Southern Africa Labour and Development Research Unit

Private Schools and Student Learning Achievements in Kenya

by

Fredrick M. Wamalwa and Justine Burns

Working Paper Series
Number 202, Version 2

About the Author(s) and Acknowledgments

Fredrick Wamalwa: School of Economics, Faculty of Commerce, University of Cape Town, Private Bag, Rondebosch 7701, Cape Town, South Africa. E-mail: fredwamalwa@yahoo.co.uk - corresponding author
Justine Burns: School of Economics, University of Cape Town, South Africa

This work is funded by the African Economic Research Consortium (AERC).

Recommended citation

Wamalwa., F.M, Burns, J. (2017). Private Schools and Student Learning Achievements in Kenya. Version 2 Cape Town: SALDRU, UCT. (SALDRU Working Paper Number 202).

ISBN: 978-1-928281-63-4

Working Papers can be downloaded in Adobe Acrobat format from www.saldru.uct.ac.za.
Printed copies of Working Papers are available for R25.00 each plus vat and postage charges.

Orders may be directed to:
The Senior Administrative Officer, SALDRU, University of Cape Town, Private Bag, Rondebosch, 7701,
Tel: (021) 650 1808, Fax: (021) 650 5697, Email: tania.hendricks@uct.ac.za

Private Schools and Student Learning Achievements in Kenya*

Fredrick M. Wamalwa† & Justine Burns‡

April 10, 2017

This paper examines the effect of private schools on literacy (language) and numeracy (maths) skill acquisition among children drawn from lower primary grades in Kenya. We use a comprehensive household survey data that allows us to apply a number of econometric techniques to deal with the challenge of the endogeneity of private school choice. We begin with the OLS as a baseline model. We then estimate the village and household fixed effects (FE) models that control for unobservables at the village and household levels, respectively. We supplement the OLS and FE models with the propensity score matching (PSM) technique. We find positive and significant private school effect throughout all these methodologies. However, assessing the impact of omitted variable bias on the estimated coefficient of private schools by use of recent techniques, we find that the estimated bias in household FE is quite small in magnitude relative to the bias based on other estimation techniques. Since (private) schooling decision is made at the household level, it is likely that a substantial part of the unobservable component is pertaining to the household.

Key words- Private schools, student learning achievements, Kenya.

---

*This work is funded by the African Economic Research Consortium (AERC)

†*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. Tel: +254 726 406 371

‡School of Economics, University of Cape Town, South Africa

1 Introduction

In a number of sub-Saharan African countries, evidence shows that the elimination of user fees in public primary schools was followed by dramatic increases in private schools (Dixon and Tooley 2012, Dixon 2012, Tooley and Dixon 2005, Tooley et al. 2008, 2011, Tooley 2013, Tooley and Longfield 2015, Oketch and Ngware 2010, Oketch et al. 2010, 2012, Larbi et al. 2004). The rise in private schools has been associated with high demand for school places in the face of limited supply of quality schools from the government. Majority of these schools have been born out of community or private initiatives to establish schools mainly within the urban informal settlements, schools that levy low fees, referred in the literature as low-cost private schools (Rose 2006, Tooley et al. 2008, 2011, Tooley and Longfield 2015).

The effectiveness of private schools has also been discussed in the recent literature (see for example Javaid et al. (2012), Andrabi et al. (2008), French (2008), Pal (2010), Bold et al. (2011, 2013a), Desai et al. (2009)). Majority of these studies find that relative to public counterparts, private schools are better at promoting student achievements, mainly measured in terms of test scores. The validity and magnitude of private school effect is however still debated, questioned and subject to further research. Researchers such as Goldberger and Cain (1982), Newhouse and Beegle (2006) and Altonji et al. (2000, 2005) urge that such private school advantage may be due to spurious correlations between private school attendance and unobserved student and family characteristics. Children who attend private schools may already have high academic potential or even access to complementary educational resources in a manner that is not easily observable to the researcher.

Endogenous selection into private schools is quite evident in sub-Saharan African countries. In Kenya, studies have shown that poor parents bypass *free public primary schools* and send their children to otherwise *fee-paying low-cost private* schools due perceived better quality in private schools (Tooley et al. 2008, Oketch and Ngware 2010, Oketch et al. 2010). Such parents, concerned with their children's education, despite being poor, are more likely to ensure that the home environment is favorable for learning for their children. For instance, they are likely to invest in efforts such as helping their children with homework. These are factors that are not easily observed to researchers. Evidence shows that accounting for unobservable factors and selection bias can wash away or dramatically reduce such private school advantage. In Indonesia, Newhouse and Beegle (2006) accounts for selection effects and finds that private schooling has in fact significant negative effects on test scores.

In this paper, we use a rich household survey data to estimate the effect of private schools on literacy (language) and numeracy (maths) skill acquisition among children mainly drawn from lower primary grades in Kenya. Our main contribution lies in accounting for the endogeneity of private school choice while estimating the effect of private schools. We do so by using different econometric techniques. We begin with the OLS as a baseline model. We then estimate the village and family fixed effects[1] models that control for unobservables at the village and household levels, respectively. Using a methodology advanced by Altonji et al. (2005), we estimate the potential size of any bias on the estimated coefficient of the private school variable due to unobservable selectivity in the OLS and the fixed effects models. We supplement the OLS and fixed effects models with a non-parametric estimation technique, that is, propensity score matching (PSM). Following Rosenbaum (2002), we check the extent to which our estimates based on PSM suffer from hidden bias (unobservables).

As pointed out by French et al. (2010), each of the above strategies has its own strengths and weaknesses and no strategy on its own can yield convincing estimates of the private school premium. It is for this reason that we apply different methods to see if they lead to similar conclusions about the effect of private schools on children literacy and numeracy skill acquisition. This comparative analysis also provides us with a sense of the range of the size of the private school effect.

Private schooling in Kenya has expanded dramatically over the past decade (Tooley et al. 2008, Heyneman and Stern 2014, Tooley and Longfield 2015, Edwards Jr. et al. 2015, Piper and Mugenda 2010, Oketch et al. 2010, 2012, Piper et al. 2015). However, there is a dearth of studies that look at the effectiveness of these schools. Research has mostly focused on understanding why households, mainly poor households, choose to enroll their children in low-fee paying private schools and not in free public schools. The reasons are varied. Tooley et al. (2008) and Oketch and Somerset (2010) find that perceived better quality of private schools (in terms of teaching, teacher attendance, school performance, small class size and discipline) is a key driver of parents' choice of private schools. Oketch et al. (2012) focusing in urban areas and Nishimura and Yamano (2013) focusing in rural areas both find that an increase in household wealth is likely to lead to the household enrolling a child in a private school.

To our knowledge, only one study in Kenya by Bold et al. (2013a) has estimated the effect of private schools on student scores while addressing the endogeneity of school choice in Kenya. The authors find a large private school premium, equivalent to one standard deviation based on grade eight tests scores (end of primary cycle examinations). Their study however suffers

[1] The term family and household are used interchangeably throughout this thesis.

from two shortcomings. First, in a typical country like Kenya, the sample of children in school becomes more and more self-selective as one advances to higher grades due to relatively high drop-out rates. Second, the fact that a pupil sits for end of primary cycle examinations (grade eight tests) at a private school does not mean that he or she has been in that school from grade one.[2] The estimates in Bold et al. (2013a) based on end of primary cycle examinations (grade eight tests) are, likely to suffer from sample selection bias (due to relatively high drop-out rates) and mis-attribution (due to possible transfer from private to public schools or vice versa). To deal with these challenges, we restrict our sample to children in lower primary grade two to grade four. We are not aware of any study that focuses on the importance of private schools on cognitive development of children from *lower primary grades* in Kenya and by extension sub-Sahara Africa.

Here is a preview of our results. We find a positive and significant private school advantage across all the estimated methods. Our results however seem to suggest that the OLS, village fixed effects and propensity score matching regressions overestimate the true primary school effect while the household fixed effects model underestimates the private school effect. From this, we have a sense of the range of the private school premium. In maths, the premium ranges from 0.13 to 0.21 score standard deviation, based on the household and village fixed effects models, respectively. In the case of language, it ranges from 0.20 to 0.29 score standard deviation, based on the household and village fixed effects models, respectively. Results from Altonji et al. (2005) shows that the size of the bias on the estimated coefficient of the private school variable due to unobservable selectivity, especially based on the household fixed effects model, is quite small.

The rest of the paper unfolds as follows. Section 2 provides the context of private sector education provision in Kenya while in section 3, we review literature on private school effectiveness. In section 4, we provide a description of our survey data and the summary statistics. Section 5 outlines the estimation techniques. In section 6, we discuss the empirical results and finally issue concluding remarks in section 7.

[2]Evidence in Kenya shows that since the majority of private schools are not formally registered, most parents send their children to these schools but transfer them to public schools for their end of primary cycle exams in grade eight (Edwards Jr. et al. 2015).

2 The Context: Private School Provision in Kenya

Kenya has a long history of private sector education provision. Private sector education providers include non-governmental organizations, faith-based organizations, community-based providers and private-for-profit agents (Tooley et al. 2008, Heyneman and Stern 2014, Tooley and Longfield 2015, Edwards Jr. et al. 2015). Faith-based organizations and community-based providers have supported education provision since the early 1960s. Through the 1980s and 1990s, private education provision expanded owing to structural adjustment programs (SAPs) that led to the reduction in public education funding (Nishimura and Yamano 2013). This period saw the entry of private-for-profit agents. Despite these developments, private education remained out of reach for children from poor and rural households as many community-financed school projects were not financially sustainable (Olembo 1985).

Currently, Kenya's primary education provision is characterized by free public provision of education and a huge market for private fee-charging schools. The fall in the quality of education offered in public schools mainly following the introduction of free primary education led to an increase in private school provision (Bold et al. 2013b, Oketch and Somerset 2010, Oketch et al. 2010). Private school provision accounts for about 25 percent of the total primary school sector in Kenya (KIPPRA 2016). This does not however account for the highly unregulated and unregistered non-formal private sector schools in urban informal settlements.

Today, private schools in Kenya reflect a diverse range of institutions, ranging ranging from: (a) highly unregulated and sometimes unregistered non-formal schools mainly located in informal settlements[3]; (b) formal private academies in middle and high-income urban areas and (c) very few old traditionally exclusive private schools offering foreign curriculum such as the General Certificate of Secondary Education (GCSE) (Piper and Mugenda 2010, Piper et al. 2014).

The main development in the sector during the post-free public primary school era has been the mushrooming of non-formal schools located in urban informal settlements whose goal has been to meet the high demand for school places in those urban informal settlements (Tooley et al. 2008, Heyneman and Stern 2014, Edwards Jr. et al. 2015, Tooley and Longfield 2015, Piper et al. 2014, 2015). These schools levy low fees to make them affordable for children from poor urban informal settlements. They are the main source of education for children in urban

[3]Our conversation with staff at the Ministry of Education shows that most of these schools are mainly in urban areas of big municipalities like Mombasa, Eldoret, Nairobi, Thika, Nakuru, Kisumu and Kitale.

informal settlements and for some families, the choice may not be between a government primary school and a non-formal school, but between the non-formal school and no school at all (Oketch and Somerset 2010).

Such low-fee private schools generally lack school infrastructure and facilities, trained teachers and adequate teaching and learning resources. They are characterized by high student and teacher turnover. Parents with children in these institutions pay tuition fees that average less than USD. 10 per month (Piper and Mugenda 2010, Piper et al. 2014). Admission to these low cost private schools is granted at the discretion of the head teacher and only a few schools conduct interviews for new students as part of the selection process. Although some are registered and receive some form of support from the government[4], the majority run with limited engagement with the government. In fact, low cost private schools operate within the same constraints as households living in their catchment area (Edwards Jr. et al. 2015). In this regard, they have little or no security of tenancy, are highly unregulated, lack space or sanitation facilities and are vulnerable to the challenges that characterize densely populated and volatile urban informal settlements in Kenya (Tooley et al. 2008, Heyneman and Stern 2014, Edwards Jr. et al. 2015, Piper et al. 2014, Tooley and Longfield 2015, Piper et al. 2015, Piper and Mugenda 2010).

While a lot has been documented with regard to low cost non-formal private schools in urban informal settlements, there is little information regarding private schools in rural areas as well as private academies in middle and high-income urban areas. There are indications of remarkable penetration of such schools in rural areas. We are only aware of a study by Nishimura and Yamano (2013) which looked at determinants of school choice in rural Kenya. This study is based on a panel survey of 76 randomly selected rural sub-locations from Western and Central provinces. The authors found that between 2003 and 2007, 35 (out of 119) new private schools were established in these regions relative to only six (out of 318) public schools, reflecting a clear increasing demand for private schools in rural Kenya.

