119

Working Paper

Young Lives

An International Study of Childhood Poverty

The Effect of Lengthening the School Day on Children's Achievement in Ethiopia

Kate Orkin

www.younglives.org.uk

DECEMBER 2013

Working Paper 119

Young Lives
An International Study of Childhood Poverty

The Effect of Lengthening the School Day on Children's Achievement in Ethiopia

Kate Orkin

This paper was presented at a conference on Inequalities in Children's Outcomes in Developing Countries hosted by Young Lives at St Anne's College, Oxford on 8-9 July 2013.

http://www.younglives.org.uk/news/news/children-inequalities-younglives-conference-2013

The data used come from Young Lives, a longitudinal study of childhood poverty that is tracking the lives of 12,000 children in Ethiopia, India (in Andhra Pradesh), Peru and Vietnam over a 15-year period. **www.younglives.org.uk**

Young Lives is funded from 2001 to 2017 by UK aid from the Department of International Development and co-funded by the Netherlands Ministry of Foreign Affairs from 2010 to 2014.

The Effect of Lengthening the School Day on Children's Achievement in Ethiopia

Kate Orkin

First published by Young Lives in December 2013

ISBN: 978-1-909403-32-1

A catalogue record for this publication is available from the British Library.

Printed on FSC-certified paper from traceable and sustainable sources.

Funded by

UKaid from the British people

Ministry of Foreign Affairs of the Netherlands

Young Lives, Oxford Department of International Development (ODID), University of Oxford, Queen Elizabeth House, 3 Mansfield Road, Oxford OX1 3TB, UK

Tel: +44 (0)1865 281751 • E-mail: younglives@younglives.org.uk

.

Contents

1. Introduction 3

2. Full-day schooling in Ethiopia 7

3. Data 9

4. Changes in instructional time in Young Lives sample 10

5. Research design 11

6. Plausibility of the equal time trends assumption 14

6.1. Incomparable treatment and control groups 14

6.2. Changes in group composition 16

6.3. Correlated shocks or policy changes 17

7. Results 18

8. Conclusion 20

References 23

Author Biography. Kate Orkin is a Research Fellow at Clare Hall and the Faculty of Economics, Clare Hall, University of Cambridge. She is a Student Associate at Young Lives, where she worked in the past on the design of school surveys in Ethiopia.

Acknowledgements. The paper uses data from Young Lives (www.younglives.org.uk). Young Lives is core-funded from 2001 to 2017 by the UK Department for International Development (DFID), and co-funded by the Netherlands Ministry of Foreign Affairs from 2010 to 2014. Thanks to Young Lives children, parents, teachers and principals for their continued participation. My qualitative fieldwork was funded by the Wingate Foundation, the Skye Foundation and St Antony's College, University of Oxford. Ato Solomon Shiferaw and Ato Tayachew Ayalew (Ethiopian Federal Ministry of Education); Ato Tizazu Asare and Ato Setotaw Yimam (former Ethiopian Federal Ministry of Education); Chris Berry and Belay Addise (DFID Ethiopia); Marianne Kujala-Garcia (Embassy of Finland, Addis Ababa); Margherita Lulli (Development Cooperation Office of the Italian Embassy, Addis Ababa); Allyson Wainer (USAID, Addis Ababa); Alula Pankhurst, Tassew Woldehanna and Workneh Abebe (Young Lives); Catherine Dom (Mokoro); and Nicola Berry (Save the Children) provided invaluable direction, information and guidance. Stefan Dercon, Laura Camfield, Paul Glewwe, Francis Teal, Andreas Georgiadis, Jo Boyden, Martin Woodhead, Alula Pankhurst and Abhijeet Singh provided valuable comments.

