Teacher human Capital, teacher Effort and Student Achievements in Kenya

Fredrick M. Wamalwa & Justine Burns

ERSA working paper 723

January 2018

Economic Research Southern Africa (ERSA) is a research programme funded by the National Treasury of South Africa.

Teacher Human Capital, Teacher Effort and Student Achievements in Kenya

Fredrick M. Wamalwa* & Justine Burns†

Although research generally shows that teachers matter for student achievements, it is not clear what attributes make a teacher more successful in enhancing students' performance. Much of the empirical work has focused on observable teacher characteristics. In this paper, we go further than that. We examine how teacher *subject knowledge*, *pedagogical skill, effective instruction time* and *classroom teaching practices* influence student test scores in maths and language. Results indicate that teachers who spend more time on instruction and can keep student on-task during the lesson are associated with higher student test scores. A number of classroom teaching practices, including the practice of challenging students by asking questions during the lesson, have an effect on student test scores although the effect differs between language and maths.

Key words: Teachers, Kenya, Student Achievements.

1 Introduction

Over the last two decades, efforts at national and global levels were devoted to ensuring that by 2015, all children have access to free quality primary education. At the global level, the United Nations' Millennium Development Goals provided overarching political commitments that guided the achievement of this goal. The overwhelming evidence that user-fees were a major barrier to access (Riddell, 2003; Bruns and Rakotomalala, 2003) provided the basis for the key reform of making public primary education free (World-Bank, 2009). Kenya enacted the policy in 2003 while other sub-Saharan African countries did the same between 1994-2005.[1]

The free primary education policy reversed hitherto declining primary school enrollments in most developing countries. Across the developing world, the post free primary education era has been characterized by huge investments in school infrastructure (Pritchet and Banerji, 2013). In Kenya, recent studies and school surveys show that schools universally equipped with teaching resources and there is no much difference between public and private schools. In 1999, nine students shared a textbook in grade 3 close (Hardman et al., 2009; MoEHRD, 1999). A few years later, government has

* *Corresponding author*: School of Economics, Faculty of Commerce, University of Cape Town, Private Bag, Rondebosch 7701, Cape Town, South Africa. E-mail: fredwamalwa@yahoo.co.uk. Cell +254 726 406 371

† School of Economics, Faculty of Commerce, University of Cape Town, Private Bag, Rondebosch 7701, Cape Town, South Africa. E-mail: justine.burns@uct.ac.za. Tel: +27 21 6503506

[1] Others include Ethiopia and Malawi (1994), Uganda (1997), Lesotho (2000), Mozambique (2004) and Ghana (2005) (Bold et al., 2013; World-Bank, 2009).

nearly achieved its policy of three students per text book at primary level in English and maths, at least according to the data for this study.

While strides have been made in terms of investment in school infrastructure and access to schooling opportunities, evidence shows that a large share of children in developing countries learn little in school and complete their education lacking even basic reading, writing and arithmetic skills (Pritchet and Banerji, 2013). Annual evidence gathered by Twaweza East Africa[2] finds that a good proportion of children in the final grade of primary school cycle in Kenya, Uganda, and Tanzania, are unable to read a simple passage or perform simple addition – skills that should have been mastered by the end of the second year of primary school.

Such poor learning outcomes, coming in the wake of increasing public spending on education inputs raises an important research question as to whether school inputs matter for student achievement and if they do, which inputs matter. A consistent finding in the literature, in both developed (Hanushek, 2006) and developing countries (Pritchet and Banerji, 2013; Glewwe et al., 2011), is that school inputs explain very little of the variation in student learning. Despite this, studies show that one school input, namely, the teacher input, matters for student achievement. Nevertheless, research is not clear on what makes an effective teacher (Aaronson et al., 2007). Most of the studies in developed (Boyd et al., 2006; Clotfelter et al., 2007; Rivkin et al., 2005; Rockoff et al., 2010) and developing (Glewwe and Jacoby, 1994; Aslam and Kingdon, 2011; Glewwe, 2002) countries have looked at the effect of characteristics of teachers that can easily be measured (observed) on student achievement. The majority of these studies find that these common observable characteristics such as certification, education, experience (as well as gender and age) often used for teacher remuneration and promotion, explain little of the variation in student performance(Rockoff, 2004; Rivkin et al., 2005; Clotfelter et al., 2007; Kukla-Acevedo, 2009; Dee, 2005).

Latest literature has begun to emphasize the importance of two aspects of teacher input: teacher competence (knowledge) and teacher effort. According to this literature, when teachers, exert the necessary effort and are appropriately skilled, increased resource flows in schools can lead to better student outcomes (Spence and Lewis, 2009; Swanson et al., 2012; Hanushek, 2008; Pritchet and Banerji, 2013; Glewwe et al., 2011). An increasing number of studies have taken an empirical look at the effect of teacher knowledge (such as Shepherd et al. 2015; Shepherd 2013) and teacher effort, measured by *what teachers do in class* (such as Kane et al. 2011; Zakharov et al. 2014; Lavy 2011, 2012; Bietenbeck 2011; Hidalgo Cabrillana and Lopez-Mayan 2015; Jacob and Lefgren 2005; Kingdon 2006) on student achievements.[3]

These studies generally find that teacher knowledge and teacher classroom practices matter for student achievement. However, the findings are also not conclusive. For instance, Schwerdt and Wuppermann (2011) along with Van Klaveren (2011) study the effect of the percentage of time spent in lecture-style teaching using the Trends in International Mathematics and Science Study (TIMSS) wave of 2003 for the United States and the Netherlands, respectively. While Schwerdt and Wuppermann (2011) find that shifting time from problem solving to lecturing results in an increase in student achievement,

[2]Twaweza East Africa (https://www.twaweza.org) is a regional non-governmental organization that carries out household level literacy and numeracy assessments among children aged 6-16 years in Kenya, Uganda and Tanzania

[3]The terms student test scores and student achievements will be used interchangeably throughout this thesis.

Van Klaveren (2011) finds no relationship. In addition, most of these studies are based on data from developed countries.

This paper contributes to this body of literature by examining the effect of *teacher human capital* and *teacher effort* on student achievement in maths and language in Kenya. We define *teacher human capital* by two measures: *what teachers know* (teacher subject knowledge) and *whether the teachers know how to teach* (teacher pedagogical skill). *Teacher effort* is measured by three indicators, namely; (i) teachers' effective instruction time, (ii) teacher's ability to keep students engaged during the teaching lesson and (iii) a number of teacher classroom practices. We use a recent schools survey that allows us to directly link student test scores to the human capital and effort of their specific teachers while controlling for a number of child, teacher, classroom and school characteristics. This allows us to look at the education production function with greater depth than has been done before in the context of sub-Saharan Africa, and in particular Kenya.

Here is a summary of our results. We find that student test scores in maths and language are partially influenced by our measures of teacher human capital and teacher effort. A one standard deviation increase in the teacher's knowledge in language (maths) increases student test score in language (maths) by 0.075 (0.126) of a standard deviation. An additional hour of teacher effective instruction time increases student achievement in language and maths, respectively, by 0.051 and 0.059 of a standard deviation. There is evidence that a number of classroom teaching practices, including the practice of challenging students by asking questions during the lesson, have an effect on student test scores although the effect differs between language and maths.

The chapter proceeds as follows. In section 2 we describe the data. Section 3 provides a detailed description of our measures of teacher human capital and teacher effort. Section 4 discusses our main results including robustness checks and finally section 6 provides the conclusion.

2 The Service Delivery Indicators (SDI) Survey

We use the 2012 Education Service Delivery Indicators Survey for Kenya. Launched in 2009, the Service Delivery Indicators (SDI) is an African-wide initiative of the World Bank, the African Development Bank (AfDB) and the African Economic Research Consortium (AERC). It aims at collecting data on how services are delivered at front-line primary service provider facilities in health (health facilities) and education (schools).

The SDI survey of education in Kenya conducted in 2012 is a cross-sectional survey covering 306 primary schools. The survey used a multistage cluster sampling strategy where counties[4] were categorized as (i) rural or urban, based on the 1999 national census data; (ii) relatively rich or poor, based on the 1999 national census data;[5] and (iii) high or low-performing based on county-level 2010 Kenya Certificate of Primary Education (KCPE) pass rates. These three binary distinctions yielded 8 strata within which schools were sampled. Within each stratum, counties were randomly selected, followed by a random selection of locations within each county, and then a random selection of schools within

[4] Counties are units of devolved government as envisioned in the 2010 Constitution of Kenya..

[5] The 1999 census was the latest data set on urbanization and poverty rates available to the researchers at the time of the sampling process for the SDI schools survey.

the locations. The probability of selecting a county or a location was proportional to the population within it.[6] The data collection involved *three* major components briefly described below.

2.1 Classroom Observation

One of the unique features of the SDI survey is the fact that it involved classroom observation. In every school, a trained surveyor observed a teacher delivering a 35-minutes grade 4 lesson in either language (English)[7] or mathematics but not both. In the cases where there was more than one class of grade 4 (e.g grade 4A and 4B), the surveyor randomly selected one class using a methodology taught during the training. [8] The observations were based on an adapted version of the Stallings Classroom Snapshot instrument (Stallings, 1977; Stallings and Knight, 2003).[9] For every minute, the surveyor scanned the class in a 360-degree circle starting with the teacher and recorded what was happening. This *snapshot,* took about *15 seconds* and captured *what was happening at that instant* but not *what took place during the entire minute interval.*

Of the 306 schools, classroom observation took place in 276 schools where in each school, only *one teacher was observed in either a language or mathematics lesson*. In 28 schools, teachers were absent from class and in 2 schools, classroom observation took place for creative arts and science subjects. We exclude these 30 schools from the analysis. Of the 276 schools where classroom observation took place, language lessons were observed in 144 schools while mathematics lessons were observed in 132 schools.

2.2 Pupil and Teacher Assessments

After the classroom observation, approximately 10 pupils from the class that had been observed were randomly selected and assessed in mathematics, language and non-verbal reasoning. The tests were largely based on materials up to grade 3 and were aimed at assessing the child's basic reading, writing, and arithmetic skills. They were designed by experts in international pedagogy and based on a review of primary curriculum materials from 13 African countries, including Kenya (Martin and Pimhidzai, 2013). They were administered as a one-on-one interaction, where the surveyor read out instructions to pupils. Where necessary, the surveyor read the instructions to the pupil in their mother tongue.

The student language test consisted of a number tasks ranging from alphabet and word recognition, to more challenging tasks like sentence and paragraph reading, and eventually tasks involving comprehension of written material. The mathematics test included tasks ranging from number identification and sequencing, to single and double digit addition and subtraction, to single digit multiplication and division. The non-verbal reasoning test consisted of tasks related to pattern recognition based on Raven matrices test (see Martin and Pimhidzai (2013) for a detailed description).

[6] Of the 47 counties in Kenya, 3 counties of the region previously known as North Eastern province were excluded due to security concerns at the time. North Eastern Province was one of eight administrative regions that preceded creation of the counties.

[7] Throughout this document, instead of calling this an English test, we refer to it as language test.

[8] Schools were not informed of their selection and were not advised of the visits in advance. Teachers did not receive advance notice that they would be observed.

[9] The Stallings instrument uses a standardized coding grid to register minute by minute activities and materials being used by the teacher and students over the course of a single lesson(Bruns and Luque, 2014).

In addition, the survey assessed the knowledge of 1,679 primary teachers in language, mathematics and pedagogy. All current grade 4 mathematics and language teachers (including those observed in class) and those who taught grade 4 mathematics and language during the previous year (2011) were assessed. The language test administered to teachers consisted of 22 items involving grammar, cloze[10] and composition tasks. The mathematics test consisted of 15 items related to addition, subtraction, multiplication, division, fractions, interpreting graphs and data. The pedagogy test was designed to capture skills teachers would routinely be asked to apply when teaching.[11] As a courtesy to teachers, the teacher tests were designed as a marking exercise, in which teachers was asked to mark and correct a hypothetical student's exam (Martin and Pimhidzai, 2013).

Out of 276 teachers in 276 schools (one teacher per school) who were observed in class, only 222 teachers took part in the teacher assessments. The rest, 54 teachers, did not take part in the teacher assessments. We do not know why these teachers opted not to participate in the teacher assessment exercise. We however know that participation in classroom observation and teacher tests was voluntary. This effectively means that our sample has 222 schools where one teacher was concurrently observed in class (either in language or mathematics) and assessed in the teacher tests. Of these, 109 teachers were observed in language and the rest, 113 teachers, were observed in mathematics.

One might be concerned with the presence of systematic differences, on account of school and teacher characteristics, between the different samples: the full sample, the sample of 222 schools where the language or maths teacher was observed and tested, the sample of 54 schools where the language or maths teacher was observed and not tested and the sample of 30 schools where no classroom observation took place. In Table A.1 (in the appendix), we do not find systematic differences between these samples based on a range of school characteristics. Table A.2 (in the appendix) shows that there is no systematic differences, on a range of teacher classroom practices, between the 222 teachers, who were observed and tested and 54 teachers, who were observed and not tested.

2.3 Teacher Absenteeism

Besides class observation and test assessments, the SDI survey involved an assessment of absenteeism of 2,960 teachers. During the first visit, the surveyor randomly selected a maximum of ten teachers from the list of all teachers in the school. In schools where there were less than 10 teachers all teachers were selected. The whereabouts of the selected teachers were then verified during *a second unannounced visit* based on five mutually exclusive options: (a) *teacher was in class teaching*; (b) *teacher was in class but not teaching*; (c) *teacher was present in school but not in class*; (d) *teacher was in school but teaching outdoors* and; (e) *teacher was absent from school.* We define absenteeism as the ratio of teachers (out of ten) who were *present in school but not in class (c above)* and *were absent from school (e above).* On average, the second unannounced visit took place about one or two weeks after the first visit.

The final part of the survey protocol was a structured interview with the school headteachers (or

[10] The cloze consisted of a passage with certain words removed (cloze text), where the teacher was asked to replace the missing words.

[11] Details about the mathematics, language and pedagogy test items administered to teachers are provided later in the text.

substitute) to gather information about the school demographics including information about the teaching staff, school infrastructure and teaching resources.

3 Descriptive Statistics

3.1 Overview of School Inputs

Table 1: Selected Descriptive Statistics

		All Mean	Public Mean	Private Mean	Public-Private Mean Diff.
(a)	**School Characteristics**				
	School is rural	0.68	–	–	–
	Classroom has a blackboard and a chalk (%)	0.99	0.99	1.00	-0.01
	Share of pupil with a pen & an exercise book (%)	0.98	0.97	0.99	-0.01
	Average number of students per textbook (maths)	2.59	2.81	1.91	0.9***
	Average number of students per textbook (English)	3.61	4.03	2.09	1.9
	School has toilets (%)	1.00	1.00	0.99	0.01
	Sufficient light for reading from the back of the class (%)	0.85	0.85	0.88	-0.04
	School has electricity connection (%)	0.14	0.09	0.34	-0.26***
	Pupils per Teacher	30.2	34.9	19.7	15.2***
(b)	**Teacher Characteristics**				
	Female	0.54	0.53	0.59	-0.06***
	Age (in years)	38.11	40.53	29.13	11.40***
	Experience (in years)	13.90	16.13	5.65	10.48***
	Highest Level of Education Completed				
	Secondary Complete	0.30	0.28	0.35	-0.06***
	Diploma/Certificate	0.63	0.62	0.50	-0.01
	University degree	0.08	0.09	0.02	0.07***
	Highest Teacher Training Completed				
	None	0.13	0.10	0.24	-0.14***
	Certificate in Early Childhood Education	0.10	0.07	0.23	-0.16***
	Primary 1 and 2 certificate	0.59	0.62	0.47	0.15***
	Degree	0.18	0.22	0.07	0.15***
(c)	**Child Characteristics**				
	Student is female	0.51	0.50	0.47	0.03
	Student's age (in years)	10.43	10.60	9.74	0.9***
	Student attends a private school	0.22	–	1.00	–
	Student had breakfast	0.87	0.85	0.94	-0.1***
	Number of children	**2,953**	**2,378**	**575**	
	Number of schools	**306**	**239**	**67**	

Source: Own calculations based on SDI 2012. Notes: (1) ***significance<1 percent, ***significance<5 percent, ***significance<10 percent.