[4]Starting in 2005, the Ministry of Education extended support (for instructional materials) to the low-fee paying private schools in informal settlements. The support comes through the Instructional Materials initiative, in which schools receive funding each academic term to pay for instructional materials, such as chalk, erasers and books. The amount allocated per pupil is intended to equip schools with key textbooks on a shared basis of one book for every two children. In order to qualify for this government support, the school should: (i) be registered, (ii) be assessed by Ministry of education officials in terms of location, sanitation, safety etc and (iii) have a School Management Committee comprising of teachers and two representatives of the parents (Edwards Jr. et al. 2015). The number of schools under government support had increased from 59 in 2004 to 410 in 2009.

3 Do Private School Children achieve better Learning Outcomes than counterparts in Public Schools?

To estimate the effect of private schools, the earliest studies estimated the effect of private schools using a dummy variable indicating whether a child is attending a private school or not alongside other characteristics related to students and their families. Some of these include; Hammersley et al. (1981) for the United Kingdom, Psacharopoulos (1987) for Colombia and Tanzania, Govinda and Varghese (1993) for India, McEwan and Carnoy (2000) for Chile and McEwan (2002) for Argentina and Chile, among others. Using OLS estimation, these studies generally find that attending private schools is associated with an increase in the student test scores. The main challenge with all these studies is that they treat private school choice as exogenous which is not likely to be the case.

Most recent studies have dealt with the endogeneity of school choice through different approaches. One such approach is the use of experimental data where students are randomly assigned to schools, mainly through school vouchers. Angrist et al. (2002) examines the impact of a program that used a lottery to distribute vouchers that covered the cost of private secondary school education in Colombia. Students were randomly awarded or denied private school vouchers and as such, vouchers provided a convenient method of *randomly assigning* students to private schools. The study finds that after three years, voucher winners attending private schools, scored 0.2 score standard deviations higher than their public school counterparts in achievement tests comprising tasks in maths, reading and writing.

Despite its attractiveness, the use of experimental data has weaknesses. It depends on the accuracy of the randomization process and secondly, it is quite expensive to roll out randomization programs. As a result, most studies are based on non-experimental data and have employed a number of approaches in an attempt to estimate the true effect of private schools on learner test scores. The instrumental variable (IV) approach is an example. In this approach, certain variables (instruments) which affect private school attendance but not student achievement are used to estimate the casual effect of private schooling (Goldhaber and Eide 2003, Wooldridge 2010).

Different instruments have been explored in the literature. Angrist et al. (2002) uses *receipt of a voucher* as an instrument for attending private school to determine the effect of private schools in Colombia. The argument is that those who obtain a voucher actually attend private school, but receiving the voucher in itself does not guarantee that one excels in school. The study finds that attending a private school increases the probability of finishing eighth grade by 13 to 15 percentage points, and increases test scores by 0.29 score standard deviations in combined tasks involving maths, reading and writing. In Pakistan, Andrabi et al. (2008) use *distance to the private school* as an instrument for private school enrolment. They find that attending a private school is associated with increase in student test scores by 0.8 to 1 standard deviations (depending on the subject). Evans and Schwab (1995) use *religious affiliation* as an exogenous source of variation in Catholic school (private school) attendance. They find that attending Catholic schools increases high school graduation by 12 percentage points and college attendance by 14 percentage points[5].

Although the use of instruments is appealing, finding a credible instrument is challenging. As a result, other studies adjust for selection bias using propensity score matching to estimate the causal effect of private schools. Using the data from a nationwide survey of 452 schools and 22,500 secondary level students in Nepal, Thapa (2015) adjusts for selection bias using propensity score matching and still finds that attending a private school increases student scores by about 8 percentage points in school leaving certificate exams. In India, Azam et al. (2015) use propensity score matching to examine the effect of attending private secondary schools on student test scores in standardized maths test in two Indian states, Orissa and Rajasthan. They find a private school advantage of 0.4 and 1.3 score standard deviations in rural and urban areas of Rajasthan. The results in Orissa depend on the location where the household is located. For instance, they find a private school advantage of 0.3 score standard deviations in rural Orissa and no evidence of private school advantage in urban Orissa.

Most recently, studies have used other estimation approaches such as the family/household fixed effects models, which control for household level observable and unobservable factors, in estimating the causal impact of private schools. In India, French et al. (2010) estimate a household fixed effects model and finds a private school premium of 0.17 score standard deviations in combined student maths and language test scores. Similarly, Javaid et al. (2012) estimates a household fixed effects model using household survey data from Pakistan and reports a private school premium of 0.04 score standard deviations in combined student maths and language test scores. Besides family fixed effects models, both studies estimate cluster fixed effects at different levels (province, district and village). In comparison, the

[5] Availability of private school in the child's residential area has also been used as an instrument. Studies measure this instrument differently. In Nepal, Shama (1999) cited in French (2008) instruments private school attendance using the number of private schools in the child's residential area. In India, Desai et al. (2008) uses a dummy variable on whether a child's region has a private school or not as an instrument for private school attendance.

household fixed effects models produce the lowest estimates; showing it is effective in reducing the unobservables associated with private school choice.

In summary, there is evidence of a private school advantage documented in the literature and studies have adopted different methods to find the true effect of private schools. As noted by Ashley et al. (2014), much of this literature in developing countries is based on South Asian countries, India and Pakistan in particular. We add to this literature by estimating the effect of private schools using the case study of Kenya, a country from Sub-Sahara Africa.

4 Data and Descriptive Statistics

4.1 The Uwezo Survey Data

We use the third round of Uwezo survey for Kenya collected in 2012. The Uwezo[6] initiative has been implementing large-scale household surveys that assess literacy and numeracy competencies of school age children since 2009. A detailed description of the sampling strategy is provided in Jones et al. (2014). The third round of Uwezo survey was based on a two-stage random sampling design. First, 30 primary sampling units (enumeration areas and/or villages) from each district were selected with the probability of selection proportional to population size. Second, about 20 households in each enumeration area were selected via systematic random sampling.[7] Uwezo targets children aged 6-16 years who are regular residents of the household. Households without such children were therefore excluded.

4.1.1 Household, School and Village Surveys

In each enumeration area, data collection involved three steps. First, data was collected from one randomly selected local public primary school within the enumeration area.[8] Information was gathered on school enrolment, teachers, classroom facilities as well as school facilities among others. Close to 4,465 schools were covered. Second, a questionnaire was administered

[6]Uwezo which means 'capability' in Kiswahili, is a non-governmental organization that aims to improve competencies in literacy and numeracy among children aged 6-16 years in Kenya. More details about Uwezo can be found at: http://www.uwezo.net/.

[7]The sample design was provided by the Kenya National Bureau of Statistics (KNBS).

[8]In cases where there was no public primary school in the sampled enumeration area, the nearest public school attended by the majority of children in the sampled enumeration area was selected. When more than one school was available in the enumeration area, the school that attended by a majority of the students was selected. *Only public schools* and not private schools were targeted.

to the village head of the sampled enumeration areas (villages). Among others, it gathered information on availability of: (i) social amenities (chief's office, shopping center and police post), (ii) infrastructure (roads, all-weather roads, protected water points and electricity), and (iii) the number of educational and health facilities in the village.[9]

Finally, households were visited. A questionnaire was then administered to the head of the household (or representative). For children aged 6-16 years, information was gathered about their age, gender, disability, school grade, whether they were enrolled in school and for those enrolled, the type of school they were enrolled in (private or public) and the time taken to arrive at school. The household questionnaire also collected information on parental age and education as well as indicators of household socioeconomic status.

4.1.2 Children (Student) Assessments

Each child of school age (6-16 years), whether or not attending school, was assessed in language (English and Kiswahili) and mathematics. The tests were based on the grade two level curriculum. In this study, we do not consider Kiswahili tests. The language (literacy) tests were designed to assess five principal competencies, namely: (1) letter recognition, (2) word recognition, (3) ability to read a paragraph, (4) ability to read a (short) story and (5) ability to comprehend information in the story (Uwezo 2012, 2014, Jones et al. 2014, Wakano 2016).

As noted in Jones et al. (2014), each competence level was assessed by a separate test item. However, due to the ordered nature of the competencies,[10] not all children are assessed on each competency item (see Jones et al. (2014)). In this regard, the literacy assessments began with level 3 competency (reading a paragraph) and either stepped up (to comprehension) or stepped down (to letter level) in difficulty depending on the child's initial response. Overall, each child was classified into one of these five ascending categories: (1) knows nothing; (2) can identify a letter; (3) can identify a word; (4) can read a paragraph; (5) can read a short story and (6) can do comprehension.

[9]Information gathered included the number of primary and secondary schools (public and private) as well as the number of village polytechnics. Data was collected on the number of health facilities run by government and non-governmental organizations.

[10]Ordered nature of the competencies means that *comprehension of the paragraph* (highest competence) requires ability to read story, process it, and understand its meaning. Knowing how to read a *story* implies ability to read a *paragraph* which in turn implies ability to recognize *words*. Ability recognize *words* implies ability to recognize *letters* (lowest competence).

The numeracy tests, structured and administered in a similar way to the literacy tests, assessed the following competencies: (1) counting, (2) number recognition (two digits), (3) rank ordering of two numbers, (4) addition, (5) subtraction, (6) multiplication and (7) division (Wakano 2016, Uwezo 2012, 2014, Jones et al. 2014). It began with the level 5 item (subtraction) and the level of difficulty was either stepped up (to division) or stepped down (to counting) depending on the child's initial response. Similarly, each child was classified into one of these ascending categories: (1) knows nothing; (2) can count; (3) can identify a number; (4) can discriminate numbers; (5) can add (6) can subtract; (7) can multiply and (8) can divide.

4.2 Descriptive Statistics

The 2012 Uwezo survey covered about 145,564 children aged 6-16 from 72,000 households residing in close to 4,000 villages. In this paper, we focus on children who were currently enrolled and were in grades two to four. We impose this restriction for various reasons.[11] First, there is scant information on the effect of private schools on children learning in lower grades in Kenya and by extension sub-Sahara Africa. Second, in a typical developing country like Kenya, the sample of primary school children is likely to become more self-selective as one goes higher up due to drop-out rates. Focusing on grade two to four allows us to minimize such potential self-selection problems. Imposing this restriction reduces the number of children to 52,709 distributed in close to 20,000 households.

We have two dependent variables, that is, student scores based on the language (literacy) and maths (numeracy) tests. As already mentioned, the Uwezo literacy (language) tests assessed five principal competencies, ranging from letter recognition to comprehension of the paragraph. The numeracy (mathematics) tests assessed seven principal competencies, ranging from counting and matching numbers to simple division tasks. To define the dependent variables based on these competencies, we closely follow two recent studies by French et al. (2010) and Wakano (2016). French et al. (2010) estimated the effect of private schools in rural India based on the Annual Status of Education Report (ASER) survey. The Uwezo and ASER surveys are in fact quite similar in structure.[12] Wakano (2016) used the Uwezo survey for Kenya collected in 2010 to estimate the effects of locally hired teachers on student

[11] We exclude grade one since the assessments were based on grade two syllabus.

[12] Just like Uwezo survey, the ASER survey collects data at three levels: households, school (one public schools per village) and village levels. Children in both surveys (Uwezo and ASER surveys) are assessed in similar literacy and numeracy competencies.

achievements in Kenya.

Table 1: Student Test Score Outcomes in Language and Maths

Language	Mark	Maths	Mark
Could do nothing	0	Could do nothing	0
Could read letters	1	Could count and match numbers	1
Could read words	2	Could identify numbers	2
Could read a paragraph	3	Could discriminate quantities	3
Could read a story	4	Could do addition	4
Could do comprehension	5	Could do subtraction	5
		Could to multiplication	6
		Could to division	7

Source: Own calculations based on Uwezo 2012. Notes: Based on Uwezo (2012), Uwezo (2014) following French et al. (2010) and Wakano (2016).

In the footsteps of French et al. (2010) and Wakano (2016), we allow scores in language to range from 0 (student could not manage any of the tasks) to 5 (student could manage all tasks up to comprehension level). In the same vein, maths scores range from 0 (student could not manage any of the tasks) to 7 (student could manage all tasks up to division level) (see table 1). In the regression analysis, we follow French et al. (2010) by standardizing the scores in table 1 to the mean of 0 and standard deviation of 1 to make interpretation easy. We do so separately for language and mathematics scores.