Abstract. Many developing country schools have four-hour school days and teach two groups of children each day. Governments are considering lengthening the school day, at great expense, to improve school quality. Advocates of the shift system argue the reform is unnecessary, as evidence from developed countries suggests increasing instructional time only improves achievement scores by small amounts. This paper is the first study of the effect of a large increase in instructional time in a low-income country. In 2005, the Ethiopian federal government directed school districts to abolish teaching in shifts and lengthen the school day from four to six hours. Districts implemented the reform at different times, creating exogenous variation in instructional time. I use a difference-in-difference specification controlling for time-invariant unobservables at school level on a unique longitudinal dataset. For eight-year-old children, a longer school day improved writing and mathematics scores, but had no significant effect on reading. However, effects are larger among better-off children: children who are not stunted, children from richer households and children in urban schools. The exception is that the reform has larger positive effects on girls than boys. The reform thus improves achievement on average, but may exacerbate gaps between wealthier and poorer children.

1. Introduction

Many developing country schools teach using a shift system, where schools teach one group of children in the morning and another group in the afternoon. Each group has a short school day of three to four hours and the school can teach roughly double the number of children with the same number of teachers and spending on land and buildings. Many governments implemented shift schooling when facing shortages of classrooms or teachers, often as they expanded enrolment rapidly to make primary education universally available. Now, as many Latin American and African countries are exploring lengthening the school day as they have reached near-universal enrolment and are focusing on improving the quality of education (Bray, 2008, 55).

However, advocates of shift schooling, such as the World Bank and UNESCO, argue that lengthening the school day is an inefficient use of limited resources (Bray, 2008). They draw on evidence from developed and middle-income countries, which finds that increases in instructional time have only very small effects on achievement, of between 0.009 and 0.12 standard deviations. In addition, lengthening the school day has been relatively costly when it has been applied in other developing countries: Table 2 show that moving from shift to full-day schooling can increase the costs of primary education by between 25 and 60 per cent. More classrooms and facilities must be built. If there are teacher shortages and teachers teach both shifts, more must be trained and hired. If teachers teach only one shift before the school day is lengthened, salaries will probably increase to reflect more hours of work.

I provide the first evidence from a low-income country on the effect of lengthening the school day. I use a natural experiment in Ethiopia. In 2005, the federal government directed districts to abolish teaching in shifts, but districts phased in the reform over time. Schools increased the length of the day from four to five and a half hours, an increase of roughly 30 per cent. Using a difference-in-difference specification to control for trends in achievement and school fixed effects to control for time-invariant unobservables at school level, I find that lengthening the school day has large positive effects on achievement. In 2009, children in "reform" schools which implemented the full-day reform are 2.17 times more likely to be numerate than children in those same schools in 2002, before the full-day reform was applied. They are also 3.51 times more likely to be able to write with difficulty or write easily, versus not being able to write. They are also 1.12 times more likely to be able to read, but the difference is not statistically significant.

These results are larger than others in the literature. Studies of the effect of instructional time on achievement from developed and middle-income countries find that increases in instructional time result in positive but small increases in achievement. Table 3 shows that significant effect sizes range between 0.009 and 0.12 standard deviations.[1] However, most research examines small changes in instructional time of around 10 per cent in schools with days of six or more hours in developed or middle-income countries. No research to date has examined reforms to lengthen the day in a low-income country where the initial school day is short, so the increase in time is large in proportion to the total time spent in school.

The first contribution of this paper is that it considers a reform in a low income country which increased the length of the school day from a short school day, of four hours, to between five-and-a-quarter and six-and-a-half hours, a change of between 30 and 60 per cent. The Ethiopian experience is more informative than existing evidence for low-income countries considering lengthening the school day. Table 1 shows the length of school day in some low-income countries, where moving from shift to full-day schooling would increase time in school by between 20 and 40 per cent.

The second contribution of the paper is that there is limited research on the effects of increased instructional time for younger children, even though studies which compare effects in different grades find larger effects for younger children (Marcotte, 2007; Lavy, 2010). I examine children in the first three grades of school and find large effects. The data does not enable examination of the effect of the reform on older children, so I cannot disentangle whether this effect size occurs because of the age of the children or the large increase in the amount of time spent in school.