We begin by showing a general picture of school inputs that support student learning. Table 1 (a) shows summary statistics of the school environment. Majority of the schools are based in rural areas and are public. Generally, schools are almost universally equipped with minimum teaching resources, which we define as availability of: (a) functioning blackboard[12], (b) chalk (c) writing materials (pens

[12]The blackboard is defined as functioning if a text written on it could be read at the front and back of the classroom.

and exercises books), and (d) number of students per textbook. For instance, over 90 percent of schools have classes with a blackboard and piece of chalk and students provided with writing materials (pencil/pen and an exercise book). Although there are more students per textbook in public relative to private schools, the government has nearly achieved its policy of three students per textbook at primary level, especially in mathematics. Schools also do well in terms of minimum infrastructure, defined as availability of: (a) functioning toilets (toilets that are clean, private, and accessible) and (b) sufficient light to read the blackboard from the back of the classroom. For instance, all schools have toilets with over 90 percent of them having toilets that are designated for boys and girls, accessible (unlocked and not overflowing) and private (had doors or separating entryway wall). Access to electricity, although not defined as a minimum functioning infrastructure requirement, is generally low, especially among public schools.

Table 1 (b) shows selected teacher *observable characteristics* based on 2,960 teachers who were assessed in teacher absence. There were slightly more female teachers. On average, teachers in the sample are aged 38 and their overall mean years of experience in the teaching profession is 13. A public school teacher has more than 10 years of experience relative to his/her private school counterpart. Majority of the teachers have a diploma/certificate as their highest education attainment and hold a primary 1 and 2 certificate in teacher training. Lastly, Table 1 (c) summarizes children's characteristics.

Generally, table 1(a) to (c) shows that children who attend private schools do seem to have disproportionately high academic potential and access to complementary educational resources relative to their counterparts who attend public schools. For instance, their parents have more years of education and they come from households with higher wealth index. They are also taught by more knowledgeable teachers and teachers are also more likely to be in school and in class. This might raise concerns as to whether the higher performance in private schools is a reflection of endogenous selection of students from better backgrounds.

3.2 Student Performance in Test Scores

When children enroll in school and progress through specific grades, they are expected to master specific competencies in each grade. Children are expected to master skills and competencies outlined in the curriculum from the earliest grades. As we have already noted, the child tests in SDI survey were set at grade 3. These tests provide an effective way of assessing whether children are mastering competencies necessary for further learning. There are two ways we can proceed with assessing learning competencies. We can calculate the overall mean score for each tested competency or calculate the share of children achieving a specific competency level. Since SDI assessments are anchored in the content of the national curriculum, the second metric is more meaningful because it helps to assess whether children meet the expected numeracy and literacy competency requirements.

We begin describing our data by showing the proportion of children who meet specific competence levels for language and maths. This is shown in table 2 and table 3 respectively. Slightly more than

Table 2: Student Test Scores in Language

Sub-task	(1) Percent of students (combined public and private schools)	(2) Percent of students (public schools only)	(3) Percent of students (private schools only)
Could identify letters[a]	0.91	0.90	0.98
Could identify words[b]	0.89	0.87	0.98
Could read sentence[c]	0.72	0.67	0.94
Could read a paragraph[d]	0.16	0.11	0.40
Factual comprehension [e]	0.39	0.29	0.79
Analytical comprehension [f]	0.45	0.38	0.74
Sample	**2,953**	**2,378**	**575**

Source: SDI 2012. Notes: (1) We show the share of students who managed to perform each literacy items tested

[a] The child was shown 9 letters and asked to identify 3.
[b] The surveyor read out 9 words to the student and asked him/her to identify any 3 of them.
[c] In this sub-task, a student was presented with a 10 word sentence and asked to read.
[d] The student was presented with one paragraph of 58 words and asked to read.
[e] The factual comprehension question measured the learner's ability to identify stated facts in the passage they had read.
[f] The analytical comprehension question required the learner to at least infer or draw conclusions from opinions or ideas implied or suggested in the text.

Table 3: Student Test Scores in Maths

	(1)	(2)	(3)
Sub-task	Percent of students (combined public and private schools)	Percent of students (public schools only)	Percent of students (private schools only)
Could identify a number[a]	0.98	0.97	0.99
Could discriminate quantities[b]	0.73	0.71	0.83
Could do addition[c]	0.89	0.88	0.96
Could do subtraction[d]	0.63	0.58	0.82
Could do multiplication[e]	0.54	0.49	0.74
Could do division[f]	0.41	0.35	0.64
Multiplication (word problem) [g]	0.16	0.11	0.40
Complete sequence[h]	0.26	0.23	0.34
Sample	**2,953**	**2,378**	**575**

Source: SDI 2012. Notes: (1) We show the share of students who managed to perform each numeracy items tested

[a]The surveyor read out 9 numbers (mixture of single and double numbers) to the student and asked him/her to identify any 3 of them.

[b]The student was presented with 6 numbers of different quantities and asked to order them in a descending order.

[c]This tasked involved a student being asked to solve three addition problems involving respectively, single, double and triple digits. Here, we define a student as *'could do addition'* if he/she answered any of two problems correctly.

[d]This task involved one single and one double digit subtraction. Here, we define a student as *'could do subtraction* if he/she answered the two problems correctly.

[e]This tasked involved a student being asked to solve three multiplication tasks involving respectively, single, double and triple digits. Here, we define a student as *'could do multiplication* if he/she answered any of two problems correctly.

[f]This task involved three one division tasks involving both single and double digits. We define a student as *'could do division'* if he/she answered any of two problems correctly.

[g]In this task, a student was presented with an arithmetic (single digit) multiplication task presented in a word format that required some level of thinking to solve.

[h]In this task, a student was present with a descending number series following some pattern involving division by three with one last missing last value. He/she was asked to fill in the last missing value.

half of the sample are female students. Students generally do well in simple literacy tasks such as letter and word identification. Nevertheless, about 10 percent of grade 4 students could not identify or read correctly simple words presented to them. Oral reading fluency, which we measure by the ability to read connected text accurately, is still low among learners.[13] Although 72 percent of students could read all the 10 words in a basic sentence, only 16 percent managed to read correctly and accurately all the 58 words in a paragraph even after three years of primary schooling.

We find weak literacy skills when it comes to comprehension, that is the ability to implicitly and explicitly understand simple information that is stated in the text. Less than half of the students (39 percent could answer a question testing their ability to identify directly stated facts in the passage (factual comprehension). Equally, less than half of the students could answer questions that required them to infer or draw conclusions from opinions or ideas implied or suggested by the author of the text (analytical comprehension). Surprisingly, as column 1 in table 2 shows, more students could answer questions about a text (e.g. 39 percent answered correctly a factual comprehension question) than could read correctly all the words in the same text (e.g 16 percent could read all words in a paragraph). Furthermore, more students did better in analytical comprehension than in factual comprehension.

Table 3 shows student scores in maths. More than 70 percent of the learners generally do well in items that require procedural fluency. They can comfortably count and match numbers, identify numbers, discriminate quantities and add. Learners however faced difficulty dealing with relatively complex tasks involving subtraction, multiplication and division. We see that 37 percent of students could not solve two subtraction problems, one involving single digits and another involving double digits.

When presented with three multiplication problems involving single, double and triple digits, only 54 percent of grade 4 students managed to solve at least two problems. Basic numeracy skills in division are particularly low among students; less than 30 percent of the students were able to answer at least two out of three division problems involving mixed single and double digits. Learners are generally weak at solving problems that require some level of analytical reasoning, especially arithmetic word problems; only 16 percent could solve a simple single digit multiplication problem presented in word format and only 26 percent could solve a simple division problem involving completing a series, both problems requiring some level of reasoning.

For schools to properly function, they need teachers who are knowledgeable and skilled. First, teachers *need to know the subject they teach* (teacher subject knowledge) and second, they *need to know how to teach* (teacher pedagogy). We refer *teacher subject knowledge* and *teacher pedagogy* to as *teacher human capital.* Second, schools need teachers *who exert the necessary effort in applying their knowledge and skills* (*teacher effor*t). In the two sub-sections that follows, we provide a quantitative description of the *teacher human capital* and *teacher effort* before empirically linking them to the student test scores.

[13]Oral reading fluency is the ability to read a connected text *accurately*, *quickly*, and with *expression*. Oral reading fluency therefore assesses three aspects: *speed*, *accuracy*, and proper *expression*. The test in the SDI largely assessed student's ability to read correctly (accurately) every word in the text. Aspects of speed and expression were not assessed. We therefore proxy for oral reading fluency by the student ability to read correctly (accurately) every word in a text presented to them.

3.3 Teacher Subject and Pedagogical Knowledge

Next, discuss the aspect of *teacher human capital* based on the test score results for the 1,679 teachers who were assessed in language, maths and pedagogy.[14] In table 4, we show the teacher test scores by subject. The language test comprised of two sets of tasks. The first task, *spelling and simple grammar exercises*, involved tasks that were similar to student language tasks and largely covered lower primary curriculum materials.[15] To effectively teach a grade four student, the teacher needs to have knowledge that goes beyond the lower primary curriculum. As a result, the next two tasks, *cloze passage*[16] *and composition*,[17] covered materials that were slightly beyond the lower primary curriculum and hence a grade 4 child would have found these tasks relatively difficult.

In line with Martin and Pimhidzai (2013), we show an indicator, *the share of teachers with minimum knowledge,* defined as the proportion of teachers who marked 80 percent or more of *all the language tasks correctly.* Two main observations stand out from the language scores (table 4). First, content knowledge among Kenyan teachers is relatively low; only 13.16 percent of teachers can be classified as having *minimum knowledge* in language (column 1). Second, an increasing number of teachers struggle with tasks requiring some level of knowledge beyond lower primary curriculum. For instance, while nine out of ten teachers marked tasks involving *spelling and simple grammar exercises* correctly, almost half of them could not mark the *composition* sub-task, involving correcting spelling, grammar as well as punctuation mistakes in a child's letter.

Similarly, the maths tests contained tasks that were common across the student and teacher tests (such as addition and subtraction) as well as tasks that were slightly beyond the lower primary curriculum. In table 4, we dis-aggregate the scores by selected sub-tasks. Slightly more than half of the teachers, 56.34 percent, can be classified as having *minimum knowledge* in maths. Nevertheless, a high proportion of teachers also faced difficulties in relatively advanced tasks. For instance, nearly 32 percent of the teachers could not solve a problem involving division of fractions and close to 38 percent could not solve a problem involving interpreting information in a graph.

[14] As mentioned, teacher tests were administered to teachers as a marking exercise.

[15] The purpose of the *spelling and simple grammar exercises* was to assess if the teacher had knowledge equivalent to a grade 4 learner.

[16] The cloze consisted of a passage with certain words removed (cloze text), where the teacher was asked to replace the missing words.

[17] In this sub-task, the teacher was given a sample letter written by a grade 4 child and asked to correct the letter for grammar, punctuation (between sentences and within sentences), spelling, syntax, and salutation.

Table 4: Teacher Test Scores by Subject

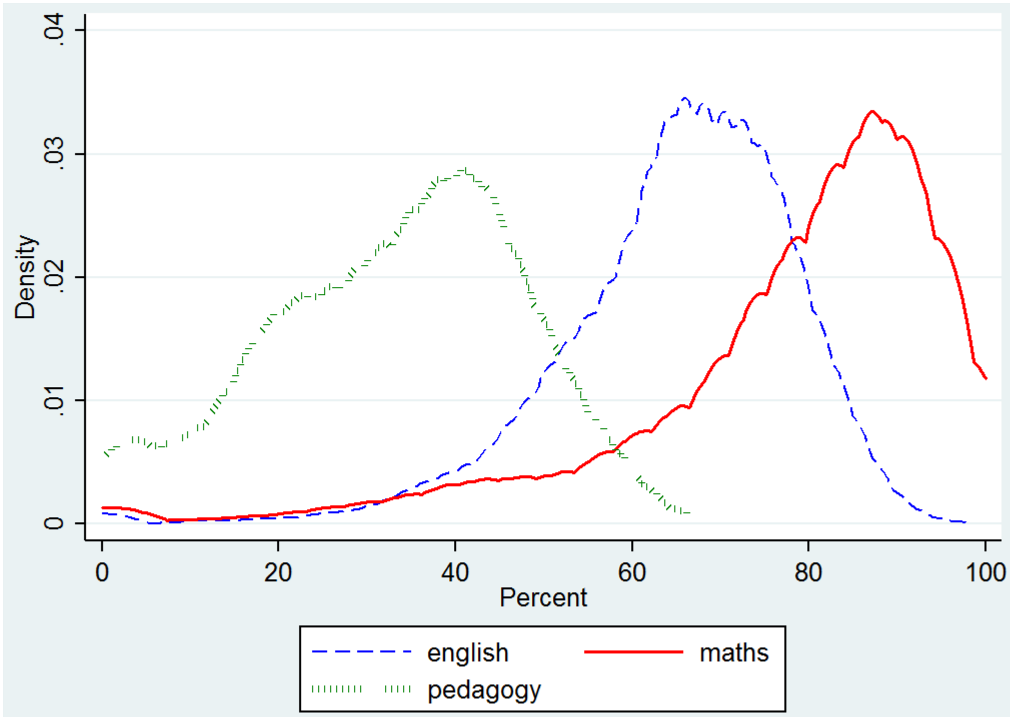
	(1) All Mean	(2) Public Mean	(3) Private Mean	(4) Public-Private Difference
Language Test				
Minimum knowledge in Language: 80 % correct	13.16	11.97	18.60	-6.63***
English Section mean score (% correct)	64.97	64.55	66.87	**-2.26*** **
% correct in spelling and grammar exercises	92.00	91.81	92.89	-1.01
% correct in cloze	67.25	66.70	69.77	-3.06***
% correct in composition	45.95	49.52	51.91	-2.39***
Maths Test				
Minimum knowledge in Maths: 80 % correct	56.34	54.14	66.45	-12.31***
Maths Section mean score (% correct)	77.59	76.79	81.24	**-4.44*** **
% correct in double digit addition	97.50	97.39	98.01	-0.62
% correct in double digit subtraction	87.19	87.01	88.04	-1.03
% correct computing a perimeter of a rectangle	80.05	78.59	86.71	-8.12***
% correct in one variable algebra	72.66	71.48	78.07	-6.59***
% correct in division involving fractions	68.43	66.40	77.74	-11.34***
% correct in interpreting data on a graph	62.48	61.39	67.44	-6.05***
Pedagogy Test				
Minimum knowledge in Pedagogy: 80 % correct	0	0	0	0
Pedagogy mean score (% correct)	**32.85**	32.24	35.65	**-3.41*** **
% correct in lesson plan formulation	43.60	43.19	45.46	-2.27*
% correct in assessing child ability	34.27	33.47	37.96	-4.49***
% correct in assessing child progress	29.91	28.78	35.08	-6.30***
Sample	**1,679**	**1,378**	**301**	

Source: Own calculations based on SDI 2012. ***significance<1 %, **significance<5 %, *significance<10 %

The pedagogy test consisted of three sub-tasks and was designed to capture skills teachers would normally use when teaching. In the first sub-task, teachers were asked to come up with a lesson plan based on a topic that was given to them on the spot allowing no time for prior preparation. In this sub-task, teachers were supposed to formulate (a) aims and objectives of the lesson, (b) questions to check student understanding and (c) questions to check student's to application of what they have learned to other contexts. In the second sub-task, teachers were given samples of writing from two grade 4 level students and asked to mark and comment on the strength and weakness of each student. Lastly, teachers were given raw scores for 10 students and asked to calculate the averages and comment on the performance.

As table 4 shows, teacher's pedagogical knowledge is exceptionally low. From column 1, we can see that no teacher could be considered as having minimum knowledge in pedagogy (none managed to mark correctly 80 percent of the pedagogy tasks). Furthermore, less than half of them (43.60 percent) could adequately prepare the lesson plan. On the other hand, only 34.27 percent of teachers could contrast between sample writing of two students and only 29.91 percent could turn raw scores into averages and comment on student performance. Private teachers performed better than public teachers in almost all subject categories although the general lack of pedagogical skills appear to be similar in public and private schools. Figure 1 shows the distributions of the test scores.

Figure 1: Teacher Test Scores in Pedagogy, Language and maths.



Source: Own calculations based on SDI 2012.

3.4 Teacher Effort

Education delivery in most parts of the world requires the physical presence of teachers in school and in class. Put differently, for students to learn, teachers need to exert the necessary effort in applying their knowledge and skills. They need to turn up to school and to class and once in class, spend time on instruction. Teacher's *presence in school (or absence from school)* has dominated literature as a measure of teacher effort (Chaudhury et al., 2006; Alcazar et al., 2006). However, even when in school, teachers are not necessarily in class. As such, teacher effort has also been measured by the presence of teachers *not just in school but in class*. Most recently, studies have begun to measure teacher effort by the *amount of time teachers spend on instructional activities while in class* (Bruns and Luque, 2014). Next, we discuss these three aspects of teacher effort (*absence from school, absence from class and time-on-task)* and show how they correlate with student achievements.

3.4.1 Teacher Absence from School and Class

As mentioned in Chapter 2, up to 10 teachers in each school were randomly selected from the teacher roster. A total of 2,960 teachers were selected. After 1-2 weeks, an unannounced visit was conducted during which surveyors identified whether the sampled teachers were in the school, and if they were, whether they were in class teaching. We derive two indicators of teacher effort based on this information. The first is *absence from school*, defined as the share (out of 10) of teachers who could not be found on the school premises during the second unannounced visit. The second indicator is *absence* *from class,* defined as a combined share (out of 10) of teachers who were *absent from school* and *absent from class but present on the school premises.*[18]

In addition, during the second visit, the surveyor physically counted the number of classrooms with students and noted whether a teacher was inside the classroom or not. We divide *the number of classrooms that had students but no teacher by the total number of classrooms that contained students* to obtain another complimentary measure of teacher effort, which we call o*rphaned classrooms.*

Table 5: Teacher Effort Indicators: Teacher Absence from School and Class.

	All	Public	Private	Public.-Private
	Mean	Mean	Mean	Mean Difference
Absence from class (percent)	44.93	47.97	34.07	13.90***
of which				
Absence from school (percent)	16.68	17.51	13.74	3.79
Absent in class but present in school (percent)	28.25	30.46	20.33	10.13***
Orphaned classes (percent)	34.88	38.46	22.10	16.36***
Sample	**306**	**239**	**67**	

Source: Own calculations based on SDI 2012. Notes: (1) Results collapsed at school level; (2) We calculate absence from class as the sum of the indicators (a) absence from school and (b) present in school but absent in class; and (3) ***significance<1 %, **significance<5 %, *significance<10 %

As table 5 shows, on average, 16.68 percent of teachers were absent from school. Although absence from school is slightly higher among teachers in public schools, we do not find any significant differences between public and private school teachers on this indicator. The table further shows that a larger proportion of teacher's lost time is actually lost in schools. We see that 28.25 percent of teachers *were present in school but absent from the classroom* and there are significant differences in this indicator between public and private schools. Recall that the indicator absence from class is defined as a combined share (out of 10) of teachers who were *absent from school* and *absent from class but present on the school premises*. In relation to table 5, it means that 44.93 percent of the teachers were absent from class and a large proportion of these teachers are from public schools. On the orphaned classrooms indicator, we find that about 35 percent of classes had pupils with no teacher. The mean differences between public and private is quite significant for this indicator.[19]

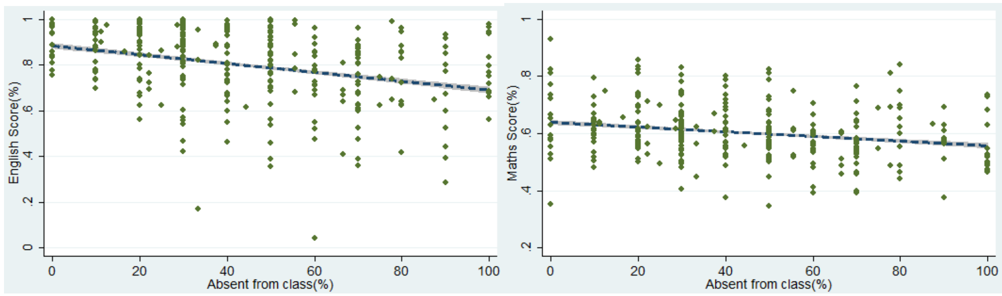
In figure 2a and figure 2b, we plot student test scores (in language and maths) against *teacher absence in class* indicator and *percent of orphaned classes* indicator, respectively. Both are calculated at the school level. The results consistently show that students perform worse in test scores in schools where teachers are more likely to be absent from class and in schools with high proportion of orphaned classes. The correlation is particularly stronger for language scores.

[18] It is possible that some teachers could have been on the school premises and not in class because it was not their teaching time. As a result, the surveyor used the official school timetable to verify that the teachers who were present on school premises and not in class were indeed supposed to be in class teaching at that time.

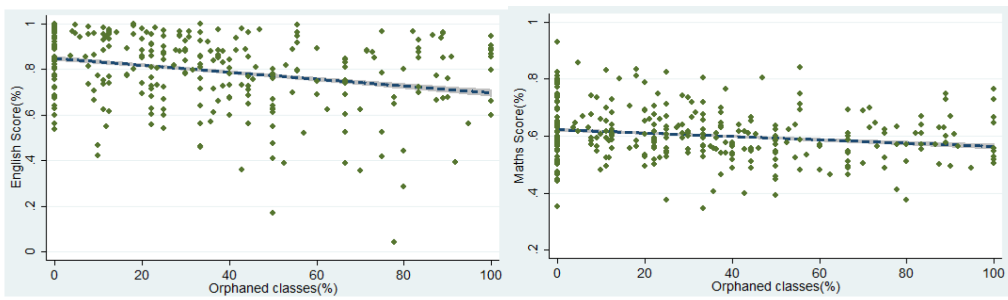
[19] The percent of orphaned classes in table 5 is not an average of (i) absence from school and (ii) absent in class but present in school. This is because the indicators (i) absence from school and (ii) absent in class but present in school are only based on 10 teachers who were sampled for teacher absence.

Figure 2: Correlations between Teacher Effort Indicators and Student Test Scores

(a) Correlation between Absence from Class and Student Scores

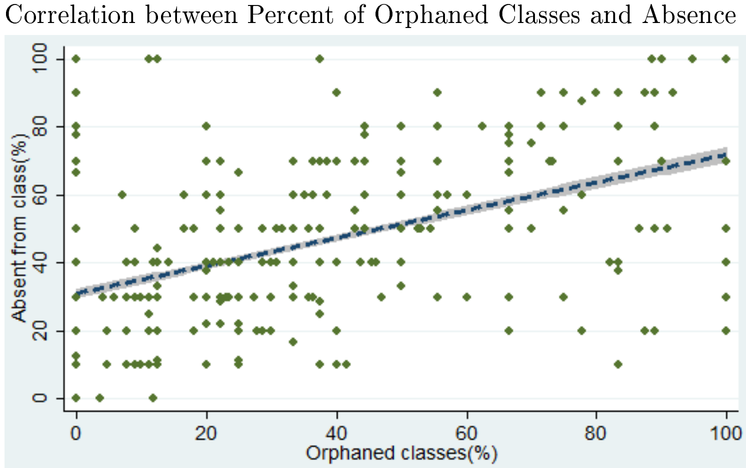


(b) Correlation between Percent of Orphaned Classes and Student Test Scores



Source: Own calculations based on SDI 2012.

Figure 3: Correlation between Percent of Orphaned Classes and Absence from Class



Source: Own calculations based on SDI 2012.

In calculating absence from school and class, we do not account for the fact that some teachers could have been absent from school and/or class for legitimate reasons. This is likely to over-estimate teacher absence rate, in school and in class. The absence of a teacher from class, is nevertheless a missed learning opportunity for children in most Kenyan schools given evidence of lack of substitute teachers generally in sub-Saharan Africa and the developing world (Alcazar et al., 2006). The possibility of lack of substitute teachers is evident in figure 3 where we find that when a large share of teachers were absent from class, unsurprisingly, a large share of classrooms were only occupied by students.

3.4.2 Teacher Effort Inside the Classroom

Even when teachers show up in class, a certain percentage of teaching time is likely to be lost to non-instruction activities (Bruns and Luque, 2014). We use the classroom observation snapshots to quantify how much time is spent inside the classroom between instruction and non-instruction related activities. This section is based on observation of 276 teachers who participated in the classroom observation.[20] In taking the snapshots, the surveyor coded two key aspects of the classroom dynamics: (i) how the teacher was using the class time between two mutually exclusive activities: *instruction activities*[21] and *non-instruction activities*[22] (*off-task* activities) and (ii) the number of students who were visibly not following the teacher, taken after every 5 minutes. Table 6 shows our approximation of how teachers spent their time in class.

On average, the lessons lasted 35 minutes. When calculating how teachers spent their time in class, we however, consider the first 30 minutes.[23] On average, teachers spent about 86.01 percent of the teaching time on instructional/teaching activities. Application of Stallings instrument in U.S schools over several decades led Stallings and Knight (2003) to observe that high-performing schools achieve an average of 85 percent of class time spent on instruction (Bruns and Luque, 2014). Private school teachers spent slightly more time on instruction related activities although there is no significant difference between public and private schools. In the previous section, we found that public school teachers are about 16 percent points more likely to be absent from class relative to their counterparts in private schools. This means that the challenge in Kenya is getting public school teachers into classrooms. Once in class, there seems to be no discernible difference in performance between public and private teachers.

Following Stallings and Knight (2003), we break down the expenditure of instructional time into *active* and *passive* related activities as shown in table 6. The most common instructional activities are those related to *active instruction* (taking 61.13 percent of total instructional time) involving the teacher directly engaging students, mainly as a whole class, through lecture and explanation.[24] Correspondingly, 38.87 percent of total instructional time is spent on passive instruction methods,

[20] As reported in Chapter 2, of the 306 schools, classroom observation for maths and language took place in 276 schools. That is, 276 teachers (144 language and 132 maths teachers) were observed. In 28 schools teachers were absent from class and in two schools classroom observation took place for creative arts and science subjects, which we exclude from the analysis. Using the t-test, we do not find systematic differences in terms of schools and teacher characteristics between schools where classroom observations took place and where it did not take place (see details in the last Chapter).

[21] If the time was being used for instruction, what instructional activities were happening and how was the teacher-pupil interaction (whole group; small group or one-on-one).

[22] If the time was being used for off-task activities, what activities was the teacher involved in.

[23] There are few instances (in about 4 schools) where the surveyor arrived relatively late when the lesson had started. This does not affect our estimates of how teachers spent their time in class since we are taking percentage use of time based on the length of the lesson that was observed.

[24] This is often punctuated by pupil copying from the board, teacher asking pupils questions, learners solving problems on the board and choral answer responses. This kind of classroom interaction is similar to what has been reported by others in the context of Kenya (Ackers and Hardman, 2001; Hardman et al., 2009; Pontefract and Hardman, 2005; MoEHRD, 1999; Ngware et al., 2014; Trudell and Piper, 2013).

Table 6: How Time is Spent in the Classroom

			by School Type			by Subject		
	Task[a]	All subjects	Public	Private	Diff	Maths	Language	Diff
Teachers	**1. Instruction Activities (as % of total lesson time)**[b]	**86.01**	**85.64**	**87.28**	-1.65	**84.35**	**87.51**	-3.15**
	of which:							
	Active Instruction[c]	61.13	61.92	58.43	3.4	58.80	63.22	-4.41
	Passive Instruction [d]	38.87	38.08	41.57	-3.49	41.20	36.78	4.41
	2. Off-task Activities (as % of total lesson time)	**13.99**	**14.36**	**12.72**	1.65	**15.65**	**12.49**	3.15**
	of which:							
	Classroom Management (percent) [e]	65.75	63.15	74.79	-9.00	62.00	69.43	-9.12
	Teacher reporting late and/or leaving earlier (percent)[f]	34.25	36.85	25.21	9.00	38.00	30.57	9.12
Students	Students off-task (as a % of the class size) [g]	**12**	**14**	**7**	7**	**13**	**13**	

Source: Own calculations based on SDI 2012. Notes: (1) Calculations based on Stallings and Knight (2003) framework; and (2) ***significance<1 percent, **significance<5 percent, *significance<10 percent

[a] Here we are interested how the teacher spent their time in class based on the classroom observation data. As such we only consider 276 teachers who were in class and were observed. Unfortunately, we do not have any other reference source on the use of instruction time in the classroom in Kenya from which to compare our results.

[b] We account for the first 30 minutes of the lesson.

[c] Activities under *active instruction* time include among others: teacher reads or lectures to the pupils; teacher supervises pupil(s) writing on the board; teacher leads kinesthetic group learning activity (students carrying out physical activities, rather than listening to a lecture or watching demonstrations); teacher writing on blackboard and teacher listening to pupils recite/read.

[d] *Passive instruction* activities include among others: teacher waiting for pupils to complete task and teacher testing students in class/giving assignment to students and teacher going round the class to monitor learners.

[e] Classroom management activities mainly include teacher maintaining discipline in class among others.

[f] Recorded qualitative data from the surveyor teams show various reasons for teachers reporting late and/or leaving early: sometimes, teachers were attending to other administrative matters in or outside the school premises. Sometimes, they waited for learners to settle. Sometimes, they were socializing with fellow teachers and students. In cases where teachers taught only some part of the lesson and left earlier, some left learners unattended while others gave them assignments.

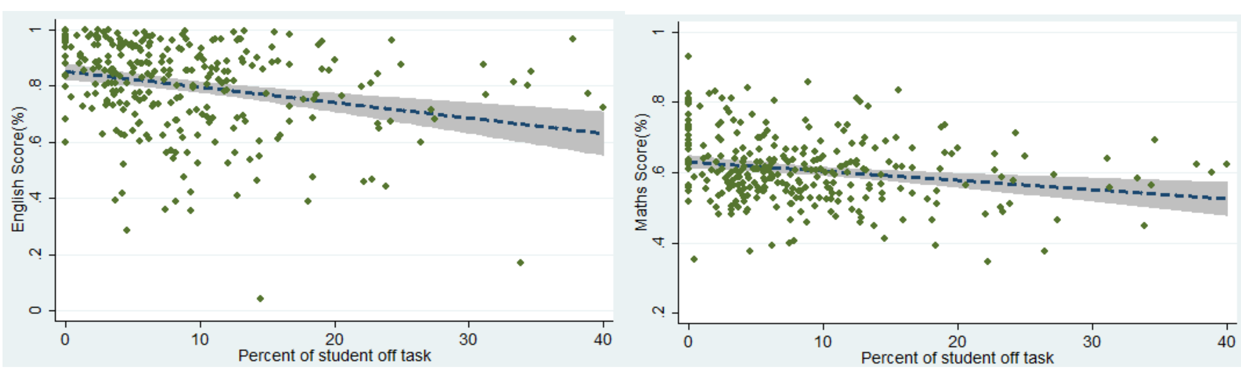
[g] For every 5 minutes, the surveyor scanned through the room in a 360-degree circle and recorded the number of students who were visibly not following the teacher. In total, 6 snapshots were taken during the 35 minutes lesson. To arrive at this indicator, we take the average number of students who were off-task over the 6 snapshots as a share of the total number of students in class.

which according to (Bruns and Luque, 2014) can be characterized by students doing assignments at their seats with teachers moving around the classroom monitoring progress as they wait for students to finish the tasks.

Table 6 further shows that at least 13.99 percent of the lesson time is lost to off-task activities or non-instructional activities.[25] A major part of the time here, 65.75 percent, is tied up in classroom management activities involving teachers doing paper work, arranging seating positions and maintaining discipline among other tasks. The rest of the off-task activities are absorbed by teachers arriving late to class and/or leaving early.

As table 6 shows, about 12 percent of students were not engaged over the 35 minutes lesson. These students were visibly not following the teacher with some talking to each other and others looking outside the classroom. The percentage is higher in public schools relative to private schools probably due to large classes in public schools. There are significant variations across schools in terms of the percentage of students not engaged. Outcomes range from all students fully engaged to as much as over 60 percent of students not engaged. Research based on schools in the USA shows that highly effective teachers keep the share of students off-task below 6 percent (Stallings and Knight, 2003). Figure 4 shows that students perform worse, especially in classes where a higher proportion of students are not following the teacher.

Figure 4: Correlations between Student Test Scores and Percent of Students Off-task



Source: Own calculations based on SDI 2012.

Following Martin and Pimhidzai (2013), we adjust for the time teachers are absent from the classroom at the school level, and for the time the teachers teach while in the classrooms to derive a measure of the amount of time teachers spend teaching in a school during a normal day, known as *teacher effective instruction time or time-on-task*. Table 7 shows how this indicator is computed. First, the reported average length of the school day devoted to teaching excluding all the breaks (short breaks and lunch breaks) as provided by the school head during the structured interview with the surveyor is 5 hours and 38 minutes. However, on average four out of 10 teachers were either absent from school or from class at any given time (see table 5). This reduces the school level scheduled teaching time to 3 hours and 23 minutes. Based on the case of grade 4, we have shown that roughly 86 percent of any

[25] Our estimate of off-task activities is quite conservative and in fact on the lower bound. We do not include teachers who were totally absent. Including them will definitely shift the estimated value of off-task activities upwards.

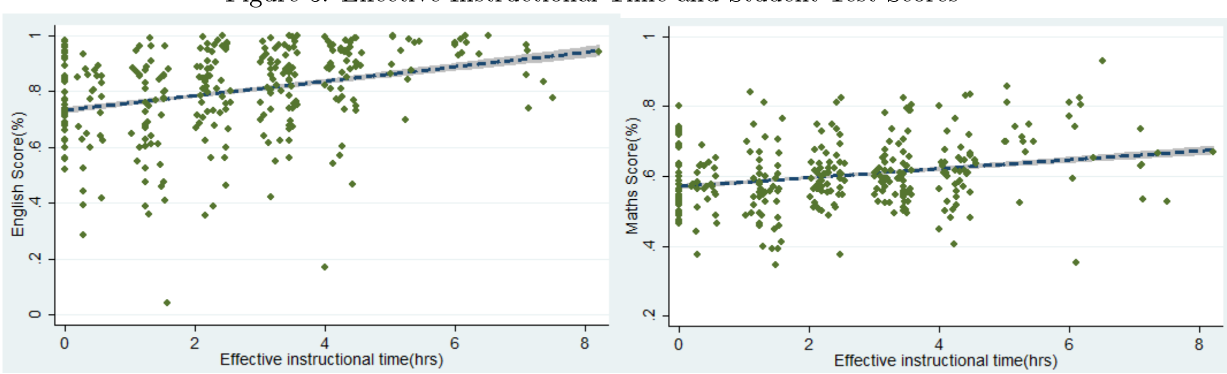
typical lesson is devoted to teaching while the rest is lost to non-instructional related activities. This further reduces teaching time to 2 hours 54 minutes, which we call *teacher effective instruction time or time-on-task*. Figure 5 shows that students perform better in tests in schools with higher effective instruction time.

Table 7: Effective Instruction Time or Time-on-Task

Item		Final Time
Reported school level teaching time excluding breaks per class (hrs)		5hrs 38min
of which		
Lost due to absence from class (hrs)	0.4*(5 hours 38 minutes)	2hrs 15min
Remaining school level teaching time (hrs)	0.6*(5 hours 38 minutes)	3hrs 23min
of which		
Lost to off-task activities (hrs)	0.14*(3hrs 23min)	28min
Effective instruction time (hrs)	0.86*(3hrs 23min)	2hrs 54min

Source: Own calculations based on SDI 2012.

Figure 5: Effective Instructional Time and Student Test Scores



Source: Own calculations based on SDI 2012.

In addition to the minute to minute classroom recording, the surveyor answered *a set of questions* designed to capture various aspects of classroom teaching practices based on their own observations, except for questions that required teacher answers. These included teacher demeanor, interaction and feedback to students, use of teaching aid, introducing and summarizing the lesson, assigning and review of homework as well as instructional language usage. The questions required a *yes* or *no* answer and they were filled out just after the classroom observation. Danielson (1996) has developed a framework that identifies aspects of an effective teacher, which she calls *domains*, that are known to promote improved student learning.[26] We end this section by classifying the classroom teaching practices based on Danielson (1996) Framework for Teaching, as shown in table 8.

[26] A comprehensive review of this framework can be found at: http://tpep-wa.org/the-model/framework-and-rubrics/instructional-frameworks/danielson-framework/

Table 8: Teacher Classroom Teaching Practices

	All Schools	Public Schools	Private Schools	Public-Private
	Mean	Mean	Mean	Mean Diff.
Domain 1: Planning and Preparation[a]				
Teacher had a well planned lesson plan	0.86	0.85	0.87	-0.02
Teacher had a well planned scheme of work	0.82	0.84	0.75	0.09
Domain 2: Creating an environment for learning[b]				
Teacher called children by name	0.87	0.87	0.87	0.00
Teacher gave feedback of praise and moral encouragement	0.85	0.85	0.84	0.01
Teacher gave student feedback that was scolding at a mistake	0.19	0.19	0.19	0.00
Teacher hit or slapped children	0.03	0.03	0.02	-0.02
Domain 3: Teaching for learning[c]				
Teacher challenged students (through questions)	0.43	0.42	0.48	-0.06
Teacher used local language and local information	0.33	0.36	0.20	0.16***
Teacher reviewed homework	0.38	0.35	0.47	-0.13
Teacher assigned homework	0.69	0.65	0.83	-0.18***
Number of observations	**276**	**213**	**63**	

Source: Own calculations based on SDI 2012. Classification of the teaching practices is based on Danielson (1996) Framework for Teaching, ***significance<1 percent, **significance<5 percent, *significance<10 percent

[a] The components in Domain 1 outline how a teacher organizes the content of what students are expected to learn, in other words, how the teacher designs instruction.

[b] The components in domain 2 consist of interactions that occur in the classroom that are generally non-instructional. These consist of creating an environment of respect and rapport among the students and the teacher, establishing a culture for learning, managing classroom procedures, managing student behavior, instilling some level of trust among students.

[c] The components in Domain 3 constitute the core of teaching – the engagement of students in the learning contest. These include communicating clearly and accurately, using questioning and discussion techniques, engaging students in learning, providing feedback to students, and demonstrating flexibility and responsiveness.

4 Methodology

4.1 Theoretical Framework

Our work is based on the production function framework, specified by among others, Glewwe (2002), Glewwe and Kremer (2006) and Orazem and King (2007). In this framework, learning outputs are as a result of a combination of various inputs. Here, a school is viewed as a *production firm* where inputs interact to produce outputs (Glewwe and Kremer, 2006; Todd and Wolpin, 2003; Glewwe, 2002). The output of this education production process can be measured by, for instance, student test scores.

The framework assumes that each household (in particular, the parents of the child) maximizes a utility function subject to constraints such as budget and credit constraints. The main arguments of this utility function are: (i) the consumption of goods and services including leisure and (ii) each child's learning (Glewwe and Kremer, 2006; Glewwe, 2002). Formally, utility is expressed as:

$$U_i = u(A_i, G_i) \tag{1}$$

where A_i is child's academic achievement and G_i is the household consumption possible after sending a child to school including leisure. In its general form, child i's academic achievement, A_i, is hypothesized to depend on: the *child's characteristics* (e.g. age, gender, child's motivation and innate ability); *family background characteristics and home inputs into education* (e.g parent's education, preferences for their children education, books at home, time spent helping with homework); *school characteristics* (such as school infrastructure and instructional inputs including teacher input); and *community background characteristics*. This can be formally expressed as:

$$A_i = f\,(X_i,\, F_i,\, Q_i,\, V_i) \tag{2}$$

where: X_i, F_i, Q_i and V_i represent child, family, school and community background characteristics, respectively. Theoretically, teachers are part of the school based inputs as specified in equation (2). Because of market failure, a child's academic achievement, A_i cannot be purchased, it has to be produced. Since equation equation (2) is a production function, its arguments must be proper inputs. By proper inputs we mean those inputs that *define* the frontier of production function but not *shift* it. In the case of equation (2), proper inputs into the child's education production function are contained in Q_i and as we clarify in equation (4) (in the next sub-section), they include inputs as teacher's subject knowledge and pedagogical skills. Other inputs, such as those captured by X_i, F_i and V_i in are not proper inputs, rather shift parameters of the production function.

4.2 Estimation Strategy

Following Glewwe (2002); Todd and Wolpin (2003); Glewwe and Kremer (2006); Orazem and King (2007) and many other empirical studies, we estimate the linear functional form of equation (2), formally expressed as:

$$A_{ijkd} = \alpha + \delta X_{ijkd} + \xi_{ijkd} + \gamma T_{jkd} + D_d + \mu_{ijkd.} \quad (3)$$

where: A_{ijkd} is the test score (in language or maths) of student i taught by teacher j in school k located in region (division) d; X_{ijkd} is a vector representing observable student, teacher, school and village characteristics; ξ_{ijkd} is a measure of unobserved student ability captured by the non-verbal reasoning scores;[27] T_{jkd} is our variable of interest and is a vector representing teacher j's *human capital* and *effort*; D_d is the division fixed effects[28] and μ_{ijkd} is the idiosyncratic unobservable factors affecting achievements.[29]

Naturally, the full elements of a teacher's true human capital and effort (T_{jkd}) are not easy to collect and some are not easy to observe. In our case, teacher human capital is measured by teacher's *subject knowledge* and *pedagogical skill* which were collected through the SDI survey teacher assessments. Similarly, we measure teacher effort by *effective instructional time, percent of student off-task during the lesson* and *a vector of classroom practices* obtained through the SDI survey classroom observations. We therefore replace the vector T_{jkd} with information about teacher j's *subject knowledge* denoted as TK_{jkd}, *pedagogical skill* denoted as TP_{jkd}, *effective instructional time* denoted as TI_{jkd}, *percent of student off-task* denoted as TT_{jkd} and *classroom practices* denoted by the vector CP_{jkd}. We add the term η_{jkd} to capture the error in using TK_{jkd}, TP_{jkd}, TI_{jkd}, TT_{jkd} and CP_{jkd} as measures of T_{jkd}. As a result, we estimate the empirical model shown in Equation (4):

$$A_{ijkd} = \alpha + \delta X_{ijkd} + \xi_{ijkd} + \gamma_1 TK_{jkd} + \gamma_2 TP_{jkd} + \gamma_3 TI_{jkd} + \gamma_4 TT_{jkd} + \gamma_5 CP_{jkd} + D_d + \eta_{jkd} + \mu_{ijkd.} \quad (4)$$

Equation (4) remedies the lack of clarity in equation (2) with regards to proper and improper inputs into the child's education production function. The variables TK_{jkd}, TP_{jkd}, TI_{jkd} and TT_{jkd} are proper inputs from free from distraction**.** They are the inputs into a child's human capital production and in our case they are our treatment variables, i.e., the factors of policy interest. As mentioned, these inputs define the frontier of the function but do not shift it. Observe that CP_{jkd}, measuring teacher classroom practices, is an innovations that shifts the production function although is of policy interest.

We estimate equation (4) using Ordinary Least Squares (OLS) and correct for the correlation of

[27] Recall that part of the SDI survey student tests involved a non-verbal reasoning test that consisted of four items related to pattern recognition based on Raven matrices test. Aslam and Kingdon (2011) and Glewwe and Jacoby (1994) are among studies that measured unobserved student ability using non-verbal reasoning scores.

[28] A division is a forth tier administrative unit in Kenya after a location. A group of villages constitute a sub- location, which in turn constitute a location and then a division. Ideally, we should control for unobservables at the lower tiers than the division such as village, sub-location and location level. However, in the SDI survey, there is only one school per village (as well as per sub-location and location) and as a result, we can not also estimate a village (and/or sub-location or location) fixed effects model. We provide a more detailed explanation of this issue in details shortly.

[29] Among others, μ_{ijkd} accounts for variables in equation (2) that are not in the data. For instance, the general form of the production function as specified in equation (2) contains family characteristics. However, the SDI survey did not collect information related to home environments apart from a question asking whether the child had breakfast at home or not. Such omitted variables are absorbed in μ_{ijkd}.

students within class/school level by clustering standard errors at the class/school level (Wooldridge, 2003).[30] We also assume a contemporaneous production function when estimating equation (4). In this context, we assume that it is current inputs that mainly determine the observed learning achievements (see Todd and Wolpin (2003) for details about contemporaneous education production function).

As observed by previous studies (Glewwe (2002); Todd and Wolpin (2003); Glewwe and Kremer (2006); Orazem and King (2007); Case and Deaton (1999)), empirical estimation of equation equation (4) is extremely challenging due to a number of challenges: omitted variable bias, sample selection bias, and measurement error. Shortly, we discuss how we deal with these challenges in the light of the limitations of our data. Before that, in the section that follows, we present a detailed description of the measures of *teacher human capital* and *teacher effort*, our independent variables of interest.

5 Education Production Function Estimates

Next, we link measures of teacher human capital and teacher effort to student literacy and numeracy test scores by way of an education production function as outlined in equation (4). The estimates we present are based on the sample of 222 schools where in each school, one grade 4 teacher was *(concurrently) observed in class* and *assessed in the teacher tests.* In 109 schools covering a total of 1,034 students, a language lesson was observed and in the rest, 113 schools covering a total of 1,077 students, a maths lesson was observed. [31]

For our teacher effects γ in equation (4) to be causal, the measure of teacher human capital and teacher effort need to be uncorrelated with unobserved student (μ_{ijkd}) and teacher (η_{jkd}) traits. The challenge of using cross-sectional data is how to deal with the non-random sorting of students and teachers into schools. True random-assignment of teachers and students in schools and in classes within schools is rare in any education context and more so, in the developing country context (Lavy, 2010). Due to this, we make no claim that our teacher estimates represent causal effects. Nevertheless, in what follows, we offer some discussion as to why we think our teacher estimates will not be completely driven by *within* and *between-school* sorting and suggest strategies to deal with these challenges.

5.1 Within-School Sorting

One natural concern is the possible non-random allocation of students into classes based on observed and unobserved factors such as student motivation or ability (*within-school* sorting). This results from the influence of teachers, school administrators or parents (Dieterle et al., 2015; West and Wobmann, 2006; Clotfelter et al., 2006; Aaronson et al., 2007; Todd and Wolpin, 2003). This problem rises if, for instance, a school has more than one stream of grade 4 (say grade 4A and 4B) and brighter or highly motivated students are assigned to a certain stream and matched to better or highly motivated teachers, who in turn adopt certain teaching practices based on their student ability.[32] Within-school sorting

[30] Recall that clustering standard errors at the school level in our case is similar to clustering at the class level since for each school, we are only interested in the subject teacher who was *observed* and *assessed in the teacher tests*.

[31] See Chapter 2, section 2.1.1

[32] In this case our measures of teacher knowledge and effort in Equation (4) will be biased because they are correlated with student unobservables (μ_{ijkd}) (Hidalgo Cabrillana and Lopez-Mayan, 2015).

also arises if there are unobserved teacher attributes that have a direct impact on student performance, while they are correlated with the teaching style[33] (Aslam and Kingdon, 2011; Hidalgo Cabrillana and Lopez-Mayan, 2015).

We think that our estimates will not be so much influenced by within-school sorting because of the following reasons. First, our sample is restricted to one stream of grade 4 and all the students in the sampled stream are taught by the same language or maths teacher within the school.[34] Second, in determining student achievements in Equation (4), we include teaching process variables (how teachers spend their time in the classroom and what teachers do in class), collected during the classroom observation, alongside observable teacher characteristics (such as education, experience or training). In most studies, these teaching process variables are generally considered to be part of the teacher unobservables, η_{jkd} (Aslam and Kingdon, 2011). Their inclusion therefore helps to reduce some of the bias associated with the possible correlation of unobserved teaching traits with observed teacher and student characteristics (Aslam and Kingdon, 2011).

The next reasons are largely institutional and contextual. First, there is evidence to show that parents in Kenya, especially those in rural areas, have little influence on choices made in schools, including those related to teacher allocation (Spernes, 2011; Echaune et al., 2015; Makori et al., 2015). Bold et al. (2011) observe that the free public primary education in Kenya created the notion that schools belong to the government and not the local communities, thereby eroding parental involvement in what goes on inside schools.[35] Second, in our sample, 85 percent of the schools can be classified as *public* and *private-not-for-profit* while the rest, 15 percent, are *private-for-profit* and mostly located in urban areas.[36] In Kenya, within-school sorting is likely to be more prevalent in *private-for-profit* schools than in *public* or *private-not-for profit schools* for a number of reasons. For public schools, the Basic Education Act No. 14 of 2013[37] calls for equal treatment of students in all public schools and forbids public schools from admitting students based on previous academic records and tracking children in class by ability. Actually, one of the well documented reasons for the declining quality of education in public schools especially following the free public primary education is the unrestricted and unlimited entry into public schools of low aptitude students (Onsomu et al., 2005; Piper and Mugenda, 2012; Wasanga et al., 2010).

For *private-not-for profit schools*, their operational environment makes it difficult for them to track students by ability or even require entry exams, an issue we treat in detail in the next chapter. Generally, most private-not-for profit schools in Kenya are concentrated in informal settlements whose

[33] This would happen if unobserved teacher ability or motivation affect the choice of the teaching style, while they have a direct effect on student test scores, aside from the effect through the teaching style. In this case the measures of teacher knowledge and effort in Equation (4) will be biased because they are correlated with teacher unobservables (η_{jkd}).

[34] As observed by Aslam and Kingdon (2011), within-school sorting would have been a serious problem if data was collected from more than one section (stream) of grade 4 within a school. In each school, we are only looking at students from one stream of grade 4.

[35] We come back on this point in Chapter 6.

[36] *Private-for-profit entities* are educational institutions operated by private agents and are primarily profit-seeking businesses mainly located in middle-and high-income urban areas as well as in peri-urban areas (Piper and Mugenda, 2010). *Private-not-for-profit entities* are educational institutions mainly operated by non-governmental organizations, majority of whom are faith-based organizations and are primarily not for profit-seeking businesses, targeting children from poor households in urban informal settlements (Heyneman and Stern, 2014; Tooley and Longfield, 2015; Tooley et al., 2011; Dixon and Tooley, 2012; Larbi et al., 2004; Oketch and Somerset, 2010; Oketch et al., 2012; Dixon, 2012).

[37] More details about the Act can be found at: http://www.kenyalaw.org.

populations are largely poor and comprise labor migrants leading to a high teacher and student turnover (Heyneman and Stern, 2014; Tooley and Longfield, 2015; Tooley et al., 2011; Dixon and Tooley, 2012; Larbi et al., 2004; Oketch and Somerset, 2010; Oketch et al., 2012; Dixon, 2012). Evidence shows that admission to these schools is granted at the discretion of the head teacher and not mainly based on previous academic record and once in school, children are less likely to be tracked by ability(Edwards Jr. et al., 2015).

Lastly, the fact that we are focusing on grade 4 also alleviates much of the concern regarding *within-school* sorting. We noted from our conversation with staff at the Ministry of Education, Science and Technology that the incentive to track students in class by ability is likely to happen as children advance to higher grades such as grade 7 or 8 when schools are preparing learners for the end of primary cycle examinations, often viewed as a strategy to improve the schools performance. Also, in a typical developing country like Kenya, the sample of children in schools is likely to become more and more self-selective with the rise in grades due to high drop-out rates. Focusing on grade 4 learners minimizes such selection biases that arise due to high dropouts in later grades.

5.2 Between-School Sorting

The second concern is *between-school* sorting, which is the endogenous selection of teachers and students across schools, mainly based on factors that are unobservable to us (Todd and Wolpin, 2003; Aaronson et al., 2007). An example of between-sorting in Kenya is parental or teacher preference of certain regions given their socioeconomic conditions such as better schools. Related to this, some parents may prefer certain schools, for instance, private schools, based on characteristics such as past student performance or teaching philosophy. Table 4 and table 5 show that private school teachers are more knowledgeable and invest more effort than public school counterparts. Since private schools are not free in Kenya, fee paying parents in private schools may take greater interest in the school and be involved in monitoring the quality of education in schools (Bold et al., 2011).

There exists a number of estimation strategies in a cross-sectional data setup to deal with such selection biases. Some of these include school fixed effects (Aslam and Kingdon, 2011; Altinok and Kingdon, 2012), within-teacher within-student variation (Metzler and Woessmann, 2012) and within-pupil across-subject approach (Shepherd, 2013; Shepherd et al., 2015). The application of any of these methods depends on the nature of the data at the researcher's disposal. For our case, we cannot estimate a school fixed effects model since there is *no within school variation* in the variables capturing *teacher human capital* and *effort* as only *one teacher was concurrently observed* and *assessed in the teacher tests* in every school.[38] To estimate a within-teacher within-student variation or a within-pupil across-subject approach in our context, data would have to be gathered from schools where: (i) *students are taught by the same teacher in both subjects (math and language)* and (ii) *that this teacher was*

[38] Recall that we are directly linking student test scores to the human capital and effort of their teachers who were observed and observed. However, in each school, only one teacher was concurrently observed and tested. In other words, there is only one single observation for indicators of teacher human capital and effort per school. Such lack of *within school variations* hampers estimation of a school fixed effects model. A school fixed effects model in our context would have been possible if: (i) two or more teachers from different streams (e.g 4A and 4B) and/or from different grades (e.g grade 3 and grade 4) were observed and assessed in the teacher tests. Recall that in our case, only one teacher was concurrently observed and tested in every school. So our measures of teacher effort do not vary within a school to allow estimation of a school fixed effects model.

concurrently assessed in the teacher tests and *observed in both subjects* (see Shepherd (2013); Shepherd et al. (2015); Metzler and Woessmann (2012)).[39] Estimating a school fixed effects model and/or within-teacher within-student variation is therefore not possible with the data we have.

To minimize between-school sorting, we include division fixed effects, D_d, within the OLS estimation as shown in Equation (4). Through the division-level fixed effects, we are able to remove all sources of observed and unobserved heterogeneity at the division level.[40] Including division fixed effects does not, however, effectively deal with the endogenous selection of students in public and private schools as there are private and public schools within a division. Since we cannot run a school fixed effects, we undertake separate estimation for public schools and see how our estimates compare with those in the main regression.[41]

While teachers and students are likely to sort across schools mainly on the basis of unobservable characteristics (Todd and Wolpin, 2003; Aaronson et al., 2007), we are of the opinion that such teacher and student across school sorting is in fact more likely to be on the basis of observed factors such as education, training, experience and gender among others. With this in mind, we follow a method employed by Aaronson et al. (2007) to provide an idea of the extent of *between-school* teacher and student sorting in our data on the basis of observable characteristics. We use the 1,679 teachers who were assessed in teacher tests and all the 2,954 students in our sample.

In this method, we begin by computing the within-school standard deviation for selected observable characteristics of teachers based on the actual data as it is. Low values of within-school standard deviation means that there is a lot of sorting, hence similar teachers are in the same schools. Second, using a simulation[42], we compute the average within-school standard deviation assuming that teachers were randomly sorted into schools. Basically, we sort teachers using a random variable which is essentially a white noise and based on this random variable compute the average within-school standard deviation (see Aaronson et al. (2007) for more details). Thirdly, we compute the average within-school standard deviation assuming that teachers were perfectly sorted into schools. Here, instead of sorting teachers using a random variable (white noise), teachers are sorted based on actual data (the respective variable) (see Aaronson et al. (2007) for more details). We perform a similar operation for selected observable characteristics of children. We do this for all schools and then separately for public and private schools.

For each variable, we compare the average within-school standard deviation for actual distribution,

[39] The idea behind the within-student variation approach and/or a within-pupil across-subject approach is that by testing and observing the same teacher taking the students in both subjects, we are able to estimate whether the *same student* taught by the *same teacher* in *two different academic subjects* (e.g. maths and language) performs better in one of the subjects (e.g. maths) if the teacher's knowledge and efforts are relatively better in this subject (e.g. maths) thus allowing us to infer causality. The within-teacher within-student variation and within-pupil across-subject approaches can identify the effect based on within-teacher within-student variation by controlling for student fixed effects, teacher fixed effects, and subject fixed effects (see Shepherd (2013); Shepherd et al. (2015); Metzler and Woessmann (2012) for a detailed exposition about these approaches).

[40] Ideally, we should control for unobservables at the lower tiers than the division such as village, sub-location and location level. However, in the SDI survey, there is only one school per village (as well as per sub-location and location) and as a result, there are no variations in the variables measuring teacher human capital and effort at the village, sub-location and location level.

[41] The sample comprising only private schools (both for-profit and not-for-profit schools) is quite small. As a result, we do not undertake a separate estimation for private schools.

[42] We are greatly indebted to Deon Filmer and Stacy Brian of the World Bank for the assistance and inputs regarding the STATA routine for running the simulations based on methodology proposed by Aaronson et al. (2007).

random distribution and perfect/non-random distribution. The results reported in table 9 to table 11 show that the actual distribution is closer to the random than non-random distribution. There is some level of evidence to show that teachers in our data, in all schools combined and in public and private schools, are not perfectly sorted across schools on the basis of teacher knowledge, experience and age.

Table 9: Sorting of Teachers and Students Across Schools (Public and Private)

Teacher sorting				Students sorting			
	Actual	Random	Non-random		Actual	Random	Non-random
Math Score	0.738	0.899	0.008	NVR* score	0.898	0.970	0.006
Language Score	0.850	0.909	0.010	Maths Score	0.797	0.964	0.009
Pedagogy Score	0.802	0.947	0.006	Language Score	0.681	0.909	0.005
Experience	0.789	0.975	0.006	Age	0.840	0.967	0.008
Teacher Age	0.777	0.978	0.007	Breakfast status	0.221	0.289	0.001
No. of Teachers		**1,679**		No. of Students		**2,954**	
No. of Schools		**306**		No. of Schools		**306**	

Source: Own calculations based on SDI 2012. * Non-Verbal reasoning

Table 10: Sorting of Teachers and Students Across Schools (Public)

Teacher sorting				Students sorting			
	Actual	Random	Non-random		Actual	Random	Non-random
Math Score	0.746	0.903	0.009	NVR* score	0.924	0.968	0.008
Language Score	0.866	0.942	0.012	Maths Score	0.850	0.971	0.012
Pedagogy Score	0.824	0.962	0.008	Language Score	0.740	0.933	0.006
Experience	0.888	0.976	0.008	Age	0.879	0.967	0.013
Teacher Age	0.876	0.963	0.008	Breakfast status	0.684	0.871	0.006
No. of Teachers		**1378**		No. of Students		**2378**	
No. of Schools		**239**		No. of Schools		**239**	

Source: Own calculations based on SDI 2012. * Non-Verbal reasoning

Table 11: Sorting of Teachers and Students Across Schools (Private)

Teacher sorting				Students sorting			
	Actual	Random	Non-random		Actual	Random	Non-random
Math Score	0.722	0.867	0.048	NVR* score	0.879	0.956	0.032
Language Score	0.787	0.907	0.050	Maths Score	0.844	0.979	0.028
Pedagogy Score	0.716	0.963	0.320	Language Score	0.571	0.722	0.033
Experience	0.658	0.793	0.045	Age	0.798	0.971	0.040
Teacher Age	0.727	0.777	0.042	Breakfast status	0.565	0.639	0.038
No. of Teachers		**301**		No. of Students		**575**	
No. of Schools		**67**		No. of Schools		**67**	

Source: Own calculations based on SDI 2012. Notes. (1) This include for-profit and not-for-profit private schools, (2) * Non-Verbal reasoning

Similarly, students are also not perfectly sorted on the basis of subject knowledge, age and socioeconomic status (measured by an indicator for whether a student had breakfast before attending school).

Of particular interest is the fact that students are not sorted across schools based on the unobserved student ability (ξ_i) in equation (4) which is captured by non verbal reasoning scores. As noted by

Filmer et al. (2015), if one is to take the non-verbal reasoning scores of students as a measure of innate intelligence quotient (IQ) that is not immutable directly by schooling, then there is no evidence that more innately gifted students are concentrated in certain schools. To conclude, we do not in any way claim that our estimates of teacher human capital and efforts that we present next are causal. These estimates should be interpreted in the context of the limitations we have raised. There is however good reason to think that selection bias is fairly small. That said, our results shed some light on important teacher factors that influence student achievements and provides an opportunity to take this work further especially after considering the SDI data limitations we have raised.

5.3 Model Estimation Results

5.3.1 Effect of Teacher Human Capital and Effort

In table 12 and table 13, we present the estimated effects of our measures of the human capital and effort of teachers on student achievements for language and maths respectively. The dependent variable is the student score, which is standardized to the mean of zero and standard deviation of one. We run separate regressions for maths and language. For language, we have 109 schools covering a total of 1,034 students. These are schools where a teacher was observed (teaching a language class) and assessed (in teacher tests) thereafter. For maths, we have 113 schools covering a total of 1,077 students. Similarly, these are schools where a teacher was observed (teaching a maths class) and assessed. We control for a comprehensive set of teacher, school, classroom and pupil related variables. This is one credible way to deal with endogeneity when exclusion restrictions are impossible to find (Stock, 2010).

We also control for the village socioeconomic conditions measured by *village wealth index* based on the 2009 Kenya Population and Housing Census. The SDI survey does not contain a village level module. However, with the support of staff at the National Bureau of Statistics in Kenya, we mapped the schools to the 2009 Kenya Population and Housing Census enabling us to extract village level information corresponding to the respective schools.[43] The census data includes questions related to the household ownership of durable and livestock assets, type of material used to construct the wall of the dwelling unit, type of lighting regularly used by the household and household sanitation status among others. We construct a *village wealth index,* using the principal components analysis (PCA), based on these individual and household characteristics *aggregated at the village level.*[44] The index has 21 continuous variables based on indicators shown in table A.3 in the appendix.

In model 1, we enter our measures of teacher human capital, effective instruction time and teacher's

[43]The extensive matching process was made possible by the use of the school GPS coordinates. Staff at the the National Bureau of Statistics in Kenya helped us to link every school to its respective enumeration area/village. The SDI and the Kenya Population and Housing Census data are only two years apart. We are mainly controlling for village level conditions for which schools are located. To large extent, this also captures socioeconomic condition of villages (communities) where children come from especially in rural areas where children are likely to attend schools within their villages. This might not necessarily be the case for children in urban areas who are more likely to attend schools outside their villages. This survey does not allow us to know whether the child attends a school in the village or not.

[44]The practice of aggregating individual and household socioeconomic characteristics to construct measures of welfare (wealth/poverty) at the higher levels (e.g village, district etc) is quite common in the literature. For instance, in Costa Rica, Cavatassi et al. (2004) uses individual and household socioeconomic characteristics from the census to construct district level poverty index using principal components analysis (PCA).

ability to keep students engaged. In model 2, we add our indicators of teacher classroom practices. These indicators are based on the SDI classroom observations and closely match measures of classroom practice explored in other studies (such as Ackers and Hardman (2001); Hardman et al. (2009); Ngware et al. (2014); MoEHRD (1999); Piper and Mugenda (2010); Aslam and Kingdon (2011)). Each practice is a binary indicator of whether or not the teacher engaged in it. In model 3, we control for other teacher observable characteristics. In model 4, school and classroom related variables are entered while model 5 includes student attributes. In table 12 and table 13, we only show the coefficients for the measures of the human capital and effort of teachers as well as teachers' observable characteristics.

Although they are not causal, what do these results tell us? First, even though we have not controlled for family background information, all models exhibit a relatively strong goodness-of-fit, as shown by the value of R-squared statistics. Moreover, much of the variations in the student test scores is explained by our measures of teacher human capital and effort. Jones et al. (2014) observe that R-squared statistics from estimated education production functions in most developed countries are often in the bracket of 10 percent and 30 percent while those from developing countries range around 30 percent. Second, addition of extra controls, including those related to teacher observables, does not dramatically change the direction of the effect and magnitude of the measures of teacher human capital and effort.

Table 12: Teacher Human Capital, Teacher Effort and Student Language Test Scores

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Teacher subject knowledge	0.053*	0.079**	0.103***	0.080**	0.075**
	(0.027)	(0.032)	(0.029)	(0.031)	(0.031)
Teacher pedagogical knowledge	0.001	-0.038	-0.033	-0.011	-0.007
	(0.043)	(0.042)	(0.043)	(0.048)	(0.053)
Effective instruction time (in hours)	0.051*	0.045*	0.071***	0.056**	0.051**
	(0.026)	(0.023)	(0.022)	(0.024)	(0.023)
Percent of student off-task (average)	-0.005	-0.018***	-0.018***	-0.029***	-0.030***
	(0.007)	(0.006)	(0.006)	(0.008)	(0.008)
Teacher Classroom Practices					
Teacher reviews and assigns homework		0.181*	0.353***	0.451***	0.383***
		(0.097)	(0.094)	(0.113)	(0.135)
Teacher uses local language to illustrate learning		-0.243***	-0.177**	-0.171*	-0.161**
		(0.089)	(0.084)	(0.094)	(0.081)
Teacher challenges students by asking questions		-0.121	-0.189***	-0.225***	-0.190**
		(0.079)	(0.062)	(0.073)	(0.074)
Teacher keeps a lesson plan and scheme of work		-0.251*	-0.296**	-0.242*	-0.263*
		(0.139)	(0.126)	(0.132)	(0.112)
Teacher instills discipline in students		-0.392***	-0.329***	-0.365***	-0.298***
		(0.113)	(0.108)	(0.100)	(0.101)
Teacher Controls					
Teacher is female			0.201**	0.136	0.078
			(0.084)	(0.099)	(0.097)
Teacher experience (in years)			0.035**	0.035**	0.035**
			(0.017)	(0.016)	(0.015)
Teacher experience squared (in years)			-0.001**	-0.001***	-0.001***
			(0.000)	(0.000)	(0.000)
Teacher is on contract (Ref: government)			-0.457***	-0.220	-0.250
			(0.137)	(0.162)	(0.172)
Teacher's highest education level is a diploma or a degree (Ref: Secondary)			0.054	0.055	0.105
			(0.094)	(0.093)	(0.088)
Teacher has ECD or primary certificate in teaching (Ref: Diploma or degree)			-0.075	-0.183	-0.175
			(0.087)	(0.121)	(0.137)
Controls					
School and Classroom Controls	N	N	N	Y	Y
Student Controls	N	N	N	N	Y
Village Controls	Y	Y	Y	Y	Y
Division fixed effects	Y	Y	Y	Y	Y
Observations	1,077	1,077	1,077	1,077	1,077
R-squared	0.376	0.396	0.407	0.413	0.495

Notes: (1) The estimates are based on 113 schools; (2) The measures of teacher human capital and teacher effort are based on the single teacher who was observed (teaching a language class) and assessed in the teacher tests; (3) School controls include: *a set of dummies which indicate whether the school is public, rural and located near a tarmack.* Other school characteristics are *classroom size, number of pupils per teacher, an index of school infrastructure* (based on the following items, given equal weight: (a) presence of toilets that were judged as designated for boys and girls, accessible, private and clean, (b) availability of electricity and (c) sufficient light for reading from the back of the class) and *an index of classroom equipment* (based on the following items, given equal weight: (a) proportion of students with pens and exercise books, (b) number of students per text book (in language), (c) whether the classroom had the following: piece of chalk, a black board, a corner library, children's work displayed on the walls of the classroom); (4) Student controls include: *student age, age squared, whether the student is female, student score in maths, student non-verbal reasoning ability and whether the student ate breakfast*; (5) Standard errors are in parenthesis and are clustered at the class (school) level and (6) ***1 percent significance level, **5 percent significance level and *10 percent significance level.

Table 13: Teacher Human Capital, Teacher Effort and Student Maths Test Scores

	Model 1	Model 2	Model 3	Model 4	Model 5
Teacher subject knowledge	0.173***	0.172***	0.175***	0.166***	0.126***
	(0.055)	(0.051)	(0.057)	(0.046)	(0.045)
Teacher pedagogical knowledge	0.039	0.005	-0.023	0.111***	0.112***
	(0.053)	(0.051)	(0.045)	(0.038)	(0.035)
Effective instruction time (in hours)	0.150***	0.137***	0.199***	0.087**	0.059*
	(0.039)	(0.041)	(0.038)	(0.034)	(0.030)
Percent of student off-task (average)	-0.018**	-0.020***	-0.010	-0.005	-0.002
	(0.008)	(0.007)	(0.008)	(0.006)	(0.005)
Teacher Classroom Practices					
Teacher reviews and assigns homework		-0.208*	-0.217*	-0.463***	-0.377***
		(0.113)	(0.114)	(0.087)	(0.084)
Teacher uses local language to illustrate learning		-0.262**	-0.162	-0.106	-0.073
		(0.129)	(0.135)	(0.089)	(0.096)
Teacher challenges students by asking questions		0.211	0.219	0.348***	0.273***
		(0.150)	(0.134)	(0.114)	(0.102)
Teacher keeps a lesson plan and scheme of work		0.012	-0.038	-0.243*	-0.248*
		(0.152)	(0.163)	(0.120)	(0.115)
Teacher instills discipline in students		0.037	0.111	0.338**	0.217
		(0.178)	(0.174)	(0.156)	(0.156)
Teacher Controls					
Teacher is female			-0.267**	-0.091	-0.165*
			(0.134)	(0.091)	(0.085)
Teacher experience (in years)			-0.065***	-0.078***	-0.082***
			(0.018)	(0.016)	(0.016)
Teacher experience squared (in years)			0.002***	0.002***	0.002***
			(0.001)	(0.000)	(0.000)
Teacher is on contract (Ref: government)			0.019	0.243*	0.247*
			(0.185)	(0.133)	(0.131)
Teacher's highest education level is a diploma or a degree (Ref: Secondary)			0.284*	-0.041	-0.008
			(0.155)	(0.132)	(0.129)
Teacher has ECD or primary certificate in teaching (Ref: Diploma or degree)			-0.096	0.080	0.090
			(0.160)	(0.118)	(0.103)
Controls					
School and Classroom Controls	N	N	N	Y	Y
Student Controls	N	N	N	N	Y
Village Controls	Y	Y	Y	Y	Y
Division fixed effects	Y	Y	Y	Y	Y
Observations	1,034	1,034	1,034	1,034	1,034
R-squared	0.302	0.313	0.332	0.370	0.430

Notes: (1) The estimates are based on 109 schools; (2) The measures of teacher human capital and teacher effort are based on the single teacher who was observed (teaching a maths class) and assessed in the teacher tests; (3) School controls include: *a set of dummies which indicate whether the school is public, rural and located near a tarmack.* Other school characteristics are *classroom size, number of pupils per teacher, an index of school infrastructure* (based on the following items, given equal weight: (a) presence of toilets that were judged as designated for boys and girls, accessible, private and clean, (b) availability of electricity and (c) sufficient light for reading from the back of the class) and *an index of classroom equipment* (based on the following items, given equal weight: (a) proportion of students with pens and exercise books, (b) number of students per text book (in maths), (c) whether the classroom had the following: piece of chalk, a black board, a corner library, children's work displayed on the walls of the classroom); (4) Student controls include: *student age, age squared, whether the student is female, student score in maths, student non-verbal reasoning ability and whether the student ate breakfast;* (5) Standard errors are in parenthesis and are clustered at the class (school) level and (6) ***1 percent significance level, **5 percent significance level and *10 percent significance level.

We begin by interpreting results for language in table 12. Consistent with findings of Metzler and Woessmann (2012), Shepherd (2013) and others, our results show that teacher's subject specific knowledge has a positive and statistically significant effect on student language scores. For instance, model 5 shows that a one standard deviation increase in teacher's language knowledge increases student test score in language by 0.075 of a standard deviation.[45] The estimated effect of teacher pedagogical skill on student language test scores is effectively zero.

The estimated coefficient of the first measure of teacher effort, that is, *teacher effective instruction time,* is 0.051 and is statistically significant. This result means that *an additional hour of effective instruction* increases the test score in the language exam by 0.051 of a standard deviation. The variable *percent of students off-task* is negatively related to student scores in language meaning that teachers who keep students engaged (on-task) are likely to produce students with higher language test scores. Results show that a one percent increase in the number of students off-task reduces the language test score by 0.030 of a standard deviation.

Turning to teacher classroom practices, we find that the practice of reviewing and assigning homework increases the language test score by 0.383 of a standard deviation. However, the use of local language and local information reduces it by 0.161 of a standard deviation. There is vast literature on how local language influences learning concepts in second (foreign) language. One branch of this literature argues that illustrating concepts to students in local language and testing them in foreign language can be detrimental. Corder (1982) argues that students make errors when applying rules learned in the first language (local language) to the second language (English). He defines errors as deviations from correct usage of the second language since the student does not know its correct rules.[46]

The estimated effects of some classroom teaching practices are however puzzling. For instance, the practice of challenging students intellectually by asking them questions reduces reduces the test score in the language exam by 0.190 of a standard deviation. Unfortunately, our measures of classroom practices are binary indicators and simply indicate whether or not the teacher observed in the classroom engaged in the specific practice or not. In this regard, we do not know the nature of questions that teachers posed to learners and the type of learners that the questions were directed to. However, a number of qualitative studies (Ackers and Hardman, 2001; Hardman et al., 2009; Ngware et al., 2014; MoEHRD, 1999) report that classroom interaction in Kenyan schools is mainly characterized by teacher-initiated questions often calling for choral responses. Studies by MoEHRD (1999) and Hardman et al. (2009) report that open-ended questions and questions initiated by pupils are rare, estimated at less than 1 percent of the questioning exchanges. Furthermore, boys are nearly twice as likely to be asked a question by teachers than girls (Hardman et al., 2009; MoEHRD, 1999). It is therefore likely that such environment of teacher-student classroom question interface is unlikely to facilitate student performance, more so, for girls.

Strangely, we also find that having a lesson plan and scheme of work does not translate to better student scores. This result is contrary to Aslam and Kingdon (2011) who found that being taught by a teacher who plans for the lesson raises scores (language and maths pooled together) by 0.23 standard deviations in Pakistan. Again, this is a binary indicator as to whether the teacher had a lesson plan and scheme of work, simply judged by the surveyors to be well prepared. We do not have a way to further ascertain the quality of these teaching tools. Nevertheless, as shown in table 4, teachers

[45] Unless otherwise, we concentrate on model 5, which we consider as our headline/preferred regression since our variables of interest do not change substantially, in sign and level of coefficient, with the progressive addition of controls.

[46] McLaughlin (2013) further observes that errors due to second language problem should not be confused with mistakes, arguing that the latter can be corrected by the student him/herself while the former can only be corrected by the teacher, a person well conversant with the second language.

performed poorly in the pedagogy sub-task that asked them to prepare a lesson plan.[47] With such low levels of knowledge in lesson planning among teachers, it is unsurprising that merely having a lesson plan and/or scheme of work is unlikely to promote student learning. In fact, a recent study by Piper and Mugenda (2010) based on 220 schools randomly selected from three regions (Nairobi, Thika and Nakuru) of found that regions where teachers were judged to have well prepared lesson plans were associated with lower student achievements in both maths and language.

It is not surprising that the practice of instilling discipline in students by hitting and/or scolding them while giving feedback during a lesson leads to low student performance. We find that teachers who engage in this practice reduce the test score in the language exam by 0.298 of a standard deviation.

Turning on the estimates for the maths regression as shown in table 13, we find a significant and positive effect of teacher's subject knowledge on student maths scores (model 5). Unlike in language regression, there is evidence that teacher's pedagogical skill has a significant effect on student maths test scores. A one standard deviation increase in teacher's pedagogical knowledge increases the test score in the maths exam by 0.112 of a standard deviation. Similarly, teacher's effective instruction time positively and significantly affects student test scores in maths. Teacher's effort, measured by teacher's ability to keep students engaged becomes insignificant once we control for teacher observables.

The results on classroom teaching practices seem to suggest that students learn differently between language and maths. Unlike the case of language, the practice of assigning and reviewing homework is associated with lower student maths test scores. Aslam and Kingdon (2011) finds similar results in Pakistan and explains that homework could be taking away time that could be spent learning new materials. The use of local language has a negative effect which turns insignificant when we control for teacher observables. The practice of challenging students intellectually by asking questions, which had a negative and significant effect on language regression, now has a positive and significant effect on maths test scores. Similarly, classes where a teacher had a lesson plan are associated with lower student maths test scores. Finally, the practice of instilling discipline in students by hitting and/or scolding them while giving feedback during a lesson does not seem to have any effect on student test scores.

5.3.2 Robustness Checks

In this section, we implement a number of robustness tests to check the validity of our results. The first robustness check is methodological. Some authors argue that OLS produces statistically unbiased estimates of the relationships among variables but its standard errors are biased downward because it does not take account of the nesting of students in schools or classes (Gelman and Hill, 2006; Hox, 2010; Fielding, 2010). As a result, hierarchical linear modeling has become a popular way to analyze data with such statistical dependency.[48] We therefore estimate all the models in table 12 and table 13 using hierarchical linear modeling estimation strategy.

[47] A lesson plan is a teacher's detailed description of the course of instruction, or 'learning trajectory' for a lesson. It is therefore a tool through which teachers demonstrate how they can translate their subject knowledge into meaningful teaching (pedagogical skills).

[48] Hierarchical linear modeling does not correct bias in the regression coefficient estimates compared with an OLS model but it produces unbiased estimates of the standard errors associated with the regression coefficients when the data are nested (Gelman and Hill, 2006; Fielding, 2010; Bryk and Raudenbush, 1992; Hox, 2010).

Table 14: Teacher Human Capital, Teacher Effort and Student Language Test Scores

	Language				Maths			
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
	HLM Model Estimation	Pure Public Schools	Public and non-profit private Schools	Rural Schools	HLM Model Estimation	Pure Public Schools	Public and non-profit private Schools	Rural Schools
Teacher subject knowledge	0.075**	0.196***	0.072	0.111**	0.126**	0.127***	0.136***	0.154***
	(0.038)	(0.049)	(0.048)	(0.048)	(0.049)	(0.046)	(0.043)	(0.021)
Teacher pedagogical knowledge	-0.007	0.010	0.121**	0.187***	0.112**	0.088**	0.069*	0.022
	(0.052)	(0.050)	(0.050)	(0.065)	(0.045)	(0.040)	(0.036)	(0.023)
Effective instruction time (in hours)	0.051*	0.043*	0.029	0.048**	0.059	0.092**	0.113***	0.321***
	(0.026)	(0.022)	(0.020)	(0.021)	(0.040)	(0.045)	(0.028)	(0.027)
Percent of student off-task (average)	-0.030***	-0.020***	-0.028***	-0.050***	-0.002	0.014	0.011	-0.033***
	(0.008)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)	(0.006)	(0.003)
Teacher Classroom Practices								
Teacher reviews and assigns homework	0.383***	0.535***	0.639***	0.967***	-0.377***	-0.419***	-0.338***	-0.135***
	(0.124)	(0.126)	(0.115)	(0.158)	(0.103)	(0.105)	(0.093)	(0.033)
Teacher uses local language to illustrate learning	-0.161*	-0.188***	-0.146*	-0.075	-0.073	0.008	0.010	0.213**
	(0.089)	(0.057)	(0.080)	(0.073)	(0.108)	(0.095)	(0.075)	(0.082)
Teacher challenges students by asking questions	-0.190**	-0.305***	-0.138**	-0.255***	0.273**	0.356***	0.411***	0.377***
	(0.088)	(0.066)	(0.065)	(0.077)	(0.122)	(0.110)	(0.104)	(0.040)
Teacher keeps a lesson plan and scheme of work	-0.263*	-0.228*	-0.152	-0.110	-0.248*	-0.156	-0.158	-0.533***
	(0.140)	(0.115)	(0.121)	(0.142)	(0.129)	(0.163)	(0.119)	(0.062)
Teacher instills discipline in students	-0.298**	-0.387***	-0.225**	-0.242**	0.217	0.053	0.020	-0.552***
	(0.139)	(0.089)	(0.087)	(0.097)	(0.181)	(0.199)	(0.163)	(0.111)
Controls								
Teacher Controls	Y	Y	Y	Y	Y	Y	Y	Y
School and Classroom Controls	Y	Y	Y	Y	Y	Y	Y	Y
Student Controls	Y	Y	Y	Y	Y	Y	Y	Y
Village Controls	Y	Y	Y	Y	Y	Y	Y	Y
Division fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1,077	877	944	769	1,034	806	875	674
R-squared		0.473	0.481	0.474	-	0.375	0.414	0.495

Notes: (1) The estimates are based on 113 schools for the language regressions and 109 schools for the maths regressions; (2) The measures of teacher human capital and teacher effort are based on the single teacher who was observed (teaching a language class for the language regressions and a maths class for the language regression) and assessed in the teacher tests; (3) School controls include: *a set of dummies which indicate whether the school is public, rural and located near a tarmack, classroom size, number of pupils per teacher, an index of school infrastructure* (based on the following items, given equal weight: (a) presence of toilets that were judged as designated for boys and girls, accessible, private and clean, (b) availability of electricity and (c) sufficient light for reading from the back of the class) and *an index of classroom equipment* (based on the following items, given equal weight: (a) proportion of students with pens and exercise books, (b) number of students per text book, (c) whether the classroom had the following: piece of chalk, a black board, a corner library, children's work displayed on the walls of the classroom); (4) Student controls include: *student age, age squared, whether the student is female, student score in maths, student non-verbal reasoning ability and whether the student ate breakfast*; (5) Standard errors are in parenthesis and are clustered at the class (school) level and (6) ***1 percent significance level, **5 percent significance level and *10 percent significance level.

Models 1 in table 14 show results based on HLM for language and maths (based on model 5 of table 12 and table 13). As can be seen from the tables, the patterns of estimates based on hierarchical linear modeling are similar to those based on OLS. We still find that teacher subject knowledge and effective instruction time are positively related to student achievement in maths and language. As before, teacher pedagogical skill promotes student test scores in maths and not language. Similarly, we find that an increase in the percent of students off-task is negatively related to student test scores in maths and language. We also do not see dramatic changes in the effect of teacher classroom practices.

The next robustness checks attempt to address whether the estimates in our main regressions (in table 12 and table 13) reflect some level of endogenous selection into schools. Next, we test whether our estimates are driven by possible endogenous selection into private schools. We proceed as follows. First, we undertake separate estimation for a sample comprising *pure public* schools. Second, we undertake another separate regression for a sample comprising of *pure public* schools and *private-not-for-profit* schools. As noted, *private-not-for-profit* schools are similar to pure public schools in many ways. Like public schools, these schools are not driven by the profit motive. They also do not admit learners based on previous academic record and once in school, children are less likely to be tracked by ability (Edwards Jr. et al., 2015).[49]

If our estimates are driven by between-school sorting, we expect the estimates based on these two restricted samples to be dramatically different from the estimates in our main regressions in table 12 and table 13. Models 2 of table 14 report estimates based on the sample of pure public schools for language and maths. In model 3 of the same table, we report estimates based on the sample of combined *pure public* and *private-not-for-profit* schools for language and maths.

Looking at the results based on both pure public schools (models 2) and the combined *pure public* and *private-not-for-profit* schools (models 3), the estimates of the indicators of teacher human capital and teacher effort for the two restricted samples are not dramatically different from those in our regressions in model 5 of table 12 and table 13. However, the variables teacher's subject knowlegde and effective instruction time, which were positively and significantly related to student language scores in the main regression, now have zero effects in the combined sample of *pure public* schools and *private-not-for-profit* schools (models 3). The results in the maths regression are practically similar to those in the main regressions in table 13 (model 5).

Another way to check if our estimates in table 12 and table 13 reflect some level of endogenous selection into schools is to undertake a separate estimation for rural schools. As we discuss in the next chapter, the prevalence of private schools is quite low in rural areas. Unlike urban areas, parents in rural do not have a wide variety of private school choices and are less likely to influence school choice for their children. Besides, parents in rural areas are less likely to influence choices made in schools (Spernes, 2011; Echaune et al., 2015; Makori et al., 2015). Models 4 of table 14 reports the estimates based on the rural schools for language and maths (based on model 5 of table 12 and table 13). Again, the results are closely comparable to those in the main regressions.

In chapter 2, we noted that classroom observations took place in 276 schools. In each of these schools, one grade 4 teacher was observed in either language or maths lesson. Of the 276 teachers who were observed, 222 teachers (in 222 schools) took the teacher assessment tests. The rest, 54 teachers (in

[49] We do not undertake a separate regression for pure private schools since the sample is too small.

54 schools), did not take the teacher tests for reasons we are not aware. Since our interest has been to estimate the effect of teacher human capital and effort, the estimates we have presented so far are based on a sample of 222 schools where a teacher was observed and tested.

As mentioned, one might be be concerned about the presence of a systematic pattern with respect to the teachers who did not write the test. In table A2.2 in appendix A, we show that there are no significant differences between the 222 teachers who were observed and tested and the 54 teachers who were only observed. We further account for the 54 teachers to see if our estimates change. To proceed, we estimate the regressions in table 12 and table 13 for the 276 teachers but account only for the measures of teacher effort since we do not have data on teacher test scores for 54 teachers. Results are shown in table 15. The estimated coefficients of our measures of effective instruction time, teacher ability to keep students engaged and teacher classroom practices are similar to those we present in the main regressions in table 12 and table 13. This means that our headline results do not reflect a systematic pattern with respect to the teachers who did not write the test.

Lastly, we calculated our measure of *effective instruction time* by accounting for teacher absence from class and how teachers spend their time while teaching. In order to address potential reservations of how this variable is defined, especially the aspect of adjusting for teacher absence from class, we check the direct effect of *teacher absence from class* on student test scores. In this case, we estimate our models in table 12 and table 13 by replacing the variable *teacher effective instruction time* with the variable teacher absence.

Results are shown in table 16. First, the results show that schools with higher *teacher class absence rate* are associated with lower student test scores in both subjects. For instance, we find that after accounting for teacher, school, class and student controls, a one percent increase in teacher absence in class reduces the test score in the language and maths by 0.004 and 0.002 of a standard deviation, respectively. Second, the estimates of our measures of teacher human capital and teacher effort for the two restricted samples are not dramatically different from those in our main regressions in table 12 and table 13. The estimated effect of teacher subject knowledge on student language test scores however losses its significance level but retains its sign as before.

6 Conclusion

In this paper, we examine the effect of *teacher human capital* and *teacher effort* on student achievement in maths and language among grade 4 students in Kenya. We let teacher subject knowledge and teacher pedagogical skill to represent teacher human capital. We measure teacher effort by *effective instruction time, teacher's ability to keep students engaged during the lesson* and by *a number of teacher classroom practices*. These measures of teacher effort were collected through teacher classroom observation. We control for a detailed set of variables related to teachers, schools and children as well as information capturing the socioeconomic conditions of the regions where schools are located. Our estimates are based on the OLS model with division-level fixed effects to control for potential sources of common unobservables at the divisional level.

We find that student test scores in maths and language are partially influenced by our measures of teacher human capital and teacher effort. We find that a one standard deviation increase in teacher

Table 15: Teacher Effort on Student Test Score (excluding measures of Teacher Human Capital)

	Language					Maths				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Effective instruction time (in hours)	0.043*	0.046*	0.047**	0.050**	0.048**	0.130***	0.107***	0.132***	0.083**	0.059*
	(0.022)	(0.021)	(0.020)	(0.020)	(0.020)	(0.038)	(0.037)	(0.036)	(0.037)	(0.033)
Percent of student off-task (average)	-0.007	-0.009**	-0.014***	-0.016***	-0.014***	-0.016**	-0.016**	-0.011**	-0.014**	-0.012**
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.007)	(0.007)	(0.005)	(0.006)	(0.005)
Teacher Classroom Practices										
Teacher reviews and assigns homework		0.107	0.163**	0.186**	0.160**		-0.125	-0.097	-0.231**	-0.194**
		(0.076)	(0.068)	(0.074)	(0.069)		(0.105)	(0.098)	(0.106)	(0.088)
Teacher uses local language to illustrate learning		-0.089	-0.080	-0.055	-0.057		-0.196*	-0.147	-0.055	-0.079
		(0.081)	(0.073)	(0.079)	(0.074)		(0.103)	(0.105)	(0.097)	(0.085)
Teacher challenges students		-0.172**	-0.159**	-0.186**	-0.182**		0.335***	0.234**	0.404***	0.325***
		(0.076)	(0.064)	(0.071)	(0.070)		(0.116)	(0.115)	(0.117)	(0.108)
Teacher has a lesson plan and scheme of work		-0.249*	-0.270**	-0.269**	-0.365***		0.311**	0.262*	0.224*	0.220**
		(0.133)	(0.111)	(0.115)	(0.121)		(0.131)	(0.149)	(0.128)	(0.108)
Teacher instills discipline in students		-0.149	-0.204**	-0.260***	-0.208**		-0.051	-0.098	-0.033	-0.027
		(0.122)	(0.101)	(0.095)	(0.092)		(0.125)	(0.091)	(0.099)	(0.108)
Controls										
Teacher Controls	N	Y	Y	Y	Y	N	Y	Y	Y	Y
School and Classroom Controls	N	N	Y	Y	Y	N	N	Y	Y	Y
Student Controls	N	N	N	Y	Y	N	N	N	Y	Y
Village Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Division fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1,365	1,365	1,365	1,365	1,365	1,256	1,256	1,256	1,256	1,256
R-squared	0.331	0.340	0.357	0.363	0.461	0.285	0.303	0.327	0.343	0.412

Notes: (1) The estimates in the language regression are based on 144 schools where classroom observation involved a language lesson. Similarly, in the maths regression, the estimates are based on 132 schools where classroom observation involved a maths lesson; (2) The measures of teacher human capital and teacher effort are based on the single teacher who was observed (teaching a language class for the language regressions and a maths class for the language regression) and assessed in the teacher tests; (3) School controls include: *a set of dummies which indicate whether the school is public, rural and located near a tarmack, classroom size, number of pupils per teacher, an index of school infrastructure* (based on the following items, given equal weight: (a) presence of toilets that were judged as designated for boys and girls, accessible, private and clean, (b) availability of electricity and (c) sufficient light for reading from the back of the class) and *an index of classroom equipment* (based on the following items, given equal weight: (a) proportion of students with pens and exercise books, (b) number of students per text book, (c) whether the classroom had the following: piece of chalk, a black board, a corner library, children's work displayed on the walls of the classroom); (4) Student controls include: *student age, age squared, whether the student is female, student score in maths, student non-verbal reasoning ability and whether the student ate breakfast*; (5) Standard errors are in parenthesis and are clustered at the class (school) level and (6) ***1 percent significance level, **5 percent significance level and *10 percent significance level.

Table 16: Teacher Human Capital, Teacher Effort and Student Scores: Using Teacher Absence from Class

	Language					Maths				
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
Teacher subject knowledge	0.047	0.005	0.031	0.029	0.013	0.160***	0.142**	0.132**	0.150***	0.149***
	(0.036)	(0.032)	(0.033)	(0.032)	(0.035)	(0.057)	(0.055)	(0.062)	(0.047)	(0.048)
Teacher pedagogical knowledge	0.012	-0.055	-0.003	0.022	0.027	0.041	-0.001	-0.024	0.121***	0.121***
	(0.043)	(0.039)	(0.043)	(0.050)	(0.051)	(0.054)	(0.053)	(0.050)	(0.036)	(0.037)
Teacher absence from class (percent)	-0.003*	-0.002	-0.004***	-0.003**	-0.004***	-0.005*	-0.003	-0.006**	-0.004**	-0.003*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Percent of student off-task (average)	-0.005	-0.017***	-0.019***	-0.028***	-0.029***	-0.023***	-0.025***	-0.013	-0.006	-0.005
	(0.006)	(0.005)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.006)	(0.005)
Teacher reviews and assigns homework to students		0.155	0.360***	0.478***	0.382***		-0.201	-0.201*	-0.472***	-0.419***
		(0.100)	(0.103)	(0.126)	(0.142)		(0.121)	(0.115)	(0.083)	(0.084)
Teacher uses local language to illustrate learning		-0.228**	-0.173*	-0.155	-0.136		-0.235*	-0.183	-0.103	-0.105
		(0.088)	(0.090)	(0.097)	(0.088)		(0.140)	(0.157)	(0.092)	(0.092)
Teacher challenges students		-0.092	-0.142**	-0.202***	-0.167**		0.319**	0.308**	0.362***	0.317***
		(0.077)	(0.068)	(0.073)	(0.076)		(0.144)	(0.144)	(0.121)	(0.118)
Teacher has a lesson plan and scheme of work		-0.416***	-0.257**	-0.146	-0.306**		0.075	0.063	-0.318***	-0.336***
		(0.141)	(0.129)	(0.140)	(0.144)		(0.145)	(0.161)	(0.112)	(0.110)
Teacher instills discipline in students		-0.372***	-0.302***	-0.324***	-0.233**		0.022	0.051	0.341**	0.308*
		(0.118)	(0.114)	(0.103)	(0.101)		(0.193)	(0.186)	(0.164)	(0.161)
Teacher Controls	N	N	Y	Y	Y	N	N	Y	Y	Y
School and Classroom Controls	N	N	N	Y	Y	N	N	N	Y	Y
Student Controls	N	N	N	N	Y	N	N	N	N	Y
Village Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Division fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1,077	1,077	1,077	1,077	1,077	1,034	1,034	1,034	1,034	1,034
R-squared	0.377	0.394	0.405	0.411	0.495	0.288	0.303	0.318	0.370	0.388

Notes: (1) The estimates in the language and maths regression are based on 113 and 109 schools respectively; (2) The measures of teacher human capital and teacher effort are based on the single teacher who was observed (teaching a language class for the language regressions and a maths class for the language regression) and assessed in the teacher tests; (3) School controls include: *a set of dummies which indicate whether the school is public, rural and located near a tarmack, classroom size, number of pupils per teacher, an index of school infrastructure* (based on the following items, given equal weight: (a) presence of toilets that were judged as designated for boys and girls, accessible, private and clean, (b) availability of electricity and (c) sufficient light for reading from the back of the class) and *an index of classroom equipment* (based on the following items, given equal weight: (a) proportion of students with pens and exercise books, (b) number of students per text book, (c) whether the classroom had the following: piece of chalk, a black board, a corner library, children's work displayed on the walls of the classroom); (4) Student controls include: *student age, age squared, whether the student is female, student score in maths, student non-verbal reasoning ability and whether the student ate breakfast*; (5) Standard errors are in parenthesis and are clustered at the class (school) level and (6) ***1 percent significance level, **5 percent significance level and *10 percent significance level.

knowledge in language increases the student language test score by 0.075 of a standard deviation. The effect for maths is 0.126 of a standard deviation. Teacher's pedagogical skills matter only for maths test scores. A one standard deviation increase in teacher pedagogical skill increases the maths test score by 0.112 of a standard deviation. Our results further show that an additional hour of effective teaching is associated with an increase in the language and maths test score by 0.051 and 0.059 of a standard deviation, respectively. Teachers ability to engage students during the lesson also matters for language and maths achievement.

The effects of classroom teaching practices on student tests scores are not uniform across the two subjects. For example, using local language to illustrate learning reduces student achievement. It reduces maths and language scores by 0.161 and 0.073 standard deviations, although the effect is insignificant for maths. The practice of challenging students by asking them questions has a statistically significant positive effect on maths but a statistically significant negative effect on language. Reviewing and assigning homework has a statistically significant positive effect on language but a significant negative effect on mathematics.

7 Appendix

Table A.1: School Characteristics for different Samples

	Full Sample (N=306)	Observed Sample (N=222)	Observed & not Tested Sample (N=54)	Not Observed Sample (N=30)
School Characteristics				
Rural	0.68	0.68	0.61	0.79
Public	0.78	0.77	0.76	0.86
Class size	33.67	32.28	36.63	38.39
Pupil per teacher	31.27	31.13	31.28	31.85
School located near tarmac road	0.34	0.32	0.46	0.21
Minimum Teaching Resources				
Classroom has a blackboard (percent)	0.99	0.99	1.00	0.96
Classroom has a piece of chalk (percent)	0.92	0.97	1.00	0.43
Share of pupil with a pencil (percent)	0.98	0.98	0.98	0.97
Share of pupil with an exercise book	0.99	0.99	0.98	0.97
Average number of students per textbook (maths)	2.62	2.61	2.58	2.72
Average number of students per textbook (English)	3.64	3.64	3.07	4.73
Infrastructure Availability (percent)				
School has toilets	1.00	1.00	0.98	1.00
School toilets are designated for boys and girls	0.97	0.97	0.98	0.96
School toilets are private (have doors)	0.93	0.94	0.93	0.89
School toilets are accessible	0.95	0.95	0.94	0.93
School toilets are clean	0.71	0.70	0.69	0.79
Sufficient light for reading from the back of the classroom	0.85	0.86	0.83	0.82
Students Characteristics				
Number of students	2,953	2,142	520	271
Female	0.50	0.50	0.47	0.51
Age	10.43	10.45	10.22	10.64
Had breakfast	0.87	0.86	0.90	0.89
English	79.74	79.03	83.10	79.60
Maths	60.28	60.28	59.90	61.36
Non verbal reasoning	58.88	58.87	60.38	56.27

Source: Own computation from SDI 2012. ***significance<1 percent, **significance<5 percent, *significance<10 percent

Table A.2: Teacher Classroom Practices for the 222 Schools Sample and 54 Schools Sample

	Observed & Tested (N=222)	Observed & not Tested (N=54)	Mean Difference
Effective Instruction Time	2hrs 59min	3hr 01min	2min
Percent of student off-task	9.13	11.44	-2.31
Teacher classroom practices			
Teacher had a well planned lesson plan	0.85	0.89	-0.04
Teacher had a well planned scheme of work	0.79	0.93	-0.13**
Teacher called children by name	0.87	0.87	0.00
Teacher gave feedback of praise and moral encouragement	0.85	0.85	0.00
Teacher gave student feedback that was scolding at a mistake	0.79	0.91	-0.12**
Teacher hit or slapped children	0.98	0.94	0.03
Teacher introduced the lesson	0.89	0.98	-0.09
Teacher summarized the lesson	0.65	0.69	-0.04
Teacher challenged students (through questions)	0.43	0.46	-0.03
Teacher used local language and local information	0.33	0.28	0.06
Teacher reviewed homework	0.38	0.37	0.01
Teacher assigned homework	0.68	0.71	-0.03
Other teacher characteristics			
Female	0.41	0.56	-0.15**
Experience (in years)	13.17	14.46	-1.38
Teacher has post-secondary education+	0.72	0.56	0.17**

Source: Own computation from SDI 2012. ***significance<1 percent, **significance<5 percent, *significance<10 percent

Table A.3: Indicators for Constructing Village Level Wealth Index

No.	Individual or household characteristic
1	Average number of years of education per adult (person aged 18 years and above).
2	Percent of people in the village aged 6 years & above who got a service from the following in the last one month of the census survey:
	(a) Radio
	(b) TV
	(c) Mobile phone
	(d) Computer
3	Percent of households in the village who own the following assets:
	(a) Radio
	(b) TV
	(c) Mobile phone
	(d) Computer
	(e) Bicycle
	(f) Motorcycle
	(g) Car
	(h) Refrigerator
4	Percent of households in the village whose main water source is:
	(a) Pond, dam, lake, rain, jabia and water vendor
	(b) unprotected well, unprotected spring, stream/river and borehole
	(c) protected well and protected spring
	(d) piped into dwelling place
	(e) piped
5	Percent of households in the village whose dwelling materials (walls) is made up of:
	(a) Stones
	(b) Bricks/block
	(c) Mud/wood
	(d) Mud/cement
	(e) Wood only
	(f) Corrugated iron sheets
	(g) Other materials (grass and tin)
6	Percent of households in the village whose main source of lighting is:
	(a) Electricity and solar
	(b) Pressure and gas lamp
	(c) Lantern and
	(d) Fuel wood
7	Percent of households in the village who own the following livestock:
	(a) Cattle
	(b)Sheep
	(c) Goat
	(d) Camel
	(e) Chicken

References

Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in the chicago public high schools. *Journal of labor Economics*, 25(1):95–135.

Ackers, J. and Hardman, F. (2001). Classroom interaction in kenyan primary schools. *Compare*, 31(2):245–261.

Alcazar, L., Rogers, F. H., Chaudhury, N., Hammer, J., Kremer, M., and Muralidharan, K. (2006). Why are teachers absent? probing service delivery in peruvian primary schools. *International Journal of Educational Research*, 45(3):117–136.

Altinok, N. and Kingdon, G. (2012). New evidence on class size effects: A pupil fixed effects approach. *Oxford Bulletin of Economics and Statistics*, 74(2):203–234.

Aslam, M. and Kingdon, G. (2011). What can teachers do to raise pupil achievement? *Economics of Education Review*, 30(3):559–574.

Bietenbeck, J.-C. (2011). Teaching practices and student achievement: evidence from timss. *Yayımlanmamış yüksek lisans tezi. CEMFI*, (114).

Bold, T., Kimenyi, M., Mwabu, G., and Sandefur, J. (2011). Why did abolishing fees not increase net enrollment in kenyan public schools?. *Center for Global Development, Working Paper*, (271).

Bold, T., Kimenyi, M. S., and Sandefur, J. (2013). Public and private provision of education in kenya. *Journal of African Economies*, 22(suppl 2):ii39–ii56.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education*, 1(2):176–216.

Bruns, B. and Luque, J. (2014). *Great teachers: How to raise student learning in Latin America and the Caribbean*. World Bank Publications.

Bruns, B. and Rakotomalala, R. (2003). *Achieving universal primary education by 2015: A chance for every child*, volume 1. World Bank Publications.

Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical linear models: applications and data analysis methods*. Sage Publications, Inc.

Case, A. and Deaton, A. (1999). School inputs and educational outcomes in south africa. *Quarterly journal of economics*, pages 1047–1084.

Cavatassi, R., Davis, B., Lipper, L., et al. (2004). Estimating poverty over time and space: construction of a time-variant poverty index for costa rica. *Rome: Food and Agriculture Organization*.

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

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4):778–820.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6):673–682.

Corder, S. P. (1982). *Error analysis and interlanguage*, volume 198. Oxford University Press.

Danielson, C. (1996). A framework for teaching. *Alexandria, VA: Association for Supervision and Curriculum Development*.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, pages 158–165.

Dieterle, S., Guarino, C. M., Reckase, M. D., and Wooldridge, J. M. (2015). How do principals assign students to teachers? finding evidence in administrative data and the implications for value added. *Journal of Policy Analysis and Management*, 34(1):32–58.

Dixon, P. (2012). Why the denial? low-cost private schools in developing countries and their contributions to education. *Econ Journal Watch*, 9(3).

Dixon, P. and Tooley, J. (2012). A case study of private schools in kibera: an update. *Educational Management Administration & Leadership*, page 1741143212456908.

Echaune, M., Ndiku, J, M., and Sang, A. (2015). Parental involvement in homework and primary school academic performance in kenya. *Journal of Education and Practice*, 6(9).

Edwards Jr., D. B., Klees, S., and Wildish, J. (2015). Dynamics of low-fee private schools in kenya: Governmental legitimation,school-community dependence,and resource uncertainty. Technical report, Teachers College Record, forthcoming.

Fielding, A. (2010). Module 8: Multilevel modelling in practice: Research questions, data preparation and analysis. *LEMMA VLE, University of Bristol, Centre for Multilevel Modelling. Retrieved December*, 16:2014.

Filmer, D., Molina, E., and Stacy, B. (2015). What goes on inside the classroom in africa? assessing the relationship between what teachers know, what happened in the classroom, and student performance.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of economic literature*, 40(2):436–482.

Glewwe, P. and Jacoby, H. (1994). Student achievement and schooling choice in low-income countries: Evidence from ghana. *Journal of Human Resources*, pages 843–864.

Glewwe, P. and Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2:945–1017.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., and Ravina, R. (2011). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. Technical report, National Bureau of Economic Research.

Hanushek, E. A. (2006). School resources. *Handbook of the Economics of Education*, 2:865–908.

Hanushek, E. A. (2008). Incentives for efficiency and equity in the school system. *Perspektiven der Wirtschaftspolitik*, 9(s1):5–27.

Hardman, F., Abd-Kadir, J., Agg, C., Migwi, J., Ndambuku, J., and Smith, F. (2009). Changing pedagogical practice in kenyan primary schools: the impact of school-based training. *Comparative Education*, 45(1):65–86.

Heyneman, S. P. and Stern, J. M. (2014). Low cost private schools for the poor: What public policy is appropriate? *International Journal of Educational Development*, 35:3–15.

Hidalgo Cabrillana, A. and Lopez-Mayan, C. (2015). Teaching styles and achievement: student and teacher perspectives. *Available at SSRN 2569020*.

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

Jacob, B. A. and Lefgren, L. (2005). Principals as agents: Subjective performance measurement in education. Technical report, National Bureau of Economic Research.

Jones, S., Schipper, Y., Ruto, S., and Rajani, R. (2014). Can your child read and count? measuring learning outcomes in east africa. *Journal of African economies*, 23(5):643–672.

Kane, T. J., Taylor, E. S., Tyler, J. H., and Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, 46(3):587–613.

Kingdon, G. G. (2006). Teacher characteristics and student performance in india: A pupil fixed effects approach.

Kukla-Acevedo, S. (2009). Do teacher characteristics matter? new results on the effects of teacher preparation on student achievement. *Economics of Education Review*, 28(1):49–57.

Larbi, G., Adelabu, M., Rose, P., Jawara, D., Nwaorgu, O., and Vyas, S. (2004). Non-state providers of basic services. *Country Studies, Nigeria. Birmingham: University of Birmingham*.

Lavy, V. (2010). Do differences in schools' instruction time explain international achievement gaps? evidence from developed and developing countries. Technical report, National Bureau of Economic Research.

Lavy, V. (2011). What makes an effective teacher? quasi-experimental evidence. Technical report, National Bureau of Economic Research.

Lavy, V. (2012). Expanding school resources and increasing time on task: effects of a policy experiment in israel on student academic achievement and behavior. Technical report, National Bureau of Economic Research.

Makori, A., Chepwarwa, J., Jepkenei, E., and Jacob, P. (2015). Admission into public secondary schools in kenya:understanding parental preferential limitations. *Journal of Educational and Social Research*, 5:3–12.

Martin, G. H. and Pimhidzai, O. (2013). Education and health services in kenya: Data for results and accountability.

McLaughlin, B. (2013). *Second language acquisition in childhood: Volume 2: School-age Children*. Psychology Press.

Metzler, J. and Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics*, 99(2):486–496.

MoEHRD (1999). Classroom interaction in kenyan primary schools: A national primary baseline special study. Technical report, Nairobi, Kenya: Ministry of Education and Human Resource Development.

Ngware, M. W., Oketch, M., and Mutisya, M. (2014). Does teaching style explain differences in learner achievement in low and high performing schools in kenya? *International Journal of Educational Development*, 36:3–12.

Oketch, M., Mutisya, M., and Sagwe, J. (2012). Do poverty dynamics explain the shift to an informal private schooling system in the wake of free public primary education in nairobi slums? *London Review of Education*, 10(1):3–17.

Oketch, M. and Somerset, A. (2010). Free primary education and after in kenya: Enrolment impact, quality effects, and the transition to secondary school.

Onsomu, E., Nzomo, J., and Obiero, C. (2005). The sacmeq ii project in kenya: A study of the conditions of schooling and the quality of education. Technical report, ResearchTriangle Park. NC, USA: RTI International.

Orazem, P. F. and King, E. M. (2007). Schooling in developing countries: The roles of supply, demand and government policy. *Handbook of development economics*, 4:3475–3559.

Piper, B. and Mugenda, A. (2012). The primary math and reading (primr) initiative:baseline report. Technical report, RTI International,Research Triangle Park, NC, USA.

Piper, B. and Mugenda, O. (2010). The primary math and reading (primr) initiative: Baseline report. Technical report, Kenya National ExaminationsCouncil, Nairobi.

Pontefract, C. and Hardman, F. (2005). The discourse of classroom interaction in kenyan primary schools. *Comparative education*, 41(1):87–106.

Pritchet, L. and Banerji, R. (2013). Schooling is not education! using assessment to change the politics of non-learning. Technical report, Center for Global Development Washington, D.C.

Riddell, A. (2003). The introduction of free primary education in sub-saharan africa. *Background paper prepared for the Education for All Global Monitoring Report*, 4.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, pages 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, pages 247–252.

Rockoff, J. E., Staiger, D. O., Kane, T. J., and Taylor, E. S. (2010). Information and employee evaluation: Evidence from a randomized intervention in public schools. Technical report, National Bureau of Economic Research.

Schwerdt, G. and Wuppermann, A. C. (2011). Is traditional teaching really all that bad? a within-student between-subject approach. *Economics of Education Review*, 30(2):365–379.

Shepherd, D. (2013). The impact of teacher subject knowledge on learner performance in south africa: A within-pupil across-subject approach. In *International Workshop on Applied Economics of Education, Cantanzaro*, pages 1–32.

Shepherd, D. et al. (2015). Learn to teach, teach to learn: A within-pupil across-subject approach to estimating the impact of teacher subject knowledge on south african grade 6 performance. Technical report.

Spence, M. and Lewis, M. A. (2009). *Health and growth*. World Bank Publications.

Spernes, K. (2011). I buy paraffin so he can read in the evening–a study from kenya about parental involvement in school. *International Journal About Parents in Education*, 5(1).

Stallings, J. (1977). *Learning to Look: A Handbook on Classroom Observation and Teaching Models*. Belmont, CA: Wadsworth Publishing.

Stallings, J. and Knight, S. (2003). Using the stallings observation system to investigate: Time on task in four countries. Technical report, Unpublished paper for the International Time on Task Project, World Bank, Washington, DC.

Stock, J. H. (2010). The other transformation in econometric practice: Robust tools for inference. *The journal of economic perspectives*, 24(2):83–94.

Swanson, R. C., Cattaneo, A., Bradley, E., Chunharas, S., Atun, R., Abbas, K. M., Katsaliaki, K., Mustafee, N., Meier, B. M., and Best, A. (2012). Rethinking health systems strengthening: key systems thinking tools and strategies for transformational change. *Health Policy and Planning*, 27(suppl 4):iv54–iv61.

Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement*. *The Economic Journal*, 113(485):F3–F33.

Tooley, J., Bao, Y., Dixon, P., and Merrifield, J. (2011). School choice and academic performance: Some evidence from developing countries. *Journal of School Choice*, 5(1):1–39.

Tooley, J. and Longfield, D. (2015). The role and impact of private schools in developing countries: A response to the dfid-commissioned rigorous.

Trudell, B. and Piper, B. (2013). Whatever the law says: Language policy implementation and early-grade literacy achievement in kenya. *Current Issues in Language Planning*, 1(15).

Van Klaveren, C. (2011). Lecturing style teaching and student performance. *Economics of Education Review*, 30(4):729–739.

Wasanga, P. M., Ogle, A. M., and Wambua, R. M. (2010). Report on monitoring of learner achievement for class 3 in literacy and numeracy. Technical report, Nairobi: Kenya National Examinations Council. World Bank.

West, M. R. and Wobmann, L. (2006). Which school systems sort weaker students into smaller classes? international evidence. *European Journal of Political Economy*, 22(4):944–968.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *The American Economic Review*, 93(2):133–138.

World-Bank (2009). Abolishing school fees in africa: Lessons from ethiopia, ghana, kenya, malawi and mozambique. Technical report, Washington DC, US: World Bank.

Zakharov, A., Carnoy, M., and Loyalka, P. (2014). Which teaching practices improve student performance on high-stakes exams? evidence from russia. *International Journal of Educational Development*, 36:13–21.