Table 2: Student Learning Outcome by School Type

	(1)	(2)	(3)	(4)
	Whole sample	Public sample	Private sample	Public-Private
	Mean	Mean	Mean	Mean Dif.
Language	2.72	2.62	3.47	-0.84***
Maths	4.75	4.67	5.39	-0.71***
Sample	52,196	44,373	6,256	

Source: Own calculations based on Uwezo 2012. Notes: *** 1% significance level, ** 5% significance level and * 10% significance level.

Before turning to regression results, we show some descriptive statistics. In table 2, we show the *overall mean score* for language and maths based on score allocations as described in table 1. The mean score for language is 2.72. Looking at table 1, this means that the majority of students assessed could manage tasks ranging between reading words and reading a paragraph. Similarly, the mean score for maths is 4.75 meaning that the majority of students could manage tasks ranging between addition and subtraction. In both subjects, private school students do better and are able to handle higher level tasks relative to their counterparts in public schools.

In table 3, we report the means for child, family and village related variables used in the estimation. Panel A shows that 15 percent of grade two to grade four children in the survey attend private schools. We know that these students attend private schools based on the response to the question that asked about the type of school a child attends. The analysis in section 2 shows that private schools in Kenya are quite heterogeneous, ranging from highly fragmented non-formal schools (mainly located in informal settlements) and formal private academies (in middle and high-income urban areas). The manner in which the Uwezo survey was collected does not allow us to know whether the child attends a non-formal or formal private school.[13] The inability to account for such heterogeneities is likely to bias our estimates. As noted by Ashley et al. (2014), these are common challenges affecting analysis of the effectiveness of private schools in developing countries.

We report results for the whole sample, for the public sample and the private sample. In column 4, we show the mean differences between public and private samples. Since we are interested in the effect of private schools relative to public counterparts, our interest here lies mainly in column 2, 3 and 4. Panel A shows that relative to their counterparts from public schools, private school students are more likely to be young, female and with no disability. Similarly, private school students are also more likely to attend tuition classes.

Panel B and C of table 3 show that the mean for measured family and village background characteristics are substantially higher for students who attend private schools. For instance, private school learners are more likely to be born of young parents, with higher educational attainments relative to their counterparts in public schools. They also come from homes with a higher index of asset ownership[14] as well as homes that are more likely to have their own water source and toilet facility. In addition, they come from households that are more

[13] Our conversation with the Uwezo survey team however shows that majority of the private schools in the sample include non-formal and formal private academies. The sample does not include old traditionally exclusive private schools that offer foreign curriculum such as the General Certificate of Secondary Education (GCSE).

[14] We use the ordinary principal component analysis (PCA) to construct the index of household assets. The index is based on household ownership of the following assets: durables (TV, radio, car, computer, mobile phone, bicycle, motorbike and cart) and livestock (cattle, donkey, camel, sheep/goat).

likely to use electricity as a source of lighting (as opposed to sources such as paraffin) and whose dwelling place is made of bricks and stones (as opposed to materials such as mud and polythene).

It therefore appears that pupils who attend private schools already have disproportionately higher academic potential and access to complementary educational resources relative to their counterparts attending public schools. As noted by Altonji et al. (2005) and Goldberger and Cain (1982), this raises the possibility that part or even all of the gap in student test scores between private and public students as observed in table 2 column 2 and 3 may be a reflection of who attends private schools, thus indicating some level of selection. In other words private schools could be attracting students who are already advantaged given their home environments. This provides the motivation for this paper: to estimate the private school premium while addressing such potential selection into private schools based on the methods we have presented in the next section.

Table 3: Mean Statistics for Children, Households and Villages

	(1) Whole Sample	(2) Public School	(3) Private School	(4) Public-Private diff
Panel A: Student Characteristics				
Student attends private school	0.15	0.00	1.00	–
Age of student	9.31	9.41	8.55	0.86***
Student is female	0.48	0.48	0.49	-0.01**
Student has some disability	0.03	0.03	0.02	0.01**
Student goes for paid tuition	0.36	0.33	0.69	-0.36***
Panel B: Household Characteristics				
Age of the mother	35.34	35.57	34.09	1.48***
Education level of the mother				
None	0.24	0.26	0.09	0.17***
Has primary education level	0.54	0.56	0.47	0.09***
Has secondary education level	0.20	0.17	0.38	-0.22***
Has post-secondary education level	0.02	0.01	0.06	-0.05***
Education level of the father				
None	0.20	0.22	0.06	0.15***
Has primary education level	0.47	0.50	0.37	0.13***
Has secondary education level	0.29	0.27	0.47	-0.20***
Has post-secondary education level	0.04	0.02	0.10	-0.08***
Number of household members	5.87	5.95	5.32	0.63***
Household has source of water at home	0.21	0.17	0.39	-0.22***
Household has toilet/latrine at home	0.77	0.75	0.91	-0.15***
Index of household assets	0.13	-0.01	1.07	-1.09***
Meals taken per day				
Less than three meals	0.25	0.27	0.14	0.25***
Three meals	0.75	0.73	0.86	-0.12***
Wall material for dwelling place				
Mud	0.57	0.61	0.30	0.31***
Polythene and iron	0.08	0.08	0.09	-0.01***
Timber	0.11	0.11	0.12	-0.02***
Bricks and/or Stone	0.24	0.20	0.49	-0.28***
Regular source of lighting				
Paraffin	0.74	0.79	0.50	0.29***
Electricity	0.19	0.13	0.47	-0.34***
Other	0.07	0.08	0.03	0.05***
Time taken to reach at school	0.52	0.53	0.42	0.11***
Panel C: Village Characteristics				
Village has electricity	0.47	0.75	0.43	-0.32***
Village has tarmac road	0.21	0.41	0.18	-0.23***
Village has all-weather road	0.81	0.88	0.81	-0.07**
Village has a protected water point	0.43	0.49	0.42	-0.07
Village has chief's office	0.64	0.75	0.62	-0.13***
Village has police post	0.28	0.45	0.26	-0.19***
Village has an education committee	0.31	0.31	0.32	-0.01
Number of children	**52,709**	**44,754**	**6,343**	
Number of households	**20,180**	**16,659**	**2,993**	
Number of villages	**1,296**	**1,098**	**164**	

Source: Own calculations based on Uwezo 2012. Notes: (1) We do not include father's age because of large missing values; (2) We use the ordinary principal component analysis (PCA) to construct the index of household assets. The index is based on household ownership of the following assets: durable assets (TV, radio, car, computer, mobile phone, bicycle, motorbike and cart) and livestock assets (cattle, donkey, camel, sheep/goat); and (3) ***1 percent significance level, **5 percent significance level and *10 percent significance level.

5 Econometric Issues and Estimation Strategies

Our intention is to triangulate the effects of private schools on student achievement using methods that try to control for endogeneity of school choice by households. These include OLS, village and household FE models and propensity score matching. As mentioned, each of this method has its strengths and weakness. It is for this reason that we try multiple methods to get a sense of the range of the private school premium in Kenya. In this section, we lay out these methods.

5.0.1 Parametric Estimation: Ordinary Least Squares Method

We begin with the basic ordinary least squares technique. Here, the educational outcome of student i in household j located in village k, denoted as A_{ijk}, depends on the type of school attended, (where $PRIV_{ijk}$ is a dummy variable that equals 1 if the child attends a private and 0 if a child attends a public school) and the vector of: (a) individual characteristics, X_{ijk}, (b) family characteristics, ψ_{ijk}, (c) village characteristics, ϕ_{ijk}. ε_{ijk} is a random error with the mean 0 and variance σ^2. Formally,

$$A_{ijk} = \beta_0 + \beta_1 PRIV_{ijk} + \beta_2 X^{'}_{ijk} + \beta_3 \psi^{'}_{ijk} + \beta_4 \phi^{'}_{ijk} + \varepsilon_{ijk} \tag{1}$$

The main concern of this paper is the endogeneity of $PRIV_{ijk}$. A major assumption of OLS is that $PRIV_{ijk}$ is uncorrelated with ε_{ijk} conditional on X, ψ and ϕ. However, even after controlling for a comprehensive set of individual, family, village controls in the Uwezo survey data, it is possible that in equation (1), there are factors that determine student achievement but have been omitted, mis-measured or unobserved. If such factors are correlated with private school attendance, then the estimates of $PRIV_{ijk}$ will be biased either upwards or downwards. In what follows, we explore how to deal with this potential endogeneity of $PRIV_{ijk}$.

5.0.2 Village Fixed Effects Approach

The first step towards refining the estimates of $PRIV_{ijk}$ is controlling for any sources of observed and unobserved heterogeneity at the village level. In this regard, following Dostie and Jayaraman (2006), Javaid et al. (2012), Andrabi et al. (2008), French et al. (2010) and Mani et al. (2013), we estimate a village fixed effects model as outlined in equation (2):

$$A_{ijk} = \beta_0 + \beta_1 PRIV_{ijk} + \beta_2 X^{'}_{ijk} + \beta_3 \psi_{ijk} + \phi_k + \varepsilon_{ijk} \tag{2}$$

ϕ_k is the village fixed effects. Through the village fixed effects, we are able to (a) remove all sources of observed and unobserved heterogeneity at the village-level and (b) address cluster-related issues in the standard errors since common village-level unobservables are also cluster effects (Wooldridge 2003, Mani et al. 2013).

5.0.3 Household Fixed Effects Approach

Equation (2) does not address the potential observed and unobserved heterogeneity at the household level. To address this identification challenge, we follow Desai et al. (2008), Vegas and Devercelli (ND), De Haan et al. (2014) and French et al. (2010) by estimating a household/family fixed effects model shown in equation (3):

$$A_{ij} = \beta_0 + \beta_1 PRIV_{ij} + \beta_2 X^{'}_{ij} + \psi_j + \varepsilon_{ij} \tag{3}$$

ψ_j is the household fixed effects. Similarly, through the household fixed effects, we are able to remove all sources of observed and unobserved heterogeneity at the household level. Each household in our data is considered as a cluster since there are individual children within the household who are genetically linked. We cluster standard errors at the household level (Nichols and Schaffer 2007, Vegas and Devercelli ND, Maitra et al. 2016).

5.0.4 Assessing the Potential Bias from Unobservables

The OLS, village fixed effects and family fixed effects methods cannot deal with all sources of endogeneity. For instance, even after using the household fixed effects, within-household child-varying factors (such as child ability and/or motivation) [15] which play a role in influencing parental education decisions (Behrman et al. 1994, Andrabi et al. 2008, Maitra et al. 2016) are likely to remain in the error term, and may be correlated with attendance to private schools. Put differently, when estimating a household fixed effects model, the estimates might

[15] These factor influence parental decision to send a child to either private or public schools. However, they are not easy to observe or gathered in a typical survey like Uwezo.

be free from within-household child-specific factors (such as age, gender etc) and household unobserved heterogeneity. However, the possible presence of omitted variables mainly originating from within-household child-varying factors that are likely to be correlated with both private school attendance and student achievement means that we should be careful with interpreting our private school effects as being causal.

Acknowledging this limitation, we turn to a methodology proposed by Altonji et al. (2005) to assess the potential size of any bias on the estimated coefficient of the private school variable due to unobservable selectivity.[16] In their paper, Altonji et al. (2005) propose the idea *that selection on observables is the same as selection on unobservables,* which is equivalent to the condition that:

$$\frac{Cov(\varepsilon\,,\,PRIV)}{Var(\varepsilon)} = \frac{Cov(\beta_2 X\,,\,PRIV)}{Var(\beta_2 X)} \tag{4}$$

where X is a vector of all observable (child, household and village) characteristics, and ε is the error term potentially correlated with $PRIV$. Intuitively, we can assess the strength of the evidence of private school effect by checking how plausibly large the quantity on the left must be relative to the quantity on the right to explain the whole β_1 estimate in the OLS and fixed effects models under the null hypothesis that there is no private school effect (e.g. $\beta_1 = 0$).

The bias from OLS and the fixed effects models is $\frac{Cov(\varepsilon,\widetilde{PRIV})}{Var(\widetilde{PRIV})}$ and this is equivalent to $\frac{Cov(\varepsilon,PRIV)}{Var(\widetilde{PRIV})}$ if ε and X are orthogonal. The tildes denote the residuals from a regression of $PRIV$ on X. This bias can be assessed by the following equation.[17]

$$\frac{Cov(\varepsilon,PRIV)}{Var(\widetilde{PRIV})} = \frac{Cov(\varepsilon\,,\,PRIV)}{Cov(\beta_2 X\,,\,PRIV)}\frac{Var(\beta_2 X)}{Var(\varepsilon)}\frac{Cov(\beta_2 X\,,\,PRIV)}{Var(\beta_2 X)}\frac{Var(\varepsilon)}{Var(\widetilde{PRIV})} \tag{5}$$

$$\frac{Cov(\varepsilon,PRIV)}{Var(\widetilde{PRIV})} = \frac{Cov(\beta_2 X\,,\,PRIV)}{Var(\beta_2 X)}\frac{Var(\varepsilon)}{Var(\widetilde{PRIV})} \tag{6}$$

[16] This approach has been widely used in the context of problems similar to ours. Altonji et al. (2005) uses the procedure to study the effectiveness of Catholic schools in the USA. Others include Kingdon and Teal (2010) whose study examines the impact of teacher unionization on student achievement in India, Cavalcanti et al. (2010) who study the effect of private school attendance on public university entrance examinations in Brazil and most recently Simumba (2013) who investigates the effect of child labour on children's school attendance in rural Zambia.