Finally, I explore heterogeneity of treatment effects by child characteristics and household socio-economic status. Most research only has data at district or school level and can only examine how the effects of reforms differ between schools which have, on average, better- or worse-off parents. The limited research comparing children of wealthier and poorer families in the same school finds increased instructional time has much larger effects for children from poorer families (Lavy, 2010, 2012). The Young Lives dataset is one of very few in Africa with linked information on individual children's achievement, households and the quality of school they attend (the data used by Glick and Sahn (2009) in Senegal and Lam, Ardington, and Leibbrandt (2011) in South Africa are two other examples). This enables analysis of the effects of school policies on different types of children. Effects are larger among better-off children: children who are not stunted, children from richer households and children in urban schools. The exception is that the reform has larger positive effects on girls than boys. The reform thus improves achievement on average, but may exacerbate gaps between wealthier and poorer children.

The major econometric difficulty in identifying the casual effect of instructional time on achievement is that instructional time may be assigned to students non-randomly. If schools control the

[1]These effect sizes are small compared to other education interventions: a review of 22 recent randomised controlled trials finds that, of 14 education interventions with significant effects, 13 had effects of 0.1 to 0.46 standard deviations of test scores (Kremer and Holla, 2009, C-3).

amount of instructional time, there are probably omitted school-level variables correlated with their choice. The covariance between instructional time and unobserved school-level variables is non-zero, rendering inferences from regression models to the population invalid. Linked to this are two selection problems. Firstly, parents choose better schools if they can. If more instructional time is thought to signal a better quality school, there may be a non-zero correlation between omitted variables (such as parental motivation or child ability) and the amount of instructional time a child receives. Secondly, certain types of teachers may sort into schools based on the amount of instructional time.

Time in school could be varied randomly in a field experiment, but recent reviews (Kremer and Holla, 2009) suggest this has not been done. Studies have adopted four other strategies to deal with omitted variable bias and selection issues. Some authors compare gains in achievement over time for children receiving different amounts of instructional time.DeCicca (2007) finds white and Hispanic children in the US who attended full-day kindergarden learn more in first grade than those who attended half day kindergarden. Eide and Showalter (1998) find that the achievement scores of high school students whose schools had longer terms improved more between tenth and twelfth grade. However, there results could reflect selection bias (if better students both select into schools with longer days and improve more quickly) or bias from omitted school-level variables (if the schools which choose to teach a longer day are also better in unobservable ways).

Wößmann (2003) and Lavy (2010) exploit variation in instructional time across different subjects to investigate whether the same child does better in subjects in which they receive more teaching time. This solves selection issues to some extent: students are less likely to select into schools on the basis of subject-specific instructional time, although teachers may select into schools which prioritise their subject. Examining within-child variation also controls for time-invariant school-level unobservables if children stay in the same school. However, this strategy assumes the effect of instructional time is the same across subjects and cannot account for instructional time spillovers (more time in English may affect maths scores).

Thirdly, authors have examined substantial cross-country variation in instructional time. Lee and Barro (2001) find a small positive association between the length of the school year and achievement. However, they only have one year of data on term length and cannot control for country fixed effects to account for unobservables at country level correlated with both term length and achievement.

Finally, studies use natural experiments which provide exogenous variation in instructional time. These also address concerns about selection of children or teachers into treatment schools to some extent. By definition, one cannot tell if selection into schools on unobservable characteristics is occurring. But if "treatment" and "control" schools are similar at baseline on observable characteristics and children and teachers do not appear to change schools in response to the reform, one can argue that selection into treatment schools is of limited concern.

