Journal of Public Economics 167 (2018) 69–90

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Pay for locally monitored performance? A welfare analysis for teacher attendance in Ugandan primary schools

Jacobus Cilliers[a], Ibrahim Kasirye[b], Clare Leaver[c,d], Pieter Serneels[e,f], Andrew Zeitlin[a,g,*]

[a] McCourt School of Public Policy, Georgetown University, United States
[b] Economic Policy Research Centre, Uganda
[c] Blavatnik School of Government, University of Oxford, United Kingdom
[d] Centre for Economic Policy Research, United Kingdom
[e] School of International Development, University of East Anglia, United Kingdom
[f] IZA, Germany
[g] Center for Global Development, United States

ARTICLE INFO

Article history:
Received 12 March 2017
Received in revised form 27 March 2018
Accepted 18 April 2018
Available online 27 September 2018

JEL classification:
D61
H52
I25
I26
O15

Keywords:
Performance pay
Monitoring
Campbell's law
Field experiment
Education
Welfare
Uganda

ABSTRACT

To achieve the twin objectives of incentivizing agent performance and providing information for planning purposes, public sector organizations often rely on reports by local monitors that are costly to verify. Received wisdom has it that attaching financial incentives to these reports will result in collusion, and undermine both objectives. Simple bargaining logic, however, suggests the reverse: pay for locally monitored performance could incentivize desired behavior and improve information. To investigate this issue, we conducted a randomized controlled trial in Ugandan primary schools that explored how incentives for teachers could be designed when based on local monitoring by head teachers. Our experiment randomly varied whether head teachers' reports of teacher attendance were tied to teacher bonus payments or not. We find that local monitoring on its own is ineffective at improving teacher attendance. However, combining local monitoring with financial incentives leads to both an increase in teacher attendance (by 8 percentage points) and an improvement in the quality of information. We also observe substantial gains in pupil attainment, driven primarily by a reduction in dropouts. By placing a financial value on these enrollment gains, we demonstrate that pay for locally monitored performance passes both welfare and fiscal sustainability tests.

1. Introduction

Public sector organizations around the world rely on reports by local monitors that are costly to verify. Typically, these reports serve two objectives: to incentivize desired behavior, and to provide information for planning purposes. To these ends, in many education systems head teachers submit pupil enrollment and attendance figures, and schools (sometimes even pupils) receive financial transfers based on these reports. In health systems, it is common for hospital administrators to submit performance indicators, such as the number of patient visits or hospital waiting times, and for healthcare professionals to be rewarded based on these reports. Governments use such reports not only to incentivize agents but also to make policy decisions in aggregate, for example relating to facility construction, human resource transfers, the taxation of unhealthy habits, and public health campaigns.

When stakes (whether pecuniary or reputational) are attached to these reports, there is a clear risk of misreporting. Across 21 countries in Africa, head teacher over-reporting of pupil enrollment figures increased dramatically when countries introduced school funding on a per-pupil basis (Sandefur and Glassman, 2015). Veterans Affairs hospitals in the US kept patients off official waiting lists in order to meet targeted 14-day waiting times for appointments

* Corresponding author at: Department of Public Policy, Georgetown University, Old North, Suite 100, Washington, DC 20057, United States.
E-mail address: az332@georgetown.edu (A. Zeitlin).

https://doi.org/10.1016/j.jpubeco.2018.04.010
0047-2727/© 2018 Published by Elsevier B.V.

(VA Office of Inspector General, 2014). These distortions not only weaken incentives for providers, but also undermine governments' ability to plan and allocate resources effectively.

Administrative monitoring alone does not resolve these conflicts of interest. In Kenya, head teachers were asked to monitor teacher attendance and reward teachers based on these reports. Head teachers systematically overstated teacher presence and there was no improvement in teacher performance (Chen et al., 2001). Similarly, in India, teachers could reward their pupils for attending school and were found to manipulate student presence figures (Linden and Shastry, 2012). Environmental auditors, when hired by the firms they investigated, systematically understated the extent of pollution (Duflo et al., 2013). These examples point to collusion, with the local monitor lying about agent performance in return for a share of the reward.

Is collusion between local monitors and the targets of bureaucratic incentive schemes inevitable? Conventional wisdom suggests as much. Campbell's Law states that "the more any quantitative social indicator is used for decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor", and has guided much thinking on accountability in schools and other domains of public sector organization (Campbell, 1979; Rothstein, 2011; Neal, 2013).

However, Campbell's law need not always hold. Absent transaction costs, parties interested in service delivery outcomes (parents, head teacher, teaching staff and government officials, say, in an education context) can bargain to an efficient allocation of delivery effort. Side payments allow frontline agents to internalize the social benefit of service provision, alongside their private cost of effort. As others have observed (e.g. Dixit, 1996), if service delivery outcomes are inefficiently low, then transaction costs must be preventing the interested parties from bargaining effectively. Such frictions are widespread in low-income settings: financial constraints may limit the scope for transferable utility, while physical distances and/or a lack of comprehension may impede information flows and efforts to coordinate (Banerjee et al., 2010). And observable measures of effort, such as presence, are correspondingly low (Chaudhury et al., 2006). It follows that a policy that reduces these frictions, for instance by making payments based on local monitoring and thus putting transferable resources on the table, could improve the efficiency of service delivery, precisely because of—not despite—the role played by side payments. When this is the case, pay for locally monitored performance (hereafter referred to as P4LMP) may improve learning outcomes and have positive welfare and fiscal consequences.

This paper sets out to answer three related questions at the heart of P4LMP in the context of public service delivery. Can P4LMP induce improvements in service providers' behavior? Does P4LMP reduce or improve the quality of reported information for planning purposes? And what is the overall welfare and fiscal impact?

To answer these questions, we conducted a randomized controlled trial in Ugandan public primary schools, where we explored how incentives for primary school teachers can be effectively designed when based on local monitoring by head teachers. This is an important issue in Ugandan education: teacher absenteeism levels are such that pupils in rural, northern Uganda receive only 50 effective days of instruction in the entire school year (Wane and Martin, 2013).[1] Remote school locations and limited resources for inspections make local monitors a particularly important source of information on school inputs in this context.

Our experiment lasted for three school terms and varied the existence of financial stakes attached to local monitoring reports. In one treatment (20 schools), our *Info* arm, head teachers were requested to submit reports of teacher attendance using mobile technology. This information was then collated and relayed back to the community. The second treatment, our *Info & Bonus* arm (25 schools), was exactly the same, except that teachers received a bonus payment of UShs 40,000 if they were reported as present regularly over a month. This bonus payment was equivalent to 12% of an average teacher's monthly salary and was paid monthly. Another forty schools were randomly assigned to a control. We conducted our own independent spot-checks of teacher presence (both prior to the intervention and during every term that the intervention took place), which we then compared to headteacher reports. A school survey captured basic school and teacher characteristics. We also measured learning outcomes and grade attainment for a cohort of students that we tested before and after the intervention.

The key results are as follows. P4LMP improves teacher attendance but local monitoring alone does not: there is a positive and significant treatment effect on teacher attendance in the Info & Bonus arm, but not in the Info or Control arms. This translates into student enrollment gains over the period of the study. Enrollment impacts are observed across all grades, but are highest in grades where school dropouts are a serious problem. While these large compositional effects preclude tight bounds on learning impacts,[2] they are consistent with economically substantial impacts on schooling attainment. P4LMP also improves the quality of information available to district-level administrators relative to local monitoring alone: there are significantly fewer instances of unreported absence, and no more instances of absence falsely reported as presence, in the Info & Bonus arm compared to the Info arm.

We use these results to undertake a welfare analysis of moving from unincentivized to incentivized locally monitored performance, using data from a representative household survey and Uganda Revenue Authority tax receipts to estimate welfare and fiscal consequences. We place a financial value on the expected total pupil benefit from improved teacher performance in three stages. First, we calculate the impact on net enrollment, using data reported by school head teachers and data from a tracked cohort of pupils. Second, we back out gains in grade attainment implied by the enrollment figures. Third, we combine data from the Uganda National Panel Survey with estimates from the literature on the causal return to schooling to calculate the increase in the net present value (NPV) of future lifetime earnings due to higher grade attainment. We report estimates for four scenarios based on the two data sets used to calculate enrollment gains and two discount rates. Our preferred estimate is USD 1649. This figure exceeds the school-level bonus cost of USD 597, implying that there is a welfare gain from attaching bonus payments to local monitoring reports even before we consider the value of information. Since the quality of information in fact improved with the introduction of financial incentives, we conclude that it *is* welfare-enhancing to pay for locally monitored teacher attendance. We also show that moving from unincentivized to incentivized local monitoring is fiscally sustainable: the sum of the additional tax revenue per school from increased lifetime earnings, combined with the amount that government has revealed it is willing to pay for improved information, exceeds the per-school bonus cost.

We interpret these results through the lens of a theoretical model of P4LMP that illustrates how attaching incentives to third-party reports can improve teacher performance and informational outcomes. To begin, we model how the preferences of both teacher (agent) and head teacher (monitor) affect teacher attendance and head-teacher monitoring and reporting, and how these equilibrium outcomes depend on the financial stakes attached to the reports. To

[1] Comparable problems exist in schooling systems across the developing world (Chaudhury et al., 2006).

[2] Specifically, estimated Lee (2009) bounds for the impact of P4LMP on student test scores span a zero impact.

evaluate the potential trade-off between performance and quality of information, the model also considers the welfare of a bureaucracy that values teacher presence but also places a value on holding correct beliefs about teacher absence. P4LMP introduces a source of transferable utility between head teachers and teachers, who use this to bargain to locally efficient outcomes. Consistent with received wisdom, there are parameter regions where P4LMP delivers no benefit in terms of either information or teacher presence. However, we demonstrate that there are also parameter regions of positive impact, namely when the cost of attendance is intermediate and the cost of monitoring is not too high, so that it is mutually beneficial for the parties to agree on a side contract where the head teacher effectively 'pays' the teacher to attend. Here, P4LMP incentivizes desired behavior and can also provide unbiased information for planning purposes. Contrary to received wisdom, our study shows that P4LMP can improve both service delivery and the quality of information, and that this dual objective can be met sufficiently cheaply to pass both welfare and fiscal sustainability tests.

This paper contributes to three aspects of the literature on state effectiveness in poor countries. First, a number of papers have sought to understand how incentives—both pecuniary and non-pecuniary—impact the effort levels of frontline service providers. Researchers have typically collected the performance metric themselves, whether administering student assessments to measure outcomes of provider effort (Muralidharan and Sundararaman, 2011), or administering tamper-proof disposable cameras to measure teacher presence (Duflo et al., 2012). Such experiments provide proof of concept, demonstrating a necessary condition for impacts: that agents respond to the performance incentive when ideally administered. Recent attempts to extend 'automated' monitoring to the public sector have, however, proven challenging, notably in health in India (Banerjee et al., 2007; Dhaliwal and Hanna, 2014) and in education in Haiti (Adelman et al., 2015). These experiences underscore the importance of studying, as we do, how monitoring contracts and technologies interact with the preferences of local parties.

Second, there is growing interest in applying the lens of public finance to experimental and quasi-experimental evaluations of public policies in developing countries. Baird et al. (2016), for example, calculate the long-term financial gain due to improved health of children that received de-worming in Kenya. They argue that the additional tax revenue from future income alone is sufficient to pay for the program. Similar approaches have been taken in recent work on tax policy (Best et al., 2015) and unemployment benefits (Gerard and Gonzaga, 2014). Our paper speaks to this interest. In addition to cost-benefit analysis, we also consider the fiscal consequences of the intervention and show that moving from unincentivized to incentivized locally monitored performance generates sufficient additional tax revenue to be fiscally sustainable.

Third, a small but growing body of literature documents the prospects of digital technologies to improve public service delivery.[3] As with mobile money, such technologies offer opportunities to circumvent frictions that otherwise lead to market failures (Suri et al., 2012; Suri and Jack, 2016; Muralidharan et al., 2016). Callen et al. (2016) demonstrate that information collected by smartphones (in place of paper forms) on health worker absence 'crowds in' central inspections in politically competitive constituencies. Aker and Ksoll (2015) find that phone calls by government officials to local parties (teachers, community representatives, and a random sub-sample of students) improve learning outcomes of an adult education program in Niger. Our paper contributes to this hitherto empirical literature by using theory to study how digital technologies interact with the preferences of local actors to determine responses to the incentive environment.

The remainder of the paper proceeds as follows. Section 2 outlines the field experiment and data. Section 3 reports estimates of impacts on our outcomes of interest: teacher attendance, student enrollment and learning outcomes, and the quality of information available to district-level school administrators. Section 4 considers implications for welfare and fiscal sustainability. Section 5 interprets the experimental results through the lens of a theoretical model. Section 6 offers concluding remarks, including scaling up the intervention in Uganda and external validity.

2. Field experiment

2.1. Context

The study took place in 85 rural, government primary schools drawn from six different districts of Uganda: Apac, Gulu, Hoima, Iganga, Kiboga, and Mpigi. These districts span the four regions of Uganda. The first column in Table 1 shows some basic descriptive statistics of the teachers and schools in our sample. At the time of the baseline survey in July 2012, these study schools were experiencing challenges typical of education delivery in low-income countries. The teacher presence rate of 74% is comparable to previous estimates for rural, government schools in Uganda, and is also consistent with rates documented across the developing world (Chaudhury et al., 2006; Bold et al., 2017, Wane and Martin, 2013).[4] Table 1 further shows that 59% of the teachers are male, and 79% have at least completed primary school. Their average monthly salary is UShs 326,049 (roughly USD 120, or 2.3 times Uganda's per capita GDP in 2012), which is slightly lower than in most developing countries.[5] The average pupil-teacher ratio of 45 : 1 is comparable to previous estimates in Uganda (Wane and Martin, 2013) , and not much different to the average across all low-income countries of 42 : 1 (World Bank, 2017). Uganda is also similar to many developing countries, in that it has succeeded in obtaining near universal primary school enrollment (94% primary school enrollment in 2013, compared to 90% globally), yet pupils' learning trajectory in primary schools remains low (Bold et al., 2017).

2.2. Experimental design

The field experiment compared the impact of two local monitoring schemes, under which head teachers were prompted to submit daily reports of teacher attendance. The two schemes were identical except for one key feature: in the *Info & Bonus* arm, these reports triggered bonus payments for teaching staff, whereas in the *Info* arm, no such financial incentives were attached.

Both local monitoring schemes were built on a mobile-based platform developed by software engineers at the Makerere University School of Computing and Informatics Technology. The Java-based platform, accessible from low-cost phones, provided a customized form to the assigned monitors in each school, which was pre-populated with the names and unique identifiers for all teachers.

[3] As the 2016 World Development Report notes, digital technologies can help improve service delivery by: informing citizens; streamlining processes; receiving feedback; and "improving service provider management through better monitoring so that government workers both show up at work and are productive" (WDR, 2016, p. 157). Our paper is part of the literature documenting the fourth of these so-called "digital dividends ".

[4] The World Bank's *Service Delivery Indicators* for Uganda reports a teacher absenteeism rate in rural, government schools of 30% (Wane and Martin, 2013), and average teacher absence rate of 23% across six different African countries, ranging from 15% in Kenya and 45% in Mozambique (Bold et al., 2017).

[5] In 2017, primary school teachers in India, for example, earn roughly 3.2 times India's GDP per capita; in Tanzania this ratio is 3.8 : 1. As of 2017, in Uganda in the ratio for regular government teachers ranges between 2.1 and 2.5. For head teachers, the ratio ranges between 3.1 and 4.1. Note that we were only able to collect teacher salary data at endline and have it for a reduced sample of teachers.

Table 1
Balance statistics.

	(1) Mean	(2) Obs	(3) Control mean	(4) Info v Control	(5) Bonus v Control	(6) Bonus v Info
Teacher characteristics						
Teacher attendance	0.74	809	0.76	−0.04	−0.04	0.01
	(0.44)		(0.43)	(0.06)	(0.06)	(0.07)
Female teacher	0.41	809	0.42	−0.05	0.02	0.07
	(0.49)		(0.49)	(0.04)	(0.04)	(0.04)
Teacher age	34.92	786	35.36	−0.30	−1.09	−0.79
	(8.41)		(8.42)	(0.83)	(0.91)	(1.03)
Government contract	0.95	788	0.96	0.00	−0.02	−0.02
	(0.22)		(0.20)	(0.02)	(0.03)	(0.03)
Completed primary	0.79	795	0.78	−0.00	0.02	0.03
	(0.41)		(0.41)	(0.05)	(0.05)	(0.05)
Left school	0.17	809	0.17	0.01	−0.02	−0.03
	(0.38)		(0.38)	(0.04)	(0.03)	(0.04)
Years at school	4.36	804	5.12	−1.09**	−1.42***	−0.33
	(5.03)		(5.70)	(0.50)	(0.48)	(0.49)
Salary	326,077	634	333,910	−13,206	−14,600	−1394
	(146,148)		(206,083)	(15,755)	(13,952)	(9694)
School characteristics						
Pupil enrollment	538.06	85	556.25	−76.80	−20.08	56.72
	(298.85)		(295.44)	(67.31)	(71.22)	(76.10)
Pupil-teacher ratio	45.34	85	47.80	−3.72	−6.42*	−2.70
	(16.36)		(17.01)	(4.12)	(3.81)	(4.03)
Ethno linguistic frac.	0.92	79	0.95	0.20	−0.26	−0.46
	(1.22)		(1.29)	(0.43)	(0.26)	(0.41)
Exam pass rate	0.85	79	0.98	−0.21	−0.27	−0.06
	(0.91)		(1.28)	(0.21)	(0.22)	(0.10)

Note: Salary data is from the endline survey in November 2013; all the other data are from the baseline survey in July 2012. Column (1) shows mean values, and Column (2) the number of observations, for the full sample. Column (3) shows mean values for the control group only. Columns (4) to (6) report differences in means across the stated arms. Standard errors are in parentheses. For teacher characteristics, standard errors are clustered at the school level and corresponding *p*-values are calculated using the cluster wild bootstrap resampling method.
*** is significant at the 1% level.
** is significant at the 5% level.
* is significant at the 10% level.

(A comprehensive teacher list for every school was collected during baseline data collection, and updated during every round of spot-checks.) To ensure availability and installation, this platform was added to a phone that we provided, with instructions that it be kept in the administrative offices of the school.[6]

In both treatment arms, head teachers were asked to submit daily reports of the attendance of each teacher on their staff, alongside their own attendance. If the head teacher was absent, then a deputy could submit a report.[7] At the end of each month, we broadcast a summary report to school stakeholders via SMS that collated the presence of teachers on four randomly chosen days — one day from each week of the month. In the Info & Bonus arm, reported teacher attendance triggered a bonus payment of UShs 40,000 (roughly USD 15) if teachers were marked present on all four randomly selected days that month.[8] In the months with fewer than four weeks of school, the bonus payment was calculated in proportion to the number of school weeks in that month (e.g. UShs 10,000 if only one week of school in that month). Fig. A.1 shows that the median and modal monthly salary of regular government teachers in our sample was UShs 320,000 (roughly USD 120), with the majority of salaries falling between UShs 300,000 and UShs 400,000. The bonus therefore typically ranged between 10 and 13% of monthly salary.

Stratifying by district, we randomly assigned 40 schools to a control arm in which no monitoring intervention took place, 20 schools to the Info arm, and 25 schools to the Info & Bonus arm.[9] The intervention was implemented in September 2012 at the beginning of the third school term, and lasted for a year, until the end of the second school term of 2013.

2.3. Implementation

We worked with World Vision to train head teachers and their deputies in the use of this platform, and to explain its purpose to the broader school community. Training took place during September 2012. Prior to this date, District Education Officers sent a letter to each school to inform the head teacher of the training and when it would happen. The officers also called head teachers and all members of the school management committee in advance, and asked the head teacher to invite all parents and a selection of pupil representatives to the training meetings. These meetings typically took place over two afternoons. On the first afternoon, the trainers talked with school stakeholders—teachers, head teachers, members of the

6 This design feature, in combination with the norm that head teachers should directly observe presence when monitoring, is consistent with our interpretation that submitting reports is costly to the monitor, as discussed in Section 5 and the theoretical model of Appendix B.

7 Although two reports could be submitted on the same day, this happened on just 14 of the 6525 possible reporting school days (0.21 %); the second report was treated as an update and correction of the first for analytical and award purposes in these cases. Deputy head teachers took on the monitoring role in addition to their teaching responsibilities at the school.

8 To mitigate equity concerns, and because absenteeism was typically equally prevalent among all staff, head teachers and their deputies were eligible for the bonus based on their self-reported attendance. Head teacher and deputy head teacher outcomes are excluded from the main analysis of the paper.

9 An additional 95 schools were also allocated to other monitoring schemes, which are not the focus of this paper.

school management committee, parents, and pupil representatives—and explained to them the basic functioning of the program. World Vision staff also collected mobile phone numbers from every school stakeholder who declared an interest in receiving monthly updates of teacher attendance via SMS. On the second afternoon, trainers showed the head teachers and their deputies how to submit reports on the phone and also assigned them unique IDs. The monitors were asked to keep these IDs private because they were required to log onto the platform. Schools were told that the program would start in October 2012.

The intervention ran through August 2013. Monitor engagement and the content of reports submitted, which we discuss as an outcome of payment for locally monitored performance in Section 3, were consistent across terms, showing no evidence of decline in interest. In the Info & Bonus arm, accumulated bonuses were paid to teachers' bank accounts at the end of each school term. The average total payout per school over the course of this year-long implementation was USD 597. Fig. A.2 (a) shows the distribution of cumulative bonus payouts for the sample of teachers that were in the Info & Bonus treatment arm at the beginning of the study. The red line indicates that the mean payout was UShs 98,069 (USD 37). The majority of teachers in our sample, 58 %, received at least two bonus payouts, and 92% received at least one bonus payout. Four teachers received the maximum payout of UShs 290,000 (USD 109).[10] Fig. A.2 (b) shows that the probability of a day being selected to trigger the bonus payment was equal across the working week, as one would expect given our randomization.[11]

2.4. Outcomes of interest

There are two categories of outcome that could be impacted by the local monitoring schemes described in Section 2.2. The first category covers school behavior. Naturally, our hope is that local monitoring improves *teacher attendance*, and that financial incentives strengthen this effect rather than undermine it. To the extent that local monitoring improves teacher attendance, we may also see an impact on students. Having a teacher present more often should improve the *student learning outcomes* of a given cohort of students. Anticipating this, parents may be more willing to keep their children in school, thereby increasing *student enrollment*.

The second category of outcome relates to the information available to district-level school administrators. Local monitoring reports submitted by SMS could supplement, or even replace, the information collected in district-led school inspections. It seems natural to look at the frequency of reporting here, and we do include this outcome in Section 3 below. Arguably, however, district-level administrators are principally interested in the volume of reporting, in that this will provide them with more accurate beliefs. For this reason, we focus on a measure of the *quality of information* that captures both reporting frequency and accuracy. Specifically, guided by the statistical decision theory literature, we think in terms of a (Bayesian) district-level school administrator's ability to correctly predict teacher attendance. Suppose such an administrator is asked to predict whether a given teacher is present or absent on a given day. The administrator will predict the teacher to be present if, reflecting on any local monitoring report received, he/she believes this is more likely than the teacher being absent. If the teacher is indeed present this prediction is correct, and if the teacher is absent this prediction is incorrect. The theoretical model set out in Appendix B shows that an incorrect prediction can occur in the following circumstances: (i) the teacher is absent but falsely reported as present; (ii) the teacher is present, but falsely reported as absent; and (iii) the teacher is absent, but no monitoring report is filed.[12] In Section 3, we compare the rate of each of these outcomes, as well as their sum (our measure of the quality of information) across the two treatment arms.

2.5. Data collection

Our analysis draws from four sources of data: head teacher reports of teacher attendance submitted by mobile phone, our own independent spot-checks of teacher attendance, a school survey, and student scores on independently administered numeracy and literacy tests.[13] We conducted random spot-checks of teacher attendance, both before the intervention started and during every term that the intervention took place: July 2012 (pre-intervention baseline), November 2012 (Term 1), April/May 2013 (Term 2), and August 2013 (Term 3).[14] This data is at the teacher-day level: each observation is a different spot-check for a different teacher. We then matched this data set of teacher attendance with the monitoring reports for the same teacher on the same day. A school survey was conducted in July 2012 and November 2013 (post-intervention endline), providing additional information about school and teacher characteristics. Finally, we track the outcomes of a sample of 20 pupils who were in grade three in a pre-intervention assessment taken two years prior to the start of our study (hereafter the *tracked cohort*). For these students, we observe their enrollment status at endline, and conditional on enrollment, their rate of grade progression and learning outcomes.

To minimize risk of Hawthorne effects, we went to great length to ensure that data collection was independent. Field workers employed for data collection were not the same as the World Vision staff that implemented the program, and they carried identification issued by the Uganda Bureau of Statistics; spot-check visits to schools were conducted on dates unrelated to the training and implementation of the intervention; and field workers communicated clearly that their data collection was independent, not shared with government, and without consequences for school staff. We did not announce when field workers would visit schools, or that they would be conducting multiple visits, so head teachers did not know if or when to expect them.

Table 1 shows the balance of variables at baseline. The third column shows the mean values for the control group. The final three columns report differences in means across the stated arms (based on coefficients from regressing each characteristic on the set of treatment dummies, controlling for district fixed effects, clustering the standard errors at the school level and constructing p-values using the cluster wild bootstrap method). The sample is balanced across all arms for most characteristics. Importantly, there is no statistically significant difference in the teacher attendance rate or pupil enrollment figures, two key outcome variables for this paper. Statistical

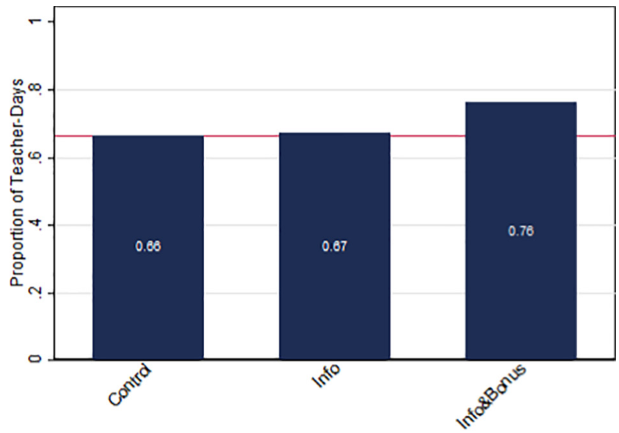
[10] In months in which schools were not open for all weeks, bonuses were computed as a weekly piece rate, reflecting the number of weeks in which the teacher qualified as present. Alternative combinations of these weekly values explain the intermediate values in the support of the distribution of teachers' total bonus payouts.

[11] This figure does not include data from the first month of the study period (November 2012). In the first month, we randomly selected a day from among the days in a week when a report was actually submitted (rather than randomly selecting a day of the week). We did this at the start of the program to build credibility.

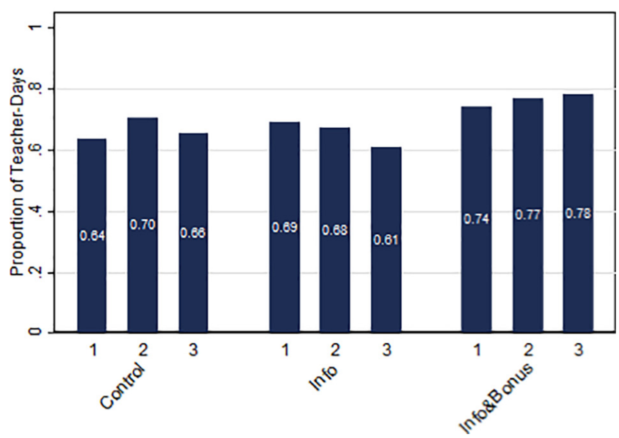
[12] Strictly speaking, our model predicts that only outcomes (i) and (iii) will occur *in equilibrium*.

[13] These tests were administered by the Uganda National Examinations Board, using papers from the preceding year's National Assessment of Progress in Education (NAPE). The NAPE is an exam sat by students in a nationally representative sample of schools for purposes of tracking learning progress in the education system as a whole; none of these schools were in our sample, so students would not have had prior sight of the questions.

[14] The school year coincides with the calendar year in Uganda. To avoid confusion, we refer to terms based on the chronology of our intervention. For example, the third term of the 2012 school year is our Term 1, etc.



(a) All terms



(b) By term of exposure

Fig. 1. Teacher attendance. *Note:* The figure is based on 3363 teacher-days with independent spot-checks.

significance is observed for three out of the 36 comparisons, which is no more than would be expected by chance.

Table 1 also shows that 17% of teachers sampled at baseline are no longer at the school at endline. The most common reasons for leaving the school are routine transfer to another school, and retirement (61 and 10% respectively). Importantly, this attrition rate is balanced across treatment arms, so differential changes in teacher composition across treatment arms are not biasing any results. Moreover, the replacement rate is constant, so the average number of teachers per school is almost exactly the same at the end of the study. We took care to update the teacher list during every round of independent spot-checks, ensuring that teachers are not incorrectly recorded as absent after they have left the school. Teachers who joined the school after the start of the program are excluded from the analysis.

2.6. Empirical strategy

To estimate impacts on *teacher attendance*, we use two specifications. The first is a simple cross-sectional comparison across all treatment arms

$$Y_{i,s,t} = \sum_{t=1}^{3} \delta_t + \gamma_1 (\text{Info})_s + \gamma_2 (\text{Info \& Bonus})_s + \rho_d + \varepsilon_{i,s,t}, \quad (1)$$

Table 2
Teacher attendance and student enrollment.

	Attendance		Enrollment	
	(1)	(2)	(3)	(4)
Info	0.0039	0.0082	−23.52	0.0469***
	(0.04)	(0.04)	(24.83)	(0.06)
Info & Bonus	0.0903**	0.0924**	46.50*	0.1383**
	(0.04)	(0.04)	(26.21)	(0.05)
Strata indicators	Yes	Yes	Yes	Yes
Baseline control	No	Yes	Yes	No
Obs	3363	3363	85	1140
R-squared	0.039	0.044	0.896	0.046
Control mean	0.665	0.665	556.3	0.344
Info mean	0.671	0.671	493.7	0.396
Info & Bonus mean	0.762	0.762	587.5	0.485
Info=Info&Bonus: *p* value	0.084	0.076	0.016	0.190

Note: The dependent variable in Columns (1) and (2) is teacher attendance measured in our independent spot-checks. Column (1) reports regression results on post-treatment data only, while Column (2) controls for baseline teacher attendance. The dependent variable in Column (3) is total student enrollment measured at school-level in our endline survey. We control for baseline student enrollment using our baseline survey data. The dependent variable in Column (4) is a binary indicator of enrollment at follow-up for a sample of pupils tracked from the pre-intervention period. The final row reports *p*-values for a test of the equality of the Info and Info & Bonus treatment arms. Standard errors are in parentheses and, except for the school-level regression in Column (3), are clustered at the school level. When standard errors are clustered, *p*-values are estimated using the wild bootstrap resampling method.
*** is significant at the 1% level.
** is significant at the 5% level.
* is significant at the 10% level.

where: $Y_{i,s,t}$ is a binary indicator of attendance for teacher i in school s in post-treatment time period t; δ_t are time dummies for each of the three rounds of post-treatment data collection; ρ_d refers to district fixed effects; $(\text{Info})_s$ and $(\text{Info \& Bonus})_s$ are the two treatment dummies; and $\varepsilon_{i,s,t}$ is an error term.[15]

Our second, preferred specification makes use of baseline data, as recommended by McKenzie (2012)

$$Y_{i,s,t} = \theta Y_{i,s,PRE} + \sum_{t=1}^{3} \delta_t + \gamma_1 (\text{Info})_s + \gamma_2 (\text{Info \& Bonus})_s + \rho_d + \varepsilon_{i,s,t}, \quad (2)$$

where $Y_{i,s,PRE}$ is baseline attendance for teacher i in school s. In both specifications, we pool treatment impacts across post-treatment rounds of data collection. Robustness checks confirming the absence of time-varying treatment effects on teacher attendance are reported in Section 3.2.1.

To estimate impacts on *student enrollment*, we use the following specifications

$$Y_s = \theta Y_{s,PRE} + \gamma_1 (\text{Info})_s + \gamma_2 (\text{Info \& Bonus})_s + \rho_d + \varepsilon_s \quad (3)$$

$$Y_{i,s} = \gamma_1 (\text{Info})_s + \gamma_2 (\text{Info \& Bonus})_s + \rho_d + \varepsilon_{i,s}, \quad (4)$$

where: in the first school-level regression Y_s and $Y_{s,PRE}$ are counts of total enrollment in school s at endline and baseline respectively; in the second pupil-level regression $Y_{i,s}$ is a binary indicator of enrollment at endline for tracked cohort pupil i in school s; and all other independent variables are defined as above. For inferential purposes, we allow the error terms from Eqs. (1), (2) and C.1 to be arbitrarily correlated within schools. Given the small number of clusters in the

[15] Since we stratified our sample at district level, in the tables below we refer to the presence of strata indicators rather than district fixed effects.

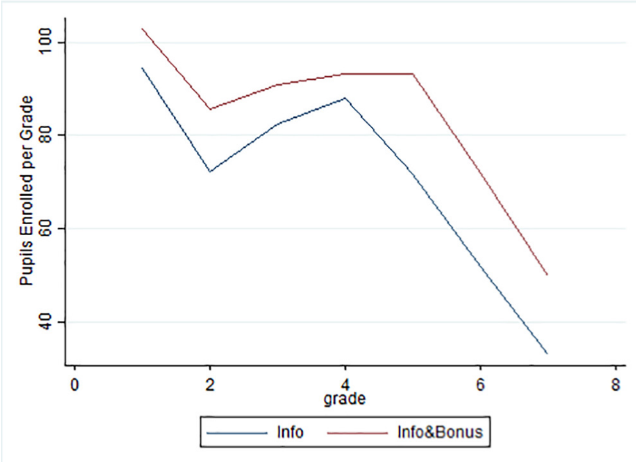


Fig. 2. Average enrollment at endline, by grade. *Note:* The figure plots the average number of pupils enrolled per grade in Info and Info & Bonus schools, as reported in the endline survey conducted in November 2013.

study, we further estimate *p*-values using the cluster wild bootstrap method (Cameron et al., 2008).

Turning to *student learning outcomes*, the analysis is complicated by differential sample selection across experimental arms. There is less drop out in the Info & Bonus arm, and data from the tracked cohort indicate that students who dropped out performed worse (although not statistically significantly so) in both numeracy and literacy tests sat prior to the intervention, relative to their peers who remained enrolled. Because the bias that arises from potentially non-random sample selection that differs across study arms cannot be signed a priori, we take a Lee Bounds approach (Lee, 2009) and examine test score levels at endline for a trimmed sub-sample of the tracked cohort. For any pairwise comparison of treatment arms, this approach places bounds on the treatment effect experienced by the subset of students who would have remained in the sample under either treatment condition. To calculate the lower bound, this estimator drops the best-performing pupils from the group with the lower attrition rate (here, the Info & Bonus arm) such that the attrition rate in each experimental arm is equal. For the upper bound, the sample selection assumption is reversed, and the estimator drops the worst-performing pupils from the group with the lower attrition rate. Unconditional means (of endline test scores for students remaining in the sample) are then compared across experimental arms.

Finally, when examining impacts on the *quality of information*, we estimate five different models using the specification

$$Y_{i,s,t} = \sum_{t=1}^{3} \delta_t + \gamma_2(\text{Info \& Bonus})_s + \rho_d + \varepsilon_{i,s,t}, \tag{5}$$

where $Y_{i,s,t}$ is a binary variable for teacher i in school s in post-treatment time period t that is, in turn, coded to 1 if: (1) a local monitoring report was submitted; (2) the teacher was absent and a report was submitted indicating he/she was present; (3) the teacher was present and a report was submitted indicating he/she was absent; (4) the teacher was absent and no report was submitted; and (5) any of events (2) to (4) occurred.

3. Experimental impacts

3.1. Results

In this subsection, we report results for our main outcomes of interest: teacher attendance, student enrollment, student learning, and quality of information. To summarize, both teacher attendance and the quality of information improved with the introduction of bonus payments. Due to substantial increases in enrollment in the Info & Bonus arm, we cannot make any definitive claims on impacts on student learning, due to the possibility of non-random student attrition.

3.1.1. Teacher attendance

Fig. 1 (a) shows that teacher attendance increased when financial incentives were attached to local monitoring. On the days when we conducted independent spot-checks, teachers were 9 and 10 percentage points more likely to be present in the Info & Bonus schools compared to Info and Control schools respectively. Table 2 Column (1) confirms that the difference between Info & Bonus and Control schools is statistically significant at the 5% level, and (in the final row) that the difference between Info & Bonus and Info schools is statistically significant at the 10% level. Teacher attendance was not significantly higher in Info schools relative to Control schools.

3.1.2. Student enrollment

Fig. 2 plots average enrollment by grade in Info and Info & Bonus schools, as reported in the endline survey. Two facts stand out. First, in both treatment arms there is a downward trend in enrollment. This is consistent with the prevailing view that school dropouts are a serious concern in Uganda.[16] Second, at each grade, average enrollment is higher in Info & Bonus schools relative to Info schools, suggesting that paying for locally monitored performance may have been more successful at averting dropouts.

Table 2 verifies that the enrollment gain (or rather reduced loss) in Info & Bonus schools is statistically significant. Column (3) reports results from estimating the school-level model in Eq. (3) using our baseline and endline survey data.[17] Schools in the Info & Bonus arm report on average 47 more pupils enrolled across all grades compared with Control schools (8% increase), and 70 more pupils compared to Info schools (13% increase). This finding is corroborated in Table 2 Column (4), which reports results from estimating the pupil-level model in Eq. (4) using data for a cohort of 20 pupils surveyed in 2010 as part of a previous study and representative of those enrolled in Primary 3. We tracked the enrollment outcomes of these children during our endline survey in November 2013. In the Control schools, only 34% of these children were still enrolled in the same school three years later. In Info & Bonus schools, the percentage of the tracked cohort still enrolled in 2013 was 14 percentage points higher than in Control schools, and 9 percentage points higher than in Info schools. The similarity of results across the two different data sets is reassuring and suggests that the enrollment impacts are due to the introduction of financial incentives.

3.1.3. Student learning outcomes

Table 3 shows results from constructing Lee Bounds on student learning outcomes, comparing unconditional means of the school-level change in test scores calculated using the trimmed sub-samples. The odd-numbered columns show results for the literacy test, and the even-numbered columns the results for numeracy. The first two columns compare the difference in learning outcomes between the Info & Bonus arm and the Control schools; the next two columns compare Info and Info & Bonus schools; and the final two

[16] Official records indicate that only 30% of pupils nationwide enrolled in Grade 1 make it to Grade 7 (Ministry of Education and Sports, 2014:121). In our Control schools, the number of pupils in Grade 7 is on average 40% of the number of pupils in Grade 1.

[17] Baseline data are missing in two schools due to enumerator error, prompting us to use EMIS 2012 data. We feel confident doing this because our enrollment figures correspond closely to the EMIS data. In fact, the 2013 figures were exactly the same for the two schools with absent 2012 data. Results hold when we drop those two schools.

Table 3
Lee bounds on student learning outcomes.

	Info & Bonus vs Control		Info & Bonus vs Info		Info vs Control	
	(1)	(2)	(3)	(4)	(5)	(6)
	Literacy	Numeracy	Literacy	Numeracy	Literacy	Numeracy
Lower	−7.519***	−5.860***	−3.890	−4.424	−4.607*	−2.623
	(2.02)	(1.67)	(2.89)	(2.83)	(2.70)	(2.19)
Upper	9.637***	9.499***	8.936***	6.875**	0.436	1.982
	(3.09)	(2.91)	(3.15)	(2.89)	(2.50)	(2.14)
Total Obs	860	860	620	620	800	800
Selected Obs	282	282	224	224	236	236
Ratio	0.288	0.288	0.199	0.199	0.111	0.111

Note: Each column reports a separate regression of pupil-level learning on treatment assignment using the Lee Bounds estimator for the tracked cohort of pupils. The odd-numbered columns report results from literacy tests and the even-numbered columns report results from numeracy tests. Standard errors are in parentheses.
*** is significant at the 1% level.
** is significant at the 5% level.
* is significant at the 10% level.

Table 4
Quality of information.

	(1)	(2)	(3)	(4)	(5)
	Report submitted	Absent, rep. present	Present, rep. absent	Unreported absence	Any of (2)–(4)
Info & Bonus	0.1780**	0.0006	0.0209	−0.0883***	−0.0668*
	(0.07)	(0.02)	(0.01)	(0.03)	(0.04)
Strata indicators	Yes	Yes	Yes	Yes	Yes
R-squared	0.186	0.014	0.034	0.073	0.024
Obs	1854	1854	1854	1854	1854
Info mean	0.582	0.076	0.028	0.158	0.262
Info & Bonus mean	0.743	0.073	0.049	0.076	0.198

Note: The data are based on all independent spot-checks of teacher attendance, matched with local monitoring reports that were submitted for the same teacher on the same day. Each observation is for a different teacher-day. The binary dependent variables are coded to 1 if: (1) a local monitoring report was submitted; (2) the teacher was absent and a report was submitted indicating he/she was present; (3) the teacher was present and a report was submitted indicating he/she was absent; (4) the teacher was absent and no report was submitted; and (5) any of events (2) to (4) occurred. Standard errors are in parentheses and clustered at the school level. *p*-Values are estimated using the cluster wild bootstrap resampling method.
*** is significant at the 1% level.
** is significant at the 5% level.
* is significant at the 10% level.

columns compare Info and Control schools. The first and third row indicate the lower and upper bounds respectively. The third-last row shows the total number of observations for the full sample, including those who dropped out (i.e. 20 pupils per school); the second-last row shows the number of pupils still at the school; and the final row shows the proportion of non-missing observations dropped from the treatment arm with higher retention. It is clear from Table 3 that large differentials in rates of retention across treatment arms invalidate any assessment of learning outcomes. Depending on the assumption relating to sample selection, one could infer that the Info & Bonus arm had a statistically significant positive or negative impact on learning gains.[18]

3.1.4. Quality of information

We begin by comparing the frequency of reporting across the two treatment arms. Table 4 Column (1) shows results from estimating Eq. (5) using our data on teacher days with independent spot-checks, and shows that the availability of financial incentives increased reporting frequency.[19] The probability of a local monitor submitting a report was 18 percentage points higher in the Info & Bonus arm compared to the Info arm, a difference that is statistically significant at the 5% level.

To capture the accuracy as well as the frequency of reporting, we compare our spot-checks of teacher attendance with the local monitoring report for the same teacher on the same day (if one was submitted). Fig. 3 shows this graphically. The dark red part of the bars shows the rate of absence falsely reported as presence. Consistent with common intuition, this source of mistake occurs in the Info & Bonus arm (on 8% of teacher-days with independent spot-checks) but, interestingly, also in the Info arm (on 7% of teacher-days with independent spot-checks). The grey shading shows the rate of presence falsely reported as absence. In both treatment arms this outcome is rare, and we strongly suspect it is due to measurement error.[20] Table 4 Columns (2) and (3) confirm that there is no statistically significant difference in either form of false reporting across the two treatment arms. The light blue shading shows the rate of unreported absence. This number is 9 percentage points higher in Info

[18] The range is largest in Table 3 Columns (1) and (2). This is because attrition was largest in the Control arm, so a larger proportion of observations in the Info & Bonus group need to be dropped.

[19] We use the sub-sample of teacher days on which we conducted independent spot-checks for consistency across columns. Table 5 below shows results for the full sample of days.

[20] There are two possibilities here. First, head teachers may have made reporting errors, either hitting the wrong button for a given teacher, or filing their report on the wrong day. We did our best to clean the latter (e.g. reports on weekends or public holidays) but cannot rule out the former. Second, there may have been late arrivals, i.e. teachers who arrived between the time that the head teacher monitored and our independent enumerators visited the school. Our field teams endeavored to arrive at the start of the school day, but this was not always possible.

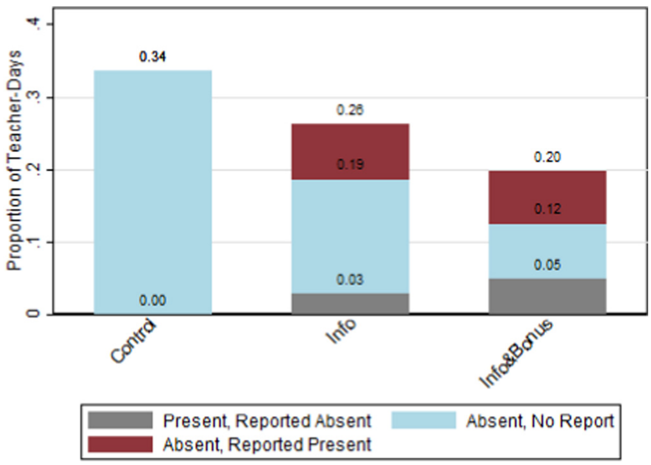


Fig. 3. Quality of information. *Note:* The figure is based on 3363 teacher-days with independent spot-checks. The overall height of the bar shows the rate at which a (Bayesian) district-level school administrator (the 'bureaucracy' in our theoretical framework) would make an incorrect prediction of teacher attendance, given local monitoring. It follows that the quality of information is highest in the Info & Bonus arm and lowest in the Control arm.

schools relative to Info & Bonus schools (16% versus 7 %); a difference that is statistically significant at the 1% level as shown in Table 4 Column (4). The overall height of the bars in Fig. 3 depicts our measure of the quality of information — the probability with which a (Bayesian) district-level school administrator would make an incorrect prediction of teacher attendance. This rate is 6.7 percentage points higher in the Info arm than in the Info & Bonus arm (and, unsurprisingly, substantially higher still in the Control arm). Since Table 4 Column (5) confirms that the difference between treatment arms is statistically significant at the 10% level, we conclude that P4LMP generated higher quality information relative to local monitoring alone.

3.2. Robustness checks

In this subsection, we perform three robustness checks: we test for dynamic impacts on teacher attendance, we look at reporting behavior throughout the period of the intervention including days on which independent spot-checks did not take place, and we employ two different strategies to test for Hawthorne effects. All these checks yield results that are consistent with our main estimates.

3.2.1. Does teacher attendance change over time?

One threat to the policy implications of our study is that teachers and their local monitors may change behavior over time, such that the effects reported above might not be expected to persist. To address this, we allow for a more general specification where the treatment effect on teacher attendance can vary across terms. Specifically, we estimate the following equation

$$Y_{i,s,t} = \theta Y_{i,s,PRE} + \sum_{t=1}^{3} \delta_t + \sum_{t=1}^{3} \gamma_{1,t}\,(\text{Info}_{i,s,t}) + \sum_{t=1}^{3} \gamma_{2,t}\,(\text{Info \& Bonus}_{i,s,t}) + \rho_d + \varepsilon_{i,s,t} \quad (6)$$

where $\text{Info}_{i,s,t} = \text{Info}_{i,s} \times \delta_t$. The treatment effects $\gamma_{1,t}$ and $\gamma_{2,t}$ are thus allowed to vary across time. We cannot reject the null hypotheses that the coefficients on each treatment arm remain the same over time; in other words, we cannot reject the null of joint equality: $\gamma_{1,1} = \gamma_{1,2} = \gamma_{1,3}$ ($p = 0.33$); and $\gamma_{2,1} = \gamma_{2,2} = \gamma_{2,3}$ ($p = 0.98$). When comparing only the Info and Info & Bonus treatment arms, we estimate

$$Y_{i,s,t} = \theta Y_{i,s,PRE} + \sum_{t=1}^{3} \delta_t + \sum_{t=1}^{3} \gamma_{2,t}\,(\text{Info \& Bonus}_{i,s,t}) + \rho_d + \varepsilon_{i,s,t}. \quad (7)$$

Again, we cannot reject the null of joint equality: $\gamma_{2,1} = \gamma_{2,2} = \gamma_{2,3}$ ($p = 0.96$). Fig. 2 (b) depicts teacher attendance by treatment arm and term of exposure, and shows graphically that impacts are remarkably stable across the duration of the program. At least during the period of our study, we saw no evidence of evolution in reporting or teacher behavior that would suggest a threat to sustainability of impacts.

Table 5
Reporting behavior using all potential-report days.

	Report submitted
Info & Bonus	0.0978**
	(0.05)
Strata indicators	Yes
Date fixed effects	Yes
Obs	6525
Info mean	0.369
Info & Bonus mean	0.466

Note: The data are based on a sample expanded from Table 4 to include all days on which a report could have been submitted during the intervention. The dependent variable indicates whether at least one monitoring report was submitted on a given day for a given school or not. Standard errors are in parentheses and clustered at the school level; *p*-values are estimated using the cluster wild bootstrap resampling method.
** is significant at the 5% level.

3.2.2. Are results based on spot-check days representative?

Since our independent spot-checks of teacher attendance are collected only during specific weeks of each term, one might be concerned that behavior during these periods is somehow unrepresentative. For instance, proximity to holiday periods or exam dates might affect teacher attendance. For the sake of consistency, all of the results in Section 3.1 are based on teacher days with independent spot-checks, even when the outcome under consideration depends only on reports generated by the intervention. Although by construction we cannot estimate actual teacher attendance outside of spot-check periods, we can test if the impacts on reporting behavior hold over the whole duration of the program. Table 5 expands the sample to include all days on which a report could have been submitted. The dependent variable indicates whether at least one monitoring report was submitted on a given day for a given school or not. The coefficient shows that, on average over the duration of the program, reporting was 10 percentage points higher in Info & Bonus schools than in Info schools. This confirms that the result based only on spot check days is robust to using the full sample of reports from the entire duration of the program.

3.2.3. Are the experimental impacts due to Hawthorne effects?

Even though we took steps to maintain independence between the program and the unannounced spot-checks (as discussed in Section 2), one might still be concerned that our measurement activities had a direct impact upon teacher attendance. If such a Hawthorne effect varied between treatment arms (for example, if teachers in the Info & Bonus arm were more responsive to our visits because they mistakenly believed that spot checks had an implication for their bonus payments), then this would necessitate a different interpretation of results. In this subsection we discuss two

Table 6
Testing for Hawthorne effects on teacher attendance.

	Period 2		Periods 2 and 3	
	(1)	(2)	(3)	(4)
Info	−0.0324	−0.0486***	−0.0378	0.0411**
	(0.05)	(0.15)	(0.05)	(0.13)
Info & Bonus	0.0619	0.0191	0.0836*	0.125
	(0.04)	(0.13)	(0.04)	(0.11)
No. spotchecks in period 1		0.0309		0.0528
		(0.05)		(0.04)
Info × no. spotchecks in period 1		0.007		−0.0407
		(0.07)		(0.06)
Info & Bonus × no. spotchecks in period 1		0.0292		−0.0174
		(0.07)		(0.06)
Strata indicators	Yes	Yes	Yes	Yes
Baseline control	Yes	Yes	Yes	Yes
Period fixed effects	No	No	Yes	Yes
Obs	1273	1273	1924	1924
Info = Info & Bonus: p value	0.080	0.687	0.012	0.521

Note: The data are based on a sub-sample from Table 2, excluding the first term of intervention. The dependent variable in all columns is teacher attendance. Column (1) reports results for the specification in Table 2 based on Eq. (2) using data from the second term only. Column (2) includes the number of spot-checks during the first term (1, 2 or 3 visits per school) and the interaction between this variable and treatment status. Columns (3) and (4) repeat this exercise for the pooled sample of the second and third terms. Standard errors are in parentheses and clustered at the school level; p-values are estimated using the cluster wild bootstrap resampling method.
*** is significant at the 1% level.
** is significant at the 5% level.
* is significant at the 10% level.

different tests that we use to rule out any evidence of Hawthorne effects.

First, in anticipation of the need to test for Hawthorne effects, we randomly varied the frequency of measurement (1, 2 or 3 visits) in our first round of spot-checks. If our spot-checks did induce a behavioral response that is correlated with treatment, then we would expect a stronger impact in schools that received more visits. Table 6 reports results from a regression of subsequent teacher attendance on the number of visits during the first round of spot-checks, together with an interaction term between treatment status and the number of spot-checks. The first two columns report results for the second term only, since a Hawthorne effect, if one exists, should be larger immediately after the measurement activity. Column (1) reports results for the specification in Table 2 based on Eq. (2) and gives average impacts for the second term. Column (2) adds in the number of spot-checks and an interaction between treatment status and spot-checks. We observe that subsequent teacher attendance is no different for schools that received more/fewer visits in the first term. Moreover, the coefficients on the interaction terms are small and statistically insignificant at conventional levels; that is, there is no evidence of a stronger treatment impact in schools that received more visits. To increase statistical power, the final two columns report results for the second and third terms combined. Even in this expanded sample, the coefficients on the interaction terms remain insignificant.[21]

Second, even though by construction it is impossible to test if actual teacher attendance differed across spot-check and non-spot-check days, we can test if reporting was any different. Table 7 shows that the treatment impact of introducing teacher bonus payments on reporting behavior (both the probability of submitting a report and reporting a teacher as present) was not different on days when we conducted an independent spot-check. Column (1) reports data at the school-day level, restricted to the dates when we were conducting independent spot-checks in at least one school.[22] The dependent variable indicates whether at least one monitoring report was submitted on a given day for a given school or not. Column (2) reports data at the teacher-day level. Here, the dependent variable indicates whether the teacher was reported present on that day or not (either because no report took place, or because the teacher was reported absent). For both columns, we regress the outcome on the Info & Bonus treatment dummy, a dummy indicating whether we conducted an independent spot-check at the school on the given day or not, and the interaction between the two variables. The coefficients on the treatment dummy show that there was a statistically significant impact on the days when we did not conduct spot-checks. Reporting behavior on average was no different on days when we conducted spot-checks, compared to days when we did not. The coefficients and standard errors on the interaction term mean that we cannot reject the null hypothesis that the impact of treatment was the same on spot-check and non-spot-check days. Once again, we find no evidence that our measurement activities in any way changed the impact of the program.

4. Welfare analysis

Section 3 delivered estimates of impacts on teacher attendance, student enrollment, student learning outcomes, and the quality of information. Although these results established that attaching financial incentives to local monitoring reports was effective at increasing teacher attendance and student enrollment, and actually improved the quality of information available to district-level school administrators, this does not answer the welfare question: what should a policymaker do? That answer depends not only on the magnitude of behavioral responses to the intervention, but also the cost of bonus payments—in our experiment, an average of USD 597 per school—

[21] We repeated this analysis for the five dependent variables related to the quality of information in Table 4. Reassuringly, we also found that subsequent informational outcomes are no different for schools that received more/fewer visits in the first term. These results are available upon request.

[22] We restrict the sample to the spot-check period because we only want to pick up differences that are due to our school visits, rather than different times of the year. As noted above, Table 5 confirms that our results hold when looking at the whole duration of the program. We define a spot-check period as days when an independent spot-check was conducted for at least one school.

Table 7
Testing for Hawthorne effects on reporting behavior.

	(1)	(2)
	Submited report	Reported presence
Info & Bonus	0.115**	0.169***
	(0.05)	(0.05)
Spotcheck day	−0.0538	0.0273
	(0.04)	(0.04)
Info&Bonus x spotcheck day	0.0706	0.000784
	(0.05)	(0.05)
Strata indicators	Yes	Yes
Date fixed effects	Yes	Yes
Observations	1350	13,624

Note: The data are based on a sample expanded from Table 4 to include all days on which a report could have been submitted during the period when we were conducting spot-checks. Column (1) reports data at the school-day level, for every day during the spot-check period. The dependent variable indicates whether at least one monitoring report was submitted on a given day for a given school or not. Column (2) reports data at the teacher-day level. Here, the dependent variable indicates whether the teacher was reported present on that day or not. Standard errors are in parentheses and clustered at the school level; *p*-values are calculated using the cluster wild bootstrap resampling method.
*** is significant at the 1% level.
** is significant at the 5% level.

and on the social values placed on school behavior and the quality of information.

We focus on what we consider to be the more interesting welfare comparison: moving from an unincentivized to an incentivized local monitoring scheme, i.e., from Info to Info & Bonus. We quantify the expected total pupil benefit from the introduction of bonus payments in two steps. First, we back out gains in *grade attainment* moving from Info to Info & Bonus implied by the enrollment figures reported in Table 2. Second, we combine data from the Uganda National Panel Survey with estimates from the literature on the causal return to schooling to calculate the increase in *NPV of future lifetime earnings* due to this higher grade attainment, valued in USD. These two steps are summarized in Sections 4.1 and 4.2 .[23]

We answer the welfare question in Section 4.3. Specifically, we calculate the average gain in NPV per school of future lifetime earnings due to higher grade attainment, minus the average bonus cost per school for four scenarios (based on different datasets and discount rates). We find that, in the most reasonable scenario, this sum is positive. Since the quality of information was also higher under Info & Bonus than Info, we conclude that it is welfare-enhancing to pay for locally monitored teacher attendance. As an extension, we also consider whether moving from Info to Info & Bonus is *fiscally sustainable*, in the sense that the additional tax take per school from the increased lifetime earnings exceeds the bonus cost per school . We report estimates for eight scenarios (based on different datasets, discount rates, and tax evasion rates) and find that in the most conservative scenario the additional tax take per school falls just short of the bonus cost per school. After accounting for how much government values higher quality of information (as revealed by its current spending on school inspections), however, we conclude that it is also fiscally sustainable to pay for locally monitored performance. Below, we summarize the analysis underlying these conclusions, relegating a detailed discussion to Appendix B. For convenience and clarity, inputs into and outputs from each of these steps are summarized in Table 8.

4.1. Moving from enrollment to grade attainment

In Section 3, we reported that P4LMP increased student enrollment compared to monitoring alone. However, this finding does not necessarily imply a causal impact on grade attainment, as grade repetition and inbound transfers from other schools are also possibilities. To model the impact of increased enrollment on grade attainment, we back out the portion of the enrollment gain that is due to 'averted dropouts' for each grade, rather than grade repeaters, using two separate empirical strategies (see Appendix B.1). Our first approach combines our survey data on enrollment with administrative data on grade repetition. Our second approach derives annual dropout and repetition rates from differences in reduced-form dropout and repetition probabilities observed for the tracked sample surveyed in 2010 and 2013. These estimates are reported in the fourth and sixth columns of Table 9 respectively. For example, using the tracked sample of students (our cohort data) we estimate that on average 12.56 fewer pupils per school dropped out in Grade 1, 9.76 fewer pupils in Grade 2, and so on. Results are qualitatively similar from the two approaches, despite their differences in identifying assumptions, as discussed further in Appendix C.

To translate averted dropouts into impacts on grade attainment, we conservatively assume that each averted dropout progresses only one additional grade before dropping out. To the extent that the returns to additional years of schooling, net of the cost of delayed labor-market entry, are positive for all students who would have dropped out prior to Grade 7 in the absence of P4LMP, this assumption provides a lower bound. Other assumptions are possible: one could, for example, assume that averted dropouts go on to follow the attainment profile typical of students observed to complete at least one more year of schooling in the absence of treatment.

Table 8
Schematic of welfare comparison and fiscal sustainability.

Step	Inputs	Output
1. Attainment impacts		
1.	Reduced-form enrollment impacts (administrative and cohort data)	Averted dropouts
	Repetition rates (administrative and cohort data)	
2. NPV earnings impacts		
2.a.	Earnings data in wage sector (UNHS)	NPV_s^0
	Earnings data in agric sector (UNHS)	
	Probability of wage employment (UNHS)	
2.b.	NPV_s^0 (Step 2a)	NPV_s^1
	Causal return to schooling Duflo (2001)	
2.c.	Averted dropouts (Step 1)	Impacts on NPV lifetime earnings
	NPV_s^0, NPV_s^1 (Step 2a, 2b)	
3. Fiscal sustainability		
3.a.	PAYE 2014	Tax compliance rate for wage employees
	Wage employees 2014 (WDI)	
3.b.	Averted dropouts (Step 1)	NPV tax revenue
	NPV_s^0, NPV_s^1 (Step 2)	
	Tax compliance rate (Step 3a)	

Note: UNHS refers to Uganda National Panel Survey 2011/12 (Uganda Bureau of Statistics, 2012); WDI 2014 refers to World Development Indicators for Uganda (World Bank, 2017). As defined in Section 4.2, NPV_s^0 and NPV_s^1 refer to the series of potential net present values of expected earnings—without and with an additional year of schooling causally induced by P4LMP, respectively—for students indexed by the grade $s \in \{1, \ldots, 6\}$in which they *would have* left school in the absence of P4LMP. PAYE 2014 refers to the number of wage employees found in the Pay As You Earn tax filings for 2014, as held by the Uganda Revenue Authority; and WDI refers to the World Bank's World Development Indicators dataset, which reports the number of wage employees in Uganda for the year 2014 (World Bank, 2017).

[23] We do not incorporate our results on student learning outcomes into the welfare analysis for three reasons. First, as discussed in Section 3.1, the differential enrollment induced by the intervention means that we cannot credibly causally identify impacts on learning. Second, placing a contextually relevant economic value on learning gains is difficult given the paucity of studies relating these to labor market outcomes in developing countries: studies of the relationship between cognitive skills and earnings tend to focus on OECD economies (Hanushek et al., 2015). Third, given a set of estimates of grade attainment impacts, a general problem is that these partly embody learning gains, creating a risk of double-counting benefits.

Table 9
Calculating the average difference between per school Info and Info & Bonus in the gain in NPV of future lifetime earnings due to treatment.

			Administrative data		Cohort data	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Grade	NPV	Gain	Averted dropouts	Total gain	Averted dropouts	Total gain
1	6602.69	−11.59	14.02	−162.50	12.56	−145.58
2	6733.43	4.18	7.09	29.62	9.76	40.77
3	6878.53	20.31	5.86	118.98	10.62	215.74
4	7038.55	36.77	27.64	1016.44	10.80	397.16
5	7214.03	53.50	12.31	658.57	10.44	558.53
6	7405.44	70.42	7.12	501.19	8.27	582.38
Totals				2162.30		1649.00

Note: This table gives the underlying calculations for one set of parameter assumptions, and the two data sets, reported in Table 6. Column (2) shows the NPV of future lifetime earnings, given each grade attainment, assuming a discount rate of 3.5%. Column (3) shows the *gain* in NPV due to an additional year of schooling achieved, assuming a causal impact of 6.8% and 1% respectively on wage earnings and probability of formal employment. Columns (4) and (6) indicate treatment effects—the average number of averted dropouts per grade per school due to the program—calculated using the two different data sets. Columns (5) and (7) show the average financial gain per grade per school.

Because we do not know how unobserved correlates of dropout for students completing at least one more year relate to the characteristics of those causally induced to obtain an additional year of school, further assumptions are required to place an upper bound on these impacts. In particular, we would require post-primary grade attainment data outside of our sample. We do not report such estimates here to focus attention on a lower bound in which we have greater confidence. As it turns out, this lower bound will be sufficient to guide the policy decision on both welfare and fiscal criteria.

4.2. Moving from grade attainment to earnings

The next step in the welfare analysis is to place a financial value on the increase in grade attainment due to averted dropouts. As set out in Appendix B.2, we use a simple NPV model based on the following assumptions. There are two sectors: formal wage employment where wages evolve with years of experience, and subsistence agriculture where earnings are constant, both over the lifetime and with respect to education. All pupils start school aged 7, do not repeat a grade, obtain no more than Grade 7,[24] and leave formal employment aged 60. Grade attainment has a causal effect on both the probability of formal sector employment and the associated formal wage, but not on earnings from subsistence agriculture. Given these assumptions, for each grade $s = 1, \ldots, 6$, we write down two *potential* NPVs: an outcome with *s* years of schooling, denoted NPV_s^0, and an outcome with a causally induced additional year of schooling, for an individual who would otherwise have obtained *s* years of schooling, denoted NPV_s^1. Since the observational earnings profile $\{NPV_1^0, NPV_2^0, \ldots\}$ embodies both selection and treatment effects for each level of schooling, it will not be the case in general that $NPV_s^1 = NPV_{s+1}^0$. The difference between NPV_s^1 and NPV_s^0 is the grade-specific NPV gain associated with being induced to attain an additional year of schooling by moving from Info to Info & Bonus. Since an additional year of schooling delays but uplifts expected future earnings, this net gain could in principle be positive or negative.

To calculate these grade-specific NPVs, we require a variety of numbers: a discount rate, two dimensions of the causal impact of schooling (an effect on the probability of formal sector employment and a rate of return on the formal sector wage), two time series (formal sector employment probabilities and wages), and constant agricultural earnings. As discussed in Appendix B.2, we take the first three from the prior literature, and estimate the latter three using data from the 2011/12 Uganda National Panel Survey. Table 9 summarizes the calculation for one configuration of parameters, specifically a discount rate of 3.5 %, causal effects of schooling of 1% and 6.8% on the probability of formal sector employment and the formal sector wage respectively, time series of formal sector employment probabilities and wages predicted from Table C.2, and an agricultural wage of USD 228. Table 9 Column (2) reports our estimate of the grade-specific outcome NPV_s^0, and Column (3) our estimate of the gain $NPV_s^1 - NPV_s^0$.[25] Column (5) reports the total NPV gain for each grade, multiplying by the number of averted dropouts estimated (as described in Appendix B.1) using our administrative data; Column (7) does likewise for our cohort data. Summing over grades we therefore arrive at two estimates of the NPV gain at *school-level*. Each estimate puts a USD financial value on the average increase in grade attainment per school due to treatment under Info & Bonus rather than under Info alone.

4.3. Welfare comparison and fiscal sustainability

To make a statement about welfare, we compare the average gain in NPV per school of future lifetime earnings due to higher grade attainment (calculated as set out in Sections 4.1 and 4.2) with the average bonus cost per school of USD 597. The first row in Table 10 reports the average NPV earnings gain per school for four scenarios, based on the two datasets used to calculate enrollment gains and two discount rates. Our preferred estimate is USD 1649, shown in the third cell. We choose a discount rate of 3.5% because this is generally viewed as the appropriate social time preference rate of discount, and we feel that the social time preference method (as opposed to the social opportunity cost method) is appropriate for the question of *welfare*.[26] We focus on the estimate based on the cohort data because this is more conservative. Since USD 1649 exceeds the average bonus cost per school of USD 597, it follows that there is a welfare gain from attaching bonus payments to local monitoring reports even before we consider the value of information.[27] With the

[24] Because we do not have data to allow estimates of attainment impacts beyond Grade 7 (our study sample comprises primary schools only), we effectively assume that P4LMP never induces students who would not otherwise do so to obtain more than seven years of education. Consequently, assumptions about the NPV of earnings for levels of schooling above Grade 7 do not affect our results.

[25] Even though we assume constant returns to education, the estimated gain is not simply a proportional increase in the NPV. This is because the lifetime evolution of wages and expected remaining working years varies depending on when a student drops out of school. Note further that the proportional increase is far smaller (between 1 and 2 %) than the return to wage earnings. This is because few in our sample have access to formal employment.

[26] For a discussion of alternative discount rates, see Appendix B.2.

[27] Although there are differences across cells in Table 10, the internal rates of return (computed putting no value on gains in the quality of information) are similar (3.80% for the admin data and 3.74% for the cohort data), and both exceed the STP rate of discount.

Table 10
Gain per school gain in NPV of future lifetime earnings, and tax revenue, from Info & Bonus relative to Info treatment.

	Administrative data		Cohort data	
	3.5%	5%	3.5%	5%
NPV lifetime earnings	2162.30	−4372.20	1649.00	−3806.43
NPV tax revenue, no evasion	5552.01	3721.56	4630.37	3105.51
NPV tax revenue, 16.12% compliance	894.98	599.91	746.42	500.61

Note: All cells assume a causal effect of one additional year of schooling on the probability of formal sector employment (formal sector wage earnings) of 1% (respectively 6.8%). The first row reports the average gain per-school in NPV of future lifetime earnings moving from the Info to the Info & Bonus treatment, using either the administrative data or cohort data to estimate enrollment gains, and assuming either a discount rate of 3.5% or 5%. Our preferred estimate is USD 1649, which exceeds the average per-school bonus cost of USD 597. The internal rates of return for the administrative and cohort data are 3.80% and 3.74 respectively. The final two rows report the average per-school gain in tax revenue, based on the earnings gain in the first row, and assuming either a 100 or 16.12% tax compliance rate. Our preferred estimate is USD 500.61, which is less than the average per-school bonus cost. Hence, moving from Info to Info & Bonus is fiscally sustainable only if the financial value attached to improved information is sufficiently high. The internal rate of return for NPV tax revenue with 16.12% compliance is 5.02% using the administrative data and 4.33% using the cohort data.

quality of information estimated to be higher under Info & Bonus than Info, we therefore conclude that it *is* welfare-enhancing to pay for locally monitored teacher attendance.

It is also of interest to consider whether moving from Info to Info & Bonus is fiscally sustainable, in the sense that the NPV of the additional tax revenue per school from the increased lifetime formal sector earnings exceeds the bonus cost per school.[28] The final two rows of Table 10 report the average per-school NPV tax gain for eight scenarios, based on the two datasets used to calculate enrollment gains, two discount rates, and two tax evasion rates. Our preferred estimate is USD 501, shown in the final cell. As discussed in Appendix, we choose the higher discount rate of 5% to reflect the cost of government borrowing in the Ugandan context, and feel that this social opportunity cost consideration is appropriate for the question of *fiscal sustainability*. We focus on the cohort data and the higher rate of tax evasion because this is more conservative. Since USD 501 falls short of the average bonus cost per school by USD 96, we must also place a financial value on the higher quality of information. To do so, recall from Table 4 Column (5) that our estimate of the impact on the quality of information was −0.07. For P4LMP to be fiscally sustainable, the minimum value the government must place on making a correct prediction is therefore USD 96/0.07 ≈ USD 1371 per school or, since schools are open 180 days per year, just under USD 8 per school day. The Government of Uganda has stated that it aims to spend no more than UShs 150,000 (or roughly USD 56) per school inspection (Ministry of Education, Science, Technology and Sports, 2014).[29] If we take USD 56 as the Government's true (per school day) valuation of information, then it follows that P4LMP is also fiscally sustainable.

Although our focus in this paper is on the welfare and fiscal consequences of *paying* for locally monitored performance—i.e. layering bonus payments on top of an existing monitoring system—it is of interest to consider the overall cost of implementation. Factoring in phone purchases, registration of handsets, phone charging and airtime, two-day field visits by World Vision staff, and a contract with Makerere University School of Computing and Informatics Technology, our estimate of the cost of implementation is USD 533 per study school.[30] The total cost of P4LMP, including both implementation of the mobile monitoring system and payment of bonuses, is therefore USD 597 + 533 = 1130 per school. Since this is lower than our preferred, conservative estimate of the average NPV earnings gain per school, it follows that introducing *and* paying for locally monitored performance also passes a welfare cost-benefit test. This point becomes even more evident when considering long-run costs of implementation at scale: the steady-state cost of a nationwide scheme is an order of magnitude lower at just USD 56 per school.[31]

5. Discussion

There are two headline results in Section 3:

1. P4LMP *improves teacher attendance* but local monitoring alone does not — there is a positive and significant treatment effect on teacher attendance in the Info & Bonus arm, but not in the Info or Control arms;
2. P4LMP *improves the quality of information* available to district-level administrators relative to local monitoring alone: there are significantly fewer instances of unreported absence, and no more instances of absence falsely reported as presence, in the Info & Bonus arm compared to the Info arm.

Below we sketch a simple theoretical framework, set out more fully in Appendix B, that aids the interpretation of these results.[32]

5.1. Theoretical framework

The economy consists of a teacher (he), a head teacher (she), and a government bureaucracy (it). Pupils play no active role. In all arms, the teacher chooses whether to attend school at cost C^T. Attendance is valued as ε^H by the head teacher. In the Control arm, the head teacher plays no active role. In both treatment arms, the head teacher chooses whether to monitor at cost C^H, and then whether to submit a truthful report to the bureaucracy. If reported absent, the teacher incurs a cost, which we can think of as 'shame' δ. In the Info & Bonus arm, the bureaucracy pays the teacher a cash bonus of β if reported present by the head teacher. The costs C^T and C^H are observed by the head teacher and teacher but not by the bureaucracy. Realizations of these random variables are assumed to be drawn independently from uniform distributions, with lower and upper supports denoted by lower and upper bars respectively. The bonus β is the only source of transferable utility. All players are risk neutral. Payoffs are stated formally in the Appendix. Fig. 4 represents the predictions from theory graphically.

5.1.1. Teacher attendance

The dark blue regions in Fig. 4 Panel (a) plot realizations of the monitoring and attendance costs where the teacher chooses to

[28] Details of the underlying tax calculation are provided in Appendix B.3.

[29] It is questionable whether this target has actually been met. A 2008 audit documenting the frequency of inspections and the overall inspection budget suggests a cost closer to USD 190 per visit (Ministry of Education, Science, Technology and Sports, 2010).

[30] During the field visits, World Vision staff introduced the program, delivered the phones, and trained head teachers. The contract with Makerere University was for platform development and ongoing technical support.

[31] This estimate is based on the following assumptions: 10% of phones are replaced every year; a quarter of schools receive two-day field visits every year; and the same annual contract remains in place with Makerere University. Note that, here, we amortize platform development costs over schools but not over time. This approach is conservative but seems reasonable, since continued platform development and maintenance will be required to keep pace with technology and changes in school staffing.

[32] This theory was developed before the field experiment to guide the design of a contemporaneous "multiple monitors" treatment that is reported in a companion paper. We do not emphasize testing of the model in the current paper, and simply use it to offer an intuitive interpretation of our results.

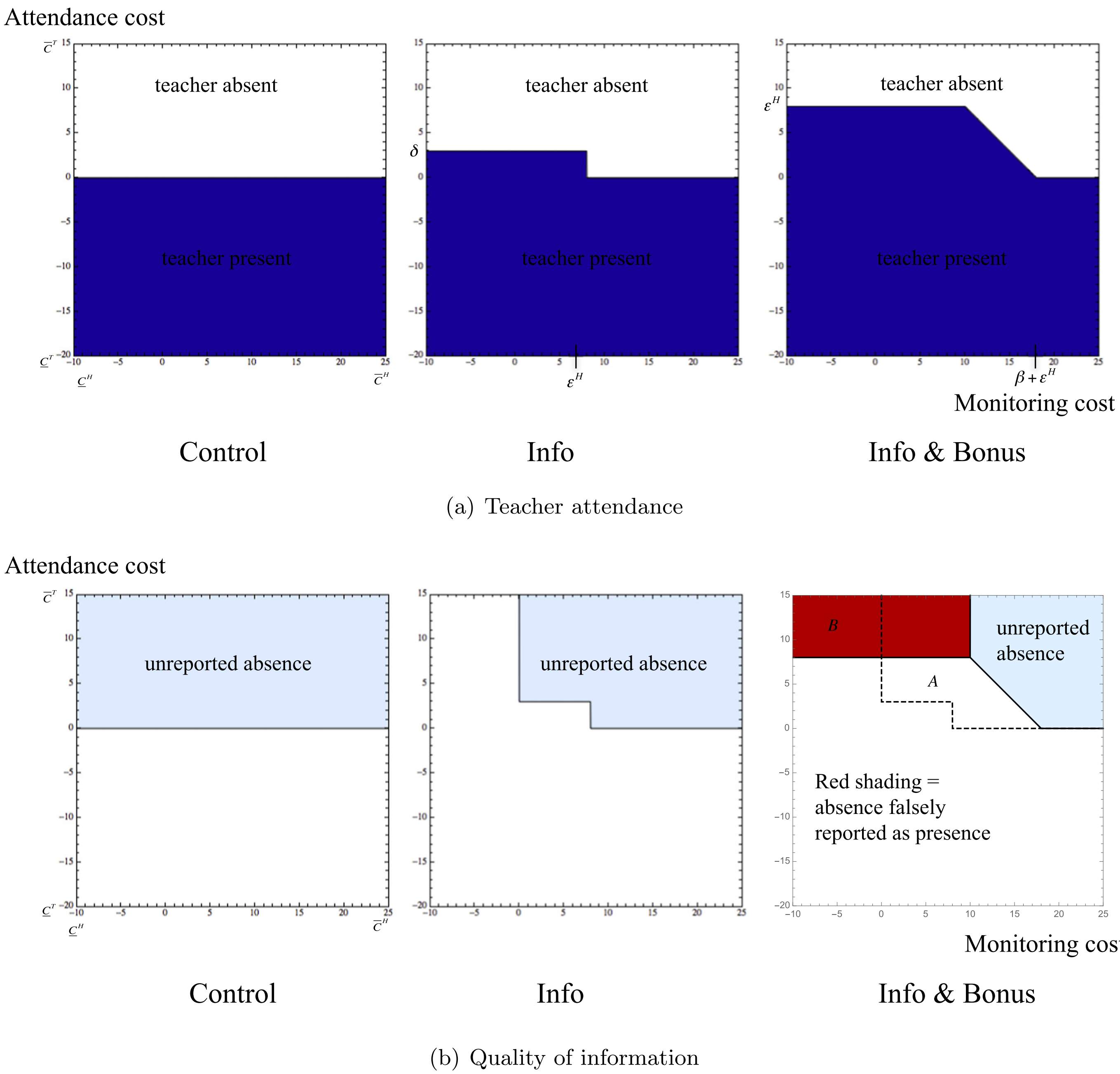


Fig. 4. Theoretical predictions. *Note:* Panel (a): the dark blue shaded regions show realizations of the monitoring cost, C^H, and attendance cost, C^T, for which the teacher chooses to attend school. Panel (b): the shaded regions show realizations of the monitoring and attendance costs where the bureaucracy would incorrectly predict the teacher to be present when in fact he is absent, due either to no report (light blue) or a false report (red). In both panels, remaining parameters are held fixed at $\varepsilon^H = 8, \delta = 3$, and $\beta = 10$. For these parameter values, region A is bigger than region B, indicating that there are fewer cost realisations resulting in an incorrect prediction (i.e. higher quality of information) in the Info & Bonus arm compared to the Info arm.

attend school. The intuition is straightforward. The area below the x−axis represents the baseline probability of attendance in the Control arm: the teacher attends because he receives an intrinsic benefit from doing so. Teacher attendance is predicted to be higher in the Info arm (larger dark blue area) because the head teacher is able to leverage the reputation cost δ to secure the benefit ε^H. If δ exceeds the attendance cost and ε^H outweighs the monitoring cost, then the players reach an equilibrium where the head teacher commits to monitor and report truthfully, and the teacher then attends.

Teacher attendance is higher still in the Info & Bonus arm because there is now transferable utility on the table, in the form of the bonus β. The head teacher can either use this transferable utility to induce the teacher to attend, or she can collude and file a false report. If the attendance cost is intermediate (higher than δ but less than ε^H) *and* β and ε^H together exceed the combined attendance and monitoring costs (a joint efficiency requirement), the head teacher opts for the former outcome. The players reach an equilibrium where they first agree on a side contract that commits the head teacher to monitor and report truthfully and the teacher to share some fraction of the bonus; the teacher subsequently attends.

5.1.2. Quality of information

The shaded regions in Fig. 4 Panel (b) plot realizations of the monitoring and attendance costs where the bureaucracy would incorrectly predict the teacher to be present when in fact he is absent. The area above the *x*-axis represents the baseline probability of an incorrect prediction in the Control arm: the bureaucracy does not receive a monitoring report and so, applying Bayes' rule, concludes the teacher is present. The quality of information is predicted to be higher in the Info arm than in the Control arm (the smaller, light blue area) for two reasons. First, there is a region (to the left of the $y-$axis and above δ) where the head teacher willingly sends a truthful report that the teacher is absent. Second, there is small region (below δ and to the left of ε^H) where the head teacher's truthful reporting induces the teacher to be present, which again ensures the bureaucracy no longer makes an incorrect prediction.

Turning to the Info & Bonus arm, the received wisdom is that attaching financial incentives to local monitoring reports will lower the quality of information. Our theoretical framework shows that this need not be the case. For the parameter values in Fig. 4 Panel (b), the area where the bureaucracy would make an incorrect prediction is actually smaller in the Info & Bonus arm than in the Info arm. To see why, compare regions A and B. In region A, the bonus β is additional transferable utility that enables the head teacher to both cover her monitoring costs and induce the teacher to attend. Without this transferable utility, there is no report and the teacher is absent, leading to an incorrect prediction in the Info arm. The flipside is that in region B, when it is too costly to 'pay' the teacher to attend but the head teacher is nonetheless intrinsically motivated to monitor, the bonus gives her an incentive to submit a false rather than truthful report. This lack of truth-telling results in an incorrect prediction in the Info & Bonus arm. Since region A is bigger than region B, we therefore have an example illustrating that attaching incentives to local monitoring can improve the quality of information.

5.2. Interpretation of empirical results

The framework above provides an explanation for our finding that P4LMP improved teacher attendance, but monitoring alone did not. A meaningful, *transferable* source of utility is necessary to enable local parties to reach a bargain that maximizes joint surplus; only then will a teacher internalize the welfare gains from his/her attendance as well as the costs. There may be shame attached to being reported absent but, in our experiment, it appears not to have been great enough (or transferable enough) to induce a significant improvement in teacher attendance. The same logic can also explain our second finding that P4LMP improved the quality of information compared to monitoring alone. Contrary to common intuition, 'collusion' is not the only relevant factor; the extensive margin of reporting also matters and is higher under P4LMP. True, information quality suffers due to absences falsely reported as presence (the dark red bar in Fig. 3) but unreported absences (the light blue bar) are lower under P4LMP, and this is the dominant effect.

It is also worth commenting on what the theory cannot explain. Our framework does not predict absence falsely reported as presence in the Info arm. Empirically, however, we observed this outcome on 7% of teacher days with independent spot checks — nearly as often as in the Info & Bonus arm. These head teachers must have expected some other adverse consequence (rather than loss of bonus) to follow from a truthful report of absence. Our framework also fails to predict presence falsely reported as absence. This outcome was rare, occurring on just 3% of teacher-days with independent spot-checks in the Info arm, and 5% of such days in the Info & Bonus arm. Rationalizing this outcome as equilibrium behavior is harder and calls for a radically different model — something that we do not undertake given our suspicion, noted in Section 3.1 above, that this is due to measurement error.

6. Conclusion

Received wisdom has it that pay for locally monitored performance (P4LMP) will fail to incentivize desired behavior and will bias information for planning purposes. Simple bargaining logic, however, suggests the reverse: P4LMP could both incentivize desired behavior and *improve* decision making.

Responding to these observations, this paper set out to answer three related questions. Can P4LMP induce improvements in service providers' behavior? Does P4LMP reduce or improve the quality of reported information for planning purposes? And what is the overall welfare and fiscal impact of P4LMP? To answer them, we used data collected during a field experiment in Ugandan primary schools to estimate impacts on teacher attendance, student enrollment and learning outcomes, and the quality of information available to district-level school administrators. We then combined our experimental estimates with additional administrative data and the Uganda National Panel Survey to undertake a welfare and fiscal analysis of alternative intervention designs. The key takeaways are that P4LMP can improve service providers' behavior—in our case, teacher attendance—as well as the quality of information for planning purposes. This dual objective can be met sufficiently cheaply to pass a welfare cost-benefit test. What is more, attaching financial incentives to local monitoring reports is fiscally sustainable; taken together, the tax revenue from increased lifetime earnings and savings from better information more than compensate for the cost of making bonus payments.

A natural question is whether the P4LMP model evaluated in this paper should be rolled out at scale. It seems clear that the high rates of teacher absenteeism observed in Uganda are at least partly due to the system's failure of to provide appropriate resources and incentives for monitoring. Districts have formal responsibility for monitoring schools but are typically under-staffed with a limited transportation budget and so find it difficult to undertake regular inspections. Across Uganda there are 87 schools per inspector and so, at most, an inspector can visit each school twice a year,[33] although in practice they see schools far less frequently.[34] Since previous research has shown that more monitoring is associated with lower teacher absence (Muralidharan et al., 2017), one policy response would be for the government to transfer additional resources to districts in the hope that this will translate into more school inspections. The available evidence for Uganda, however, indicates that this may not work: the number of inspectors has more than quadrupled since 2008 with no evidence of an improvement in teacher attendance.[35]

Our results suggest a cheaper, and arguably more incentive compatible, alternative to district-led school inspections that makes use of cheap, readily scalable digital technology. Local monitoring and reporting by mobile phone is inexpensive to set up, simply requiring the creation of a monitoring template and central database, and costs little to run as there are no transport costs or salaries to pay to inspectors. Rather than greater investment in traditional monitoring by district officials, the main policy response suggested by this paper is further decentralization with a scaling up of *local* monitoring

[33] In 2016 there were 281 inspectors and 24,419 schools, each open for 36 weeks per year.

[34] Monitoring failures are not unique to Uganda. In India, for example, "poor state capacities in terms of inadequate resources and systemic infirmities contribute significantly to ineffective monitoring" (Bhatty and Saraf, 2016).

[35] In 2008, there were 68 inspectors, compared to 281 in 2016 (Ministry of Education, Science, Technology and Sports, 2010). Chaudhury et al. (2006) estimate an absence rate of 27% for 2006, compared to our estimate of 26% in 2012 and the World Bank's Service Delivery Indicators estimate of 27% in 2013 (Wane and Martin, 2013).

and incentives to exploit one of the so-called “digital dividends” discussed in the 2016 World Development Report.[36]

Looking beyond the education context, there are many settings where public sector organizations do (or could) rely on reports by local monitors that are costly to verify. It is therefore of interest to ask whether P4LMP will generally prove as cost effective as it has in Ugandan primary education. Drawing on a simple theoretical model, we have argued that failures in public service delivery can be interpreted as a breakdown in bargaining. Seen in this light, the question of transaction costs becomes paramount (Dixit, 1996). Our theoretical results suggest that service delivery can be improved via P4LMP in settings where (i) local monitoring costs are low relative to central government; (ii) the local monitor shares, at least to some small degree, the preferences of the beneficiaries; and (iii) there is a lack of transferable utility between the local monitor and agent that prohibited bargaining in the first place. When all three conditions hold, P4LMP will put transferable money on the table and could improve service delivery precisely because of (not despite) the role played by side payments.

Acknowledgments

We are grateful for useful feedback from David Evans, James Habyarimana, Julien Labonne, and Sandip Sukhtankar. We thank numerous seminar audiences and participants at the following conferences: AEA Annual Meeting, RES Annual Conference, IGC Growth Week, World Literacy Summit, Uganda Education Research Symposium, SEEDEC, IiG, NEUDC, CMPO and Non-Profits, Governments, and Organizations. Anirvan Chowdhury and Esmeralda Sindou provided excellent research assistance. We are grateful to Nada Eissa and the IGC for access to Uganda Revenue Authority data. The ‘Improving Institutions for Pro-Poor Growth’ Research Consortium, the IGC, and the Oxford Institute for Global Economic Development provided funding. All errors and omissions are our own.

Appendix A. Additional figures

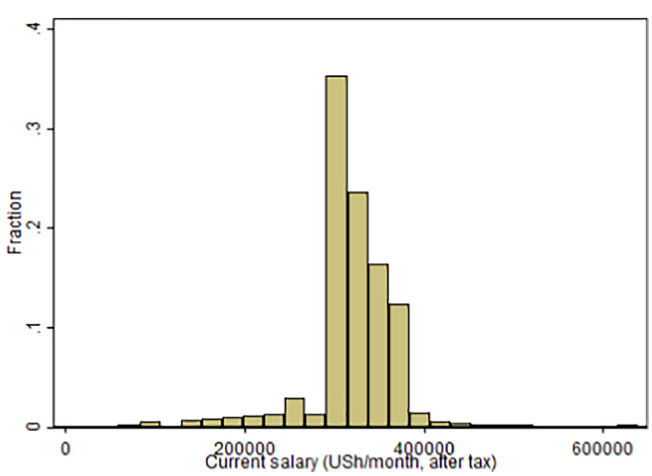
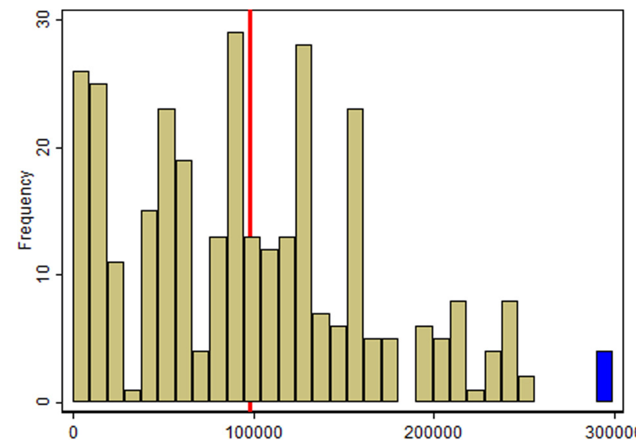
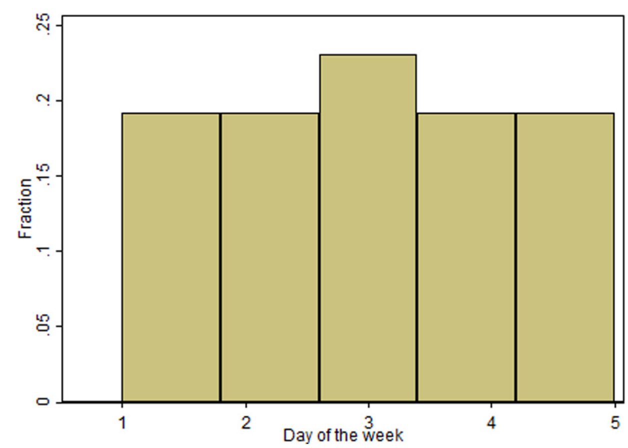


Fig. A.1. Distribution of teacher salaries. *Note:* The figure shows the distribution of monthly salaries for 629 teachers (excluding headteachers and their deputies). We only collected data on teacher salaries at endline, so the sample is restricted to teachers who remained in the sample schools for the whole duration of the program. Six observations, where reported salaries are 10 times larger or 10 times smaller than the median salary, are excluded since they are likely to be measurement errors.



(a) Distribution of total bonus payouts



(b) Distribution of days chosen for bonus payment

Fig. A.2. Implementation of bonus scheme. *Note:* Panel (a) shows the distribution of cumulative bonus payments for the sample of 316 teachers who were in the Info & Bonus treatment arm at the beginning of the study. The red line indicates the average payment received; the blue bar indicates the maximum possible payout. Panel (b) shows the probability that different days of the week were chosen as the qualifying report for bonus payouts, excluding the first month of program implementation. Within this 26 week period, Wednesday was chosen six times and the other days were chosen five times.

Appendix B. Theory

B.1. Model

We consider three variants, which we refer to as experimental arms. The basic structure in all arms is that teachers can choose between showing up for work or not. In the *Control* arm, teacher attendance remains unmonitored. Having pinned down a baseline, we then study how the introduction of local monitoring by the head teacher, who reports to a government bureaucracy, affects teacher attendance. We refer to this as the *Info* arm. Next, we investigate how combining local monitoring with financial incentives triggered by the head teacher's reporting affects teacher attendance, referring to this as the *Info & Bonus* arm. We compare equilibrium outcomes—teacher attendance, head teacher monitoring/reporting, and the bureaucracy's equilibrium beliefs—across the three experimental arms.

[36] Since our theory predicts that the welfare gain from P4LMP is increasing in the strength of the headteacher's preference for teacher attendance (ε^H in the model), complimentary policy efforts to recruit and retain pro-socially motivated headteachers could also prove important.

B.1.1. Players and actions

The economy consists of a teacher (he), a head teacher (she), and a government bureaucracy (it). In all arms, the teacher chooses whether to attend school, $a \in \{0,1\}$. In the Control arm, the head teacher plays no active role. In both treatment arms, the head teacher chooses whether to monitor $m \in \{0,1\}$. If the head teacher monitors, $m = 1$, she observes a and chooses a public report $r \in \{0,1\}$. We will say that the head teacher sends a *truthful report* iff $r = a$. If the head teacher does not monitor, $m = 0$, she cannot send a report.[37] In the Info & Bonus treatment arm, the bureaucracy pays a cash bonus β directly to the teacher iff he is reported present, $r = 1$. The bureaucracy takes no further action under any arm, other than to form a posterior belief over teacher presence.[38]

B.1.2. Payoffs

All players are risk neutral. Net of any side transfers, payoffs to the teacher and head teacher are:

$$U^T = 1_{\{m=1,r=1\}} \cdot \beta - 1_{\{a=1\}} \cdot C^T - 1_{\{r=0\}} \cdot \delta$$
$$U^H = 1_{\{a=1\}} \cdot \varepsilon^H - 1_{\{m=1\}} \cdot C^H.$$

If the teacher attends school, $a = 1$, he incurs a (possibly negative) cost of C^T. If the teacher attends, the head teacher receives a private benefit of ε^H. If the head teacher monitors, she incurs a (possibly negative) cost of C^H. Reporting entails no further cost for the head teacher but a mark of absent, $r = 0$, imposes a reputational cost of δ on the teacher. If the head teacher is indifferent, we assume that she reports truthfully.

B.1.3. Key assumptions

The costs C^T and C^H are observed by the head teacher and teacher but not by any other player. For convenience, we assume that realizations are drawn independently from uniform distributions. The lower and upper support of the distribution of C^T are denoted by $\underline{C}^T$ and $\overline{C}^T$. To calibrate the model to the baseline absenteeism rate, we assume $-\underline{C}^T > \overline{C}^T > 0$ (so that attendance in Control schools is more than 50 %). The lower and upper support of the distribution of C^H are denoted similarly, although here we simply assume $\overline{C}^H > 0 > \underline{C}^H$. Again for convenience and in the spirit of rationalizing baseline absenteeism, we assume that the bonus β is the only source of transferable utility.[39] Relatedly, we assume that side contracts sharing β are costless and enforceable, and that the head teacher can commit to monitor. Finally, we assume that parameters satisfy $\beta > \varepsilon^H > \delta > 0$.

B.1.4. Timing

To emphasize the differences across arms, it is worth spelling out the order of play. The timing in the Control arm is:

0. Nature draws a realization of C^T and reveals this cost to the teacher.
1. The teacher chooses whether to attend school, $a \in \{0,1\}$. Payoffs are realized and the game ends.

The timing in the Info arm is:

0. The bureaucracy announces the monitoring scheme. Nature draws realizations of C^T and C^H and reveals both of these costs to the teacher and the head-teacher.
1. The head teacher chooses whether to make an announcement to the teacher. An announcement $R(a)$ commits the head teacher to monitor, $m = 1$, and specifies the report r that the head teacher will send to all players following each possible action a.
2. The teacher chooses whether to attend school, $a \in \{0,1\}$.
3. If the head teacher made the announcement at Stage 1, she monitors and sends the public report $r = R(a)$. Otherwise, the head teacher takes no action. Payoffs are realized and the game ends.

The timing in the Info & Bonus arm is:

0. The bureaucracy announces the monitoring and incentive scheme. Nature draws realizations of C^T and C^H and reveals both of these costs to the teacher and the head teacher.
1. The head teacher chooses whether to make a side contract offer to the teacher. A side contract $< R(a), \tau >$ commits the head teacher to monitor, specifies the report r that the head teacher will send to all players following the action a and specifies the side transfer τ that the teacher will pay to the head teacher in the event that $r = 1$.

If the side contract is accepted at Stage 1, the game continues as follows.

2. The teacher chooses whether to attend school, $a \in \{0,1\}$.
3. The head teacher monitors and sends the public report $r = R(a)$. If $r = 1$, the bureaucracy pays β to the teacher who then transfers τ to the head teacher. Payoffs are realized and the game ends.

If the side contract is not accepted at Stage 1, the game proceeds as in the Info arm except that at Stage 3 the bureaucracy transfers β to the teacher in the event that $r = 1$.

B.2. Analysis

We now state equilibrium teacher attendance and the bureaucracy's equilibrium beliefs in the three experimental arms in turn.

B.2.1. Control

The probability of teacher attendance is

$$\Pr[a = 1|\text{Control}] = \Pr\left[C^T \leq 0\right] = \frac{-\underline{C}^T}{\overline{C}^T - \underline{C}^T}. \tag{8}$$

Anticipating the teacher's strategy, the bureaucracy deduces that the probability of teacher attendance is

$$\Pr[a = 1|m = 0] = \frac{\Pr[m = 0, a = 1]}{\Pr[m = 0]} = \Pr\left[C^T \leq 0\right] = \frac{-\underline{C}^T}{\overline{C}^T - \underline{C}^T} > 0.5,$$

[37] This assumption reflects the experimental design feature that mobile phones must be kept that the school.

[38] In forming this belief, the bureaucracy uses only its knowledge of the support of C^T together with equilibrium strategies. Clearly, this is a simplification as mid-tier bureaucrats may have other sources of information. Since the availability of such information should be orthogonal to treatment, we do not model it here. In our empirical analysis, the level of incorrect predictions when the bureaucracy receives no report (light blue bars in Fig. 3) would be biased upwards, but the difference between the two treatment arms (Table 4) should be not be affected.

[39] If all sources of utility were transferable, then the players should reach a jointly efficient outcome. The high rates of absenteeism reported above suggest this is not the case. In reality, δ might be partly transferable. We assume non-transferability to make the distinction between the Info and Info & Bonus arms as clear as possible.

and therefore predicts that the teacher is present. This prediction is incorrect in the event of an *unreported absence*, that is

$$\Pr[m=0,a=0|\text{Control}]=\Pr\left[C^T>0\right]=\frac{\overline{C}^T}{\overline{C}^T-\underline{C}^T}. \tag{9}$$

B.2.2. Info

The probability of teacher attendance is

$$\begin{aligned}\Pr[a=1|\text{Info}]&=\Pr\left[C^T\leq 0\right]+\Pr\left[0<C^T\leq\delta,C^H\leq\varepsilon^H\right]\\&=\frac{-\underline{C}^T}{\overline{C}^T-\underline{C}^T}+\frac{\delta\left(-\underline{C}^H+\varepsilon^H\right)}{\left(\overline{C}^T-\underline{C}^T\right)\left(\overline{C}^H-\underline{C}^H\right)}.\end{aligned} \tag{10}$$

The probability of monitoring and reporting of teacher attendance is

$$\begin{aligned}\Pr[m=1,r=1|\text{Info}]&=\Pr\left[C^T\leq 0,C^H\leq 0\right]+\Pr\left[0<C^T\leq\delta,C^H\leq\varepsilon^H\right]\\&=\frac{-\underline{C}^T-\underline{C}^H}{\left(\overline{C}^T-\underline{C}^T\right)\left(\overline{C}^H-\underline{C}^H\right)}+\frac{\delta\left(-\underline{C}^H+\varepsilon^H\right)}{\left(\overline{C}^T-\underline{C}^T\right)\left(\overline{C}^H-\underline{C}^H\right)},\end{aligned}$$

and the probability of monitoring and reporting of teacher absence is

$$\begin{aligned}\Pr\left[m=1,r=0|\text{Info}\right]&=\Pr\left[C^T>\delta,C^H\leq 0\right]\\&=\frac{\left(\overline{C}^T-\delta\right)-\underline{C}^H}{\left(\overline{C}^T-\underline{C}^T\right)\left(\overline{C}^H-\underline{C}^H\right)}.\end{aligned}$$

In contrast to the Control arm, the bureaucracy now reaches three information sets. The first is $m=0$. Anticipating the teacher and head teacher's strategies, the bureaucracy deduces that the probability of teacher attendance is

$$\begin{aligned}&\Pr[a=1|m=0]\\&\quad=\frac{\Pr[m=0,a=1]}{\Pr[m=0]}\\&\quad=\frac{\Pr\left[C^T\leq 0,C^H>0\right]}{\Pr\left[C^T\leq 0,C^H>0\right]+\Pr\left[C^T>0,C^H>0\right]-\Pr\left[0<C^T\leq\delta,0<C^H\leq\varepsilon^H\right]}\\&\quad=\frac{-\underline{C}^T\overline{C}^H}{\left(\overline{C}^T\overline{C}^H-\delta\varepsilon^H\right)-\underline{C}^T\overline{C}^H}>\frac{-\underline{C}^T}{\overline{C}^T-\underline{C}^T}>0.5,\end{aligned}$$

and so, at the information set $m=0$, predicts that the teacher is present. This prediction is incorrect in the event of an *unreported absence*, that is

$$\begin{aligned}\Pr[m=0,a=0|\text{Info}]&=\Pr\left[C^T>0\right]-\Pr\left[C^T>0,C^H\leq 0\right]\\&\quad-\Pr\left[0<C^T\leq\delta,0<C^H\leq\varepsilon^H\right]\\&=\frac{\overline{C}^T}{\overline{C}^T-\underline{C}^T}-\frac{-\underline{C}^H\overline{C}^T+\varepsilon^H\delta}{\left(\overline{C}^T-\underline{C}^T\right)\left(\overline{C}^H-\underline{C}^H\right)}.\end{aligned} \tag{11}$$

The second information set is $r=1$. Since the bureaucracy knows that the head teacher reports truthfully, it predicts that the teacher is present. The third information set is $r=0$. Again aware that the head teacher reports truthfully, the bureaucracy predicts that the teacher is absent. Both of these predictions are correct.

B.2.3. Info & Bonus

The probability of teacher attendance is

$$\begin{aligned}\Pr[a=1|\text{Bonus}]&=\Pr\left[C^T\leq 0\right]+\Pr\left[0<C^T\leq\varepsilon^H,C^H\leq\varepsilon^H+\beta-C^T\right]\\&=\frac{-\underline{C}^T}{\overline{C}^T-\underline{C}^T}+\frac{\varepsilon^H\left(-\underline{C}^H+\beta\right)+\left(\varepsilon^H\right)2/2}{\left(\overline{C}^T-\underline{C}^T\right)\left(\overline{C}^H-\underline{C}^H\right)}.\end{aligned} \tag{12}$$

The probability of monitoring and reporting of teacher attendance is

$$\begin{aligned}\Pr\left[m=1,r=1|\text{Bonus}\right]&=\Pr\left[C^H\leq\beta\right]+\Pr\left[\beta<C^H<\varepsilon^H+\beta-C^T\right]\\&=\frac{\beta-\underline{C}^H}{\overline{C}^H-\underline{C}^H}+\frac{\left(\varepsilon^H\right)^2/2}{\left(\overline{C}^T-\underline{C}^T\right)\left(\overline{C}^H-\underline{C}^H\right)},\end{aligned}$$

and the probability of monitoring and reporting of teacher absence is zero. The bureaucracy now reaches just two information sets. The first is $m=0$. Anticipating teacher and head teacher strategies, the bureaucracy deduces that the probability of teacher attendance is

$$\begin{aligned}&\Pr[a=1|m=0]\\&\quad=\frac{\Pr[m=0,a=1]}{\Pr[m=0]}\\&\quad=\frac{\Pr\left[C^T\leq 0,C^H>\beta\right]}{\Pr\left[C^T\leq 0,C^H>\beta\right]+\Pr\left[C^T\geq>0,C^T>0,C^H>\beta\right]-\Pr\left[\beta<C^H<\varepsilon^H+\beta-C^T\right]}\\&\quad=\frac{-\underline{C}^T\left(\overline{C}^H-\beta\right)}{\left(\overline{C}^T\left(\overline{C}^H-\beta\right)-\left(\varepsilon^H\right)^2/2\right)-\underline{C}^T\left(\overline{C}^H-\beta\right)}>\frac{-\underline{C}^T}{\overline{C}^T-\underline{C}^T}>0.5,\end{aligned}$$

and so, at the information set $m=0$, predicts that the teacher is present. This prediction is incorrect in the event of an *unreported absence*, that is

$$\begin{aligned}&\Pr\left[m=0,a=0|\text{Bonus}\right]\\&\quad=\Pr\left[C^T>0\right]\\&\qquad-\Pr\left[C^T>\varepsilon^H,C^H\leq 0\right]-\Pr\left[0<C^T\leq\varepsilon^H,C^H\leq\varepsilon^H+\beta-C^T\right]\\&\quad=\frac{\overline{C}^T}{\overline{C}^T-\underline{C}^T}-\frac{-\underline{C}^H\left(\overline{C}^T-\varepsilon^H\right)}{\left(\overline{C}^T-\underline{C}^T\right)\left(\overline{C}^H-\underline{C}^H\right)}-\frac{\varepsilon^H\left(-\underline{C}^H+\beta\right)+\left(\varepsilon^H\right)2/2}{\left(\overline{C}^T-\underline{C}^T\right)\left(\overline{C}^H-\underline{C}^H\right)}.\end{aligned} \tag{13}$$

The second information set is $m=1,r=1$. Anticipating teacher and monitor strategies (in particular that the head teacher may now send a false report), the bureaucracy deduces that the probability of teacher attendance is

$$\begin{aligned}&\Pr[a=1|m=1,r=1]\\&\quad=\frac{\Pr[m=1,r=1,a=1]}{\Pr[m=1,r=1]}\\&\quad=\frac{\Pr\left[C^T\leq\varepsilon^H,C^H\leq\beta\right]+\Pr\left[\beta<C^H<\varepsilon^H+\beta-C^T\right]}{\Pr\left[C^T\leq\varepsilon^H,C^H\leq\beta\right]+\Pr\left[\beta<C^H<\varepsilon^H+\beta-C^T\right]+\Pr\left[C^T>\varepsilon^H,C^H\leq\beta\right]}\\&\quad=\frac{\left(-\underline{C}^T+\varepsilon^H\right)\left(\beta-\underline{C}^H\right)+\left(\varepsilon^H\right)^2/2}{\left(-\underline{C}^T+\varepsilon^H\right)\left(\beta-\underline{C}^H\right)+\left(\varepsilon^H\right)^2/2+\left(\overline{C}^T-\varepsilon^H\right)\left(\beta-\underline{C}^H\right)}>\frac{-\underline{C}^T}{\overline{C}^T-\underline{C}^T}>0.5,\end{aligned}$$

and so, at the information set $m = 1, r = 1$, predicts that the teacher is present. This prediction is incorrect in the event of an *absence falsely reported as presence*, that is

$$\Pr[m = 1, r = 1, a = 0|\text{Bonus}] = \Pr\left[C^T > \varepsilon^H, C^H \leq 0\right] = \frac{-\underline{C}^H\left(\overline{C}^T - \varepsilon^H\right)}{\left(\overline{C}^T - \underline{C}^T\right)\left(\overline{C}^H - \underline{C}^H\right)}. \quad (14)$$

B.2.4. Summing up

Teacher attendance is highest in the Info & Bonus arm and lowest in the Control arm (follows from a comparison of Eqs. (8), (10), and (12)). Quality of information (i.e. the probability that the bureaucracy makes an incorrect prediction) is highest in the Control arm but there is an ambiguous comparison between the two treatment arms (following from a comparison of Eqs. (9), (11), and (13)+ (14)). In particular, the probability of an unreported absence is lower, but the probability of absence falsely reported as presence is higher, in the Info & Bonus arm relative to the Info arm.

Appendix C. Detailed welfare analysis

This section outlines the calculations and assumptions underlying the welfare analysis presented in Section 4. We proceed in three stages. First, in Section C.1 we use two different data sources—administrative data on enrollment and repetition, and a tracked cohort of pupils surveyed before and after the program—to obtain a conservative estimate of the increase in grade attainment in Info & Bonus schools relative to Info schools. Second, in Section C.2 we use prior studies and additional data sources to calculate the increase in NPV of future lifetime earnings caused by an increase in grade attainment. Third, in Section C.3 we combine this model with estimates of tax evasion among wage earners derived from Uganda Revenue Authority and the World Bank's *World Development Indicators* to project the fiscal consequences of payment for locally monitored performance.

C.1. Moving from enrollment to grade attainment

To what extent can we attribute the higher enrollment observed in Info & Bonus schools to higher grade attainment? Modelling grade attainment requires assumptions relating to: (i) persistence of the program and (ii) persistence of the program's impacts on attainment. In both cases we take the most conservative approach. On the first point, we model the welfare comparison for the actual experiment as it was conducted; i.e. a policy intervention for one year, with a return to the status quo and an end to project expenditures immediately thereafter. On the second point, we assume that grade attainment remains the same for all pupils, except for those who would have dropped out were it not for the program (the averted dropouts) and, furthermore, that these averted dropouts go on to drop out immediately after withdrawal of the treatment and so only gain one more year of education.[40] These conservative assumptions, again, allow us to estimate the lower bound for welfare analysis.

C.1.1. Estimating averted dropouts: administrative data

Our two data sources allow for two different strategies, each with different identifying assumptions on transfers. First, using our survey data and combining it with administrative data on repetition figures in 2011 and 2012, we can back out the implied number of dropouts in grade g and year t, $\Delta_{g,t}$.

[40] It is plausible that the expected future attainment of these pupils is lower than the average pupil because they were at the margin of dropping out.

Enrollment in grade g at period t can be decomposed into the following end-states

$$\pi_{g,t} \equiv \Delta_{g,t} + \rho_{g,t} + \tau_{g,t} + \lambda_{g,t},$$

where $\Delta_{g,t}$ denotes the number of pupils who dropout at the end of the year, $\rho_{g,t}$ the number who repeat the grade, $\tau_{g,t}$ the number who transition to the next grade, and $\lambda_{g,t}$ the number who transfer out to another school. Similarly, enrollment in grade $g + 1$ at the beginning of year $t + 1$ can be decomposed as

$$\pi_{g+1,t+1} \equiv \tau_{g,t} + \rho_{g+1,t} + \varphi_{g,t},$$

where $\tau_{g,t}$ denotes the number of pupils who have progressed from the previous grade, $\rho_{g+1,t}$ the number who are repeating the grade, and $\varphi_{g,t}$ the number who have transferred in from another school. Substituting in for $\tau_{g,t}$, we have

$$\Delta_{g,t} \equiv (\pi_{g,t} - \rho_{g,t}) - (\pi_{g+1,t+1} - \rho_{g+1,t}) - \mu_{g,t}$$

where we define net outbound transfers as $\mu_{g,t} \equiv (\lambda_{g,t} - \varphi_{g,t})$. We have access to administrative data on enrollment and repetition numbers per grade per school, but do not have data on inbound or outbound transfers and can therefore only estimate

$$\hat{\Delta}_{g,t} = (\pi_{g,t} - \pi_{g+1,t+1}) - (\rho_{g,t} - \rho_{g+1,t}). \quad (15)$$

For the difference in $\hat{\Delta}_{g,t}$ across treatment arms to provide a true estimate of the impact of the program, we need to assume that net transfers are on average the same across these arms. There is a risk of over-estimating the impact on averted dropouts, for example, if more pupils transfer to the Info & Bonus schools because of the program.

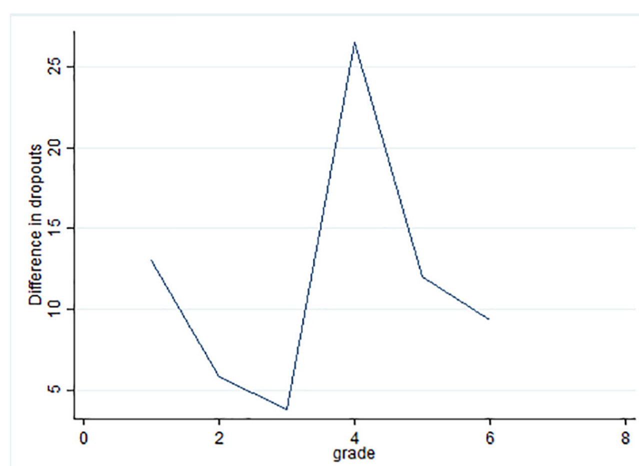


Fig. C.3. Difference in dropouts between Info and Info & Bonus schools, by grade *Note:* The figure shows the difference in the average number of dropouts between Info and Info & Bonus schools, by grade, where dropouts have been calculated based on Eq. (15) using our survey data on enrollment and administrative data on repetition. We refer to these differences as the 'averted dropouts'.

Fig. C.3 shows the difference in dropouts between Info & Bonus and Info for each grade, estimated using Eq. (15). Note that this difference is highest in Grade 4, precisely the grade after which there is a large drop in enrollment (Fig. 2 in the main body of the paper). On average 70 more pupils dropped out from Info schools relative to

Info & Bonus schools and this difference is statistically significantly different from zero at the 5% level.[41]

C.1.2. Estimating averted dropouts: cohort data

As a second strategy we can derive implied annual dropout and repetition rates from the differences in reduced-form dropout and repetition probabilities observed for the sample tracked from 2010, when the P3 pupils were first observed as part of a separate study, to 2013, when they were observed post-intervention.[42] Data are also available on the grade in which these pupils were enrolled (if any). To back out annual dropout and enrollment probabilities, we make note of the fact that three academic years were completed between the time this sample of pupils was drawn to our endline survey, but that only one of these years was spent under treatment.

Let δ_w denote the probability of dropout under treatment regime w, and recall from Table 2 that in the Control arm the fraction of the tracked cohort observed to still be enrolled at endline was 0.344. The probability of dropout in the Control arm can therefore be written as $\Pr[\text{Dropout}|w = \text{Control}] = 1 - (1 - \delta_{\text{Control}})^3 \approx 1 - 0.344$. Using the balance implied by the experimental design, we can generate observed probabilities of dropout with one period of treatment exposure in either of the treatment arms $w \in \{\text{Info, Info\&Bonus}\}$. The implied annualized dropout rates for these arms are given by setting the corresponding observed dropout probability in treatment arm w equal to $\delta_{\text{Control}} + (1 - \delta_{\text{Control}})\delta_{\text{Control}} + (1 - \delta_{\text{Control}})^2\delta_w$. The resulting implied annual transition probabilities are given in Table C.1. Relative to the estimates of Table 2, dropout rates are lower since these represent annual rather than cumulative probabilities. Moreover, differences across treatment arms are exaggerated, since (by virtue of the random assignment of treatment) in expectation all observed differences are attributable to the one year under treatment.

Table C.1
Estimated annual pupil transition probabilities, by treatment.

Treatment, w	Dropout, $\hat{\delta}_w$	Repetition
Control	0.299	0.044
Info	0.143	0.060
Info & Bonus	0.012	0.063

Note that the two different methods discussed above make use of two different identifying assumptions for estimating the true difference in dropouts. The decomposition using administrative data on enrollment and repetition requires that *net transfers* are not different between treatment arms. On the other hand, to derive dropout rates using the tracked sample, we need to assume that *outbound transfers* are not affected by the interventions.

These reduced-form experimental results estimate the short-term enrollment and dropout impacts of assignment to alternative treatment regimes, providing evidence of a statistically and economically significant impact of the Info & Bonus arm relative to the Control and Info arms. With our conservative approach, we assume that each averted dropout amounts to no more than one additional year of grade attainment. But what is the financial impact for those students who remain in school for one more year? We turn to this below.

C.2. Moving from grade attainment to earnings

As a final step in the welfare analysis, we place a financial value on the increase in grade attainment. We begin by describing the model that we use to measure the net school-level NPV gain associated with moving from Info to Info & Bonus. We then set out how we choose the numbers required to calculate this net NPV gain, drawing on prior studies and additional data sources.

C.2.1. Earnings NPV model

For simplicity, we assume that there are only two sectors (formal wage employment and subsistence agriculture), and further that earnings from agriculture do not depend on years of education and experience. We also assume that all pupils start school aged 7, do not repeat a grade, obtain no more than Grade 7, and leave formal employment aged 60.[43] Given these assumptions, the NPV of lifetime earnings for an individual who drops out after grade s can be written as:

$$NPV_s = \sum_{t=s+7}^{t=60} \left(\frac{1}{1+r}\right)^{t-s-7} (P_{s,t} \cdot w_{s,t} + (1 - P_{s,t}) \cdot A), \tag{16}$$

where $P_{s,t}$ is the probability that this individual with s years of schooling is employed in the formal wage sector at age t, $w_{s,t}$ is the associated formal wage, A is the (constant) subsistence agricultural wage, and r is a discount rate. For each grade $s = 1, \ldots, 7$, we want to obtain *potential* NPVs: one that reflects expected lifetime earnings of those who exit school at that grade in control schools, and a (counterfactual) outcome arising if individuals who would otherwise depart after grade s of schooling are induced by the treatment to obtain an additional year of schooling. To illustrate, consider $s = 1$. Lifetime earnings for those leaving school at grade s in control schools, denoted by superscript 0, can be written as:

$$\begin{aligned} NPV_1^0 &= P_{1,8} \cdot w_{1,8} + (1 - P_{1,8}) \cdot A + (P_{1,9} \cdot w_{1,9} + (1 - P_{1,9}) \cdot A) \left(\frac{1}{1+r}\right) \\ &+ \quad (P_{1,10} \cdot w_{1,10} + (1 - P_{1,10}) \cdot A) \left(\frac{1}{1+r}\right)^2 + \cdots \\ &+ \quad (P_{1,60} \cdot w_{1,60} + (1 - P_{1,60}) \cdot A) \left(\frac{1}{1+r}\right)^{52}. \end{aligned} \tag{17}$$

The year after dropping out the individual is 8 years old. With probability $P_{1,8}$ she enters the formal wage employment sector and earns $w_{1,8}$. With probability $(1 - P_{1,8})$ she earns A. Her probability of formal wage employment and the associated formal wage then evolve each year with her accumulated experience until she retires aged 60. Similarly, the outcome for such a student induced by treatment to obtain an extra year of schooling, denoted by superscript 1, can be written as

$$\begin{aligned} NPV_1^1 &= 0 + ((P_{1,8} + \pi) \cdot (1 + \rho)\, w_{1,8} + (1 - P_{1,8} - \pi) \cdot A) \left(\frac{1}{1+r}\right) \\ &+ \quad ((P_{1,9} + \pi) \cdot (1 + \rho)\, w_{1,9} + (1 - P_{1,9} - \pi) \cdot A) \left(\frac{1}{1+r}\right)^2 + \cdots \\ &+ \quad ((P_{1,59} + \pi) \cdot (1 + \rho)\, w_{1,59} + (1 - P_{1,59} - \pi) \cdot A) \left(\frac{1}{1+r}\right)^{52}, \end{aligned} \tag{18}$$

[41] Results from the regression analysis estimated using Eq. (1) are available upon request. In these regressions, the number of observations drops from 85 to 82, because the repetition number in 2012 is missing for three schools.

[42] Data on enrollment for this sub-sample of pupils are not available for the 2012 baseline to the present study.

[43] The data do not allow us to determine the age at which someone completed a grade. So we need to assume that everyone starts school aged 7 and progresses through grades at the same rate in order to determine how many working years they have available after graduation. Earnings are assumed to be zero after age 60. Because future earnings beyond that age are heavily discounted in the NPV, this assumption has little effect on our results.

where π and ρ denote the causal effects of one additional year of schooling to the probability of formal employment and the formal sector wage respectively. At age 8, the individual is still in school and earns nothing. The year after dropping out she is 9 years-old. With probability $(P_{1,8} + \pi)$ she enters the formal wage employment sector and earns $(1 + \rho)\, w_{1,8}$. With probability $(1 - P_{1,8} - \pi)$ she earns A. Her probability of formal wage employment and the associated formal wage then evolve each year with her accumulated experience until she retires aged 60. Staying in school for one additional year therefore delays her earnings by one year but uplifts her expected earnings in all future years. The difference between NPV_1^1 and NPV_1^0 is the grade 1-specific net NPV gain associated with moving from Info to Info & Bonus. Depending on parameters, this may be positive or negative. Our goal is to estimate this net NPV gain for each of grades $s = 1, \ldots, 6$.[44] To do so, we need numbers for the rates, r, π, and ρ, the series $P_{s,t}$ and $w_{s,t}$, and the constant A.

C.2.2. Choice of parameters

The standard discount rate used in social cost benefit analysis is 3.5% and is justified via the social time preference (STP) method of discounting (HM Treasury, 2011). A nominal discount rate of 10% has been used in the development literature by authors appealing to the social opportunity cost (SOC) method of discounting and, in particular, the high cost of government borrowing in low-income contexts (e.g. Ozier, 2011 and Baird et al., 2016).[45] We view the STP method, and hence $r = 3.5\%$, to be appropriate for our welfare comparison. Few studies in developing countries have looked directly at the impact of education on the probability of gaining employment in the formal wage sector. As a conservative lower bound we use a figure of $\pi = .01$. This is consistent with our own estimate from observational data (see Table C.2 below). We use $\rho = 6.8\%$ for the causal impact of education on earnings since this represents the conservative end of recent studies using plausibly exogenous variation to identify the causal impact of education on earnings in developing countries.[46]

Using the 2011/12 Ugandan National Panel Survey data, we estimate the expected probability, $P_{s,t}$, of formal sector employment for someone aged t currently residing in one of the six districts where our study took place, and who dropped out after completing grade s:[47]

$$P_{s,t} = \alpha_0 + \alpha_1 e + \alpha_2 e^2 + \alpha_3 e^3 + \alpha_4 s + \varepsilon_{s,t}, \tag{19}$$

where s is years of schooling and $e = t - s - 7$ is the number of years of experience. Using the same data, we also estimate the observational lifetime evolution of wages, $w_{s,t}$, using the standard Mincer earnings function (and further restricting the sample to individuals who have obtained at most Grade 7 and are 60 or under):

$$\ln w_{s,t} = \beta_0 + \beta_1 e + \beta_2 e^2 + \beta_3 s + \varepsilon_{s,t}. \tag{20}$$

Table C.2 shows the results of these two regressions.

Table C.2
Predicting the probability of formal sector employment and wages.

	(1)	(2)
	Prob. employed	Log wage
Years of schooling	0.00817***	0.0487**
	(0.00)	(0.02)
Years of experience	0.0114***	0.0538**
	(0.00)	(0.02)
Years of experience − squared	−0.000426***	−0.000826*
	(0.00)	(0.00)
Years of experience − cubed	0.00000452***	
	(0.00)	
Constant	−0.0326***	13.26***
	(0.01)	(0.25)
Observations	5752	282

Note: Columns (1) and (2) report regression results from Eqs. (20) and (19) respectively. In both regressions, the sample is restricted to individuals who have obtained at most Grade 7, are no longer in school, and under 60. In Column (2) the sample is further restricted to individuals who earn a salary. Standard errors are in parentheses.
*** is significant at the 1% level.
** is significant at the 5% level.
* is significant at the 10% level.

Finally, to estimate A, we take the average annual agricultural income from the sample of rural households whose main source of income is agriculture and divide this by two. This figure of USD 228 is plausibly an over-estimate of the individual earnings from agriculture (and thus leads to a more conservative estimate of the impact of the program), because more than two people per household typically work on the household's farm. Another approach would be to divide by the number of household members who claim to have worked on the household farm in the past year (roughly 4).[48]

C.3. Fiscal sustainability

C.3.1. Tax NPV model

Making the reasonable assumption that earnings from subsistence agriculture are untaxed, the NPV of the government's tax revenue collected from the lifetime formal sector earnings of an individual who drops out after grade s can be written as:

$$NPV_s\,(\text{tax}) = \sum_{t=s+7}^{t=60} \left(\frac{1}{1+r}\right)^{t-s-7} P_{s,t} \cdot (1-\epsilon)\left(\tau \cdot (w_{s,t} - \overline{w}) + \tau_0\right), \tag{21}$$

where τ is the marginal rate of income tax applied to incomes above threshold $\overline{w}$, τ_0 is the total tax on incomes below $\overline{w}$ (as discussed below, all wage earners in our sample have incomes between the first and second thresholds of the tax schedule), ϵ is the rate of tax evasion, and other parameters are defined as above. We are interested in two sets of NPVs: grade-specific outcomes with s years of schooling, NPV_s^0 (tax), and grade-specific outcomes with an additional year of schooling, NPV_s^1 (tax). Both are defined in a manner analogous to

[44] Because we do not observe children's schooling outcomes after they leave primary school, we are unable to measure impacts on the probability of completing Grade 8. Conservatively, we set the attainment impact for those who would complete Grade 7 in the absence of our study to zero, implying that our welfare estimates do not depend on NPV_7^0.

[45] The cross-sectional Mincerian earnings functions on which our estimates are built are in real terms. Inflation over the period 2010–2015 has averaged 8.6% in Uganda (IMF, 2017); combining this with a nominal discount rate of 10% implies a real discount rate even lower than 3.5 %. In addition to our preferred estimates, we also present estimates for a real discount rate of 5% as a point of comparison.

[46] Duflo (2001) uses a difference-in-difference approach based on year of enrollment and distance from newly constructed schools to estimate a return to education ranging between 6.8 and 10.6% for Indonesia. Other well-identified studies that use samples from developing countries (e.g. Card, 2001) typically find larger returns to education. Montenegro and Patrinos (2014) use observational wage and schooling data to calculate a returns to education of 15.9% in Uganda, compared to 12.4 in sub-Saharan Africa.

[47] We adopt a linear probability model specification for ease of calculating the marginal impact of an additional year of education, or experience on the probability of being employed. Nonetheless, the predicted probability of formal employment always remains between zero and one at all grades and experience levels.

[48] We note that it is plausible that productivity in the non-formal sector also improves with education, but to our knowledge there are no studies in a developing country context that credibly estimate this. Furthermore, in our sample, experience and education are not significantly correlated with household earnings from subsistence farming.

the earnings NPVs above. Again, our goal is to calculate the difference between NPV_s^1 (tax) and NPV_s^0 (tax) for each grade, and the sum of this net gain over grades $s = 1, \ldots, 6$.

C.3.2. Choice of parameters

We use the same parameters as in the earnings NPV model, with the exception of the discount rate r. Since the cost of government borrowing is central to the question of fiscal sustainability, we view the SOC method to be appropriate. Previous authors such as Ozier (2011) and Baird et al. (2016) have used a figure of 10 %, appealing to the high nominal cost of government funds. In our context, the cross-sectional earnings function from which we derive our estimates represents a *real* return. With inflation routinely running at 5 to 6% in Uganda, we opt instead for $r = 5\%$. To obtain an estimate for the rate of tax evasion, we combine World Bank estimates (World Bank, 2017) of 2.959 million wage and salaried employees in Uganda in 2014 with PAYE microdata from the Uganda Revenue Authority for the same year, in which an average of 477,118 employees pay tax in each month. This gives us a figure of $\epsilon = 83.88\%$. To calculate the tax take we apply the current PAYE schedule in Uganda. For wage earners with earnings over UShs 410,000 as distributed in our sample, this tax liability (in UShs) is $25,000 + 0.3 \cdot (w_{s,t} - 410,000)$.[49]

References

Adelman, M., Blimpo, M., Evans, D. K., Simbou, A., Yarrow, N., 2015. Can information technology improve school effectiveness in Haiti? Evidence from a field experiment. Unpublished, World Bank.

Aker, J.C., Ksoll, C., 2015. Call Me Educated: Evidence from a Mobile Monitoring Experiment in Niger. CGD Working Paper No. 406.

Baird, S., Hicks, J.H., Kremer, M., Miguel, E., 2016. Worms at work: long-run impacts of a child health investment. Q. J. Econ. 131 (4), 1637–1680.

Banerjee, A., Duflo, E., Glennerster, R., 2007. Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system. J. Eur. Econ. Assoc. 6, 487–500.

Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R., Khemani, S., 2010. Pitfalls of participatory programs: evidence from a randomized evaluation in education in India. Am. Econ. J. Econ. Pol. 2 (1), 1–30.

Best, M., Brockmeyer, A., Kleven, H., Spinnewijn, J., Waseem, M., 2015. Production vs revenue efficiency with limited tax capacity: theory and evidence from Pakistan. J. Polit. Econ. 123 (6), 1311–1355.

Bhatty, K., Saraf, R., 2016. Does Governments Monitoring of Schools Work? Technical Report, Centre for Policy Research.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Svensson, J., Wane, W., 2017. Enrolment without learning: teacher effort, knowledge, and skill in primary schools in africa. J. Econ. Perspect. 31 (4).

Callen, M.J., Gulzar, S., Hasanain, S.A., Khan, M.Y., 2016. The Political Economy of Public Employee Absence: Experimental Evidence from Pakistan. NBER Working Paper No. 22340.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-based improvements for inference with clustered errors. Rev. Econ. Stat. 90 (3), 414–427.

Campbell, D.T., 1979. Assessing the impact of planned social change. Eval. Program Plann. 2 (1), 67–90.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., Rogers, F.H., 2006. Missing in action: teacher and health worker absence in developing countries. J. Econ. Perspect. 20 (1), 91–116.

Chen, D., Glewwe, P., Kremer, M., Moulin, S., 2001. Interim report on a preschool intervention program in Kenya. Unpublished, Harvard University.

Dhaliwal, I., Hanna, R., 2014. Deal with the Devil: The Success and Limitations of Bureaucratic Reform in India. NBER Working Paper No. 20482.

Dixit, A.K., 1996. The Making of Economic Policy: A Transaction-cost Politics Perspective. The MIT Press, Cambridge, MA.

Duflo, E., 2001, September. Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment. Am. Econ. Rev. 91 (4), 795–813.

Duflo, E., Greenstone, M., Pande, R., Ryan, N., 2013. Truth-telling by third-party auditors and the response of polluting firms: experimental evidence from India. Q. J. Econ. 128 (4), 1499–1545.

Duflo, E., Hanna, R., Ryan, S.P., 2012. Incentives work: getting teachers to come to school. Am. Econ. Rev. 102 (4), 1241–1278.

Gerard, F., Gonzaga, G., 2014. Informal labor and the efficiency cost of social programs: evidence from 15years of unemployment insurance in Brazil. Unpublished, Columbia University.

Hanushek, E.A., Schwerdt, G., Wiederhold, S., Woessmann, L., 2015. Returns to skills around the world: evidence from PIAAC. Eur. Econ. Rev. 73, 103–130.

Treasury, H.M., 2011. The Green Book: Appraisal and Evaluation in Central Government. Technical Report, TSO.

IMF, 2017. International Financial Statistics. International Monetary Fund, data files last accessed online, March 2017.

Lee, D.S., 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. Rev. Econ. Stud. 76 (3), 1071–1102.

Linden, L.L., Shastry, G.K., 2012. Grain inflation: identifying agent discretion in response to a conditional school nutrition problem. J. Dev. Econ. 99, 128–138.

McKenzie, D., 2012. Beyond baseline and follow-up: the case for more T in experiments. J. Dev. Econ. 99 (2), 210–221.

Ministry of Education, Science, Technology and Sports, 2010. Value for Money Audit Report on Inspection of Primary Schools. Technical report, Government of Uganda.

Ministry of Education, Science, Technology and Sports, 2014. Inspection Grant: Planning, Implementing and Accountability Guidelines. Technical report, Government of Uganda.

Montenegro, C.E., Patrinos, H.A., 2014. Comparable Estimates of Returns to Schooling Around the World. World Bank Policy Research Working Paper,

Muralidharan, K., Das, J., Holla, A., Mohpal, A., 2017. The fiscal cost of weak governance: evidence from teacher absence in india. J. Public Econ. 145, 116–135.

Muralidharan, K., Niehaus, P., Sukhtankar, S., 2016. Building state capacity: evidence from biometric smartcards in India. Am. Econ. Rev. 106 (10), 2895–2929.

Muralidharan, K., Sundararaman, V., 2011. Teacher performance pay: experimental evidence from India. J. Polit. Econ. 119 (1), 39–77.

Neal, D., 2013. The consequences of using one assessment system to pursue two objectives. J. Econ. Educ. 44 (4), 339–352.

Ozier, O., 2011. Exploiting externalities to estimate the long-term effects of early childhood deworming. Unpublished, University of California, Berkeley.

Rothstein, J., 2011, January. Review of *Learning about Teaching*. National Education Policy Center, Think Twice Think Tank Review Project,

Sandefur, J., Glassman, A., 2015. The political economy of bad data: evidence from African survey and administrative statistics. J. Dev. Stud. 51 (2), 116–132.

Suri, T., Jack, W., 2016. The long-run poverty and gender impacts of mobile money. Science 354 (6317), 1288–1292.

Suri, T., Jack, W., Stoker, T.M., 2012. Documenting the birth of a financial economy. Proc. Natl. Acad. Sci. 109 (26), 10257–10262.

Uganda Bureau of Statistics, 2012. Uganda National Panel Survey 2011/12.

VA Office of Inspector General, 2014. Review of Patient Waiting Times, and Alleged Patient Deaths at the Phoenix Health Care System. Interim Report, Veterans Health Administration.

Wane, W., Martin, G.H., 2013. Education and health services in Uganda: data for results and accountability. World Bank, Service Delivery Indicators Launch Report, Washington DC.

World Bank, 2016. Digital Dividends. Technical report, World Bank.

World Bank, 2017. World development indicators. Source: http://data.worldbank.org. Data files last accessed online, September

[49] In this and in all other calculations, we use an exchange rate of 2648 UShs to 1 USD.