[17] The bias is given by $plim\beta_1 = \beta_1 + \frac{Cov(\varepsilon,\widetilde{PRIV})}{Var(\widetilde{PRIV})}$ and it is positive as long as the variable $PRIV$ is not orthogonal to the error term ε.

where the first equality follows if ε and X are orthogonal and the second equality follows from the fact that $\frac{Cov(\varepsilon,PRIV)}{Var(\varepsilon)} = \frac{Cov(\beta_2 X\,,PRIV)}{Var(\beta_2 X)}$. When reporting our regression results for OLS and village and household FE models, we will also show the results from applying this method based based on equation (6).[18]

5.0.5 Non-Parametric Estimation: Propensity Score Matching Approach

We further address the endogeneity of $PRIV_{ijk}$ in equation (1) by use of a non-parametric matching approach (propensity score matching).[19] We are interested in estimating the effect of private schools on student achievement. This is a clear example of a treatment evaluation study whose main pillars are *individuals* (students), *treatment* (going to private school) and *outcomes* (the tests scores). In the case of a binary treatment like ours, the treatment indicator T can be defined as $T = 1$ if a student attends a private school and $T = 0$ if a student attends a public school. If we define Y_i as the outcome of the intervention, then $Y_i(T_i)$ is the potential outcome for student i where $i = 1, ...N$. Following Roy (1951) and Rubin (1974), the treatment effect for a student i can be written as:

$$\gamma_i = Y_i(1) - Y_i(0) \tag{7}$$

However, we cannot simultaneously observe Y_i when $Y_i = 1$ and $Y_i = 0$. Since we are interested in estimating the effect of private schools, we can only observe $Y_i(1)$ and not $Y_i(0)$. The latter is called a counterfactual. Propensity score matching overcomes the problem of lack of counterfactual by matching participants (students in private schools) to non-participants (students in public schools) who are observationally similar based on pre-treatment observable characteristics, which we denote as X. The assumption here is that conditional on observables, students in private and public schools do not differ systematically along unobservables.

[18]We are grateful to Prof. Todd Elder of Michigan State University for sharing the Stata routines for estimating the potential size of any bias on the estimated coefficient of the private school variable due to unobservable selectivity.

[19]A major drawback of OLS (including the village and family fixed effects) is that it imposes a linear form on the outcome equation. In our case, as observed by Vandenberghe and Robin (2004), the effect of private schools is taken to be uniform across the distribution of all the explanatory variables and adequately captured by the (constant) coefficient of a dummy variable. Such a linear restriction does not have a theoretical justification.

Propensity score matching is founded on two key assumptions: *conditional independence* and *common support* (or *overlap condition).* Conditional independence means that given a set of observable characteristics of X which are not affected by treatment, potential outcomes are independent of treatment assignment*(Rosenbaum and Rubin 1983)*. Formally:

$$Y_i(0), Y_i(1) \perp T_i|(X_i) \tag{8}$$

The *common support* ensures that for each value of observable characteristics X, there is a positive probability of being both in the treated and untreated groups:

$$0 < Pr(T_i = 1|(X_i)) < 1 \tag{9}$$

Given that conditional independence assumption holds and assuming that there is overlap between both groups, the PSM estimator for the treatment effect on the treated (ATT) can be written in general as:

$$\gamma_{ATT}^{PSM} = E(\triangle|p(X), T = 1) = E[(Y(1)|p(X),\, T = 1)] - E[(Y(0)|p(X),\, T = 0)] \tag{10}$$

To put this in words, the propensity score matching estimator is simply the mean difference in outcomes over the common support, appropriately weighted by the propensity score distribution of participants.

For estimation, we first estimate a model of school choice (private versus public) using a probit model[20]. We then use the *propensity score*[21] from the probit model to match the treated

[20]When estimating the propensity score, two choices have to be made. The first one concerns the model to be used for the estimation (which we use probit), and the second one the variables to be included in this model. We control for a comprehensive list of individual, household and village variables that influence household school choice discussed in section 4.

[21]Matching pupils directly can be computationally demanding especially when there are a large number of characteristics of X to control (curse of dimensionality). Rosenbaum and Rubin (1983) demonstrate that matching can be done on a single-index variable, *the propensity score,* which in our case is the probability of attending a private school given the observed characteristics X estimated through the probit model. The *propensity score* considerably reduces the dimensionality problem, as conditioning is on the basis of a scalar rather than a vector. The propensity score, however, must verify the *balancing property.* That is, the function used to compute the propensity score should be such that individuals with a similar propensity to attend

(private school students) and non-treated (public school students) who are observationally similar.

5.0.6 Sensitivity Analysis using Rosenbaum Bounds

As we have mentioned, students in private schools (treated group) are matched with those in public schools (control group) on the basis of observables. The underlying assumption is that conditional on observables, private and public school students do not differ systematically along unobservables. While our matching process is based on a rich set of observable child, family and community characteristics collected by Uwezo survey (see table A.2 in the appendix), we cannot rule out the presence of selection bias resulting from unobservables. There could be a possibility of selection bias into private schools due to unobserved factors and this could happen even in cases where students in the treated group have been well matched to those in the control group. Rosenbaum (2002) has developed a procedure, known as the Rosenbaum bounds, that allows us to assess the sensitivity of our results to the selection on unobservables.[22]

We briefly explain the Rosenbaum bounds procedure drawing on a number of studies (Aakvik 2001, Bharath et al. 2011, Azam et al. 2015, Becker and Caliendo 2007) that have applied it. First, assume that student's i probability of receiving treatment (attending a private school) is given by equation (11):

$$P_i = P(X_i\,,\,u_i) = P(D_i = 1|X_i\,,\,u_i) = F(\beta X_i + \gamma u_i) \tag{11}$$

where X_i are student$i$'s observed characteristics, u_i is the unobserved variable, and γ is the effect of u_i on private school enrolment. If our treatment estimates are free from hidden bias, γ will be zero and the private school participation probability will be determined solely by X_i. However, if there is hidden bias, then it is possible that two students with the same observed characteristics X, have different chances of receiving treatment (attending private school).

Suppose we have a matched pair of students i and j, and suppose F in equation (11) follows a logistic distribution. If both students have identical observed characteristics—as implied

a private school display, on average, similar values of X_i (Vandenberghe and Robin 2004, Caliendo and Kopeinig 2008, Azam et al. 2015)

[22]This strategy is similar to the one used by Altonji et al. (2005) in the case of OLS and fixed effects models.

by the matching procedure, the odds ratio that the two students receive treatment is then given by:

$$\frac{\frac{P_i}{1-P_i}}{\frac{P_j}{1-P_j}} = \frac{P_i(1-P_i)}{P_j(1-P_j)} = \frac{e^{(\beta X_i + \gamma u_i)}}{e^{(\beta X_j + \gamma u_j)}} = e^{\gamma(u_i - u_j)} \tag{12}$$

As can be seen in equation (12), both students differ in their odds of receiving treatment by a factor that involves the parameter γ and the difference in their unobserved characteristics u. So, if there are either no differences in unobserved variables $(u_i - u_j)$ or if unobserved variables have no influence on the probability of participating $(\gamma = 0)$, the odds ratio is one, implying the absence of hidden or unobserved selection bias. Sensitivity analysis now evaluates how changing the values of γ and $(u_i - u_j)$ alters inference about the effect of private schools.

Following Rosenbaum (2002), Aakvik (2001) and Becker and Caliendo (2007), we assume for simplicity that the unobserved covariate is a dummy variable with $u_i \in 0, 1$. Rosenbaum (2002) shows that equation (12) implies the following bounds on the odds ratio that either of the two matched individuals will receive treatment:

$$\frac{1}{e^{\gamma}} \leq \frac{P_i(1-P_i)}{P_j(1-P_j)} \leq e^{\gamma} \tag{13}$$

$\Gamma = e^{\gamma}$ is a measure of the degree of departure from the case that is free of hidden bias (Rosenbaum 2002). Both matched students i and j have the same probability of participating only if $\Gamma = e^{\gamma} = 1$ and in this case the model will be free of hidden bias. Otherwise, if for example $\Gamma = e^{\gamma} = 2$, students who appear to be similar (in terms of X) could differ in their odds of receiving the treatment by as much as a factor of 2.

6 Results

6.1 OLS and Fixed Effects Estimates of Private Schools Effects

To put our results into perspective, we first present the OLS regression (equation (1)), as a baseline model, followed by the fixed effects regression. Model 1 of table 4 and table 5 presents estimated OLS results for maths and language respectively. We can see that after accounting for students, household and village characteristics, the results from OLS show that attending a private school relative to a public school is associated, on average, with an increase in maths and language scores of 0.18 and 0.27 score standard deviations respectively.[23]

We are careful not to give the OLS estimates a causal interpretation due to possible endogeneity of private school choice. We begin to address this challenge by estimating a village fixed effects model. The village fixed effects model helps us to remove all sources of village level unobserved and observed heterogeneity common to all children in the village. Through the village fixed effects, we are also able to address cluster-related issues in the standard errors since common village-level unobservables are also cluster effects (Wooldridge 2003, Mani et al. 2013). Results are shown in model 2 of table 4 and table 5. We still find a relatively large private school effect, in fact slightly higher than estimates based on OLS. The estimated private school coefficient means that attending a private school relative to a public school is associated, on average, with an increase in maths and language scores of 0.20 and 0.29 standard deviations respectively, an indication that OLS estimates are perhaps biased downwards.

The village fixed effects do not however control for the observed and unobserved heterogeneity at the household level. We explore this through the household fixed effects model which controls for household-level observed and unobserved factors, allowing us to focus on within-household child-specific factors such as gender, age and grade. Model 3 of table 4 and table 5 presents the results of the family fixed effects model. As can be seen from the table, after controlling for family-level observed and unobserved factors, we observe a reduction in the size of the private school effects. Nevertheless, there is still a sizable and significant positive effect of private schools on student language and maths test scores. Specifically, we find a private school advantage of 0.13 and 0.21 score standard deviations in maths and language respectively.

As we noted, the OLS and FE methods cannot deal with all sources of endogeneity. For instance, the household fixed effects are unlikely to control for the *within-household child-varying factors* which are likely to remain in the error term and may be corrected with attendance to private schools. In the last two rows of table 4 and table 5, we report the estimated bias using Altonji et al. (2005)'s procedure based on equation (6). This procedure estimates the potential size of any bias on the estimated coefficient of the private school variable due to unobservable selectivity (see also Cavalcanti et al. (2010) and Kingdon and

[23] In the interest of brevity, we do not interpret the coefficients for the child, household and village controls in the three models of table 4 and table 5. However, as can be seen from the two tables, these characteristics exert a significant influence on student achievement as hypothesized.

Teal (2010)).

Like Kingdon and Teal (2010), where the *sign of the private school effect* and the *bias* are *identical*, it means that we have overestimated the true effect.[24] Looking at table 4 and table 5, the estimated coefficient in the OLS regression is 0.18 and 0.27 for maths and language respectively. The estimated bias is positive and it is 0.48 and 0.74 for maths and language respectively, which suggests that the OLS regression overestimates the true effect of private schooling on test score. We find similar evidence in the village fixed effects model too. We can also see that there is a substantial reduction in the size of the bias as we move from OLS to the fixed effects models.

Since the true effect has been over-estimated in the OLS and village fixed effects, we can find the *size of the unobservables necessary to explain away the implied effect from private school attendance.* Following Altonji et al. (2005), we calculate this by dividing the estimated coefficient by the bias (that is the ratio of the estimated coefficient to the bias). For instance, in the case of the OLS results for maths (table 4), the ratio of the estimated private school effect coefficient to the bias is 0.38. This implies that the role of unobservables that determine student maths test scores would have to be more than 0.38 times the role of observables for the entire private school effect to be explained away by the unobservables. For language, it is 0.36.

[24] Recall that $E[\beta|X] = \hat{\beta} + bias$.