Studies exploit variation in the number of days before standardised tests (Sims, 2008; Fitzpatrick, Grissmer, and Hastedt, 2011; Agüero and Beleche, 2013), variation in school days because of unanticipated weather closures Marcotte (2007), or school management or funding reforms which lengthen the day as part of a package of reforms (Bellei, 2009; Cerdan-Infantes and Vermeersch, 2007; Lavy, 2012). Six out of seven studies find positive effects on achievement of between 0.009 and 0.12 standard deviations of test scores. All the studies mentioned which examine subjects separately find instructional time has positive effects on maths, but much smaller or no effects on reading or language (Sims, 2008; Brown and Saks, 1987; Lavy, 2012; Bellei, 2009; DeCicca, 2007; Marcotte, 2007). Fitzpatrick, Grissmer, and Hastedt (2011) is the only study to find larger effects for reading than for maths.

The difference-in-differences estimation strategy used in this paper makes the assumption of equal trends: it assumes that in the absence of the full-day reform, the average unobserved characteristics of shift and reform schools and children in these schools would change over time in the same way. The way in which the reform was implemented and the data used gives the study several strengths compared to others in the literature.

Firstly, the data follows two cohorts of children born eight years apart who live in the same communities and attend the same schools. The older cohort are surveyed before the reform; the younger cohort are surveyed after the reform is implemented in some schools. Rather than only controlling for differences between shift and reform schools at baseline, as most difference-in-difference specifications do, I can use school fixed effects to control for school- or district-level unobservables that are correlated with achievement and do not change over the eight-year period of the panel. These capture improve the precision of estimates by capturing variation in achievement associated with school characteristics, such as location and age, and the characteristics of the surrounding area, like climate, dominant dominant language and ethnic group. In this case, the school fixed effects estimates are very similar to estimates without school fixed effects. This suggests time-invariant school- and district-level unobservables do not have a major effect on outcomes, which may make it more likely that the assumption of equal trends holds (i.e. average unobserved time-variant characteristics of shift and reform schools change in similar ways).

Secondly, according to interviews with government officials, there was no systematic targeting of particular types of schools and districts for the reform. The reform was phased in nationally, so districts implemented the reform at slightly different times. Within a district, schools received an instruction to implement the reform and largely obeyed these instructions. Difference-in-differences does not require that treatment is randomly assigned either to districts or to schools. However, if schools had limited choice whether or not to implement the reform, it is likely that shift and reform schools and the children in them are similar at baseline on unobservable characteristics, which makes it likelier that they will change over time in similar ways. Furthermore, children in shift and reform schools are similar on observable characteristics at baseline; the schools are similar on a wide range of observable school quality characteristics at endline; achievement scores are similar in shift and reform schools at baseline before the reform once child characteristics are controlled for; and there are no observable school characteristics which predict whether the school implemented the reform. This provides further support for the assumption of equal trends.

Thirdly, it is difficult for students or teachers to select into certain types of schools in response to the reform. If more able children or teachers could select into schools that had applied the reform, this would jeopardise difference-in-differences estimation, as the average unobserved characteristics of shift and reform schools and the children in these schools might change in different ways. However, selection bias is likely to be limited due to severe constraints on children's school choice and on teacher transfers in Ethiopia.

Finally, schools which applied the full-day reform did not experience changes in the level of per-student resources or change the curriculum they taught. Schools were allocated extra resources to build classrooms or hire new teachers to keep the level of per-student resources the same. Other evaluations of changes in instructional time have usually examined changes in instructional time that are part of a broad package of school management, funding or curriculum reforms. This makes it difficult to identify effects due only to variation in instructional time.

2. Full-day schooling in Ethiopia

The Ethiopian context is very different from the developed and middle income countries where previous research on time in instruction has taken place. Eighty-four percent of the Ethiopian population of 78 million lives in rural areas, where most are sedentary farmers or pastoralists (Ministry of Education, 2011a, 10). Challenges to provision of primary education include the dispersed rural population, ethnic diversity (with over 80 ethnic groups) and political instability.

