both treatment and control schools. In fact, in both treatment and control schools, Round 1 achievement is lower in schools with P6 math books, although neither of these deficits is statistically significant.<sup>40</sup>

39. We did not ask about teacher effort at baseline. So, no teachers had given a previous answer that could serve as a reference point. Our treatment teachers are not simply reporting more effort than some Round 1 reference point because they realize that PFP is designed to increase effort.

40. Schools with books do not enjoy better resources generally. At a 5 percent significance level, none of the school characteristics in our Table 2 balance tests differ significantly between schools with and without books. Given a 10 percent significance level, schools with books have class sizes that are roughly 10 percent smaller, but they are also 10 percent more likely to have a P6 teacher with low education.

**Table 6**  
*Correlations between Books and Math Achievement*

	All Round 1 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	
<b>Panel A: Control Sample</b>										
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) $p=0.46$	1,602
<i>N</i>	3,044	3,044	3,044	1,442	1,442	1,442	1,602	1,602	1,602	1,602
Male	-0.010 (0.051) $p=0.84$	0.007 (0.052) $p=0.90$	-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) $p=0.95$	0.048 (0.069) $p=0.49$	-0.035 (0.077) $p=0.65$	758
<i>N</i>	1,370	1,370	1,370	612	612	612	758	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) $p=0.43$	844
<i>N</i>	1,674	1,674	1,674	830	830	830	844	844	844	844

(continued)

Table 6 (continued)

	All Round 1 Achievement Levels				Below R1 Median				Above R1 Median			
	Full Test	P1–P3 Subtest	P4–P6 Subtest	P4–P6 Subtest	Full Test	P1–P3 Subtest	P4–P6 Subtest	P4–P6 Subtest	Full Test	P1–P3 Subtest	P4–P6 Subtest	P4–P6 Subtest
<b>Panel B: Treatment Sample</b>												
Both	0.071 (0.041) $p = 0.09$	0.029 (0.037) $p = 0.44$	0.126 (0.054) $p = 0.02$	0.080 (0.049) $p = 0.10$	0.025 (0.043) $p = 0.56$	-0.006 (0.043) $p = 0.88$	0.094 (0.072) $p = 0.19$	0.080 (0.049) $p = 0.10$	0.118 (0.054) $p = 0.03$	0.068 (0.050) $p = 0.18$	0.171 (0.071) $p = 0.02$	1,586
$N$	3,139	3,139	3,139	1,553	1,553	1,553	671	1,553	1,586	1,586	1,586	1,586
Male	0.081 (0.050) $p = 0.11$	0.038 (0.048) $p = 0.44$	0.127 (0.061) $p = 0.04$	0.094 (0.072) $p = 0.19$	0.058 (0.062) $p = 0.35$	0.031 (0.059) $p = 0.61$	0.094 (0.072) $p = 0.19$	0.094 (0.072) $p = 0.19$	0.108 (0.066) $p = 0.10$	0.051 (0.065) $p = 0.44$	0.163 (0.082) $p = 0.05$	690
$N$	1,361	1,361	1,361	671	671	671	671	671	690	690	690	690
Female	0.061 (0.048) $p = 0.21$	0.022 (0.044) $p = 0.63$	0.124 (0.062) $p = 0.05$	0.063 (0.061) $p = 0.30$	-0.003 (0.053) $p = 0.96$	-0.036 (0.054) $p = 0.51$	0.063 (0.061) $p = 0.30$	0.063 (0.061) $p = 0.30$	0.121 (0.060) $p = 0.05$	0.076 (0.054) $p = 0.16$	0.175 (0.083) $p = 0.04$	896
$N$	1,778	1,778	1,778	882	882	882	882	882	896	896	896	896

Notes: These tables contain OLS regression coefficients from student-level regressions of Round 2 achievement measures on Round 1 achievement and an indicator for the presence of P6 math books in the student's school. The top panel presents results for the control sample. The bottom panel presents results for the PFP treatment sample. 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.

PFP treatment is associated with reported increases in teacher effort that do not depend on whether or not schools provide books for P6 students. Further, Table 6 shows that, when compared to comparable treatment students without books, treatment students who both have access to P6 math books and are prepared to use them perform much better on the P4–P6 items that resemble the exercises in P6 math books. However, in control schools, books are not correlated with any measure of Round 2 achievement among any sample of students, and books are not correlated with Round 1 achievement in either treatment or control schools. All of these patterns are consistent with the hypothesis that PFP treatment is more effective when students have access to books, especially among those with initial achievement levels that match the content level of the exercises in the books. Nonetheless, since our design did not involve random assignment of books, none of the results in Tables 3, 4, or 6 provide experimental support for this hypothesis.

## VII. Comparisons of Achievement, Attendance, and Attainment Gains

Tables 2 and 3 contain our key attainment results. PFP treatment improves Round 3 attendance by 4.2 percentage points overall, and this result is driven by even larger attendance gains among students who attend treatment schools with books. We did not present separate estimates of the impacts of PFP on attendance for students who scored above versus below the median Round 1 math score, in part because the differences we observe are not statistically significant. Still, there is some evidence that PFP may have a greater impact on attendance rates among students with lower baseline achievement. Among higher-achieving students in schools without books, PFP treatment is associated with a statistically insignificant decline in Round 3 attendance rates of 1.7 percentage points ( $p=0.601$ ). Among lower-achieving students in these schools, PFP improves Round 3 attendance rates by 4.9 percentage points ( $p=0.110$ ), but this effect is not quite statistically significant. Among higher-achieving students in schools with books, PFP raises Round 3 attendance rates by 5.7 percentages, and this impact also borders on statistical significance ( $p=0.105$ ). Finally, among lower-achieving students in schools with books, PFP raises attendance rates by 8.9 percentage points ( $p=0.004$ ). This impact is highly significant, but it is not statistically different from the 5.7 percentage point impact among higher-achieving students with books.

The 8.9 percentage point improvement in the Round 3 attendance rate for lower-achieving treatment students in schools with books may puzzle some readers. Table 4 shows that PFP did not produce statistically significant achievement gains among lower-achieving students, whether or not they enjoyed access to books. How could PFP produced an almost nine percentage point improvement in attendance a full year after PFP treatment ended among a population of students who enjoyed no measurable gains in average achievement?

We have already noted that students in PFP schools may have enjoyed school more because they received more attention from their teachers. It is also possible that, in PFP schools with books, some fraction of the students who scored below the median on our Round 1 math assessment, did make noteworthy achievement gains. This possibility

does not require that, on average, lower-achieving PFP students with books enjoyed significant achievement gains. Although PFP did raise Round 3 attendance rates among these students by almost nine percentage points, more than 40 percent of these students were not attending in Round 3. Thus, in levels, dropout rates remained high for these students.

Table 3 shows that girls in treatment schools with books did not enjoy any gains in terms of promotion rates, PLE participation, or PLE performance, and this result holds even when we restrict attention to girls in treatment schools with books who score above the median on the Round 1 achievement exam.<sup>41</sup> This too may puzzle some readers because Table 4 shows that higher-achieving girls in schools with books did enjoy significant achievement gains. We hope to explore this puzzle in future research. We would like to learn more about gender differences in employment options outside school and other factors that may create gender differences in the relationships between achievement levels and dropout decisions.

For now, we note that, although boys enjoy higher average Round 1 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 Table 3. 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 1 scores, but among treatment students with books, the corresponding gender gap in PLE participation probabilities is essentially zero at  $-0.006$ . These estimated gender gaps in PLE participation are statistically different, and we find similar but not statistically significant differences when we examine gender gaps in PLE pass rates. These results suggest that, in schools with books, PFP treatment eliminates a significant female attainment advantage. These patterns warrant future study since Table 4 shows that, in these same schools, PFP generates substantial achievement gains among higher-achieving girls.

## VIII. Related Research of Education Production

We note above that Mbiti et al. (2019a) report results from an experiment involving students in Grades 1–3 in Tanzania. They found that teacher incentives were quite effective when combined with programs that increase school resources. However, they also found that increasing school resources without providing better incentives produced no learning improvements for students.<sup>42</sup>

Several other studies have found that providing books for children in rural schools often does little to improve student outcomes.<sup>43</sup> Yet, while Glewwe, Kremer, and Moulin (2009) find that providing textbooks to schools in Kenya produced no gains for students in the bottom four quintiles of the baseline achievement distribution, students in

41. For this group, the estimated treatment impacts on P7 attendance, PLE participation, and passing the PLE are 0.024, 0.014, and  $-0.0007$  respectively, and the  $p$ -values associated with these effects are all 0.6 or greater.

42. Kerwin and Thornton (2018) report that details concerning training and resource provision interact in complicated ways that impact the effectiveness of the Mango Tree literacy program.

43. See Glewwe, Kremer, and Moulin (2009) and Glewwe and Muralidharan (2016).

the top quintile did enjoy achievement gains when they received books. These high-achieving students may have been rather self-motivated, and Glewwe et al. (2009) argue that because Kenyan education policies serve the most advantaged students, schools use textbooks that match the needs of these students. In Uganda, many of the P6 students in our sample were not prepared to use standard P6 math books, and we found that, in schools with books, PFP treatment had larger impacts on achievement growth among students who were better prepared to use these books.

Banerjee et al. (2017) review 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. Muralidharan, Singh, and Ganimian (2019) report results from another experiment 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.<sup>44</sup>

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 1 achievement levels, and this may explain why PFP produced few measurable achievement gains among students who were ill-prepared to use P6 math books.

## IX. Conclusion

In Uganda, students take their primary leaving exam (PLE) at the end of P7, and evidence suggests that schools urge weak students to drop out before P7 so that these students will never be eligible to take the PLE, and no public record of their failure to learn will exist. Because the pay for percentile (PFP) incentive system we employ provides clear incentives for each teacher to direct more attention to each student in their class, we hypothesized that PFP would not only improve achievement but also reduce dropout rates. Ex post, we find that PFP treatment did not improve overall P6 math achievement. However, PFP does appear to reduce dropout rates. A full year after our P6 PFP treatment ended, 60 percent of students in our treatment sample were still attending their P6 schools compared to only 56 percent in our control sample.

Mbiti et al. (2019a) conclude based on an experiment involving elementary school students in Tanzania that the additional effort induced by teacher incentive programs and the instructional resources that teachers have in their classrooms are complements in the production of student achievement. Just less than half of the schools in our study provide math books for P6 students. Among those that do, PFP generates a seven percentage point increase in attendance rates a year after the conclusion of PFP treatment. Further,

44. 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 on grade-level items.

among schools with books, PFP treatment improves performance on test items related to the material covered in P6 math books, and this improvement is particularly noteworthy among students who were better prepared to use P6 math books, that is, those with better baseline P6 math achievement. In contrast, among schools without books, PFP treatment does not impact achievement or attainment outcomes.

These patterns do not provide direct evidence about complementarities in education production because we did not randomly assign books to schools. However, they are consistent with the view that simultaneously providing better teacher incentives and additional instructional resources tailored to each student's baseline achievement level may yield significant learning gains and promote primary completion. Well-designed teacher incentive systems should improve teacher effort, but the extent to which more teacher effort implies more student achievement and attainment hinges on many details concerning the nature of education production in classrooms and the resources available to teachers.

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 likely do not have this capacity now, but they may soon have it. During the summer of 2017, Uganda began creating a national registry of school-aged children.<sup>45</sup>

Our results likely have policy implications for many other low-income countries. More than 30 African countries use PLE systems to ration access to all levels of secondary schooling. Other countries in Africa and Asia use similar systems to ration access to some level of upper secondary schooling.<sup>46</sup> Leaving exam results are high-stakes outcomes for students, and many governments collect little data on other student outcomes. Thus, many educators likely face incentives to encourage weak students to drop out before they become eligible to take their leaving exams. Performance pay systems like PFP that direct teacher attention to all students may mitigate these triage incentives.

Finally, although our results suggest that effective teacher incentive provision combined with policies that increase student access to vital instructional resources may promote primary completion, education authorities in Uganda cannot achieve universal primary completion simply by implementing these policies in upper primary grades, for example, P5, P6, or P7. 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.<sup>47</sup> Uganda will likely not approach universal primary completion without adopting reforms that promote much more learning and much lower dropout rates in Grades P1–P5.

45. See <https://www.nira.go.ug/wp-content/uploads/Publish/Handbook.pdf> and <https://allafrica.com/stories/201705010523.html> (accessed May 25, 2021).

46. Examples include China, Ghana, and India.

47. 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.

## References

- Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103(484):1481–95.
- Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives* 31(4): 73–102.
- Barlevy, Gadi, and Derek Neal. 2012. "Pay for Percentile." *American Economic Review* 102(5): 1805–31.
- Barrera-Osorio, Felipe, and Dhushyanth Raju. 2017. "Teacher Performance Pay: Experimental Evidence from Pakistan." *Journal of Public Economics* 148:75–91.
- Berry, James, and Priya Mukherjee. 2016. "Pricing of Private Education in Urban India: Demand, Use, and Impact." Unpublished.
- Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Christophe Rockmore, Brian Stacy, Jakob Svensson, and Waly Wane. 2013. "What Do Teachers Know and Do? Does It Matter? Evidence from Primary Schools in Africa." Technical Report. Washington, DC: The World Bank.
- Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Economic Policy* 1(4): 200–232.
- Bruns, Barbara, Deon Filmer, and Harry A. Patrinos. 2011. "Making Schools Work through Accountability Reforms." In *Making Schools Work: New Evidence on Accountability Reforms*, ed. Barbara Bruns, Deon Filmer, and Harry Anthony Patrinos, 211–51. Washington, DC: World Bank.
- Bugni, Federico A., Ivan A. Canay, and Azeem M. Shaikh. 2018. "Inference under Covariate-Adaptive Randomization." *Journal of the American Statistical Association* 113(524):1784–96.
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives* 20(1):91–116.
- Evans, David K., and Fei Yuan. 2019. "What We Learn about Girls' Education from Interventions that Don't Focus on Girls." Unpublished.
- Fryer, R., S. Levitt, J. List, and S. Sadoff. 2012. "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment." NBER Working Paper 18237. Cambridge, MA: NBER.
- Gilligan, Daniel O., Naureen Karachiwalla, Ibrahim Kasirye, Adrienne M. Lucas, and Derek Neal. 2021. "Replication Data for: Educator Incentives and Educational Triage in Rural Primary Schools." Version 1, UNF:6:9T9h8aWC5AXitExRjEjZPg==[fileUNF]. Harvard Dataverse. <https://doi.org/10.7910/DVN/FJOL7N>
- Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2(3):205–27.
- Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* 1(1):112–135.
- Glewwe, Paul, and Karthik Muralidharan. 2016. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 653–744. New York: Elsevier.
- Kerwin, Jason T., and Rebecca L. Thornton. 2018. "Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures." Unpublished.



- Lazear, Edward P., and Sherwin Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89(5):841–64.
- Loyalka, Prashant, Sean Sylvia, Chengfang Liu, James Chu, and Yaojiang Shi. 2018. "Pay by Design: Teacher Performance Pay Design and the Distribution of Student Achievement." *Journal of Labor Economics* 37(3):621–62.
- Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019a. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *Quarterly Journal of Economics* 134(3):1627–73.
- Mbiti, Isaac, Mauricio Romero, and Youdi Schipper. 2019b. "Designing Effective Teacher Performance Pay Programs: Experimental Evidence from Tanzania." NBER Working Paper 25903. Cambridge, MA: NBER.
- Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109(4):1426–60.
- Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119(1):39–77.
- Patrinos, Harry A. 2013. "The Hidden Cost of Corruption: Teacher Absenteeism and Loss in Schools." In *Global Corruption Report: Education*, ed. Gareth Sweeney, Krina Despota, and Samira Lindner, 70–73. Abingdon-on-Thames, UK: Routledge.
- Wane, Waly, and Gayle H. Martin. 2013. "Education and Health Services in Uganda: Data for Results and Accountability." Technical report. Washington, DC: The World Bank.
- World Bank. 2018. *Learning to Realize Education's Promise*. World Bank Development Report. Washington, DC: World Bank.