EDI WORKING PAPER SERIES WP/2022

TEACHER ROTATION AND STUDENT OUTCOMES: EXPERIMENTAL EVIDENCE FROM UGANDA

Isabelle Cohen
University of Washington
Ernesto Dal Bó
UC Berkeley
Frederico Finan
UC Berkeley
Kizito Omala
Makerere University
David Schonholzer
Stockholm University

[PRELIMINARY: DO NOT CITE OR CIRCULATE]

December 2021

Abstract

This report summarizes an ongoing study of the effect of teacher rotation on student outcomes in Uganda. In the status quo, teachers are transferred ("rotated") across schools at the discretion of district authorities. In randomized treatment schools, we incentivize teachers by linking upcoming transfers to student performance on standardized tests. If their students do better than comparable students at control schools, they are more likely to be transferred to a school of their choice. Otherwise, their transfer remains in the hands of district authorities (treatment "PD"), or they are randomly assigned (treatment "PR"). Preliminary findings indicate no effect of either treatment on student outcomes, teacher attendance, or teacher behavior in the classroom. We interpret the absence of treatment effects primarily as a result of (1) only modest changes in teacher beliefs at treatment schools about the effect of student achievement on their upcoming transfer; (2) surprisingly high levels of baseline teacher engagement, limiting their capacity to improve student outcomes; and (3) the COVID-19 pandemic preventing the establishment of district-wide practices and expectations regarding incentivized teacher transfers.

Cohen: University of Washington, Evans School of Public Policy & Governance (imcohen@uw.edu).
Dal Bo: UC Berkeley, Haas School of Business (edalbo@berkeley.edu).
Finan: UC Berkeley, Department of Economics (finan@berkeley.edu).
Omala: Department of Statistics and Actuarial Science, Makerere University (kizomala@bams.mak.ac.ug).
Schonholzer: Institute for International Economic Studies, Stockholm University (david.schonholzer@iies.su.se).

We express our gratitude to the district in Uganda that partnered with us for this intervention, and especially to the Chief Educational Officer of the district. We are grateful for comments by seminar audiences at UC Berkeley, UCSD, and EASST Summits. We thank IPA Uganda, especially Vianney Mbonigaba, Brian Mwesigwa and Chiara Pasquini for assistance in the _eld, and Chiman Cheung for outstanding research assistance. We gratefully acknowledge financial support from EDI, DIV, IGC, JPAL and EASST. All errors and omissions are our own.

1 Background

Low quality education is a widespread and persistent global phenomenon. Literacy rates of people aged 15 and above were only 64% in sub-Saharan Africa, and only 70% in South and West Asia (UNESCO, 2015). In the context of our innovation, Uganda, educational attainment is similarly low: only three out of ten Ugandan students can read and comprehend a simple story by grade three, and only eight out of ten can do so by grade seven (Uwezo Uganda, 2012). By the time Ugandan pupils reach adolescence, more than one in four have dropped out of school; the average Ugandan has fewer than eight years of schooling (UNESCO, 2015).

The schooling environment surrounding pupils in poor countries is a likely contributor to low attainment: school facilities are often low quality, and teachers are frequently poorly trained, poorly motivated, absent, or some combination thereof. Teacher absenteeism is a common concern in the many parts of the developing world, where teachers may show up only four out of five days a week. In Uganda, past studies documented teacher absenteeism rates of around 27% (Transparency International, 2013). Currently, more than seven million children attend school in Uganda and experience this low level of educational quality; globally, more than a billion children are subject to similar conditions (UNESCO, 2015).

Because education generates hard-to-internalize spillovers, governments play a central role in providing education services. As a result, improving government service provision is essential for improving education. Recent research in economics has identified the role of public employees and the incentives they face as a key aspect of government service provision (see Finan et al., 2015 for an overview). Teachers, specifically, have been shown to react positively to incentives, opening up a promising avenue for improvements in education (Muralidharan and Sundararaman, 2011; Duflo et al., 2012), though evidence on their efficacy in East Africa is mixed (Leaver et al, 2021). As government budgets are increasingly tightened, low-cost innovations in teacher incentives are especially worthwhile to investigate and expand.

Our innovation to improve education consists of an incentivized teacher transfer scheme. A transfer scheme organizes a periodic reallocation of public employees across job locations.

There is a long tradition of using transfer schemes to discipline or incentivize public employees, ranging from provincial assignments in ancient China all the way to country rotation among World Bank bureaucrats today. Understanding how transfer schemes can be employed successfully by governments is a current frontier of research in economics, with the only randomized evaluation to date focusing on tax collectors in Pakistan (Khan et al., 2018). That study shows promising results but does not address several of the questions motivating our study; in particular, we look at frontline workers with a much more complex function mapping effort to outcomes, and use an innovative treatment design, discussed below, that attempts to unpack the status quo (with an ex-ante design focus on connections and patronage). Despite the widespread usage of transfer schemes for teachers (including Uganda, Kenya, the Gambia, Singapore, and China), little is known about how to design and deploy teacher transfer schemes in a way that maximizes its potential for educational quality improvement.

Common ways of improving teaching quality, such as hiring additional teachers or increasing salaries are expensive. Teacher transfer schemes are a cheaper alternative, requiring only minimal support for job reallocation. In fact, given that our innovation is an incentivized teacher transfer scheme that builds on an existing transfer scheme, it may be largely cost-neutral for governments who already have any transfer scheme in place.

The core of this research program is a randomized controlled trial (RCT) to evaluate the effect of teacher transfers on educational attainment over a period of three years, located in a relatively small, rural district in Eastern Uganda. Teachers at all schoools were designated as eligible or ineligible for transfer, based on government statues mandating that five years should have elapsed at a posting prior to transfer. We then randomly allocated schools in our study district into one of two treatment groups or a control group. In the treatment schools, eligible teachers are incentivized based on student improvement in standardized test scores, using percentile value added. The higher a teacher's value added, the higher their chances of receiving a transfer to a pre-specified preferred school. By comparing the performance of eligible teachers (and their students) at treatment schools and to that of teachers at control schools, we estimate the effect of the transfer scheme on teacher and student outcomes.

In both sets of treatment schools, there are teachers who receive their desired transfers ("winners") and teachers who do not ("non-winners"). The difference between the two treatment groups is what happens to the non-winners. In one treatment group, the district transfers non-winners as they had done previously. In the second treatment group, non-winners are transferred in a randomized manner, which fully severs transfers from both connections and performance.

2 Data

The study had two full rounds of data collection in 2018, with a comprehensive baseline in February and March and a full set of follow-up surveys at the end of the school year, November (midline 1). We repeated the sequence of surveys in 2019, with midline 2 in February and midline 3 in November.

For each school, we collected data on classroom, infrastructure, and other important aspects of their environment. For teachers, we tracked attendance using a district-supported attendance book, verifying its accuracy with our own spot checks. We conducted detailed classroom observations, assessing behavior within the classroom, engagement with students, and pre-lesson preparation. We also collected data on teachers' demographics, an ordered list of their five most preferred schools, self-reported and perceived connection with the district authorities, and their understanding and belief about the program. [1] We tested all students in the district twice per year on their knowledge of literacy and numeracy (for younger students) and literacy, numeracy, science and social studies (for older students).

The 140 schools in our study district were randomly assigned to either the Control group, the Performance and District treatment group (PD schools), or the Performance and Random treatment group (PR schools) during a public lottery in late March 2018. Teachers who had stayed at a school for five years already or who would complete five years during the study were eligible for transfer, and randomly assigned to a transfer year (2019, 2020, or 2021, at the beginning of Uganda's March-December school year) conditional on having spent sufficient time in their posting. Teachers who had stayed at their current posting less than three years as of the beginning of the study were deemed ineligible.

[1] Notably, teacher preferences for schools at baseline were heterogeneous enough to ensure that, according to simulations, we would be able to place all winners to one of their preferred schools with high probability.

For all analyses, unless otherwise specified, we exclude teachers ineligible to be transferred. For baseline and year 1 results, teachers assigned to be transferred in 2019, 2020 and 2021 are included; since the 2019 cohort has already been transferred by year 2, only teachers assigned to be transferred in 2020 and 2021 are included in the year 2 results.

2.1 Matching across waves

To measure student improvement, we attempted to track every student over time, but it proved difficult in this context due to the lack of a centralized government system that uniquely identifies students. Our enumerators assigned a unique student ID to every student and we relied on the teachers to track the students in subsequent waves. As shown in Table 1, the overall matching rate between baseline and midline 1 is 66 percent. Table 2 presents the matching rate between baseline and midline 3, based on the within-study student identification number.

The overall matching rate of 41 percent is lower for four reasons: (1) many students drop out during primary school, and therefore leave the sample; (2) in practice, students who changed schools were often difficult to link to their pre-change student ID; (3) this match spans two academic years so it is likely that students are no longer taught by the same teacher, meaning younger students in particular sometimes were not successfully linked; (4) most students who were in seventh grade, the final year of primary school, at baseline had graduated by the end of the second year. Data on the number of students who took the exams in each year suggest a dropout rate of 16% on aggregate between 2018 and 2019, though this does not account for students who may have transferred in and out of the district; the gap between this and the matching rate is thus an upper bound on the true proportion of unmatched students. Importantly, we do not find any treatment effects on the propensity of students to be matched between years.

Nevertheless, conditional on having a match, baseline exam scores do predict exam scores measured in midlines reasonably well. Figure 1 is a binscatter that shows the positive relationship between baseline scores and midline 1 scores, which are correlated at 0.60. Similarly, Figure 2 relates baseline scores to midline 3 scores, which are correlated at 0.43.

3 Empirical specification

To test for baseline balance in school-level covariates measured in early 2018 ($t = 0$), we estimate the following specification.

$$y_{s0} = \alpha + \beta_1 treatment\_PD_s + \beta_2 treatment\_PR_s + \varepsilon_{s0} \tag{1}$$

The outcome variable y_{s0} is the school-level variable measured at baseline. $treatment\_PD_s$ and $treatment\_PR_s$ are the treatment indicators for PD schools and PR schools respectively. We report robust standard errors.

To test for baseline balance in exam scores, we estimate the following specification.

$$y_{is0} = \alpha + \beta_1 treatment\_PD_s + \beta_2 treatment\_PR_s + \varepsilon_{is0} \tag{2}$$

The outcome variable y_{is0} is the baseline exam score defined at the student-grade-subject level. The treatment indicators are defined the same as equation (1). We report standard errors clustered at the school level, the level where the treatment is assigned.

In the reduced-form exam score analyses in Section 5, the specification is as follows.

$$\begin{aligned} y_{ist} = {} & \alpha + \beta_1 treatment\_PD_s + \beta_2 treatment\_PR_s \\ & + \delta_1 y_{is0} + \delta_2 treatment\_PD_s \times y_{is0} + \delta_3 treatment\_PR_s \times y_{is0} + \varepsilon_{ist} \end{aligned} \tag{3}$$

We will first present year 1 results, comparing exam-level student performance at midline 1 in late 2018 ($t = 1$). They are followed by year 2 results at midline 3 in late 2019 ($t = 3$). We control for the interaction of treatment indicators and baseline scores throughout, flexibly allowing the impact of baseline scores to be different by treatment groups; exam scores are standardized so that β_1 and β_2 recover the average treatment effect. We report standard errors clustered at the school level.

In the reduced-form teacher input analyses in Section 6, the specification is as follows.

$$y_{jst} = \alpha + \beta_1 treatment\_PD_s + \beta_1 treatment\_PR_s + \delta_1 y_{js0} + \varepsilon_{jst} \quad (4)$$

The outcome variable is the input of teacher j in school s. We control for the lagged dependent variable measured at baseline. We report standard errors clustered at the school level.

4 Baseline analyses

Table 3 shows that all except two school characteristics measured at baseline are statistically indistinguishable between the two treatment groups and the control group. Relative to the control group, PR schools are 18pp less likely to provide teacher accommodation, and PD schools are 11pp more likely to use boreholes as the major water source; this level of imbalance is consistent with what one would expect given the number of outcomes tested, though the point estimates are reltaively large.

However, we do also find evidence of a strong and persistent imbalance in baseline test scores across the treatment groups, with meaningfully lower baseline test scores in PD schools. The evidence suggests that, across a variety of specifications, PD treatment students are consistently performing worse relative to the control students at baseline by between .2 and .3 standard deviations. Although we find a negative coefficient for students in PR schools around the size of .1 standard deviations, this difference is not statistically significant. Table 4, Table 5, and Table 6 document the imbalance in baseline performance by teachers' years of scheduled transfers, grades, and subjects respectively. All exam scores analyses, including baseline imbalance results and all treatment effect results in the next section, use data at the student-grade-subject level unless otherwise specified. All exam scores are normalized using the grade-subject level control schools' distribution.

Since the randomization was done after we collected data at baseline, we can rule out the potential explanation where treated teachers gamed the treatment by manipulating students' performance at baseline. To ensure such a statistically significant imbalance is not an artifact of clustering standard errors at the school level, we performed a randomization inference analysis. The randomization inference p-value, which can be interpreted as the probability that such an imbalance would have materialized, is 4.6%. Overall, after exhaustive examination, we believe the imbalance reflects a fluke occurrence due to chance rather than problems with implementation. To account for the imbalance, we control for the interaction terms between baseline scores and treatment status in all exam score analyses, flexibly allowing baseline performance to have different impacts on subsequently measured performance.

5 Null results on exam scores

Table 7 and Table 8 present year 1 and year 2 results on treatment effects on student performance. In both tables, the first two columns include all students taught by teachers scheduled to be transferred in the future. Subsequent columns separately consider teachers to be transferred in different years. Overall, the treatment effects of both arms are mostly negative but imprecisely estimated.

In Table 9 to Table 12, we cut the sample in alternative ways; we estimate year 1 and year 2 treatment effects by subjects and grades, respectively. Similar to results from Table 5 and Table 6, the point estimates are mostly negative but not statistically significant.

Table 13 presents heterogeneity analysis on the year 1 results. We focus on variation we believe ex-ante might strengthen or weaken the treatment. In particular, we look at commuting distance as a proxy for the desirability of a teacher's current posting, and different measures of connectivity to assess how undesirable being a "non-winner" might be; in this context, we expect that teachers who are better connected to district officials might expect to receive a desirable posting even in the absence of performing well. Columns (1) and (2) report the effects for teachers whose commuting distance is below median and above median. Columns (3) and (4) report the effects for teachers without and with any of their spouses or relatives working in the local committee. Columns (5) and (6) report the effects for teachers whose perceived connection with the district authorities is below and above median.[2] None of the estimates stand out to be statistically significant or economically

[2]The connection measure used in columns (3) and (4) are self-reported at baseline. The perceived connection measure used in columns (3) and (4) are elicited in midline 1. Specifically, teachers are asked to rate

meaningly in magnitude.

Table 14 presents the year 2 heterogeneity analysis. The PD treatment effect is significantly negative for teachers living farther away from school ($\hat{\beta}_1 = -0.21*$) and for teachers perceived to be more connected with the district authorities ($\hat{\beta}_1 = -0.22**$).

In Table 15 and Table 16, we use an alternative method to account for baseline imbalance. In odd columns, we implement the inverse-probability weighted procedure by inversely weighting observations by their treatment probabilities estimated with baseline score as the only explanatory variable in a multinomial logit model. To ensure correct inference, the data are first aggregated to the school level and we report robust standard errors. For ease of comparison, we also present results based on the unweighted sample. The coefficient estimates based on these two different weights always have the same sign and similar magnitudes.

6 Null results on teacher input

Table 17 presents year 1 results on treatment effects on teachers' classroom observation scores. The score is calculated as the first component in the principal component analysis that combines the following classroom observation variables: number of days that the teacher has a lesson plan, whether the teacher gives homework, whether students are allowed to ask questions, and average student engagement. Table 18 and Table 19 report year 1 results on treatment effects on whether the teacher engages in agriculture or gardening at home and non-teaching income. [3]None of the estimates are statistically significant, indicating an absence of treatment effects on input in year 1.

Table 20 and Table 21 present year 1 and year 2 results on treatment effects on teachers' presence, measured directly by the surveying team during their visits to the school.[4] The PD treatment effect is marginally significant in year 2 for teachers to be transferred in 2020. Overall, we do not find any treatment effect that is robust across specifications and

how easy it is for each of their colleagues to ask for favors from the district authorities, and these ratings are amalgamated into a teacher-specific score.

[3]Only year 1 results are available for these three outcomes as we did not measure the same variable during the 2019 surveys.

[4]In general, the survey team did notify schools that a visit would be taking place within a specified period, but did not specify the exact day a visit would take place; this was done to balance minimizing disruptions to the school while also trying to prevent Hawthorne effects on teacher presence.

estimating samples. The analysis on presence also reveals a surprising pattern: the control means imply an absenteeism rate of around 10%, in contrast to 27% as documented in an report by Transparency International in 2013. It suggests that teacher absenteeism might be less of an issue than previously thought in this context, and further may imply that teachers in Uganda may lack straightforward margins upon which to increase effort.

7 Mixed results on treatment understanding and beliefs

At midline 1, we asked five questions to test teacher understanding of the treatment design. Only 11.5 percent of the PD teachers and 5.3 percent of the PR teachers answered all five questions correctly.[5] In response to their poor understanding, we reminded all teachers of their own and their school's treatment status to reinforce their understanding of the transfer incentive environment.

Table 22 shows that treated teachers are more likely to believe that "teachers in their school are eligible to be rewarded based on performance and therefore being transferred to a school of preference", as measured at midline 1. However, considering the relatively large control mean of 0.90, the effect sizes are small in magnitude and that of the PR treatment group becomes statistically insignificant when we restrict the sample to only teachers eligible for transfer.

Table 23 presents the treatment effects on belief about transfers, as measured at midline 3. It shows that the treatments change neither the perceived relationship between performance and whether a teacher is transferred nor the relationship between performance and being transferred to a better school.[6] The sample includes teachers eligible to be transferred between 2019 and 2021.

Table 24 reports the treatment effects on the perceived most important reason of being

[5]Q1. Are teachers in this school eligible to be rewarded based on performance and therefore be transferred to a school of preference? Yes or no; Q2. What happens if a teacher of this school does not perform well? District decides where to transfer or transfer is done randomly; Q3. All pupils count equally towards a teacher's score under the pupil growth reward scheme. True or false; Q4. Pupils' grades are compared with those of other pupils. True or false; Q5. Teachers who were transferred to their current school after 2016 are not yet eligible for transfer. True or false.

[6]Note that a better school (the wording used in the survey) might be interpreted differently than school of preference (the wording of the dependent variable in the last table).

transferred to a preferred school, as measured at midline 1. Each column represents one of the four reasons teachers can rank as most important. Column (2) shows that treated teachers are more likely to think that student performance is the most important.

Table 25 and Table 26 report the effects as measured at midline 2 and midline 3. The results are similar to Table 24. A pattern worth noting is that over waves, the control teachers believe that student performance becomes less important and become more likely to believe that preferred transfers happen by chance; beliefs do not shift for treatment teachers. This may suggest some shifting of beliefs over time as teachers learned more about the new transfer system.

8 Conclusion

While we have to contend with moderate imbalance in baseline student achievement across treatment conditions, the preliminary findings presented here nevertheless indicate an absence of treatment effects on student outcomes and teacher behaviors. This may reflect several (not necessarily mutually exclusive) factors, each suggesting different approaches to the redeployment of a rotation scheme for teachers in the future. We focus on three possible explanations for the lack of treatment effects: teacher beliefs about transfers, teacher capacity to improve student achievement, and the disruption of field activities due to the COVID-19 pandemic. We discuss each of these in turn.

First, although treated teachers are consistently more likely to rate student performance as the most important factor in getting a desired post upon reassignment, the effect sizes are relatively modest and are unlikely to generate any downstream effects on teacher behavior and student outcomes. In addition, there is a district-wide trend over time indicating that student achievement is perceived to be less important by all teachers. Control teachers increasingly consider the status quo dominated by chance.[7] These issues highlight the need for (i) continuous reinforcement of treatment status to promote understanding and salience among teachers; and (ii) more frequent assessments on teacher beliefs about the treatment

[7] While contamination across treatment and control schools is in principle possible, given the absence of regular interactions of teachers at different schools, this is unlikely to drive the similarity of teacher beliefs at treatment and control schools.

design to ensure treatments work as intended.

Second, district authorities and teachers themselves do believe at baseline that teacher dedication matters for transfer decisions. In line with these beliefs, teachers exhibit a somewhat surprising level of engagement in terms of attendance even at baseline. This suggests that teachers may be failing not in terms of exerting effort, but rather in adapting their pedagogic approach towards higher student achievement. In order to do this, teachers may need not just incentives, but also a blueprint on how to translate their effort into better student outcomes. In this interpretation, a richer intervention affecting both the objectives and the means available to teachers would be required.

Finally, while the study was designed to include three cohorts of rotating teachers, we were forced to halt all field activities before the third cohort was up for transfer due to school closures on the basis of Covid-19. For this third cohort of teachers, we had planned additional activities to reinforce the mechanics of the research design, such as posters at treatment schools of testimonials of teachers transferred to their preferred school. It is possible that teachers would have had a better understanding and greater faith in the promised transfers being implemented after observing them for the previous cohorts. We hope to build on these insights in future work on teacher rotation in Ugandan schools.

9 Tables

Table 1: Matching between baseline and midline 1

	matched obs.	total obs.	matching rate
P1	14328	27142	0.53
P2	13256	20121	0.66
P3	13978	21669	0.65
P4	28359	42231	0.67
P5	25840	37242	0.69
P6	20366	29797	0.68
P7	12123	16115	0.75
Total	128250	194317	0.66

Table 2: Matching between baseline and midline 3

	matched obs.	total obs.	matching rate
P1	10334	27142	0.38
P2	9564	20121	0.48
P3	10366	21669	0.48
P4	19570	42231	0.46
P5	16480	37242	0.44
P6	12360	29797	0.41
P7	1249	16115	0.08
Total	79923	194317	0.41

Table 3: Balance table of school characteristics

	(1)	(2)	(3)	(4)	(5)	
able	Mean Control	Mean PD	Mean PR	PD vs Control	PR vs Control	PD
ber of students	589.130	523.723	542.000	-65.407	-47.130	-
	(216.332)	(206.836)	(222.830)	(43.883)	(45.554)	(4
ernment teachers per student	0.019	0.022	0.020	0.002	0.001	
	(0.007)	(0.010)	(0.008)	(0.002)	(0.002)	(
munity teacher per student	0.002	0.002	0.002	-0.000	0.000	-
	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(
ol management committee members per student	0.024	0.028	0.027	0.004	0.003	
	(0.012)	(0.013)	(0.014)	(0.003)	(0.003)	(
member per student	0.017	0.018	0.018	0.001	0.000	
	(0.012)	(0.010)	(0.010)	(0.002)	(0.002)	(
nanent classrooms per student	0.015	0.016	0.015	0.000	-0.000	
	(0.008)	(0.006)	(0.006)	(0.001)	(0.001)	(
porary classrooms per student	0.002	0.003	0.001	0.001	-0.001	0
	(0.004)	(0.004)	(0.002)	(0.001)	(0.001)	(
s per student	0.298	0.304	0.274	0.006	-0.024	
	(0.184)	(0.167)	(0.171)	(0.036)	(0.037)	(
s per student	0.016	0.018	0.016	0.002	-0.001	
	(0.009)	(0.012)	(0.007)	(0.002)	(0.002)	(
ol garden area per student (acre)	0.003	0.002	0.002	-0.000	-0.000	
	(0.002)	(0.002)	(0.003)	(0.000)	(0.001)	(
overnment fund per student)	9.100	9.173	9.211	0.074	0.112	-
	(0.419)	(0.423)	(0.335)	(0.091)	(0.082)	(
her accomodation	0.500	0.362	0.319	-0.138	-0.181*	
	(0.506)	(0.486)	(0.471)	(0.103)	(0.101)	(
her attending training by NGO in 2017	0.674	0.723	0.723	0.049	0.049	-
	(0.474)	(0.452)	(0.452)	(0.096)	(0.096)	(
ive cash or in-kind support form NGO in 2017	0.578	0.532	0.468	-0.046	-0.110	

	(0.499)	(0.504)	(0.504)	(0.105)	(0.105)	(
Hole as the major water source	0.848	0.957	0.851	0.110*	0.003	0
	(0.363)	(0.204)	(0.360)	(0.061)	(0.075)	(
Spring as the major water source	0.087	0.043	0.043	-0.044	-0.044	
	(0.285)	(0.204)	(0.204)	(0.051)	(0.051)	(
d Water as the major water source	0.022	0.000	0.043	-0.022	0.021	-
	(0.147)	(0.000)	(0.204)	(0.022)	(0.037)	(
Water as the major water source	0.043	0.000	0.021	-0.043	-0.022	-
	(0.206)	(0.000)	(0.146)	(0.030)	(0.037)	(
Water as the major water source	0.000	0.000	0.021	0.000	0.021	-
	(0.000)	(0.000)	(0.146)	(0.000)	(0.022)	(
Vater Source as the major water source	0.000	0.000	0.021	0.000	0.021	-
	(0.000)	(0.000)	(0.146)	(0.000)	(0.022)	(
ance to Lake Victoria (km)	21.983	23.397	24.337	1.414	2.354	-
	(10.775)	(8.745)	(9.256)	(2.033)	(2.082)	(
ance to highway A109 (km)	8.754	7.151	7.407	-1.603	-1.347	-
	(5.610)	(4.918)	(5.497)	(1.093)	(1.152)	(
ance to town (km)	13.285	11.963	11.909	-1.322	-1.377	
	(6.040)	(4.919)	(5.670)	(1.141)	(1.215)	(
difference between exam date and the first day of exam	9.370	10.426	10.191	1.056	0.822	
	(6.035)	(5.683)	(6.639)	(1.215)	(1.316)	(
ervations	46	47	47	93	93	

Table 4: Baseline imbalance by year of scheduled transfers

	(1)	(2)	(3)	(4)
Dependent Variable		Baseline 1 z-score		
PD treatment	-0.190**	-0.231**	-0.269**	-0.038
	(0.088)	(0.109)	(0.103)	(0.089)
PR treatment	-0.090	-0.130	-0.167	0.064
	(0.103)	(0.125)	(0.128)	(0.109)
N	86092	29182	29746	27164
Transfer year	All	2019	2020	2021
R-squared	0.006	0.009	0.013	0.002

Dependent variables are baseline exam scores standardized using subject-grade level distributions of the control schools.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 5: Baseline imbalance by grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable				Baseline z-score			
PD treatment	-0.187+	-0.210+	-0.179+	-0.170*	-0.168	-0.190	-0.268*
	(0.118)	(0.134)	(0.118)	(0.0898)	(0.135)	(0.140)	(0.140)
PR treatment	-0.00809	-0.193+	-0.102	-0.0951	-0.0686	-0.0538	-0.162
	(0.115)	(0.129)	(0.136)	(0.104)	(0.153)	(0.174)	(0.161)
N	12513	10363	9935	16483	15904	12798	8096
Grade	1	2	3	4	5	6	7
R-squared	0.007	0.009	0.005	0.006	0.005	0.006	0.012

Dependent variables are baseline exam scores standardized using subject-grade level distributions of the control schools.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 6: Baseline imbalance by subject

	(1)	(2)	(3)	(4)
Dependent Variable		Baseline z-score		
PD treatment	-0.185**	-0.167*	-0.261**	-0.176*
	(0.0869)	(0.0919)	(0.115)	(0.104)
PR treatment	-0.115	-0.0982	-0.0960	0.00332
	(0.0970)	(0.100)	(0.162)	(0.124)
N	28282	30498	14246	13066
Subject	English	Math	Science	Social Studies
R-squared	0.006	0.005	0.012	0.008

Dependent variables are baseline exam scores standardized using subject-grade level distributions of the control schools.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 7: Year 1 treatment effects on students' performance: by year of scheduled transfers

	(1)	(2)	(3)	(4)	(5)
Dependent Variable			Midline 1 z-score		
PD treatment	-0.162*	-0.057	-0.054	-0.129*	0.029
	(0.094)	(0.052)	(0.061)	(0.075)	(0.059)
PR treatment	-0.073	-0.018	-0.002	-0.079	0.048
	(0.115)	(0.063)	(0.060)	(0.103)	(0.076)
N	86092	86092	29182	29746	27164
Control	No	Yes	Yes	Yes	Yes
Transfer year	All	All	2019	2020	2021
R-squared	0.004	0.356	0.340	0.376	0.352

Dependent variables are midline 1 exam scores standardized using subject-grade level distributions of the control schools.

Control variables include baseline z-scores interacted with treatment status.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 8: Year 2 treatment effects on students' performance: by year of scheduled transfers

	(1)	(2)	(3)	(4)
Dependent Variable	Midline 3 z-score			
PD treatment	-0.123	-0.065	-0.060	-0.059
	(0.103)	(0.083)	(0.099)	(0.097)
PR treatment	-0.036	0.007	-0.085	0.108
	(0.151)	(0.111)	(0.119)	(0.125)
N	27785	27785	14918	12867
Control	No	Yes	Yes	Yes
Transfer year	All	All	2020	2021
R-squared	0.003	0.172	0.185	0.162

Dependent variables are midline 3 exam scores standardized using subject-grade level distributions of the control schools.

Control variables include baseline z-scores interacted with treatment status.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 9: Year 1 treatment effects on students' performance: by grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable	Midline 1 z-score						
PD treatment	-0.168**	-0.061	-0.040	-0.019	-0.106**	0.009	0.013
	(0.074)	(0.098)	(0.125)	(0.071)	(0.049)	(0.086)	(0.088)
PR treatment	-0.142**	-0.073	0.017	-0.033	-0.047	0.035	0.178*
	(0.069)	(0.099)	(0.139)	(0.071)	(0.069)	(0.101)	(0.101)
N	12513	10363	9935	16483	15904	12798	8096
Grade	1	2	3	4	5	6	7
R-squared	0.346	0.321	0.361	0.265	0.414	0.422	0.421

Dependent variables are midline 1 exam scores standardized using subject-grade level distributions of the control schools.

Control variables include baseline z-scores interacted with treatment status.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 10: Year 1 treatment effects on students' performance: by subject

	(1)	(2)	(3)	(4)
Dependent Variable	Midline 1 z-score			
PD treatment	-0.040	-0.106*	-0.043	0.023
	(0.058)	(0.061)	(0.064)	(0.067)
PR treatment	-0.068	-0.038	0.092	0.029
	(0.063)	(0.063)	(0.119)	(0.080)
N	28282	30498	14246	13066
Subject	English	Math	Science	Social Studies
R-squared	0.405	0.300	0.412	0.338

Dependent variables are midline 1 exam scores standardized using subject-grade level distributions of the control schools.

Control variables include baseline z-scores interacted with treatment status.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 11: Year 2 treatment effects on students' performance: by grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable				Midline 3 z-score			
PD treatment	-0.074	0.036	-0.037	-0.093	-0.059	-0.068	-0.082
	(0.154)	(0.166)	(0.153)	(0.131)	(0.122)	(0.107)	(0.131)
PR treatment	0.050	-0.110	0.040	-0.124	0.077	0.025	0.130
	(0.185)	(0.159)	(0.246)	(0.127)	(0.121)	(0.137)	(0.212)
N	1323	4111	4661	4400	5274	4856	3160
Grade	1	2	3	4	5	6	7
R-squared	0.111	0.116	0.097	0.138	0.154	0.393	0.234

Dependent variables are midline 3 exam scores standardized using subject-grade level distributions of the control schools.

Control variables include baseline z-scores interacted with treatment status.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 12: Year 2 treatment effects on students' performance: by subject

	(1)	(2)	(3)	(4)
Dependent Variable		Midline 3 z-score		
PD treatment	-0.019	-0.045	-0.132	-0.094
	(0.096)	(0.100)	(0.119)	(0.117)
PR treatment	0.017	-0.006	0.008	0.061
	(0.119)	(0.125)	(0.144)	(0.163)
N	9943	10132	3835	3875
Subject	English	Math	Science	Social Studies
R-squared	0.183	0.114	0.285	0.220

Dependent variables are midline 3 exam scores standardized using subject-grade level distributions of the control schools.

Control variables include baseline z-scores interacted with treatment status.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 13: Heterogeneity analysis: year 1 treatment effects on students' performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Midline 1 z-score					
	below median dist.	above median dist.	unconnected	connected	below median perceived conn.	above median perceived conn.
PD treatment	-0.062	-0.047	-0.063	0.009	-0.036	-0.077
	(0.067)	(0.064)	(0.054)	(0.073)	(0.065)	(0.065)
PR treatment	0.001	-0.019	-0.011	0.115	0.010	-0.033
	(0.073)	(0.081)	(0.063)	(0.088)	(0.068)	(0.082)
N	41188	33324	52905	15354	47609	34470
R-squared	0.364	0.338	0.355	0.371	0.349	0.368

Dependent variables are midline 1 exam scores standardized using subject-grade level distributions of the control schools.

Control variables include baseline z-scores interacted with treatment status.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 14: Heterogeneity analysis: year 2 treatment effects on students' performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Midline 3 z-score					
	below median dist.	above median dist.	unconnected	connected	below median perceived conn.	above median perceived conn.
PD treatment	0.088	-0.210*	-0.045	-0.025	0.032	-0.219**
	(0.107)	(0.110)	(0.082)	(0.153)	(0.097)	(0.099)
PR treatment	0.065	-0.049	0.030	0.231	0.026	-0.023
	(0.131)	(0.141)	(0.109)	(0.211)	(0.122)	(0.138)
N	12609	11713	17589	5307	17159	10556
R-squared	0.177	0.167	0.180	0.180	0.135	0.245

Dependent variables are midline 3 exam scores standardized using subject-grade level distributions of the control schools.

Control variables include baseline z-scores interacted with treatment status.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 15: Year 1 treatment effects on students' performance: school-level estimation

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable			Midline 1 z-score			
PD treatment	0.023	-0.009	-0.060	-0.047	0.048	0.024
	(0.055)	(0.058)	(0.057)	(0.066)	(0.054)	(0.062)
PR treatment	0.098*	0.081	-0.121**	-0.110^{+}	-0.041	-0.030
	(0.056)	(0.059)	(0.058)	(0.072)	(0.054)	(0.063)
N	123	123	123	123	121	121
Transfer year	2019	2019	2020	2020	2021	2021
Weight	Yes	No	Yes	No	Yes	No

Dependent variables are school-level average standardized midline 1 exam scores.

Observations in the odd columns are inversely weighted by treatment probabilites estimated with baseline scores; observations in even columns are unweighted.

Control variables include baseline scores interacted with treatment status.

Robust standard errors in parentheses.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 16: Year 2 treatment effects on students' performance: school-level estimation

	(1)	(2)	(3)	(4)
Dependent Variable	Midline 3 z-score			
PD treatment	0.056	0.013	-0.004	-0.015
	(0.091)	(0.098)	(0.074)	(0.079)
PR treatment	-0.033	-0.004	0.099	0.109
	(0.093)	(0.099)	(0.079)	(0.084)
N	117	117	113	113
Transfer year	2020	2020	2021	2021
Weight	Yes	No	Yes	No

Dependent variables are school-level average standardized midline 3 exam scores.

Observations in the odd columns are inversely weighted by treatment probabilites estimated with baseline scores; observations in even columns are unweighted.

Control variables include baseline scores interacted with treatment status.

Robust standard errors in parentheses.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 17: Year 1 treatment effects on teachers' classroom observation scores

	(1)	(2)	(3)	(4)
Dependent Variable	Midline 1 classroom obs. scores			
PD treatment	0.192	0.309	-0.135	0.347
	(0.171)	(0.251)	(0.315)	(0.252)
PR treatment	0.039	-0.158	0.017	0.206
	(0.159)	(0.234)	(0.287)	(0.253)
N	352	110	116	126
Transfer year	All	2019	2020	2021
R-squared	0.005	0.046	0.013	0.013

Dependent variables are classroom observation scores at midline 1 based on principal component analyses.

Control variables include baseline observation scores.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 18: Year 1 treatment effects on home gardening

	(1)	(2)	(3)	(4)
Dependent Variable	Agriculture or gardening at home			
PD treatment	-0.005	-0.014	-0.027	0.006
	(0.039)	(0.074)	(0.062)	(0.048)
PR treatment	0.042	0.122*	0.055	-0.064
	(0.033)	(0.063)	(0.053)	(0.055)
Control dep var. mean	0.816	0.774	0.806	0.878
SD	0.388	0.421	0.397	0.329
N	782	262	267	253
Transfer year	All	2019	2020	2021
R-squared	0.146	0.114	0.095	0.326

The dependent variable indicators whether the teacher engage in agriculture or gardening at home.

We control for the dependent variable measured at the baseline.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 19: Year 1 treatment effects on non-teaching income

	(1)	(2)	(3)	(4)
Dependent Variable	non-teaching income			
PD treatment	-13.615	-7.485	-31.251	-11.143
	(24.767)	(26.889)	(21.867)	(69.971)
PR treatment	-15.411	-12.193	-23.740	-26.911
	(16.368)	(25.979)	(20.911)	(41.149)
Control dep var. mean	119.827	107.675	109.238	147.644
SD	250.070	232.930	189.159	328.794
N	780	262	266	252
Transfer year	All	2019	2020	2021
R-squared	0.077	0.106	0.092	0.073

The dependent variable is the average monthly non-teaching income (in thousand UGX) .

We control for the dependent variable measured at the baseline.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 20: Year 1 treatment effects on teacher presence

	(1)	(2)	(3)	(4)
Dependent Variable	Midline 1 presence			
PD treatment	-0.001	0.012	0.036	-0.046
	(0.029)	(0.054)	(0.043)	(0.057)
PR treatment	-0.036	-0.016	-0.008	-0.074
	(0.029)	(0.050)	(0.043)	(0.056)
Control dep var. mean	0.873	0.885	0.871	0.861
SD	0.334	0.321	0.337	0.348
N	806	257	275	274
Transfer year	All	2019	2020	2021
R-squared	0.003	0.001	0.008	0.006

Dependent variables are teacher presence dummy at midline 1.

We control for baseline presence.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 21: Year 2 treatment effects on teacher presence

	(1)	(2)	(3)
Dependent Variable	Midline 3 presence		
PD treatment	0.056*	0.072*	0.039
	(0.032)	(0.042)	(0.053)
PR treatment	-0.025	-0.051	-0.000
	(0.036)	(0.055)	(0.052)
Control dep var. mean	0.873	0.878	0.867
SD	0.334	0.329	0.342
N	525	267	258
Transfer year	All	2020	2021
R-squared	0.011	0.028	0.004

Dependent variables are teacher presence dummy at midline 3.

We control for baseline presence.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 22: Midline 1 treatment effects on reward eligibility

	(1)	(2)
Dependent Variable	Belief on performance treatment	
PD treatment	0.042**	0.045*
	(0.020)	(0.025)
PR treatment	0.038**	0.023
	(0.019)	(0.025)
Sample	All	Only eligible
Control dep var. mean	0.888	0.895
SD	0.316	0.307
N	1224	792
R-squared	0.005	0.004

The dependent variable indicates teachers answering yes to the following question at midline 1: Are teachers in this school eligible to be rewarded based on performance and therefore being transferred to a school of preference?

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 23: Midline 3 treatment effects on belief about transfers

	(1)	(2)	(3)	(4)
Dependent Variable	Transfer		Better transfer	
PD treatment	-0.046	-0.068	0.083	0.089
	(0.106)	(0.101)	(0.064)	(0.067)
PR treatment	0.000	0.031	-0.045	-0.043
	(0.108)	(0.104)	(0.062)	(0.067)
Control	No	Yes	No	Yes
Control dep var. mean	1.995	1.973	2.633	2.632
SD	0.915	0.918	0.597	0.605
N	597	560	585	543
R-squared	0.001	0.018	0.008	0.015

The dependent variable in columns (1) and (2) is teachers' belief on relationship between performance and transfer probability.

The dependent variable in columns (3) and (4) is teachers' belief on relationship between performance and probability of being transferred to a better school.

The control variable is the dependent variable measured at the baseline.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 24: Midline 1 treatment effects on perceived reason of preferred transfer

	(1)	(2)	(3)	(4)
	effort	student performance	connection	chance
PD treatment	-0.011	0.062	-0.029*	-0.022
	(0.048)	(0.046)	(0.016)	(0.023)
PR treatment	-0.086**	0.108**	-0.016	-0.001
	(0.043)	(0.049)	(0.018)	(0.024)
Control dep var. mean	0.538	0.321	0.047	0.090
SD	0.500	0.468	0.213	0.286
N	660	660	660	660
R-squared	0.006	0.008	0.005	0.001

The dependent variables indicate the most important perceived reason of being transferred to a preferred school, as measured at midline 1.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 25: Midline 2 treatment effects on perceived reason of preferred transfer

	(1)	(2)	(3)	(4)
	effort	student performance	connection	chance
PD treatment	-0.024	0.079**	-0.011	-0.040
	(0.039)	(0.040)	(0.012)	(0.030)
PR treatment	-0.017	0.080**	-0.013	-0.047*
	(0.042)	(0.039)	(0.012)	(0.028)
Control dep var. mean	0.629	0.195	0.027	0.145
SD	0.484	0.397	0.163	0.352
N	780	780	780	780
R-squared	0.000	0.007	0.002	0.004

The dependent variables indicate the most important perceived reason of being transferred to a preferred school, as measured at midline 2.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

Table 26: Midline 3 treatment effects on perceived reason of preferred transfer

	(1)	(2)	(3)	(4)
	effort	student performance	connection	chance
PD treatment	0.005	0.085**	0.010	-0.094**
	(0.046)	(0.033)	(0.011)	(0.038)
PR treatment	-0.030	0.096***	0.005	-0.078*
	(0.053)	(0.036)	(0.010)	(0.043)
Control dep var. mean	0.509	0.094	0.009	0.259
SD	0.501	0.293	0.097	0.439
N	633	633	633	633
R-squared	0.001	0.014	0.001	0.010

The dependent variables indicate the most important perceived reason of being transferred to a preferred school, as measured at midline 3.

Standard errors clustered at the school level.

+ 0.15 * 0.1 ** 0.05 *** 0.01

10 Figures

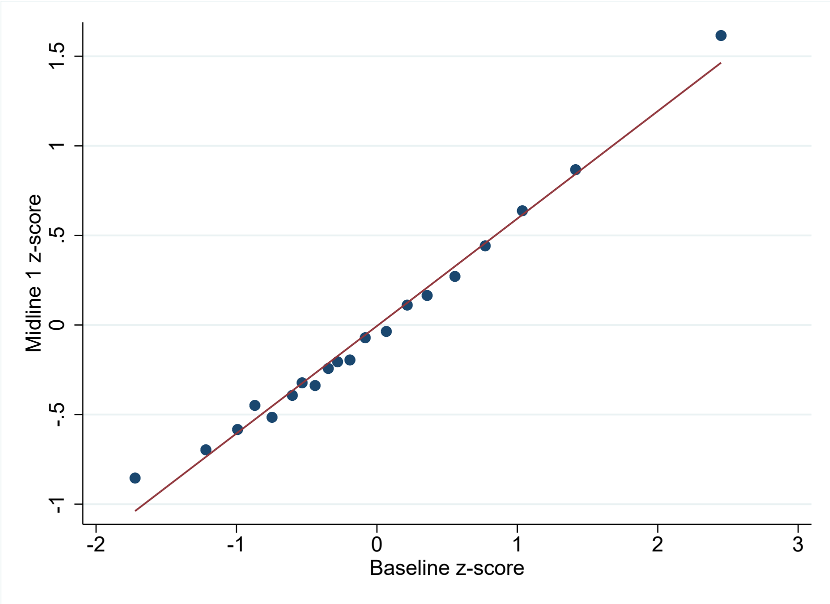

Figure 1: Binscatter of midline 1 z-scores against baseline z-scores

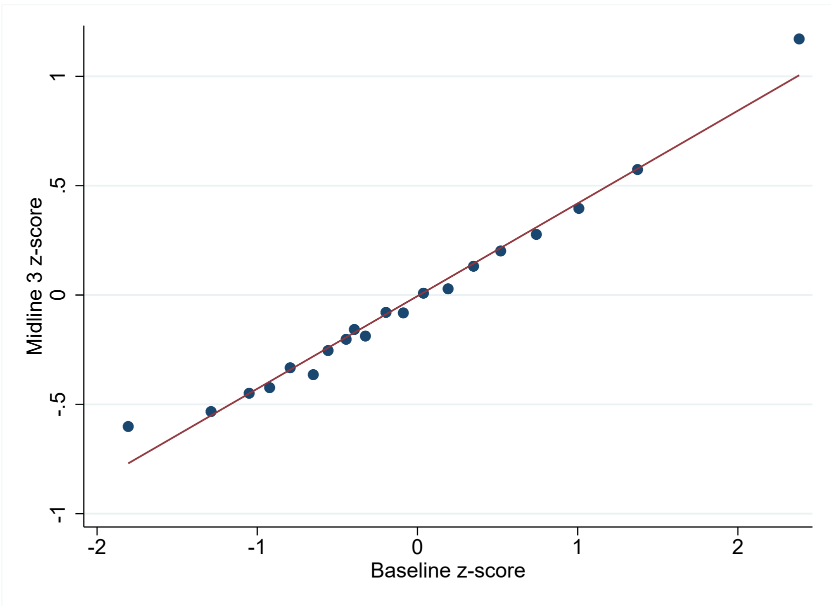

Figure 2: Binscatter of midline 3 z-scores against baseline z-scores

References

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

Finan Fred, Benjamin A. Olken, and Rohini Pande. 2015. "The personnel economics of the state." NBER Working Paper No. 21825.

Khan, Adnan Q., Asim I. Khwaja and Benjamin A. Olken. 2019. "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings." *American Economic Review*, 109(1): 237-270.

Leaver, Clare, Owen Ozier, Pieter Serneels and Andrew Zeitlin. 2021. "Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools." *American Economic Review*, 111(7): 2213-2246.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay." *Journal of Political Economy*, 119(1): 39-77.

Transparency International. 2013. "Global Corruption Report." New York: Routledge.

UNESCO. 2015. "Education for All 2000-2015: Achievements and Challenges." Paris, France: United Nations Educational, Scientific and Cultural Organization.