Under the imperial government, only the elite could access primary education . In 1961, under Emperor Haile Selassie, the net enrolment rate (NER) – the proportion of the age group who ought to be enrolled in a given level of education who are actually enrolled – was only 3.8 per cent (Kebede, 2006, 11). Soldiers deposed the Emperor in 1974 and established the Marxist *Derg* regime, which expanded formal education and increased the NER from 10 to 35 per cent between 1970 and 1989 (Negash, 2006, 19). But there were shortages of trained teachers and textbooks and pupil-teacher ratios were high. Shift schools, which had first been introduced in the 1960s, became the norm and the school day was only three hours long.

In 1991, after a civil war, the Ethiopian People's Revolutionary Democratic Front (EPRDF) and its allies overthrew the *Derg*. The NER in 1991 had dropped to 20 per cent as a result of the war (Ministry of Education, 1998, 5). Since then, improvements in access to education have been among the most dramatic in the world. In 1995/96, 3.1 million children were enrolled in primary school, an NER of 30 per cent (Ministry of Education, 1998, 5). By 2010/11, 16.72 million children were enrolled, an NER of 85 per cent (Ministry of Education, 2011b, 11). Major policy shifts which facilitated this expansion included the abolition of fees in 1995/96 and a massive school-building campaign which doubled the number of primary schools (Engel, 2010, 7).

The government recognised explicitly that there would be a trade-off between increasing access to education quickly and providing high quality schooling. Roughly-built schools were set up quickly in isolated, poorer areas, often drawing on communities to provide labour and local building materials. In the first Education Sector Development Plan (ESDP-I) in 1997, government stated that average class size would be allowed to rise from 33 to 50 and the shift system should be used to use the limited numbers of teachers and classrooms efficiently (Ministry of Education, 1998, 6). Teaching in shifts was more common in urban schools, where overcrowding was more of an issue: in 2001/02, 80 per cent of urban primary schools but only 39 percent of rural schools taught in shifts (World

Bank, 2005, 278). ESDP-II in 2002 maintained the policy of promoting teaching in shifts (Ministry of Education, 2002a, 24).

Inevitably, aggregate measures of school quality declined as the school system expanded. Successive National Learning Assessments (NLA) in 2000, 2004 and 2007 found that student scores had declined (Ministry of Education/USAID, 2008, 9). The proportion of qualified teachers in Grades 5 to 8 dropped and the pupil-teacher ratio was one of the highest in Africa (World Bank, 2005, 153).

Education quality was a major issue in the lead-up to the 2005 elections and the government subsequently began a number of quality improvement initiatives (Dom, 2010). A review of the first ten years of education provision under the EPRDF, published in 2002, stated that "the community will be organised towards gradually eliminating the shift system as well as decreasing the number of students per classroom" (Ministry of Education, 2002b, 114). ESDP-III, passed in early 2005, stated: "... the longer pupils remain in school, the more they obtain academic support from teachers, use libraries and laboratories and engage in co-curricular activities. It is therefore necessary to increase time-on-task by reducing the operation of the shift system". (Ministry of Education, 2005, 59-60)

In 2010, I interviewed the two directors of planning who had worked in the Federal Ministry of Education from 1997 to 2007, as well as the directors of planning and curriculum at the regional education bureaux in each region covered in the dataset used. The federal directors of planning during this period argued that the shift system had always been seen as a temporary compromise, and the federal ministry planned to move to full-day schooling as soon as possible. Officials in two regions, Addis Ababa and Tigray stated that they had begun planning to phase out shift schools before 2005. Officials in all five regions in the Young Lives study stated that, in the months before the 2005 election, the federal education ministry begun exerting strong pressure to convert schools to teaching a full day. As shown in Table 5, between 2002 and 2010, the proportion of schools nationally teaching in shift declined dramatically, from 44 per cent in 2002 at the time of the collection of the first round of data used in this paper, to 21 per cent in 2010 at the time of the second round of data used (Ministry of Education, 2003, 2011b).

Officials stated that the reform was supposed to be applied in all school districts and was not targeted at any particular type of school. However, regions phased the reform in in different districts, and districts applied the reform in different parts of the district at different points in time. Officials reported that there was no systematic reason that one district was selected over another. Schools could not decide whether or not to implement the reform: they received a directive from the district to lengthen the day and did so either the next school year or the year afterwards. Many schools phased in the reform by lengthening the day first in higher grades and then in lower grades or vice versa.

The regulation does not prescribe how much schools needed to lengthen the day or how they should use the extra time. A World Bank report from 2005 notes that shift schools provided four hours of instruction in six lessons of 40 minutes, while full-day schools provided five-and-a-quarter hours of instruction in seven lessons of 45 minutes (World Bank, 2005, 145). Changing from shift to full-day schooling would thus be a 31.25 per cent increase in instructional time.

3. Data

The paper uses household and school data from Young Lives, a 15-year longitudinal study in four countries. The waves of data used in this paper are shown in Table 4. The household dataset has three waves, one in 2002 before the full-day reform and two after the reform, in 2006 and 2009. The household survey follows $3,000$ children, $1,000$ in the older cohort and $2,000$ in the younger cohort. I use the household survey for children's achievement scores, controls at child and household level, and the name of the school attended by each child in each round.

The data on whether schools had implemented the full-day reform is from a survey of Young Lives children's schools. The third round of the household survey was conducted from October 2009 to March 2010. Between March and May 2010, in the same school year as the household survey, Young Lives surveyed the schools attended by both older and younger cohort children. The principal of each school was asked detailed questions about the implementation of the full-day school reform in each year between 2002 and 2010. This information enabled each school to be classified as a shift school which never applied the reform, a reform school which applied the reform after 2002, or an always full-day school, which was already a full-day school in 2002.

Table 6 shows the number of children in each cohort in the sample. This paper compares achievement among the $1,000$ older cohort children in Round 1 in 2002/3, when they were eight years old and just starting school, to achievement in the $1,886$ children in the younger cohort surveyed in Round 3 in 2009/10, when they were also eight years old. Column 1 shows all the children attending schools surveyed in the school survey. There were 652 older cohort children attending school at the time of Round 1 in 2002/3. All 652 were attending schools visited in the school survey and are included in the sample.

At the time of Round 3 in 2009/10, $1,424$ of the $1,886$ children in the younger cohort were enrolled. Only $1,270$ of these $1,424$ enrolled younger cohort children attended schools included in the school survey data.[2] The younger cohort children attended 112 schools. Column 2 shows the number of schools attended by each cohort. There are 118 schools in the sample in total. The older cohort children attended 65 schools. The younger cohort attended 112 schools.

Columns 3 and 4 describe the sample of children examined in this paper: the children who do not have any missing values on variables used in the regressions in this paper. There are eight schools attended only by older cohort children, 49 attended only by younger cohort children, and 55 attended by both older and younger cohort children.

There are only 11 children attending four "always full-day" schools which are dropped from the sample. These schools had taught a full day since the early 1990s, before the government began to argue for reduction in use of the shift system. They had either not implemented the shift system under the *Derg* or had become full-day schools before 2002. These schools, all of which were in Addis Ababa, are not considered in this analysis. The reform to lengthen the school day does not generate exogenous variation in whether schools fall into this category. It is likely that including a dummy for whether children are in a full-day school in the regression would be correlated with unobservables and bias all coefficients.

[2]Some children had migrated since Round 1 and were attending schools outside the districts surveyed. These schools were not included in the school survey.

The survey was conducted in Addis Ababa and the four most populated regions, Oromia, Amhara, the Southern Nations and Nationalities and People's Region (SNNP) and Tigray. These areas include 96 per cent of the population of Ethiopia (Woldehanna, Alemu, and Tekie, 2008). The regions which are not sampled are pastoralist and sparsely populated. In 2002, three to five districts (*woredas*) were selected in each region to capture a mix of rural and urban sites and ethnic groups and to over-sample food-insecure sites. One village (*kebele*), the lowest level of local authority, was randomly selected in each district as a survey site. Within each site, households were randomly sampled until 100 households with one child aged 6 to 18 months and 50 households with one child aged eight were identified. One can generalise to the relevant age group in the Young Lives sample and similar sites, but the sample is not nationally representative. Only 2.13 per cent of the sample has not been traced or refused to participate in later rounds (Woldehanna, Alemu, and Tekie, 2008, 16).

4. Changes in instructional time in Young Lives sample

As described in Section 2, the full-day reform was imposed at district level and was compulsory for schools within a district. However, because data was collected while the reform was being phased in, there is variation between and within districts in whether the reform has been applied. Table 7 shows the extent of the implementation of the reform by the 2009/10 school year in the 30 districts in the Young Lives survey.[3]

The reform was relatively uniformly implemented in most districts, which supports the evidence from qualitative interviews that schools could not decide whether to implement the reform. In 11 districts, shown in the first row, all schools still taught in shift, either because schools had not received a directive to increase the length of the day or because they had received the directive but not yet implemented it. These districts had probably only recently sent directives to schools. Eight districts, shown in the next two rows, had just begun implementing the reform, with either a majority of schools still teaching in shift or equal numbers teaching in shifts and for a full day. The remaining 11 districts have implemented the reform: six districts have implemented the reform in all schools and five have implemented it in a majority of schools. The reform has been implemented in districts all the five regions covered in the study and in both rural and urban schools.

The school survey collected data about whether the school taught in shifts, the duration of the school day, the amount of time allocated to lessons in each grade and the years when schools changed from shift to full-day. The data on teaching time was collected from examination of the school timetable for the relevant grades.

A World Bank report from 2005 states that shift schools provided four hours of instruction in six lessons of 40 minutes, while full-day schools provided five-and-a-quarter hours of instruction in seven lessons of 45 minutes (World Bank, 2005, 145). These numbers were verified in interviews with federal and regional officials. However, the survey finds that schools which lengthened the day

[3]In 2002, Young Lives sampled children in 20 survey sites, using the boundaries of the peasant association (*kebele*) in use at the time. Each *kebele* was part of only one district. However, district boundaries were altered in 2005, splitting some Young Lives districts into two. In addition, children from one *kebele* may attend schools in neighbouring districts. In this table and the rest of the paper, each school is allocated to a district using the school's location and the boundaries of the district in 2005.

sometimes taught for slightly longer than five-and-a-quarter hours. In 2009 in Grades 1 to 4, shift schools in the Young Lives sample taught for an average of 3:56 hours a day, reform schools for 5:49 hours and schools which were always full-day for 5:37 hours. This may reflect measurement error in the survey measures of time in lessons, or that schools structured their timetables and number of lessons in slightly different ways.

In this paper, a grade is classified as applying the reform if it changed its lesson time between 2002 and 2009 to provide class time of more than than four hours. I use the amount of time allocated to lessons rather than the length of the school day, because educational research emphasises that schools often allocate large amounts of the school day to activities in which no learning occurs (Patall, Cooper, and Allen, 2010; Abadzi, 2009). Table 7 shows that 29 schools who stated they applied the reform also lengthened teaching time to more than four hours. Four schools state that they implemented the reform in Grades 1 to 4, but did not increase the amount of teaching time to more than four hours. However, results are almost identical when I examine whether schools report implementing the reform rather than whether they actually increased teaching time.

5. Research design

I use a difference-in-differences design to compare outcomes among children attending reform schools before and after the reform was implemented, using the group of children attending shift schools as a control group to remove any trends in achievement over time. This identifies the average treatment effect on the treated (ATT) of a policy which lengthens the school day:

$$
\begin{aligned}
(5.1)\triangle^{ATT} &= \mathrm{E}\left[A(1)|\mathbf{Reform}=1,\,\mathbf{Post}=1,\,X\right]\right] - \mathrm{E}\left[A(0)|\mathbf{Reform}=1,\,\mathbf{Post}=1,\,X\right]\right] \\
&= \mathrm{E}\left[A(1)|\mathbf{Reform}=1,\,\mathbf{Post}=1,\,X\right]\right] - \mathrm{E}\left[A(0)|\mathbf{Reform}=1,\,\mathbf{Post}=0,\,X\right]\right] \\
&\quad -\mathrm{E}\left[A(0)|\mathbf{Reform}=0,\,\mathbf{Post}=1,\,X\right]\right] - \mathrm{E}\left[A(0)|\mathbf{Reform}=0,\,\mathbf{Post}=0,\,X\right]\right]
\end{aligned}
$$

Reform $= 1$ for students in schools which applied the full-day reform. **Post** $= 1$ for students in the younger cohort surveyed in 2009, when the full-day reform had been implemented in some schools. **Post** $= 0$ for students in the older cohort surveyed in 2002, before the reform was applied (the 11 children in four schools which were already teaching a full day are not included in this sample). $A(1)$ is achievement scores among students who were in reform schools when they had applied the reform and received the longer school day. The counterfactual mean $\mathrm{E}\left[A(0)|\mathbf{Reform}=1,\,\mathbf{Post}=1,\,X\right]\right]$, the scores of students in reform schools in 2009 if the reform had not occurred, is unobserved. Elsewhere, $A(0)$ is achievement for students who received a shorter school day, either because they were in schools which did not apply the reform or because they were in schools which applied the reform in the years after they were surveyed. The ATT is valid only for the group of children in reform schools and does not necessarily predict the effect that lengthening the school day would have on children in shift schools or on the sample as a whole.

The second equality follows from the "equal time trends" assumption. This assumption is that the unobserved counterfactual mean $\mathrm{E}\left[A(0)|\mathbf{Reform}=1,\,\mathbf{Post}=1,\,X\right]\right]$ in Equation 5.1 is equal to the time-adjusted mean outcome for children in reform schools in 2002:

$$\text{(5.2)} \quad \mathrm{E}\left[A(0)|\textbf{Reform}=1,\, \textbf{Post}=1,\, X]\right] = \mathrm{E}\left[A(0)|\textbf{Reform}=1,\, \textbf{Post}=0,\, X]\right] + \mathrm{E}\left[A(0)|\textbf{Reform}=0,\, \textbf{Post}=1,\, X]\right] - \mathrm{E}\left[A(0)|\textbf{Reform}=0,\, \textbf{Post}=0,\, X]\right]$$

This is equivalent to assuming that, in the absence of the reform, children's average unobserved characteristics would have changed over time in the same way for children in shift and reform schools and that unobserved characteristics at school level would also have changed over time in the same way for shift and reform schools:

$$\text{(5.3)} \quad \mathrm{E}\left[A(1)|\textbf{Reform}=1,\, \textbf{Post}=1,\, X]\right] - \mathrm{E}\left[A(0)|\textbf{Reform}=1,\, \textbf{Post}=0,\, X]\right] = \mathrm{E}\left[A(0)|\textbf{Reform}=0,\, \textbf{Post}=1,\, X]\right] - \mathrm{E}\left[A(0)|\textbf{Reform}=0,\, \textbf{Post}=0,\, X]\right]$$

This assumption can be argued to be plausible but cannot be tested directly. The next two sections discusses the plausibility of this assumption in relation to the characteristics of children and schools.

I estimate $\triangle^{ATT}$ using variants of a linear regression model:

$$\text{(5.4)} \quad A_i = \beta_0 + \beta_1 \mathrm{Post}_t + \beta_2 \mathrm{Refor}m_s + \triangle \mathrm{Post}_t * \mathrm{Refor}m_s + g(X_i) + \epsilon$$

A_{ist} is the achievement score of child i in school s at time t. $f(X_i)$ is a function of covariates of achievement at child level. $g(Y_s)$ is a function of covariates of achievement at school level. The error term ