Mobility matters!

Teacher Absenteeism and Students' Test Scores in Kwanza South, Angola*

Frederica Mendonça†

Abstract

We examine the impact of teacher absenteeism on students' educational outcomes. We collected survey data and implemented standardized test scores in Kwanza South, Angola and use them to show that there are negative returns to teacher absenteeism on student's performance. We acknowledge the existing selection problem given that absenteeism is not independent of teachers' individual characteristics. To address it we use an instrumental variable approach and compare teachers that share similar contexts but face different costs of attendance due to the distance to which they live from school – either measured in kilometers walked or minutes spent travelling from home to school. We argue that teachers that live further away or that take longer to get to school face high costs of attendance that increase the probability of them being absent. We also find that over time, teachers' absenteeism has made it impossible for students' outcomes to improve, which urges the need for policy interventions aiming at tackling this phenomenon.

Keywords: Development Economics; Teacher Absenteeism; Education; Primary Schools; Angola; Africa

JEL Classification Numbers: C26, I24, I25, I26, R42

---

* I deeply thank my supervisor, Pedro Vicente, for the continuous support and endless resilience in the search for results. I wish to give special thanks to Matilde Grácio and Mattia Fracchia, my two unofficial supervisors. I also wish to thank Stefan Leeffers and FAS field teams for the data treatment and collection. I thank Francisco Monteiro for the useful help with coding. I thank my sweet family and Sara Queirós for the continuous company and support. Finally, I wish to thank Ornella Dellacio, without whom this adventure would not have been possible.

† Nova School of Business and Economics, Universidade Nova de Lisboa. Email: 23839@novasbe.pt

1 Introduction

Goal 4 of the Sustainable Development Goals aims at obtaining quality education as it is "… the foundation to improving people's lives and sustainable development".[2] At early ages, education works as an important channel to develop ones' capabilities. The teacher should be given careful attention as she/he plays a determinant role on the quality of the education that children receive. If a teacher is repeatedly absent and students are left without being taught, there is a higher chance of lower educational outcomes (see Duflo et al., 2012). Hence, it is extremely relevant to understand the reasons behind teacher's absenteeism and to invest in the research for solutions that mitigate this phenomenon.

In our work, we use survey data and standardized test scores we collected and implemented in Kwanza South, one of the poorest regions in Angola, where 90%[3] of the population lives in rural and difficult to access communities. In order to get to work, teachers who live outside the school's neighborhood (around 73% of our sample) face damaged roads, highly sensitive to the adverse weather conditions and they cannot rely on any organized transportation network. Rather the population relies on the circulation of private vans that are usually in very poor conditions and hence highly subject to weather conditions. Plus, the inexistence of fixed schedules or proper bus stops makes it impossible for teachers to assure that they get to school every day and on time. We argue that these persisting mobility constraints impose a problem in teachers' capacity to attend school. Therefore, using observational cross-sectional data and standardized test scores in 126 primary schools, we show that teacher absenteeism is lowering students' educational outcomes, and we do this by making use of the *distance* and *time* that teachers have to face in order to get to school.

[2] United Nations, Sustainable Development Goals
[3] Preliminary census data, 2014

In this paper, we contend that the existing literature misses to analyze the impact of logistic constraints on the day-to-day routine of teachers, specifically in contexts such as ours. Given its characteristics, Kwanza South stands as the ideal scenario for studying the role of mobility in developing countries and its impact on absenteeism in economically weak communities. We argue that our findings can be extended to several African contexts, that are similar to Kwanza South in terms of infrastructure and rural dominance.

We start by estimating an OLS regression using survey measures for absenteeism and students' test scores. We run regressions at the student level, and in some specifications, we control for student's characteristics, which cover a smaller sample. We find significant negative returns to absenteeism on students' performance. However, we acknowledge the existing endogeneity in our approach, given that the probability of a teacher being absent from school is not orthogonal to his individual characteristics and his overall quality as a teacher. Hence, we address this concern by resorting to an instrumental variable approach which uses three different specifications of distance (kilometers) and time (minutes) as instruments for absenteeism. These estimates are substantially larger than the OLS estimates and once more suggest that teacher's absenteeism decreases students' test scores, as predicted by recent empirical work (see Holla, Glewwe and Kremer, 2008; Miller, Murnane and Willett, 2008; Duflo et al., 2012).

Using Baseline and Endline data, we also estimate a Differences in Differences regression at the school level, that tries to capture how the impact of teacher absenteeism in students' outcomes has evolved on average over time. We find that, despite the fact that time on its own has contributed for an improvement in the schools' average test scores, the persisting absenteeism has made it impossible for the average students' performance to improve.

The paper is organized as follows. Section 2 provides a brief literature review on the empirical work that has been conducted to address the topic of teacher absenteeism. Section 3 describes the context of our work. Section 4 describes the data collection and structure. Section 5 describes the estimation strategy used. Section 6 provides our main results, and Section 7 concludes by discussing possible policy implications of our findings.

2 Literature Review

Over the years, economists and development researchers have often addressed the issue of teacher absenteeism. A cross-country experiment based on direct observation through unannounced visits (Chaudhury et al., 2006) found that primary school teachers were absent from school 27% of times in Uganda, 19% in Indonesia and 25% in India. Also, they found that these rates tend to be higher in poorer countries.

Chaudhury et al., 2004, studied a set of possible drivers of absenteeism by testing for correlation between variables. Among others, they suggest that opportunity costs and logistic constraints may be underlying the teacher's decision to attend school. They find a significant negative correlation between living in the facility complex and living in the village with absenteeism, in India and Uganda. In this paper we argue that when a teacher is absent, students go without being taught which ultimately leads to lower educational performance. We turn to the experimental literature in order to find evidence.

Aiming at tackling teacher absenteeism, Kremer and Chen, 2001, first resorted to financial incentives in Kenya, by introducing a bonus on the teacher's salary left to the headmaster to disburse, decreasing on the number of unjustified absences. This attempt showed no significant impacts on absenteeism nor on students' outcomes, since the research team found that headmaster reports were not truthful, and most teachers ended up receiving the entire bonus. In 2003, Glewwe et al., implemented a program in Kenya which gave teachers financial incentives based on students' test scores instead of attendance. Despite an increase in students' exam performance, results show no evidence of a decrease in absenteeism. Nevertheless, the authors also found no evidence of an increase in teacher effort which suggests that these results might be attributed to an increase in “teaching for the test” sessions at the cost of regular classes.

Duflo et al., 2012, conducted an experiment in India where the teachers were assigned cameras to take photos at the beginning and end of each school day. Their salaries would then depend on the number of days present, reported through the photos. They found that this external monitoring system aligned with financial incentives significantly decreased absenteeism and also, that there were significant improvements on students’ performance. This suggests that indeed there are negative returns of absenteeism on students’ outcomes. Muralidharan and Sundararaman, 2013, studied the impact of hiring contract teachers in India and found first, that contract teachers were much less likely to be absent than civil-service teachers and second, that students in schools with contract teachers performed significantly better than those in comparison schools. Similarly, Duflo et al., 2014, examined a program in randomly selected primary schools in Kenya where new teachers were hired on annual contracts renewable conditional on performance, as to decrease teacher-pupil ratio. They found that contract teachers were considerably less absent in general whereas civil-service teachers (that cannot be fired) in treatment schools were more absent than their counterparts in comparison schools. This suggests that teachers at intervention schools took advantage of having colleagues with strong incentives not to miss school. In the end, scores increased significantly only for students assigned to contract teachers. These once more provide evidence on the relevance of teacher attendance on children learning.

The use of incentives directly targeted at the teacher has proven to be effective at times at decreasing absenteeism and improving students’ performance. Another branch of the experimental literature has attempted to tackle this issue through demand side interventions. Kremer and Vermeersch, 2004, implemented a program in Kenya that would give school meals to preschool students. These meals would be distributed by a parent, which implied higher supervision of teachers' attendance. They found an increase in students' test scores only in schools where teachers were relatively experienced before the program. Even though the teacher's income was set based on a negotiation process with the parents on how to allocate parents school contributions (which increased substantially with the program), there was no impact on attendance. Kremer et al., 2005, made three unannounced visits to 3700 schools in 20 Indian states and found evidence that suggest that whilst higher salaries do not impact absenteeism, better school infrastructure is usually associated with lower absence rates, suggesting that working conditions have an influence on teachers' behavior. Kremer et al., 2009, studied a randomized evaluation of a merit scholarship for girls in Kenya and found evidence of a substantial increase in teacher attendance as well as an increase in all students' test scores, which may suggest that teacher's effort and motivation to attend is higher when students are better prepared and more willing to learn.

In line with Chaudhury et al., 2004, we argue that there might be high opportunity and logistic costs subjacent to the teachers' decision to attend school. Parallel to this literature, Muralidharan and Prakash, 2014, studied the impact of a program in India which rewarded girl students that enrolled to secondary school with bicycles as to improve access to school. They find a significant increase in the age-participation in secondary schools, especially for girls that live more than 3km away which suggests that this program resulted in a decrease in the "distance cost" of attending school.

In this paper, we argue that there exists such a cost for teachers as well, and that this logistic constraint might be driving absenteeism, which ultimately results in lower students' test scores. Our approach consists in employing a wide variety of cross-sectional and longitudinal data into finding such causality.

3 Context

After its independence from the Portuguese colonial empire in 1975, Angola faced a big civil war, that lasted until 2002 and was responsible for the death of many and the destruction of its infrastructures. In peace at last, the country remained with poor health and education indicators, positioning itself in 147th out of 162 countries in the Human Development Index. On education matters, in 2014, the literacy rate in Angola was 66%[4] and in 2017, the country's expected years of schooling was 11.8 years[5]. Despite the fact that by 2015, gross primary school enrolment was 113.3%[6], in 2011, only 47%[7] of primary teachers were in fact trained to teach.

In this setting, Kwanza South is one of the 18 provinces of Angola, located South of Luanda by the sea. Its climate is characterized by having only two seasons: the rainy season, from September to April and the dry season, from May to August. Being the 5th most populated province in Angola (with around 1.8 million inhabitants)[8] it is also one of the poorest and almost half of the population is composed by children aged between 0-14.[9] What's more, a recent study by INE[10] and UNICEF[11] has shown that school attendance is around 27% for children aged between 5-11 and around 17% for children aged between 12-17 years in Kwanza South.

Since the 1990s, the World Bank has been actively engaged in solving these matters through the funding of FAS (Fundo de Apoio Social), a government institution which promotes sustainable social and economic development of communities[12]. Across time FAS has

---

[4] Human Development Reports, 2014
[5] Human Development Reports, 2017
[6] "Calculated by dividing the number of students enrolled in primary education regardless of age by the population of the age group which officially corresponds to primary education, and multiplying by 100." World Development Indicators, 2015
[7] Human Development Reports, 2011
[8] Preliminary census data, 2014
[9] Preliminary census data, 2014
[10] INE – Instituto Nacional de Estatística
[11] INE and UNICEF, 'Childhood in Angola – A Multidimensional Analysis of Child Poverty', 2018
[12] FAS – Fundo de Apoio Social – is an Angolan public institution, born in October 1994. Its main goal is to promote sustainable development and poverty reduction, through social projects and investments in the areas of education, health, sanitation, environment and infrastructures across the 18 provinces of Angola.

promoted the construction and improvement of many primary schools in Angola, mainly in Kwanza South. In more recent years, FAS has adopted projects aiming to promote quality asides from quantity in education. It is under this scenario that our research takes place.

4 Data

4.1 Data Collection

We use data collected in two phases during the field work conducted by FAS' teams for a Randomized Control Trial on Community Driven Development. It involves 126 school across nine out of twelve municipalities of Kwanza South province. Most of these schools were constructed or received improvements through FAS. The Baseline occurred from October 2014 to August 2015. I took part as field coordinator in the Endline, from July 2018 to December 2018. This paper uses mainly data collected during the Endline phase, apart from one specification in Section 5 that uses both.

The schools were randomly allocated in blocks by municipality and a set of school characteristics across 4 comparison groups. The first treatment, T1, consisted on door-to-door information to parents about the students, teachers and school board's performance - average scores of students, students' passing rates, teachers' absenteeism, among others – and a comic about the relevance of parental involvement in education. The second treatment, T2, consisted on general school meetings with the parents at the school to discuss the existing problems and possible solutions to address them. The third treatment T3 combined both – information and school meetings. T4 was the control group. These treatments are controlled for in all of our specifications.

The first source of data consists on multiple choice tests of both Mathematics and Portuguese language designed by the Principal Investigators' team and implemented to students of third and fourth grade in the Baseline with the addition of fifth grade in the Endline. Each test had

10 questions. These tests were completed during regular class period and each had a duration of 45 minutes. This paper uses as outcome variables the average of both tests. Due to some errors found during Endline, it was decided that the Baseline tests would have to be re-digitalized. However, part of the tests had gotten lost when this was detected and hence the FAS team only managed to re-digitalize the tests of about 80 schools.

Second, we conducted surveys to parents, teachers, directors and school administration. In the Baseline, parents were selected and interviewed using the random walk method around each school. During the Endline, 40 students of the third, fourth and fifth grade were randomly selected based on school records and their parents were asked to come to school and answer the parent's survey. These students test scores were matched with their parents' survey. All teachers and directors of the school were asked to participate in the survey in both phases. Lastly, the school administration survey was always conducted with a school board member, most times with the director herself/himself.

Relevant variables from teacher and director surveys include self-reported socioeconomic characteristics and social capital (e.g., education, income, possessions), performance outcomes (e.g., attendance), satisfaction with school, school board, parents and other teacher performances and questions concerning parental involvement of the 40 selected students.

Relevant variables from the parent survey include self-reported socioeconomic characteristics and social capital (e.g., education, income, possessions), parental involvement (e.g., participation in school meetings) and satisfaction with teacher and school performance.

From the school administration survey, relevant variables include reports on teacher, parent and school board performance (e.g., attendance) and school characteristics and possessions (e.g., number of teachers, students, classrooms, distance to closest municipality center).

Finally, administrative data on the 126 schools for 2016, 2017 and 2018 were collected from the provincial board of education of Kwanza South. These data include information on school infrastructure, facilities and administration.

4.2 Descriptive Statistics

Descriptive Statistics on characteristics of our sample are depicted in Table 1. This comprises information regarding socio-economic characteristics and performance of teachers, students, parents as well as school characteristics.

Average teacher was aged 36 years at Baseline and 40 at Endline. 52.1% of the teachers at Baseline and 64.2% at Endline were women and about one third had Superior Education at Baseline versus nearly half at Baseline. 46.4% were married or cohabiting and earning an average wage of 65.000AKZ per month (around 200USD) at Baseline and 56.1% were married or cohabiting and earning an average wage of 60.000AKZ per month (around 185USD) at Endline. On average, each teacher would skip school around half a day (0.59 days) out of ten (2 working weeks) at Baseline and almost one day (0.97 days) out of ten - according to self-reported data - and 0.57 days out of ten - according to school records - at Endline. About 77% of teachers at Endline had a Definitive contract, meaning that the School Director could not fire them. On average, each Endline teacher lives 5.64km away from school and takes around 32.4 minutes to arrive to school.

Average student's age at Baseline was 10.5 years and about half were girls (49.6%). At Endline, average age was 11.23 years old and again around half the sample were girls (50.1%). Also, around 95% of students spoke Portuguese as a first language at home in both stages. It is relevant to mention that these data only accounts for students whose parents were interviewed. Students' test score in Portuguese averaged 5.09 at Baseline and 5.96 at Endline and Mathematics test score averaged 4.4 at Baseline and 5.7 at Endline.

Regarding parents, average age was around 36 years at Baseline and 40 years at Endline. The majority of parents were women at Baseline (60.1%) and most of them with only primary education (63.1%). At Endline, around 49% of parents were women and 59% had no more than primary education. Average income was around 23.400AKZ per month (around 72USD) at Baseline and 28.000AKZ per month (around 86USD) at Endline.

Finally, regarding schools, average number of teachers per school was around 15 at Baseline and around 21 at Endline. The number of students in each school was on average 667 at Baseline and 727 at Endline. Each school distances an average of 16.7 km to the center of the closest municipality.

Table 1: Individual characteristcs - Baseline and Endline

	(A) Baseline Sample					(B) Endline Sample				
Students		**std**	**min**	**max**	**N**		**std**	**min**	**max**	**N**
Mean Test Score PT	5,09	2,61	0	10	9962	5,96	2,85	0	10	24684
Mean Test Score MAT	4,40	2,04	0	10	10199	5,74	2,61	0	10	24299
Age ⧺	10,59	1,00	5	19	2338	11,23	2,05	5	19	1549
Female ⧺	49,6%	0,50	0	1	2338	50,1%	0,50	0	1	2996
Speak PT at home ⧺	95,1%	0,22	0	1	2338	94,6%	0,23	0	1	2994
Teachers										
Abs (Self-reported) ⊥	0,59	1,12	0	10	731	0,99	1,56	0	10	1561
Abs (Director report) ⊥						0,57	1,40	0	10	1561
Age	36,46	8,27	21	63	730	38,98	8,55	18	67	1571
Female	52,1%	0,50	0	1	731	64,2%	0,48	0	1	1982
High Education ⊦	30,2%	0,46	0	1	731	46,4%	0,50	0	1	1982
Definitive Contract						77,0%	0,42	0	1	1520
Married	46,4%	0,50	0	1	731	56,1%	0,50	0	1	1571
Income (AKZ)	65222,30	52380,35	12000	680000	723	60704,36	29080,32	10000	31000	1438
Distance Home/School (Km)						5,64	6,89	0,25	20	1605
Time Home/School (minutes)						32,38	34,91	0	360	1557
Parents										
Age	36,31	12,04	18	90	2338	39,88	11,26	18	83	1927
Female	60,1%	0,49	0	1	2338	48,9%	0,50	0	1	1934
Primary Education	63,1%	0,48	0	1	2338	58,8%	0,49	0	1	1933
Income (AKZ)	23454,26	37662,75	2500	175000	2350	27854,48	38400,94	2500	200000	1474
Married	33,3%	0,47	0	1	2350	56,3%	0,50	0	1	1975
School										
No. teachers	15,40	10,43	1	55	125	20,85	34,40	1	236	123
No. students	667,10	488,40	83	2566	125	727,11	539,44	43	2782	123
No. classrooms	9,97	6,46	2	38	124	7,07	6,76	0	53	123
Distance center Mun. (km)	16,17	18,36	0	102	125	16,17	18,36	0	102	125

Note: Sample includes all individuals interviewed in Baseline and Endline. ⧺ sample includes only students whose parents were interviewed. ⊥ variable that answers how many days the teacher was absent from school in the previous two weeks. ⊦ Dummy variable equal to 1 if the teacher went to university or higher.

5 Estimation strategy

5.1 OLS regression

In this section we describe the estimation strategy used to measure the impact of teachers' absenteeism on the test score performance of students. We begin with the following OLS regression,

$$(1)\quad Y_{ijs} = \alpha + \beta A_{ijs} + \lambda \boldsymbol{W'}_s + \delta \boldsymbol{M'}_s + \eta \boldsymbol{S'}_s + \gamma \boldsymbol{Z'}_{ijs} + \varphi \boldsymbol{X'}_{ijs} + \varepsilon_{ijs}$$

where Y is the outcome of interest - performance of the student in the multiple choice test -, i,j,s are identifiers of individual students, individual teachers and schools, respectively, A is a variable that ranges from 1 to 10 and counts how many days the teacher was absent from school in the previous two weeks, $\boldsymbol{W'}$ is a vector for 3 treatment groups (the control group is the omitted indicator), $\boldsymbol{M'}$ is a vector for 8 municipality dummies, $\boldsymbol{S'}$ is a vector of school characteristics, $\boldsymbol{Z'}$ is a vector for teacher socio-economic characteristics, $\boldsymbol{X'}$ is a vector for socio-economic student characteristics and ε is the normally distributed error term. Our coefficient of interest is β, which evaluates how teacher's absenteeism is affecting the student's performance. As we are using test scores from 1 to 10, we obtain the absolute effect on the student test score. We expect the coefficient to be negative (negative returns to absenteeism). Given that our data has several sources of information on absenteeism of teachers, we chose to use two specifications of the same variable – one which is self-reported (directly asked to the teacher during the interview) and one which is reported by the school director (based on school records). In the specifications that do not include the municipality dummies, standard errors are robust and clustered at the school level. In the specifications that do include, standard errors are robust and bootstrapped with 500 replications.

5.2 Instrumental Variables Approach

Ideally, we would be able to isolate the pure effect of absenteeism on students' test score if the probability of being absent was randomly assigned. However, once we account for the motivation behind the teacher's decision, we cannot assume that our dependent variable is exogenous, since our estimates will be biased. Given our context, we argue that the decision to attend school may be affected by the existing constraints in mobility, meaning the path that the teacher has to go through in order to get to school. Our assumption is that a teacher who lives further away or a teacher that has difficult access to school faces a higher cost of attendance, thus increasing the likelihood of being absent. Hence, we turn to an instrumental variable approach to isolate the causal effect of absenteeism, using distance and time traveled between home and school as determinants of absenteeism. Going to the literature we see that Card, 1995, using a sample of the US population drawn from the National Longitudinal Survey of Young Men (NLSYM), has argued that proximity to a nearby college may be a causal determinant of schooling for students, especially among the ones which come from lower-income families. Therefore, he used geographical variation as an instrument for school choice. Our empirical approach will consist on using self-reported data on kilometers walked and minutes spent between the teacher's home and school, similarly to what has been done in recent educational literature (see Barrow et al., 2015; Carneiro et al., 2016; Groote and Declercq, 2018).

In order for these to be good instruments, they need to satisfy two conditions: (i) they should have a clear effect on attendance; (ii) they should only affect the students' test scores via its influence on teachers' attendance (Angrist and Pischke, 2008). After carefully analyzing our data, we find no evidence to suggest existing heterogeneity in our sample regarding home location: around 70% of teachers live outside the school neighborhood. Hence, we assume that the teacher's residence is given, and proximity to school is not correlated with any other determinants of absenteeism. However, the difference in the distance creates a difference in the costs of attending. Take the following model,

$$(2)\quad Y_{ijs} = \alpha + \beta A_{ijs} + \lambda \boldsymbol{W'}_s + \delta \boldsymbol{M'}_s + \eta \boldsymbol{S'}_s + \gamma \boldsymbol{Z'}_{ijs} + \varepsilon_{ijs}$$

$$(3)\quad A_{ijs} = \alpha_1 + \beta_1 \boldsymbol{D'}_{ijs} + \lambda \boldsymbol{W'}_s + \delta \boldsymbol{M'}_s + \eta \boldsymbol{S'}_s + \gamma \boldsymbol{Z'}_{ijs} + \mu_{ijs}$$

where $\boldsymbol{D'}$ is a vector of variables that measures distance and time spent from home to school. In our regressions we employed three different specifications of the first-stage regression:

- First, we used a self-reported categorical variable that answers how far in kilometers the teacher's home distances to the school. We expect β_1 to be positive, meaning that the further the teacher lives from school, the higher the number of days in which he/she is absent.
- Second, we include a self-reported continuous variable that answers how long in minutes it takes the teacher to get from home to school.
- Third, we compute an index of time and distance by computing z-scores. In order to do so, we standardize both variables of time and distance considering the transport type used by subtracting the mean and dividing by the standard deviation. Then, we combine them in a summary index by taking the average of the two normalized variables. This aggregation allows improvements in statistical power as to detect effects that go in the same direction within a domain, as suggested by Kling et al., 2007. We choose this methodology to gain transparency in the construction of the index, namely the weights attributed to each variable. Nevertheless, in *Table 3.b* in annex we present the same estimation using an index created with the principal component analysis.

Finally, μ is the normally distributed error term. Standard errors are robust and bootstrapped with 500 replications. The second-stage regression is the same as in (1) excluding student controls as to avoid missing too many observations.

5.3 Differences in Differences

Next, we use a Differences in Differences model to measure how the impact of absenteeism on the students' test score has changed from Baseline to Endline. We start with the following regression,

$$(4)\ \overline{Y_s} = \alpha_2 + \beta_2 \bar{A}_s + \beta_3 t + \rho(\bar{A}_s * t) + \eta \boldsymbol{S'}_s + \epsilon_s$$

where $\overline{Y_s}$ is the average test score of the students in each school, $\bar{A}_s$ is the average number of days in each school that the teachers were absent in the previous two weeks, t is a Dummy variable equal to one at Endline and zero at Baseline and ϵ is the normally distributed error term. Our coefficients are β_2, which measures how average absenteeism in a school affects the average performance of its students at Baseline, β_3, which measures the impact of time on the average students' performance when there is no absenteeism and ρ, our coefficient of interest, which measures on average how the persistence of teachers' absenteeism across time is affecting the students' test scores. Standard errors are robust.

Next, we also use the following regression,

$$(5)\ \Delta\overline{Y_s} = \alpha_3 + \beta_3 \Delta\bar{A}_s + \eta \Delta\boldsymbol{S'}_s + \sigma_s$$

where $\Delta\overline{Y}$ is the change in the average test score of each school from Baseline to Endline, $\Delta\bar{A}$ is the change in the average teacher absenteeism of each school from Baseline to Endline, $\Delta\boldsymbol{S'}$ is the difference in school characteristics from Baseline to Endline and ϵ is the error term. Our coefficient of interest is β_3 which measures how the average absenteeism of a school across time affects the average students' performance, also across time. σ is the normally distributed error term. Standard errors are robust.

6 Results

6.1 OLS results

This section reports the results of our OLS estimates on the effects of teacher absenteeism on students' test scores. These results are depicted in *Table 2*. Columns (A) depict results for self-reported data and columns (B) depict results for director-reported data. For each dependent variable, columns (1), (2) and (3) vary in the number of controls used. Columns a) control for 3 treatment dummies and columns b) also control for 9 municipality dummies. Column (1) is our simplest model, obtained by controlling for the already mentioned dummies, whilst in column (2) we control for school characteristics (Number of teachers, students and classrooms, Dummies for whether the school has water, electricity and a toilet) and teacher characteristics such as age, gender, a dummy variable equal to one if the teacher went to university, dummies for the teacher's home location (living in the school neighborhood – base category, living outside the neighborhood but in the school municipality and living outside the school municipality), years of teaching experience, dummy equal to one if the teacher is married or cohabiting and number of dependent children. In column (3) we also control for student and parent characteristics such as student's age and gender, a dummy equal to one if the parent studied above primary education, a dummy equal to one if the parent is married or cohabiting and the number of total dependents in the household.

Results in *Table 2* suggest that there is a negative correlation between teacher absenteeism and students' test scores. Each day absent decreases students' test scores around 0.14 points (significant at 1% level in all but one of our specifications), both when self-reported and according to school records. The consistency of the results is not affected once we control for student characteristics, even though the sample size gets drastically reduced (around 1/10 of the observations).

Table 2: OLS regressions on Teacher Absenteeism (Endline)

	(A) Self-reported Absenteeism						(B) Director-reported Absenteeism					
Dependent Variable ---->	(1)		(2)		(3)		(1)		(2)		(3)	
	a)	b)	a)	b)	a)	b)	a)	b)	a)	b)	a)	b)
Absenteeism in Past 2 Weeks	-0,138***	-0,125***	-0,147***	-0,132***	-0,151**	-0,151***	-0,143***	-0,141***	-0,151***	-0,144***	-0,141***	-0,143***
	(0,036)	(0,009)	(0,032)	(0,009)	(0,060)	(0,044)	(0,039)	(0,012)	(0,034)	(0,012)	(0,048)	(0,050)
School Controls												
No. Teachers			0,005***	0,004***	0,006**	0,005			0,007***	0,005***	0,001	-0,001
			(0,001)	(0,000)	(0,002)	(0,003)			(0,002)	(0,001)	(0,004)	(0,005)
No. Classrooms			-0,023*	-0,023***	-0,049***	-0,040***			-0,019	-0,023***	-0,050**	-0,045***
			(0,012)	(0,002)	(0,016)	(0,013)			(0,013)	(0,003)	(0,020)	(0,016)
Bathroom Dummy			-0,587**	-0,712***	-0,727*	-0,673***			-0,466*	-0,634***	-0,586	-0,514**
			(0,255)	(0,054)	(0,372)	(0,220)			(0,263)	(0,057)	(0,369)	(0,232)
Electricity Dummy			0,844***	0,845***	1,095***	1,154***			0,788***	0,868***	1,151***	1,248***
			(0,157)	(0,036)	(0,234)	(0,185)			(0,167)	(0,040)	(0,254)	(0,201)
Teacher Controls												
Age			-0,015	-0,014***	-0,000	0,004			-0,013	-0,011***	0,009	0,013
			(0,009)	(0,002)	(0,013)	(0,011)			(0,010)	(0,003)	(0,013)	(0,012)
Gender			0,284**	0,249***	0,361*	0,324**			0,291**	0,238***	0,479**	0,439**
			(0,122)	(0,034)	(0,208)	(0,159)			(0,136)	(0,038)	(0,218)	(0,171)
High Education			0,394**	0,312***	0,233	0,168			0,191	0,067*	0,042	-0,099
			(0,160)	(0,034)	(0,230)	(0,169)			(0,176)	(0,039)	(0,256)	(0,185)
No. Dependents			-0,002	-0,002	0,021	0,023			-0,025	-0,025***	0,005	0,004
			(0,005)	(0,003)	(0,035)	(0,022)			(0,023)	(0,006)	(0,040)	(0,025)
Student/Parent Controls												
Age Student					0,113***	0,117***					0,115***	0,120***
					(0,028)	(0,033)					(0,032)	(0,036)
Female Student					-0,308**	-0,341***					-0,367**	-0,400***
					(0,136)	(0,125)					(0,151)	(0,136)
High Education (Parent)					0,589***	0,577***					0,476***	0,428***
					(0,174)	(0,143)					(0,181)	(0,153)
R2 adjusted	0,010	0,066	0,123	0,135	0,127	0,138	0,020	0,061	0,108	0,124	0,115	0,132
Observations	23 414	23 414	23 237	23 237	1 319	1 319	19 668	19 668	19 530	19 530	1 125	1 125
Municipality dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes

Note: OLS regressions. Independent variable reports the number of days the teacher was absent from school in the previous 2 weeks. Sample includes all individuals interviewed at endline. Regressions without Municipality dummy have standard errors clustered at school level. All regressions include Treatment dummies. (1) no controls. (2) Includes teacher and school controls. (3) Includes student (age, gender, dummy for parent's education above primary school, dummy for parent's marital status, total dependents in the household) teacher (age, dummy equal to 1 if student is a girl, dummy for superior education, dummy for home location, years of experience, dummy for marital status, total dependents in the household) and school controls (dummies for water, electricity, bathroom, number of students, teachers and classrooms). * significant at 10%; ** significant at 5%; *** significant at 1%.

6.2 IV results

In this section we present the results of our IV estimates. Results for second-stage regressions can be found in *Table 3.* In our regressions, columns (1) control for treatment and municipality dummies whilst in columns (2) we also control for teacher and school characteristics (these controls are described in extent in section 6.1.).

In accordance with the OLS results, we can observe that there is in fact evidence of negative returns to absenteeism on the students' test scores. Results show consistent and significant negative results, that are larger in absolute value than the corresponding OLS results. This suggests that OLS regressions do not take into consideration possible selection based on unobservable characteristics in the error term. Differences in students' ability may be underestimating the true impact of absenteeism on test scores. Also, the coefficients for the schoolboard's report have greater magnitude systematically.

Table 3: 2nd stage IV regressions for Absenteeism (Endline)

	(A) Distance		(B) Distance + Time		(C) Z-score Index	
Test Scores —>	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Panel A: Self-Reported Absenteeism						
Absenteeism in Past 2 Weeks	-4,995***	-6,276	-2,240***	-0,691***	-2,571***	-0,551**
	(1,885)	(72,783)	(0,455)	(0,237)	(0,604)	(0,225)
Teacher Controls						
Age		-0,004		-0,013***		-0,013***
		(0,119)		(0,003)		(0,003)
Female		1,255		0,388***		0,370***
		(12,362)		(0,051)		(0,052)
High Education		0,734		0,357***		0,350***
		(4,299)		(0,044)		(0,039)
School Controls						
Bathroom Dummy		-2,111		-0,778***		-0,753***
		(18,401)		(0,083)		(0,075)
Electricity Dummy		1,364		0,864***		0,857***
		(-6,468)		(0,045)		(0,043)
Kleibergen-Paap	20,63	2,69	46,99	42,89	25,01	36,33
Hansen J statistic (P value)			0,00	0,00		
R2 adjusted	-10,572	-16,315	-1,932	0,002	-2,604	0,063
Observations	22 639	22 485	22 639	22 485	22 620	22 466
Panel B: Director Reported Absenteeism						
Absenteeism in Past 2 Weeks	-1,652***	-0,545***	-1,688***	-0,635***	-1,935***	-0,086
	(0,173)	(0,151)	(0,163)	(0,139)	(0,361)	(0,213)
Teacher Controls						
Age		-0,021***		-0,023***		-0,010*
		(0,005)		(0,005)		(0,006)
Female		0,309***		0,313***		0,294***
		(0,042)		(0,042)		(0,040)
High Education		-0,017		-0,041		0,109
		(0,054)		(0,056)		(0,068)
School Controls						
Bathroom Dummy		-0,685***		-0,713***		-0,548***
		(0,074)		(0,073)		(0,081)
Electricity Dummy		0,994***		1,030***		0,815***
		(0,072)		(0,065)		(0,093)
Kleibergen-Paap (F-statistics)	145,79	96,74	192,48	161,82	50,02	52,91
Hansen J statistic (P value)			0	0,2383		
R2 adjusted	-0,819	0,071	-0,862	0,042	-1,184	0,125
Observations	18 956	18 841	18 956	18 841	18 937	18 822

Note: 2nd stage IVregressions. Independent variable reports the number of days the teacher was absent from school in the previous 2 weeks. Sample includes all individuals interviewed at endline. All regressions include municipality and treatment dummies. Panel A uses Self-reported data. Panel B uses school records' data. Instrumental variables are respectively a categorical variable for distance between home and school; a categorical variable for distance and a continuous variable for time spent from home to school (both for teachers); an index constructed with z scores for time and distance; a variable that measures the distance in Km of the school to the center of the closest municipality. (1) Includes no controls. (2) Includes teacher (age, gender, dummy for superior education, dummy for home location, years of experience, dummy for marital status, total dependents in the household) and school controls. F-statistics and P-value reported are from first-stage. Standard errors are robust and bootstrapped with 500 reps. * significant at 10%; ** significant at 5%; *** significant at 1%.

6.2.1 IV Distance

Column (A) presents results for this estimation. Considering self-reported data, second-stage results suggest that each day a teacher is absent decreases students' test scores around 5 points (significant at the 1% level). F-statistics of the first-stage regression is 20.63 which is above the conventional threshold for weak instruments (Stock and Yogo, 2002). However, these findings lose all significance once we control for school and teacher characteristics.

Considering school records data, results suggest that a teacher being absent decreases students' test scores around 1.7 points. First-stage F-statistics is 145.8, which means that we reject the null for weak instruments. Once we control for school and teacher characteristics, the coefficient becomes smaller, suggesting that each day the teacher misses school decreases students' test scores around 0.5 points. F-statistics of the first-stage regression is 96.74.

6.2.2 IV Distance and Time

Results for this estimation are depicted in column (B). Looking at teacher reports, we see again evidence of negative returns to absenteeism on test scores of about 2.2 points (significant at the 1% level). F-statistics of first-stage is 46.99, meaning that we reject the null for weak instruments. However, P-value is 0, meaning that we reject the null for overidentification of instruments according to the conventional threshold (Stock and Yogo, 2005) and our instruments are not jointly significant. Once we control for student and school characteristics, results suggest that a teacher's absence decreases students test scores around 0.7 points (significant at the 1% level). Despite having a first-stage F-statistics of 42.9, the P-value is 0, indicating that our instruments are not jointly significant.

School records data suggest that these negative returns consist on a 1.7 points decrease (significant at the 1% level) in students' test scores. Even if we reject the null for underidentification (F-statistics is 192.5), a 0 P-value implies that our instruments are not jointly significant. Once we control for school and teacher characteristics, the coefficient remains negative of around 0.6 and significant at the 1% level. First-stage F-statistics is 161.8 (instruments are not weak) and P-value is around 0.24, meaning that we do not reject the null for overidentification – instruments are jointly significant.

6.2.3 Z-score Index

Results for this estimation can be found in column (C). Starting with self-reported data, results suggest that each day the teacher is absent decreases students' test scores around 2.6 points (significant at the 1% level). F-statistics for first-stage is 25.01, hence we reject the hypothesis of weak instruments. The coefficient loses magnitude once we include the already mentioned controls, suggesting that each day the teacher is absent decreases test scores around 0.6 points (significant at the 1% level). First-stage F-statistics is 36.33.

Director reported data suggests that students' test scores decrease around 1.9 points for each day the teacher is absent (significant at the 1% level). Again, we reject the null for weak instruments (first-stage F-statistics is 50.02). However, once we control for school and student characteristics, these results lose significance.

6.3 Differences in Differences results

This section reports the results of our DID estimates on the impact of teacher's absenteeism in students' test scores across time. These results can be found in *Table 4.* For each of the estimation strategies used, columns (1) and (2) vary in the added controls. Column (1) is the simplest regression, controlling for whether the data is *tainted,* using a dummy variable equal to 1 if the data was digitalized during Baseline[13]. Column (2) also controls for school characteristics (the ones already referred in OLS and IV results).

In the first estimation strategy we use an OLS regression. Results can be found in column (A) and suggest that the impact of absenteeism on students' performance has worsen over time, contributing for a decrease in the average test score in each school of around 0.37 points (significant at the 10% level). This negative impact becomes larger (0.57 points, significant at the 1% level) once we control for school characteristics. However, the positive coefficient for

[13] Section 4.1 describes the purpose of the dummy variable

the time dummy suggests that time has had a positive influence on the schools' outcomes, contributing for an increase of 1.2 points (significant at the 1% level) on the schools' average test scores. The coefficient for the average absenteeism suggests that there were positive returns of absence on the school's average test score at Baseline, although these returns are only significant (at the 10% level) and of 0.32 points when we control for school characteristics.

Table 4: DID regressions on Absenteeism using Baseline and Endline sample

Panel A: Level	(A) Level		(B) Fixed Effects		(C) Differences	
dependent variable —>	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Mean Abs in Past 2 Weeks	0,082	0,315*	0,442**	0,402*		
	(0,180)	(0,189)	(0,200)	(0,205)		
Time	1,196***	1,222***	1,341***	1,279***		
	(0,209)	(0,193)	(-0,174)	(-0,196)		
Time*Mean Abs in Past 2 weel	-0,374*	-0,567***	-0,671***	-0,633***		
	(0,196)	(0,203)	(0,207)	(0,213)		
Diff in Mean Abs (Past 2 Weel					-0,150*	-0,159*
					(0,084)	(0,083)
School Controls						
No. Teachers		0,002		0,005		0,006***
		(0,002)		(0,003)		(0,002)
No. Classrooms		-0,013		0,000		0,001
		(0,012)		(0,016)		(0,013)
Bathroom Dummy		-0,352*		0,170		0,281
		(0,180)		(0,312)		(0,285)
Electricity Dummy		0,811***		-0,270		-0,290
		(0,163)		(0,281)		(0,222)
R2 adjusted	0,149	0,328	-0,230	-0,265	0,010	-0,001
Observations	252	251	252	251	126	119

Note: Sample includes Baseline and Endline data. Level is an OLS regresion using as independent variable the mean number of days (at the school level) a teacher has skipped school in the last 2 weeks. Fixed Effects is a Fixed Effects regression and uses the same independent variable. Differences is an OLS regression which uses as independent variable the difference between the mean absenteeism (last two weeks) in Baseline with Endline. (1) Uses no controls. (2) Includes school controls (number of teachers, classrooms and students and dummies for whether the school has a bathroom, water and electricity). Standard errors are robust. * significant at 10%; ** significant at 5%; *** significant at 1%.

These results are supported by our second estimation strategy, a Panel Fixed Effects model (column (B)). We do this as to control for possible unobserved time-invariant characteristics within each school that are correlated with the error term. Results suggest that, there are negative returns on the impact of absenteeism on the average students' test scores over time of about 0.67 points (significant at the 1% level). The impact becomes marginally lower (0.63 points, significant at the 1% level) once we control for school characteristics. Once more, we see a very positive and very significant influence (at the 1% level in both specifications) of time on schools' average test scores, increasing the latter around 1.3 points and a positive influence of the school's average absenteeism at Baseline of 0.44 points (significant at the 5% level) and of 0.4 points (significant at the 10% level) once we control for school characteristics. We can only speculate about the interpretation of such result.

Lastly, we run an OLS estimation using the differences on the average absenteeism from Endline to Baseline and the differences on the average school's test score between the same periods. Results in column (C) are in accordance with the last two specifications, suggesting that the impact of the mean absence of teachers has worsen over time, contributing to a decrease in the school's mean test score of 0.15 points (significant at the 10% level). The coefficient has the same direction, the same significance and about the same size (0.16 points) once we control for school characteristics.

6.4 Alternative dependent variables – Parents Satisfaction

So far, we have been focusing on exploring a possible causal effect between the number of days that the teacher is absent and the students' performance in the form of test scores. In this section, we consider possible alternative variables that can also be explained by teachers' absenteeism. We argue that the level of satisfaction of the school parents regarding the teacher and student outcomes may also be explained by the number of days the teacher skips school. Hence, we use four dependent variables to test for this possible causal effect: a variable that reports whether parents think that the teacher cares for students (-2 "Not at all"; -1 "Not"; 0 "Neutral"; 1 "A little"; 2 "A lot") - column (A); a variable that reports parents' perception of the kids performance (-2 "Very bad"; -1 "Bad"; 0 "Reasonable"; 1 "Good"; 2 "Very Good") – column (B); a variable that reports whether the parent thinks the teacher is motivated (-2 "Not at all"; -1 "Not"; 0 "Neutral"; 1 "A little"; 2 "A lot") – column (C); a variable that reports whether the parent thinks the teacher is responsible (-2 "Not at all"; -1 "Not"; 0 "Neutral"; 1 "A little"; 2 "A lot") – column (D). Results are depicted in *Table 5.* We have chosen to only present results for when we instrument absenteeism with the z-score index, given that the conclusions do not change in the other tested specifications.[14] All regressions control for treatment and municipality dummies. Also, we control for teacher, school (the same as in previous estimates), and parents' characteristics (age, gender, a dummy variable equal to one if the parent studied above primary school, a dummy variable equal to one if the parent was born in the school's neighborhood and a dummy equal to one if the parent has an income above median). Columns (1) use as explanatory variable self-reported data and columns (2) use director-reported data.

Results suggest that there exists no causality between teacher absenteeism and parents' satisfaction, given that none of our estimated coefficients is statistically significant. Lack of information, awareness or even low education of parents may be possible explanations for the observed outcomes.

Table 5: 2nd stage IV regressions using Z-score Distance + Time as instruments on Parents satisfaction

Panel A: Parents satisfaction	(A) Teacher cares for students		(B) Students Performance		(C) Teacher is Motivated		(D) Teacher is Responsible	
dependent variable ---->	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Absenteeism in Past 2 Weeks	-0,468	-0,227	-0,901	-0,112	148,039**	1,067	-0,945	0,052
	(5,728)	(4,084)	(64,157)	(8,744)	(67,042)	(-10,092)	(6,943)	(0,137)
Teacher Controls								
Age	-0,011	-0,008	-0,023	-0,006	3,525***	0,039	-0,035	-0,021
	(0,140)	(0,166)	(1,173)	(0,354)	(1,204)	(0,384)	(0,142)	(-2,552)
Gender	0,258	0,145	0,452	0,125	-78,134**	-0,321	0,534	0,189
	(2,465)	(1,439)	(29,558)	(2,211)	(38,803)	(4,257)	(3,418)	(-26,901)
High Education	0,221	0,008	0,279	-0,075	-52,554**	0,307	0,348	-0,030
	(2,738)	(1,222)	(25,404)	(1,965)	(21,525)	(3,441)	(2,614)	(-15,634)
No. Dependents	0,004	0,020	-0,014	0,004	10,101**	-0,067	-0,069	0,032
	(0,070)	(0,331)	(0,741)	(0,510)	(5,094)	(0,310)	(0,566)	(5,885)
Parent Controls								
Age	0,003	0,003	0,004	0,002	-0,403**	0,005	0,006	0,005
	(0,044)	(0,015)	(0,090)	(0,036)	(0,175)	(0,056)	(0,042)	(0,077)
Gender	-0,033	-0,046	-0,045	-0,016	4,410	-0,065	-0,050	-0,074
	(0,543)	(0,779)	(4,241)	(1,136)	(8,630)	(1,624)	(0,617)	(3,478)
High Education	0,000	0,042	-0,020	0,080	0,044	-0,228	-0,069	-0,078
	(0,458)	(0,311)	(9,093)	(0,913)	(2,721)	(1,364)	(0,806)	(3,261)
Kleibergen-Paap	4,640	16,431	4,855	16,740	4,593	15,468	5,421	17,603
R2 adjusted	-1,253	-0,250	-2,561	-0,014	-28 577,654	-1,515	-2,134	-0,161
Observations	1 090	932	1 094	935	1 004	851	1 054	897

Note: IV regressions. Sample is Endline data. All regressions include Municipality and Treatment Dummies. All regressions include School controls (number of students, teachers, classrooms and dummies for whether the school has water, a bathroom and electricity). Standard errors are robust and bootstrapped with 500 replications. (1) uses Self-Reported data. (2) uses Director-reported data. * significant at 10%; ** significant at 5%; *** significant at 1%.

[14] Results with the other specifications can be found in table 5.b in annex

7 Concluding Remarks

Concerns for the chronic existing teacher absenteeism in education have been often explored by researchers in the field of development economics. In this paper, we address this issue using micro-level data and we find a clear negative effect of absenteeism in the students' outcomes, measured through a standardized test score. Also, we observe that the magnitude of this negative impact has increased from Baseline to Endline, leading to a decrease on the schools' average test scores over time. This underlines the relevance that teachers have on a student's overall education, thus highlighting the need for policy interventions which aim at overcoming this matter.

We find evidence to support the argument that mobility constraints are driving up absenteeism, given that teachers that live further away from the school where they teach or that take longer to get there tend to be absent more often, and in turn have students with lower test scores. Our paper does not explore this due to space constraints, but we suggest that future research should explore the heterogeneous effects that might arise within subgroups of teachers – teachers according to their age group or with a given income level. Given these results, we conclude that educational outcomes would benefit from structural policies that contribute to improve mobility of teachers (as well as students) between home and school (e.g., a transportation network for education). Further research should explore the effect of these same mobility constraints on students' absenteeism.

Sadly, we find no correlation between parent's satisfaction regarding student and teacher outcomes and teacher absenteeism. However, we do not discard the hypothesis that this seemingly inexistent causality may be due to limitations in our dataset.

We are aware of the noteworthy limitations that our dataset imposes in our approach. Our analysis is mostly static as we can only match teachers with students or parents at Endline. It would be interesting to further explore the dynamic dimension of teachers' absenteeism at the students' individual level. What is more, the use of survey measures suffers from clear limitations, especially considering that subjects may have incentives for lying (e.g., absenteeism) and there is large potential for measurement error (e.g., distance in kilometers).

Nevertheless, the fact that we use a microeconomic framework, that analyzes the socio-economic characteristics as well as preferences of a wide variety of teachers and students' households, allows us to extend our findings to other developing countries' scenarios, where there is a dominance of rural areas. We consider our paper particularly relevant in the context of Angola, where roads are poorly constructed and lack maintenance. With all this in mind, we suggest that greater emphasis from the government is given to address concerns related to infrastructure and transportation for education.

References

Abadie, A. and Imbens, G.W., 2008. On the failure of the bootstrap for matching estimators. Econometrica, 76(6), pp.1537-1557.

Angrist, J.D. and Pischke, J.S., 2008. Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Athey, S. and Imbens, G.W., 2006. Identification and inference in nonlinear difference-in-differences models. Econometrica, 74(2), pp.431-497.

Banerjee, A. and Duflo, E., 2006. Addressing absence. Journal of Economic perspectives, 20(1), pp.117-132.

Barrow, L., Schanzenbach, D.W. and Claessens, A., 2015. The impact of Chicago's small high school initiative. Journal of Urban Economics, 87, pp.100-113.

Baum, C.F., Schaffer, M.E. and Stillman, S., 2003. Instrumental variables and GMM: Estimation and testing. The Stata Journal, 3(1), pp.1-31.

Bertrand, M., Duflo, E. and Mullainathan, S., 2004. How much should we trust differences-in-differences estimates?. The Quarterly journal of economics, 119(1), pp.249-275.

Bertrand, M., Duflo, E. and Mullainathan, S., 2004. How much should we trust differences-in-differences estimates?. The Quarterly journal of economics, 119(1), pp.249-275.

Bourdon, J., Frölich, M. and Michaelowa, K., 2010. Teacher shortages, teacher contracts and their effect on education in Africa. Journal of the Royal Statistical Society: Series A (Statistics in Society), 173(1), pp.93-116.

Brown, S. and Sessions, J.G., 1996. The economics of absence: theory and evidence. Journal of economic surveys, 10(1), pp.23-53.

Burde, D. and Linden, L.L., 2013. Bringing education to Afghan girls: A randomized controlled trial of village-based schools. American Economic Journal: Applied Economics, 5(3), pp.27-40.

Card, D., 1993. Using geographic variation in college proximity to estimate the return to schooling (No. w4483). National Bureau of Economic Research.

Carlana, M., La Ferrara, E. and Pinotti, P., 2018. Goals and gaps: Educational careers of immigrant children.

Carneiro, P., Lokshin, M. and Umapathi, N., 2017. Average and marginal returns to upper secondary schooling in Indonesia. Journal of Applied Econometrics, 32(1), pp.16-36.

Chaudhury, N., Hammer, J., Kremer, M., Mularidharan, K. and Rogers, H., 2004. Roll call: Teacher absence in Bangladesh. Unpublished, World Bank.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K. and Rogers, F., 2004. Teacher and health care provider absence: A multi-country study. World Bank: DEC.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K. and Rogers, F.H., 2006. Missing in action: teacher and health worker absence in developing countries. Journal of Economic perspectives, 20(1), pp.91-116.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. H., 2004. Provider absence in schools and health clinics. Northeast Universities Development Consortium Conference, HEC Montreal.

De Groote, O. and Declercq, K., 2016. Tracking and specialization of high schools: does school choice matter?.

Duflo, E., Hanna, R. and Ryan, S.P., 2012. Incentives work: Getting teachers to come to school. American Economic Review, 102(4), pp.1241-78.

Duflo, E., Dupas, P. and Kremer, M., 2015. School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. Journal of Public Economics, 123, pp.92-110.

Glewwe, P. and Muralidharan, K., 2016. Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In Handbook of the Economics of Education (Vol. 5, pp. 653-743). Elsevier.

Glewwe, P., Holla, A. and Kremer, M., 2008. Teacher incentives in the developing world. Harvard University.

Glewwe, P., Ilias, N. and Kremer, M., 2003. Teacher Incentives" NBER Working Paper: No. 9671.

Glewwe, P., Maiga, E. and Zheng, H., 2014. The contribution of education to economic growth: A review of the evidence, with special attention and an application to Sub-Saharan Africa. World Development, 59, pp.379-393.

Glewwe, P., Maiga, E. and Zheng, H., 2014. The contribution of education to economic growth: A review of the evidence, with special attention and an application to Sub-Saharan Africa. World Development, 59, pp.379-393.

Gottfried, M.A., 2010. Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. American Educational Research Journal, 47(2), pp.434-465.

Hull, M. and Duch, K., 2016. One-to-one technology and student outcomes.

Kazianga, H., De Walque, D. and Alderman, H., 2009. Educational and health impacts of two school feeding schemes: Evidence from a randomized trial in rural Burkina Faso. The World Bank.

Khandker, S., B. Koolwal, G. and Samad, H., 2009. Handbook on impact evaluation: quantitative methods and practices. The World Bank.

Kling, J.R., 2001. Interpreting instrumental variables estimates of the returns to schooling. Journal of Business & Economic Statistics, 19(3), pp.358-364.

Kling, J.R., Liebman, J.B. and Katz, L.F., 2007. Experimental analysis of neighborhood effects. Econometrica, 75(1), pp.83-119.

Kremer, M. and Chen, D. 2001. "An Interim Report on a Teacher Attendance In- centive Program in Kenya." Mimeo, Harvard University.

Kremer, M. and Chen, D., 2001. An interim report on a teacher attendance incentive program in Kenya. Development Economics Department, Harvard University Cambridge, MA.

Kremer, M. and Vermeersch, C., 2005. School meals, educational achievement, and school competition: evidence from a randomized evaluation. The World Bank.

Kremer, M., Chaudhury, N., Rogers, F.H., Muralidharan, K. and Hammer, J., 2005. Teacher absence in India: A snapshot. Journal of the European Economic Association, 3(2-3), pp.658-667.

Kremer, M., Miguel, E. and Thornton, R., 2009. Incentives to learn. The Review of Economics and Statistics, 91(3), pp.437-456.

Miller, R.T., Murnane, R.J. and Willett, J.B., 2008. Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. Educational Evaluation and Policy Analysis, 30(2), pp.181-200.

Miller, R.T., Murnane, R.J. and Willett, J.B., 2008. Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. Educational Evaluation and Policy Analysis, 30(2), pp.181-200.

Muralidharan, K. and Prakash, N., 2017. Cycling to school: Increasing secondary school enrollment for girls in India. American Economic Journal: Applied Economics, 9(3), pp.321-50.

Muralidharan, K. and Sundararaman, V., 2013. Contract teachers: Experimental evidence from India (No. w19440). National Bureau of Economic Research.

Muralidharan, K. and Sundararaman, V., 2013. Contract teachers: Experimental evidence from India (No. w19440). National Bureau of Economic Research.

Muralidharan, K., 2017. Field experiments in education in developing countries. In Handbook of economic field experiments (Vol. 2, pp. 323-385). North-Holland.

Rindermann, H., 2008. Relevance of education and intelligence at the national level for the economic welfare of people. Intelligence, 36(2), pp.127-142.

Stock, J. and Yogo, M., 2005. Asymptotic distributions of instrumental variables statistics with many instruments (Vol. 6). Chapter.

Stock, J.H. and Yogo, M., 2002. Testing for weak instruments in linear IV regression.

Stock, J.H., 2015. Instrumental variables in statistics and econometrics.