Table 4: Effect of Private Schools on Student Test Scores in Maths

	Model (1)		Model (2)		Model (3)	
	No Fixed Effects (OLS)		Village Fixed Effects		Household Fixed Effects	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Student attends private school	0.18***	(0.01)	0.20***	(0.02)	0.13***	(0.03)
Student Characteristics						
Age of student	0.09***	(0.02)	0.08***	(0.02)	0.16***	(0.02)
Age of student squared	-0.00***	(0.00)	-0.00***	(0.00)	-0.00***	(0.00)
Student is female	0.01	(0.01)	0.01*	(0.01)	0.02**	(0.01)
Student has some disability	-0.15***	(0.03)	-0.12***	(0.03)	-0.15***	(0.04)
Student goes for paid tuition	0.10***	(0.01)	0.10***	(0.01)	0.09***	(0.02)
Student's current grade	0.41***	(0.01)	0.41***	(0.01)	0.31***	(0.01)
Household Characteristics						
Age of the mother	0.01**	(0.00)	0.00	(0.00)		
Age of the mother squared	-0.00*	(0.00)	-0.00	(0.00)		
Education level of the mother (ref: None)						
Has primary education level	0.04**	(0.02)	0.04**	(0.02)		
Has secondary education level	0.10***	(0.02)	0.10***	(0.02)		
Has post-secondary education level	0.15***	(0.05)	0.10**	(0.05)		
Education level of the father (ref: None)						
Has primary education level	-0.01	(0.02)	0.01	(0.02)		
Has secondary education level	0.03*	(0.02)	0.06***	(0.02)		
Has post-secondary education level	0.08**	(0.03)	0.13***	(0.03)		
Number of household members	-0.01***	(0.00)	-0.00**	(0.00)		
Household has source of water at home	0.07***	(0.01)	0.04***	(0.02)		
Household has toilet/latrine at home	0.02	(0.01)	0.03*	(0.02)		
Index of household assets	0.001	(0.00)	0.02*	(0.01)		
Meals taken per day(ref: less than three)						
Three meals	0.05***	(0.01)	0.02*	(0.01)		
Wall material for dwelling place (ref: Mud)						
Polythene and iron	0.06***	(0.02)	-0.00	(0.02)		
Timber	0.11***	(0.02)	0.01	(0.02)		
Bricks and/or Stone	0.12***	(0.01)	0.03	(0.02)		
Regular source of lighting (ref: Paraffin)						
Electricity	0.07***	(0.01)	0.03**	(0.02)		
Other	0.02	(0.02)	-0.04	(0.03)		
Time taken to reach at school	-0.00	(0.01)	-0.01	(0.01)		
Village Characteristics						
Village has electricity	-0.01	(0.01)				
Village has tarmac road	0.00	(0.01)				
Village has all-weather road	-0.05***	(0.01)				
Village has a protected water point	0.01	(0.01)				
Village has chief's office	0.01	(0.01)				
Village has police post	-0.03***	(0.01)				
Village has an education committee	0.01	(0.01)				
Constant	-2.08***	(0.10)	-1.99***	(0.10)	-2.06***	(0.09)
Observations (number of students)	32,689		32,689		32,689	
R Squared	0.23		0.26		0.42	
Number of villages	3,876		3,876		3,876	
Number of households	27,970		27,970		27,970	
Estimated bias based on equation (6)	0.48		0.32		-0.10	
Ratio	0.38		0.63		–	

Notes: (1) Student goes for paid tuition means student attends classes offered beyond the normal scheduled school time; (2) Student's current grade is entered as a continuous variable. Results do not change even when we enter student's current grade as a categorical variable; (3) We do not control for father's age because of large number of missing values; (4) We use the ordinary principal component analysis (PCA) to construct the index of household assets. The index is based on household ownership of the following assets: durable assets (TV, radio, car, computer, mobile phone, bicycle, motorbike and cart) and livestock assets (cattle, donkey, camel, sheep/goat); (5) Standard errors are in parenthesis; (6) In the OLS and household fixed effects models, standard errors are clustered at household level. In the village fixed effects model, standard errors are clustered at the village level; (7) Ratio (on the last row) is defined as the ratio of the coefficient on private school and the estimated bias based on equation (6). (8) ***1% significance level, **5% significance level and *10% significance level.

Table 5: Effect of Private Schools on Student Test Scores in Language

	Model (1)		Model (2)		Model (3)	
	No Fixed Effects (OLS)		Village Fixed Effects		Household Fixed Effects	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Student attends private school	0.27***	(0.02)	0.29***	(0.02)	0.21***	(0.04)
Student Characteristics						
Age of student	0.00	(0.02)	0.01	(0.02)	0.06***	(0.02)
Age of student squared	0.00	(0.00)	0.00	(0.00)	-0.00	(0.00)
Student is female	0.04***	(0.01)	0.05***	(0.01)	0.05***	(0.01)
Student has some disability	-0.14***	(0.03)	-0.09***	(0.03)	-0.12***	(0.03)
Student goes for paid tuition	0.13***	(0.01)	0.12***	(0.01)	0.11***	(0.02)
Student's current grade	0.43***	(0.01)	0.42***	(0.01)	0.38***	(0.01)
Household Characteristics						
Age of the mother	0.01***	(0.00)	0.01**	(0.00)		
Age of the mother squared	-0.00***	(0.00)	-0.00	(0.00)		
Education level of the mother (ref: None)						
Has primary education level	0.00	(0.02)	0.05***	(0.02)		
Has secondary education level	0.10***	(0.02)	0.13***	(0.02)		
Has post-secondary education level	0.18***	(0.05)	0.20***	(0.05)		
Education level of the father (ref: None)						
Has primary education level	-0.08***	(0.02)	-0.00	(0.02)		
Has secondary education level	0.01	(0.02)	0.08***	(0.02)		
Has post-secondary education level	0.05	(0.03)	0.12***	(0.03)		
Number of household members	-0.01***	(0.00)	-0.01***	(0.00)		
Household has source of water at home	0.07***	(0.01)	0.02	(0.02)		
Household has toilet/latrine at home	0.01	(0.01)	0.04**	(0.02)		
Index of household assets	0.02	(0.01)	0.03*	(0.02)		
Meals taken per day(ref: less than three)						
Three meals	0.02*	(0.01)	0.04**	(0.01)		
Wall material for dwelling place (ref: Mud)						
Polythene and iron	0.13***	(0.02)	0.02	(0.02)		
Timber	0.16***	(0.02)	0.07***	(0.02)		
Bricks and/or Stone	0.13***	(0.01)	0.07***	(0.02)		
Regular source of lighting (ref: Paraffin)						
Electricity	0.13***	(0.02)	0.05***	(0.02)		
Other	0.09***	(0.02)	-0.04	(0.03)		
Time taken to reach at school	0.01	(0.01)	-0.00	(0.01)		
Village Characteristics						
Village has electricity	0.04***	(0.01)				
Village has tarmac road	0.05***	(0.01)				
Village has all-weather road	-0.03**	(0.01)				
Village has a protected water point	0.02**	(0.01)				
Village has chief's office	0.00	(0.01)				
Village has police post	0.02	(0.01)				
Village has an education committee	0.02	(0.01)				
Constant	-1.92***	(0.10)	-2.01***	(0.10)	-1.92***	(0.08)
Observations (number of students)	32,689		32,689		32,689	
R Squared	0.27		0.28		0.47	
Number of villages	3,876		3,876		3,876	
Number of households	27,970		27,970		27,970	
Estimated bias based on equation (6)	0.74		0.48		-0.01	
Ratio	0.36		0.60		–	

Notes: (1) Student goes for paid tuition means student attends classes offered beyond the normal scheduled school time; (2) Student's current grade is entered as a continuous variable. Results do not change even when we enter student's current grade as a categorical variable; (3) We do not control for father's age because of large number of missing values; (4) We use the ordinary principal component analysis (PCA) to construct the index of household assets. The index is based on household ownership of the following assets: durable assets (TV, radio, car, computer, mobile phone, bicycle, motorbike and cart) and livestock assets (cattle, donkey, camel, sheep/goat); (5) Standard errors are in parenthesis; (6) In the OLS and household fixed effects models, standard errors are clustered at household level. In the village fixed effects model, standard errors are clustered at the village level; (7) Ratio (on the last row) is defined as the ratio of the coefficient on private school and the estimated bias based on equation (6). (8) ***1% significance level, **5% significance level and *10% significance level.

On the other hand, where the *sign of the private school effec*t and the *bias* are *not identical*, it means that we have underestimated the true effect. For instance, the estimated coefficient in the household fixed effects is 0.13 and 0.21 for maths and language respectively. The estimated bias is however negative. It is equal to -0.10 and -0.01 for maths and language respectively, suggesting that the household fixed effects regression underestimates the true effect of private schooling on test score. Since we have under-estimated the true private school effect, there is no need to calculate the *size of the unobservables necessary to explain away the implied effect from private school attendance.*

In summary, we learn the following from the application of Altonji et al. (2005)'s procedure. First, there is evidence of a bias on the estimated coefficient of the private school variable due to unobservable selectivity. This bias however reduces, substantially, as we move from OLS to village fixed effects and to household fixed effects. In the OLS and village fixed effects, the magnitude of bias, as compared to the coefficient, is quite large, suggesting that the OLS and the village fixed effects models may suffer from the problem of selection due to unobservables and hence they are not reliable for obtaining the true effect of private schools on test score. Since the private schooling decision is made at the household level, it is likely that a substantial part of the unobservable component is pertaining to the household. Consistent with this view, when we estimate a household fixed effects model, we find that the estimated coefficient is smaller in magnitude. Moreover, the estimated bias in household fixed effects model is negative and also smaller in magnitude. This in fact indicates that unlike the previous models, household fixed effects model yields a coefficient that is an underestimate of the true effect of private school. Therefore from this analysis we find that household fixed effects regression is a more reliable model that closely captures the true effect of private schooling on test scores.

6.2 Propensity Score Matching Estimates of Private Schools Effects

We now turn our attention to the estimates based on the propensity score matching approach. In the appendix, we show different indicators characterizing the level of success of the matching process. Table A.1 in the appendix shows results for the probit model[25] that produced the propensity scores which were in turn used for matching the treated and non-treated participants. Figure A.1 in the appendix shows considerable overlap between treated (private school students) group and control (public school students) group.

[25] The probit model of school choice (private versus public) is specified as follows: $Prob(Pi = 1) = F(Z\prime b)$.

Table A.2 in the appendix shows that matching balanced quite well among all the variables affecting household school choice. In column 9 and column 10 of table A.2, we report the t-tests and p-values of equality between treated (private school students) group and control (public school student) group after matching for each variable. Looking at the t-tests and the p-values, we find that almost all the values are insignificant showing a very successful level of matching. The table further shows that the matching process significantly reduces the standardized bias (SB). A standardized bias of below 5 percent after matching is widely acceptable in most empirical studies (Caliendo and Kopeinig 2008). In our case, the standardized bias, shown in column 8, for all the variables, is below 5 percent. A physical inspection of figure A.2 and A.3 further confirms that indeed matching significantly reduces standardized bias (SB).

Since a number of matching algorithms are considered (nearest neighbor matching, kernel matching, radius matching and caliper matching),[26] the match quality across these algorithms deserves attention. In this regard, table 6 provides information related to the quality of matching for the different matching algorithms based on the following indicators: pseudo R2, LR chi2 values, p-value, mean bias and variance among others. Generally, all the algorithms perform exceptionally well in terms of matching as shown by zero level of pseudo R2 and very high levels of p-values, which in most cases equal to one, after matching.[27]

Table 7 shows comparable private school effects from the different matching algorithms.[28] Estimates from different matching algorithms are within the same range. In maths, we find

[26]For a detailed exposition of these matching algorithms, we refer readers to Caliendo and Kopeinig (2008), Rosenbaum (2002) and Dehejia and Wahba (2002).

[27]Nevertheless, the nearest neighbor and caliper matching algorithms are characterized by relatively low levels of p-value.

[28]In each matching algorithm, we impose a caliper width of 0.00001 which is much lower than what most studies use. Imposing caliper widths helps avoid bad matches and hence helps to further deal with selection issues (Caliendo and Kopeinig 2008). While there is no empirical evidence on the optimal caliper width (Austin 2008), the choice of caliper width depends on the extent of variance-bias trade off the researcher wants to achieve (Heckman et al. 1997, Faries et al. 2010) and the sample size (Smith 2000). Low levels of caliper width result in the matching of more similar subjects, leading to improved comparability of groups which translates to less biased estimates. However, it may also result in the formation of fewer matched pairs, thus decreasing the precision, due to high variance, of the estimated treatment effects. In many studies, including those in education, researchers use calipers of pre-determined width (ad hoc) that are generally independent of the distribution of the propensity score. In medical literature, caliper width of 0.6 and 0.2 of the logit of the standard deviation of the logit of the propensity is predominantly used (Austin 2008).

Table 6: Propensity Score Matching Quality Test (Before and After Matching)

	Nearest Neighbor Matching ($\sigma = 0.00001$)		4-Nearest Neighbor Matching ($\sigma = 0.00001$)		5-Nearest Neighbor Matching ($\sigma = 0.00001$)		Kernel Matching ($\sigma = 0.00001$)		Radius Matching ($\sigma = 0.00001$)		Caliper Matching ($\sigma = 0.00001$)	
	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After
Pseudo R-squared	0.21	0.01	0.21	0.01	0.21	0.00	0.20	0.01	0.21	0.00	0.21	0.01
LR ch2	4466.0	49.23	4466.0	33.36	4466.0	32.06	4259.4	12.84	4466.0	32.20	4466.0	49.23
p-values	0.00	0.31	0.00	0.95	0.00	0.99	0.00	0.94	0.00	0.92	0.00	0.31
Mean bias	28.2	2.9	28.2	2.4	28.2	1.4	29.7	1.6	28.2	2.4	28.2	2.9
Medbias	26.9	2.2	26.9	1.8	26.9	1.4	24.5	1.4	26.9	2.0	26.9	2.2
ASB	129.3	25.9	129.3	21.4	129.3	20.9	126.5	8.7	129.3	21.0	129.3	25.9
R	1.26	1.35	1.26	1.29	1.26	1.29	1.22	1.14	1.26	1.03	1.26	1.35
Variance	80	30	80	20	80	30	83	0	80	30	80	30

Notes: (1) *'Before'* means before matching. *'After'* means after matching; (2) We include the following variables in the probit regression when estimating the propensity score: child variables (gender, age, aged squared, whether the child attends for extra-tuition classes and child's grade), mother's age and mother's age squared, mothers education, father's education, household size, dummy variable for whether the household has a toilet, dummy variable for whether the household has a water facility, time taken to reach school, dummy variable for whether the household owns the following assets (TV, radio, car, computer, mobile phone, bicycle, motorbike, cart, cattle, donkey, camel, sheep/goat), number of meals per day, type of household dwelling unit, regular source of lighting for the household and village characteristics (whether a village has chief's camp, a shopping center, electricity connection, tarmac road, all-weather road, education committee, protected water point and whether a village is rural); (3) The probit results are shown in table A.1 in the appendix; (4) Standard errors are in parentheses and; (5) ***1% significance level, **5% significance level and *10% significance level.

Table 7: Effect of Private Schools on Student Test Scores

	(1)	(2)	(4)	(5)	(6)	(7)
	Nearest Neighbor Matching ($\sigma = 0.00001$)	4-Nearest Neighbor Matching ($\sigma = 0.00001$)	5-Nearest Neighbor Matching ($\sigma = 0.00001$)	Kernel Matching ($\sigma = 0.00001$)	Radius Matching ($\sigma = 0.00001$)	Caliper Matching ($\sigma = 0.00001$)
Maths	0.14***	0.15***	0.15***	0.17***	0.15***	0.15***
	(0.032)	(0.029)	(0.029)	(0.017)	(0.029)	(0.033)
Language	0.27***	0.26***	0.26***	0.26***	0.26***	0.27***
	(0.032)	(0.030)	(0.029)	(0.018)	(0.029)	(0.032)
Untreated	26,598	26,598	26,598	26,598	26,598	26,598
Treated	3,428	3,428	3,428	3,428	3,428	3,428
Total	30,026	30,026	30,026	30,026	30,026	30,026

Notes: (1) We include the following variables in the probit regression when estimating the propensity score: child variables (gender, age, aged squared, whether the child attends for extra-tuition classes and child's grade), mother's age and mother's age squared, mothers education, father's education, household size, dummy variable for whether the household has a toilet, dummy variable for whether the household has a water facility, time taken to reach school, dummy variable for whether the household owns the following assets (TV, radio, car, computer, mobile phone, bicycle, motorbike, cart, cattle, donkey, camel, sheep/goat), number of meals per day, type of household dwelling unit, regular source of lighting for the household and village characteristics (whether a village has chief's camp, a shopping center, electricity connection, tarmac road, all-weather road, education committee, protected water point and whether a village is rural); (2) The probit results are shown in table A.1 in the appendix; (3) Untreated are the number of students in the untreated group (number of public school students). Treated are the number of treated students that are on the common support; (4) Standard errors are in parentheses and; (5) ***1% significance level, **5% significance level and *10% significance level.

a private school advantage ranging from 0.14 score standard deviations based on nearest neighbor matching (with replacement) to 0.17 score standard deviations based on Kernel matching. In language, we find a private school advantage between 0.26 score standard deviations based on almost all matching algorithms to 0.27 based on neighbor matching (with replacement) and caliper matching.

Lastly, we assess the extent to which estimates from propensity score matching are influenced by hidden (unobserved) bias. As discussed in section 5.0.6, if there are unobservables that influence both student assignment into private schools and student test score performance, then a bias might arise and this is likely to undermine the robustness of the estimates we present in table 7. Similar to the procedure by Altonji et al. (2005), our interest here is to determine how strongly an unmeasured bias must influence the selection process to undermine the implications of the estimates (Azam et al. 2015, Rosenbaum and Rubin 1983, Rosenbaum 2002). If there is a positive selection on unobservables, our estimated private school effects overestimate the true effects (Azam et al. 2015, Rosenbaum and Rubin 1983) because students who perform better are likely to sort into private schools.

In table 8, we show the upper bound on the p-value of the null hypothesis of no private school effect for different levels of Γ (gamma). Recall that our estimates in table 7 show that private schools are positively correlated with both language and maths student test scores. In this case, the assumption that we have under-estimated the true private school effects (e.g lower bound sig-) does not apply and therefore, as shown in table 8, we only consider the upper bound sig+ (p-value). If the p-values remain significant (for instance, less than 0.1) for reasonably large values of gamma, then our private school effects are robust to hidden unobservables (Azam et al. 2015).

The results in table 8 show that the positive effect of private schools on maths test scores withstands even at relatively high values of selection on unobservables. In other words, we can reject the null hypothesis of zero private school effect on maths test scores even in the presence of relatively high values of selection on unobservables. In the case of language, the effect of private schools disappears even under low levels of unobserved selectivity. For instance, we fail to reject the null hypothesis of zero private school effect on language test scores even with observationally similar students who differ in their relative odds of attending private school by a factor of 1.2 (see Rosenbaum and Rubin (1983) for a detailed theoretical discussion and Azam et al. (2015) for empirical application).

Table 8: Sensitivity Analysis of PSM estimates, Rosenbaum Bounds

	Maths	Language
Gamma (Γ)	upper bound sig+ (p-value)	upper bound sig+ (p-value)
1	0.0000	0.0000
1.1	0.0000	0.0000
1.2	0.0000	1.0000
1.3	0.0000	1.0000
1.4	0.0000	1.0000
1.5	0.0000	1.0000

6.3 Comparison of Estimates from the Different Estimation Approaches

Finally, we bring together estimates based on different methods as shown in table 9. For propensity score matching, we choose the 5-Nearest Neighbor Matching because this matching algorithm performs best relative to others on the basis of the matching quality indicators presented in table 6. Each of this method has its strengths and weakness. It is for this reason that we try multiple methods to get a sense of the range of the private school premium in Kenya. As it can be seen in table 9, we see that in maths, the premium ranges from 0.13 to 0.20 score standard deviation, using the household and village fixed effects model, respectively. In the case of language, it ranges from 0.20 to 0.29 score standard deviation, using the household and village fixed effects model, respectively.

Results from the Altonji et al. (2005) procedure show that the magnitude of bias, as compared to the private school coefficient, is quite large in the OLS and village fixed effects suggesting that these models may suffer from the problem of selection due to unobservables. The OLS and village fixed effects in fact overestimate the true private school effect. The propensity score matching procedure is based on an assumption that conditional on observables, private and public school students do not differ systematically along unobservables. Sensitivity analysis based on Rosenbaum (2002) however shows that the estimates based on propensity score matching, especially for language, are not free from the hidden bias and therefore do not entirely meet this assumption. Perhaps this explains why the estimates based on propensity score matching are quite close to those in OLS.

Table 9: Effect of Private School on Student Test Scores: Different Estimation Approaches

	(1) OLS Model	(2) Village Fixed Effects Model	(3) Household Fixed Effects Model	(4) PSM 5-Nearest Neighbor
Maths	0.18***	0.20***	0.13***	0.15***
	(0.01)	(0.02)	(0.03)	(0.020)
Language	0.27***	0.29***	0.21***	0.26***
	(0.02)	(0.02)	(0.04)	(0.021)
Observations	27,970	27,970	27,970	29,209

Notes: Standard errors are in parentheses, ***1% significance level, **5% significance level and *10% significance level.

Since private school choice takes place at the household level, it is likely that a substantial part of the unobservables are accounted for by the household fixed effects model. It is for this reason that the household fixed effects model yields smaller coefficients of the private school effect. Also, the estimated bias based on the household fixed effects models based on the Altonji et al. (2005) procedure are negative and also smaller in magnitude (see table 4 and table 5), indicating that unlike the other models, household fixed effects models yield coefficients that are an underestimate of the true effect of private school.

Since the OLS and village fixed effects overestimate the private school premium while the household fixed effects underestimate it, the true private school premium lies within the range we provided in table 9. However, for reasons we have already provided, the household fixed effects regressions are more reliable and closely capture the true effect of private schooling on student learning.

7 Conclusion

In this paper, we use a rich household survey data to quantify the relative contribution of private schools on cognitive achievement of lower primary (grade two to four) school children in Kenya. One of the challenges facing a study like ours is the extent to which the researcher is able to account for the endogeneity of school choice. Dealing with such a challenge depends on the quality and nature of data. In this study, we are fortunate to use the Uwezo household survey that enables use to use a number of estimation techniques that account for the endogeneity of school choice. Results from Altonji et al. (2005) shows that the size of the bias on the estimated coefficient of the private school variable due to unobservable selectivity, especially based on the household fixed effects model, is quite small. This therefore makes one confident that the range of the size of the private school premium we present could be close to the true effect.

Analyzing the effectiveness of private schools in Kenya comes with a number challenges. As discussed in section 2, private schools are quite heterogeneous. They mainly comprise highly fragmented non-formal schools (mainly located in informal settlements) and formal private academies (in middle and high-income urban areas). Heterogeneity is also reflected in the mode of funding. We have government supported private and self-reliant private schools. In India, a study by Chakrabarti and Peterson (2008) finds that the effects of private aided schools are different from those of private unaided schools. Unfortunately, the manner in which our data was collected does not allow us to account for this heterogeneity. Our estimates should therefore be interpreted in the context of these limitation.

8 Appendix

Table A.1: Probit Results for Calibrating Propensity Score

	Coefficient	Standard Error
Student Characteristics		
Age	-0.16***	(0.04)
Age squared	0.01**	(0.00)
Is female	-0.01	(0.02)
Has some disability	-0.11	(0.07)
Goes for paid tuition	0.74***	(0.02)
Current grade	-0.06***	(0.02)
Household Characteristics		
Mother's age	-0.03***	(0.01)
Mother's age squared	0.00**	(0.00)
Mother's Education level(ref: None)		
Has primary education level	-0.05	(0.04)
Has secondary education level	0.12***	(0.05)
Has post secondary education level	0.33***	(0.10)
Father's Education level(ref: None)		
Has primary education level	-0.05	(0.05)
Has secondary education level	0.11**	(0.05)
Has post secondary education level	0.24***	(0.07)
Household has less than 10 members	-0.02***	(0.00)
Household has source of water at home	0.13***	(0.03)
Household has toilet/latrine at home	0.04	(0.04)
Distance to school is less than 30 minutes	0.08***	(0.02)
Household assets		
Household has a TV	0.21***	(0.03)
Household has a radio	0.03	(0.03)
Household has a computer	0.31***	(0.07)
Household has a phone	0.13***	(0.03)
Household has a car	0.29***	(0.05)
Household has a cattle	0.09***	(0.02)
Household has a donkey	0.05	(0.04)
Household has a camel	-0.04	(0.08)
Household has a goat	-0.09***	(0.02)
Household has a bicycle	-0.07***	(0.02)
Household has a motorbike	0.09**	(0.04)
Household has a cart	0.01	(0.06)
Number of Meals taken per day(ref: Three)		
One meals	0.07	(0.07)
Two meals	-0.13***	(0.03)
Wall material for dwelling place (ref: Bricks_stone)		
Polythene and iron	0.01	(0.04)
Timber	-0.08**	(0.04)
Mud	-0.14***	(0.03)
Regular source of lighting (ref: Other)		
Electricity	0.40***	(0.07)
Paraffin	0.07	(0.06)
Village Characteristics		
Village has chief's office	-0.00	(0.02)
Village has shopping center	0.15***	(0.03)
Village has electricity	0.03	(0.03)
Village has tarmac road	0.08***	(0.03)
Village has all-weather road	-0.04	(0.03)
Village has an education committee	-0.11***	(0.02)
Village has all protected water point	-0.01	(0.02)
Village is rural	-0.14***	(0.03)
Constant	0.16	(0.25)
Observations	30,299	

Notes: (1) Standard errors in parenthesis clustered at the household level and; (2) ***1% significance level, **5% significance level and *10% significance level.

Figure A.1: Propensity Score of Observations in and off Common Support Region

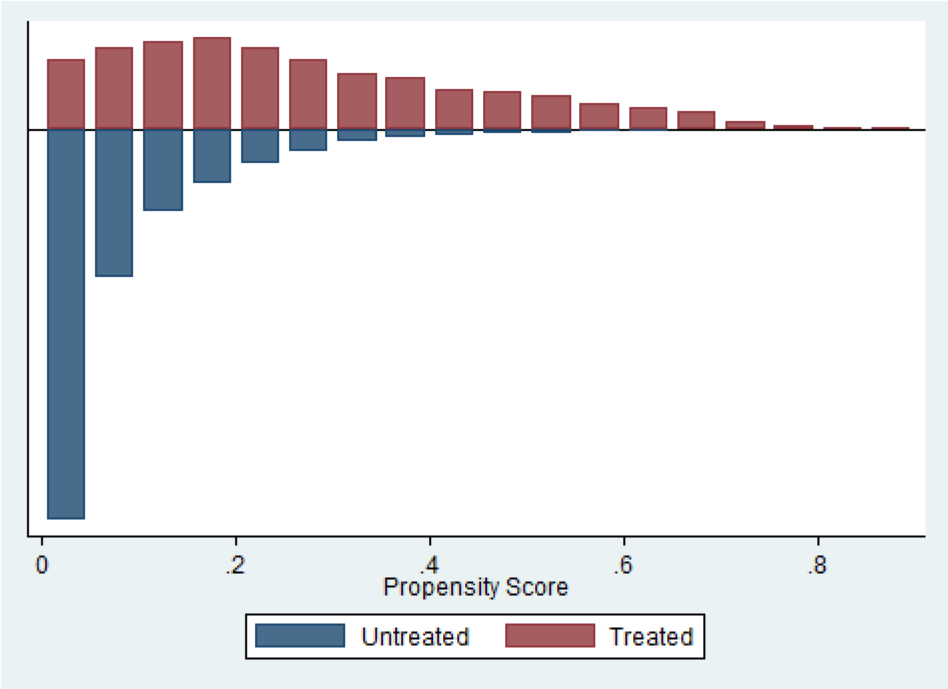


Table A.2: Testing of Balance between Private and Public Students after PSM Estimation

	Unmatched					Matched					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Treated	Control	%bias	t-test	p>|t|	Treated	Control	%bias	t-test	p>|t|	%reduct
Student Characteristics											
Student's Age	8.54	9.41	-45.2	-23.72	0.00	8.74	8.74	0.7	0.30	0.76	98.3
Student's Age squared	76.12	92.78	-44.5	-22.9	0.00	93.41	79.40	0.9	0.36	0.72	98.0
Student is female	0.49	0.47	3.8	2.10	0.04	0.49	0.48	1.3	0.49	0.63	65.4
Student has some disability	0.02	0.03	-5.8	-3.03	0.00	0.02	0.02	1.8	0.73	0.47	68.9
Students goes for paid tuition	0.72	0.34	83.6	45.28	0.00	0.68	0.68	0.1	0.05	0.96	99.8
Student grade	2.90	2.99	-11.5	-6.32	0.00	2.93	2.93	-0.9	-0.33	0.74	92.2
Mother age	33.11	35.61	-18.6	-9.91	0.00	34.29	34.34	-0.8	-0.31	0.76	95.5
Family Characteristics											
Mother's Education level											
Has no education	0.10	0.25	-38.9	-19.12	0.00	0.12	0.12	-1.0	-0.41	0.68	97.5
Has primary education	0.50	0.59	-18.4	-10.26	0.00	0.54	0.53	2.5	0.92	0.36	86.4
Has secondary education	0.36	0.16	47.5	29.40	0.00	0.32	0.33	-1.4	-0.47	0.64	97.1
Has post secondary education	0.04	0.01	22.9	19.40	0.00	0.01	0.02	-2.0	-0.88	0.38	91.3
Father's Education level											
Has no education	0.08	0.20	-36.7	-17.80	0.00	0.09	0.09	-0.9	-0.39	0.70	97.5
Has primary education level	0.39	0.52	-26.9	-14.75	0.00	0.44	0.43	0.4	0.15	0.88	98.5
Has secondary education level	0.46	0.26	42.4	24.72	0.00	0.43	0.42	2.0	0.69	0.49	95.4
Has post secondary education level	0.08	0.02	28.0	21.02	0.00	0.04	0.05	-3.8	-1.42	0.15	86.3
Household size	6.2	6.9	-27.1	-14.8	0.00	6.36	6.4	-0.5	-0.02	0.84	98.1
Household has source of water at home	0.32	0.15	41.4	25.7	0.00	0.26	0.26	0.2	0.06	0.96	99.6
Household has toilet/latrine at home	0.90	0.75	40.8	19.98	0.00	0.89	0.88	0.2	0.09	0.93	99.5
Distance to school is less than 30 min.	0.47	0.56	-12.2	-6.44	0.00	0.50	0.51	-0.5	-0.19	0.85	96.0
Household Assets											
Household has a TV	0.44	0.15	66.8	42.40	0.00	0.34	0.34	0.4	0.12	0.90	99.5
Household has a radio	0.83	0.71	28.4	14.69	0.00	0.80	0.81	-1.2	-0.47	0.64	95.8
Household has a computer	0.06	0.01	28.1	22.79	0.00	0.03	0.02	2.1	0.90	0.37	92.5
Household has a phone	0.82	0.64	42.2	21.56	0.00	0.79	0.80	-1.5	-0.61	0.55	96.4
Household has a car	0.10	0.02	31.8	24.07	0.04	0.05	0.05	1.8	0.69	0.49	94.4
Household has a cattle	0.54	0.56	-3.7	-2.06	0.00	0.56	0.58	-2.8	-1.03	0.30	25.7
Household has a donkey	0.10	0.16	-18.8	-9.65	0.00	0.11	0.11	-0.4	-0.14	0.89	98.1
Household has a camel	0.01	0.04	-16.7	-7.86	0.00	0.02	0.02	0.5	0.23	0.82	97.2
Household has a goat	0.38	0.49	-23.4	-12.77	0.00	0.40	0.40	-0.1	-0.03	0.97	99.6
Household has a bicycle	0.32	0.31	3.1	1.71	0.00	0.33	0.31	1.8	0.65	0.52	42.1
Household has a motorbike	0.11	0.05	22.2	14.25	0.00	0.09	0.09	0.9	0.32	0.75	95.8
Household has a cart	0.04	0.04	-1.9	-1.03	0.00	0.04	0.03	0.6	0.24	0.81	69.0
Number of Meals taken per day											
One meal	0.02	0.04	-8.6	-4.35	0.00	0.02	0.03	-0.5	-0.20	0.84	94.3
Two meals	0.12	0.23	-29.2	-14.74	0.00	0.13	0.13	-0.3	-0.14	0.89	98.8
Three meals	0.86	0.74	31.2	15.84	0.00	0.84	0.84	0.5	0.21	0.83	98.3
Wall material for dwelling place											
Mud	0.40	0.66	-55.4	-31.06	0.00	0.47	0.47	-0.7	-0.25	0.80	98.7
Polythene and iron	0.08	0.07	5.5	3.21	0.00	0.08	0.08	0.5	0.18	0.85	90.5
Timber	0.12	0.09	12.1	7.10	0.00	0.13	0.13	-1.7	-0.59	0.56	85.6
Bricks_stone	0.39	0.17	49.0	29.97	0.00	0.32	0.31	1.7	0.58	0.56	96.6
Regular source of lighting											
Electricity	0.37	0.10	66.4	45.16	0.00	0.26	0.26	0.1	0.02	0.98	99.9
Paraffin	0.61	0.82	-50.0	-30.71	0.00	0.71	0.71	-1.2	-0.41	0.68	97.7
Other	0.02	0.07	-22.5	-10.66	0.00	0.03	0.02	2.2	1.10	0.27	90.1
Village Characteristics											
Village has chief's office	0.70	0.62	18.2	9.83	0.00	0.68	0.69	-1.0	-039	0.69	94.3
Village has shopping center	0.35	0.24	24.0	13.88	0.00	0.32	0.31	0.6	0.20	0.84	97.6
Village has police post	0.35	0.24	24.0	13.88	0.00	0.31	0.31	0.6	0.20	0.84	97.6
Village has electricity	0.59	0.38	42.5	23.61	0.00	0.54	0.53	2.1	0.75	0.46	95.2
Village has tarmac road	0.29	0.15	33.9	20.68	0.00	0.25	0.25	0.3	0.09	0.93	99.2
Village has all-weather road	0.81	0.78	6.4	3.46	0.00	0.80	0.80	-1.5	-0.57	0.57	76.2
Village has an education committee	0.27	0.32	-11.4	-6.91	0.00	0.27	0.28	-1.9	-0.73	0.47	83.1
Village has all protected water point	0.42	0.40	2.1	1.19	0.23	0.41	0.40	3.4	1.26	0.21	-57.6
Village is rural	0.69	0.83	-36.4	-22.12	0.00	0.74	0.74	-0.6	-0.23	0.82	98.2
Village is urban	0.32	0.17	36.4	22.12	0.00	0.26	0.26	0.6	0.23	0.82	98.2

Figure A.2: Visual Inspection of Standardized Differences

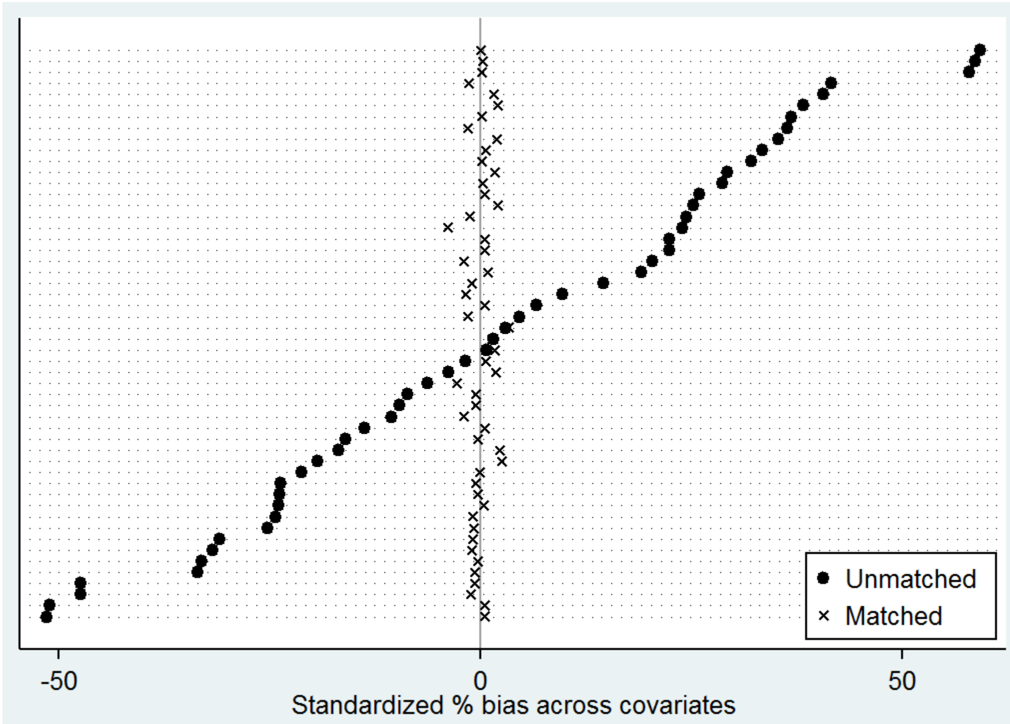
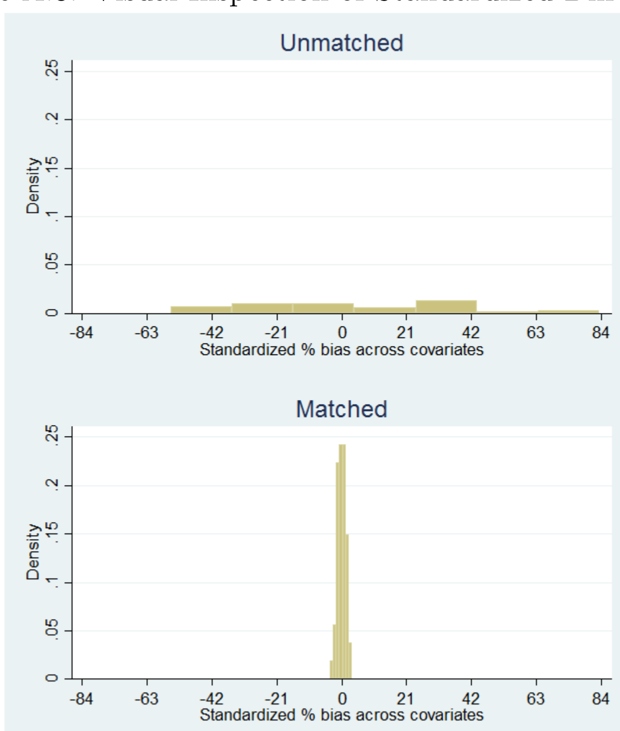


Figure A.3: Visual Inspection of Standardized Differences



References

Aakvik, A. (2001). Bounding a matching estimator: the case of a norwegian training program. *Oxford bulletin of economics and statistics*, 63(1):115–143.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2000). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. Technical report, National bureau of economic research.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(01):151–184.

Andrabi, T., Das, J., and Khwaja, A. I. (2008). A dime a day: The possibilities and limits of private schooling in pakistan. *Comparative Education Review*, 52(3):329–355.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in colombia: Evidence from a randomized natural experiment. *The American Economic Review*, 92(05):1535–1558.

Ashley, L. D., Mcloughlin, C., Aslam, M., Engel, J., Wales, J., Rawal, S., Batley, R., Kingdon, G., and Nicolai, S. (2014). The role and impact of private schools in developing countries: A rigorous review of the evidence. final report. education rigorous literature review. Technical report, Department for International Development.

Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in medicine*, 27(12):2037–2049.

Azam, M., Kingdon, G., and Wu, K. B. (2015). Impact of private secondary schooling on cognitive skills: evidence from india. *Education Economics*, pages 1–16.

Becker, S. O. and Caliendo, M. (2007). Mhbounds-sensitivity analysis for average treatment effects.

Behrman, J. R., Rosenzweig, M. R., and Taubman, P. (1994). Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment. *Journal of Political Economy*, pages 1131–1174.

Bharath, S. T., Dahiya, S., Saunders, A., and Srinivasan, A. (2011). Lending relationships and loan contract terms. *Review of Financial Studies*, 24(4):1141–1203.

Bold, T., Kimenyi, M., Mwabu, G., and Sandefur, J. (2011). The high return to private schooling in a low-income country. *Center for Global Development Working Paper*, (279).

Bold, T., Kimenyi, M., Mwabu, G., and Sandefur, J. (2013a). The high return to low-cost private schooling in a developing country. *London: International Growth Centre (IGC) London School of Economics*.

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

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1):31–72.

Cavalcanti, T., Guimaraes, J., and Sampaio, B. (2010). Barriers to skill acquisition in brazil: Public and private school students performance in a public university entrance exam. *The Quarterly Review of Economics and Finance*, 50(4):395–407.

Chakrabarti, R. and Peterson, P. E. (2008). School choice international: Exploring public-private partnerships. *MIT Press (BK)*.

De Haan, M., Plug, E., and Rosero, J. (2014). Birth order and human capital development evidence from ecuador. *Journal of Human Resources*, 49(2):359–392.

Dehejia, R. H. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, 84(1):151–161.

Desai, S., Dubey, A., Joshi, B., Sen, M., Sheriff, A., and Vanneman, R. (2008). India human development survey. *College Park, Maryland: University of Maryland*.

Desai, S., Dubey, A., Vanneman, R., Banerji, R., et al. (2009). Private schooling in india: A new educational landscape. In *India Policy Forum*, volume 5, pages 1–38. National Council of Applied Economic Research.

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.

Dostie, B. and Jayaraman, R. (2006). Determinants of school enrollment in indian villages. *Economic Development and Cultural Change*, 54(2):405–421.

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.

Evans, W. N. and Schwab, R. M. (1995). Finishing high school and starting college: Do catholic schools make a difference? *The Quarterly Journal of Economics*, pages 941–974.

Faries, D. E., Obenchain, R., Haro, J. M., and Leon, A. C. (2010). *Analysis of observational health care data using SAS*. SAS Institute.

French, R. (2008). Schooling participation and pupil achievement in india. Technical report, Department of Quantitative Social Science. London, Institute of Education, University of London.

French, R., Kingdon, G., et al. (2010). The relative effectiveness of private and government schools in rural india: Evidence from aser data. *London: Institute of Education*.

Goldberger, A. S. and Cain, G. G. (1982). The causal analysis of cognitive outcomes in the coleman, hoffer and kilgore report. *Sociology of education*, 55(2):103–122.

Goldhaber, D. D. and Eide, E. R. (2003). Methodological thoughts on measuring the impact of private sector competition on the educational marketplace. *Educational Evaluation and Policy Analysis*, 25(2):217–232.

Govinda, R. and Varghese, N. V. (1993). Quality of primary schooling in india: A case study of madhya pradesh. Technical report, Paris 07 SP, France: UNESCO.

Hammersley, M., Kozma, T., Reid, I., Yogev, A., Shapira, R., Halsey, A., Heath, A., and Ridge, J. (1981). Origins and destinations. *British Journal of the Sociology of Education*, 2(1):91–5.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, 64(4):605–654.

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.

Javaid, K., Musaddiq, T., and Sultan, A. (2012). Prying the private school effect: an empirical analysis of learning outcomes of public and private schools in pakistan. Technical report, Lahore: University of Management Sciences (LUMS), Department of Economics.

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.

Kingdon, G. and Teal, F. (2010). Teacher unions, teacher pay and student performance in india: A pupil fixed effects approach. *Journal of Development Economics*, 91(2):278–288.

KIPPRA (2016). Kenya economic report 2016: Fiscal decentralization in support of devolution. Technical report, Kenya Institute for Public Policy Research and Analysis (KIPPRA), Nairobi, Kenya.

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*.

Maitra, P., Pal, S., and Sharma, A. (2016). Absence of altruism? female disadvantage in private school enrollment in india. *World Development*, 85:105–125.

Mani, S., Hoddinott, J., and Strauss, J. (2013). Determinants of schooling: Empirical evidence from rural ethiopia. *Journal of African Economies*, 22(5):693–731.

McEwan, P. J. (2002). Public subsidies for private schooling: A comparative analysis of argentina and chile. *Journal of Comparative Policy Analysis*, 4(2):189–216.

McEwan, P. J. and Carnoy, M. (2000). The effectiveness and efficiency of private schools in chile's voucher system. *Educational evaluation and policy analysis*, 22(3):213–239.

Newhouse, D. and Beegle, K. (2006). The effect of school type on academic achievement evidence from indonesia. *Journal of Human Resources*, 41(3):529–557.

Nichols, A. and Schaffer, M. (2007). Clustered errors in stata. In *United Kingdom Stata Users' Group Meeting*.

Nishimura, M. and Yamano, T. (2013). Emerging private education in africa: determinants of school choice in rural kenya. *World Development*, 43:266–275.

Oketch, M., Mutisya, M., Ngware, M., and Ezeh, A. C. (2010). Why are there proportionately more poor pupils enrolled in non-state schools in urban kenya in spite of fpe policy? *International Journal of Educational Development*, 30(1):23–32.

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 Ngware, M. (2010). Free primary education still excludes the poorest of the poor in urban kenya. *Development in Practice*, 20(4-5):603–610.

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

Olembo, J. O. (1985). Financing primary school buildings in kenya. Technical report, Nairobi: Transafrica Press.

Pal, S. (2010). Public infrastructure, location of private schools and primary school attainment in an emerging economy. *Economics of Education Review*, 29(5):783–794.

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

Piper, B., Schroeder, L., and Trudell, B. (2015). Oral reading fluency and comprehension in kenya: reading acquisition in a multilingual environment. *Journal of Research in Reading*.

Piper, B., Zuilkowski, S. S., and Mugenda, A. (2014). Improving reading outcomes in kenya: First-year effects of the primr initiative. *International Journal of Educational Development*, 37:11–21.

Psacharopoulos, G. (1987). Public versus private schools in developing countries: evidence from colombia and tanzania. *International Journal of Educational Development*, 7(1):59–67.

Rose, P. (2006). Collaborating in education for all? experiences of government support for non-state provision of basic education in south asia and sub-saharan africa. *Public administration and development*, 26(3):219–229.

Rosenbaum, P. R. (2002). Observational studies. Technical report, 2nd ed. New York: Springer.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Simumba, J. (2013). Child labour and schooling: Evidence from rural zambia. Technical report, Zambia Institute for Policy Analysis and Research Discussion Paper.

Smith, J. (2000). A critical survey of empirical methods for evaluating active labor market policies. Technical report, Research Report, Department of Economics, University of Western Ontario.

Thapa, A. (2015). Public and private school performance in nepal: an analysis using the slc examination. *Education Economics*, 23(1):47–62.

Tooley, J. (2013). Challenging educational injustice: Grassroots privatisation in south asia and sub-saharan africa. *Oxford Review of Education*, 39(4):446–463.

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 Dixon, P. (2005). *Private education is good for the poor: A study of private schools serving the poor in low-income countries.* Cato Institute Washington, DC.

Tooley, J., Dixon, P., and Stanfield, J. (2008). Impact of free primary education in kenya a case study of private schools in kibera. *Educational Management Administration & Leadership*, 36(4):449–469.

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

Uwezo (2012). Are our children learning? annual learning assessment report 2012. Technical report, Dar es Salaam: Uwezo. National Assessment System for Monitoring.

Uwezo (2014). Are our children learning? annual learning assessment report- the state of education in kenya in 2015 and beyond. Technical report, Nairobi: Uwezo. National Assessment System for Monitoring.

Vandenberghe, V. and Robin, S. (2004). Evaluating the effectiveness of private education across countries: a comparison of methods. *Labour economics*, 11(4):487–506.

Vegas, E. and Devercelli, A. (ND). The world banks new tool to inform policy when scaling-up ecd. Technical report, Education Unit, Human Development Department, World Bank, Washington dc, usa.

Wakano, A. (2016). The effect of ratio between pta teachers and government employed teachers on education outcomes in kenya primary schools. Technical report, Graduate School of Economics and Osaka School of International Public Policy.

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

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

southern africa labour and development research unit

The Southern Africa Labour and Development Research Unit (SALDRU) conducts research directed at improving the well-being of South Africa's poor. It was established in 1975. Over the next two decades the unit's research played a central role in documenting the human costs of apartheid. Key projects from this period included the Farm Labour Conference (1976), the Economics of Health Care Conference (1978), and the Second Carnegie Enquiry into Poverty and Development in South Africa (1983-86). At the urging of the African National Congress, from 1992-1994 SALDRU and the World Bank coordinated the Project for Statistics on Living Standards and Development (PSLSD). This project provide baseline data for the implementation of post-apartheid socio-economic policies through South Africa's first non-racial national sample survey.

In the post-apartheid period, SALDRU has continued to gather data and conduct research directed at informing and assessing anti-poverty policy. In line with its historical contribution, SALDRU's researchers continue to conduct research detailing changing patterns of well-being in South Africa and assessing the impact of government policy on the poor. Current research work falls into the following research themes: post-apartheid poverty; employment and migration dynamics; family support structures in an era of rapid social change; public works and public infrastructure programmes, financial strategies of the poor; common property resources and the poor. Key survey projects include the Langeberg Integrated Family Survey (1999), the Khayelitsha/Mitchell's Plain Survey (2000), the ongoing Cape Area Panel Study (2001-) and the Financial Diaries Project.

SALDRU

www.saldru.uct.ac.za
Level 3, School of Economics Building, Middle Campus, University of Cape Town
Private Bag, Rondebosch 7701, Cape Town, South Africa
Tel: +27 (0)21 650 5696
Fax: +27 (0) 21 650 5797
Web: www.saldru.uct.ac.za

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad

SPES BONA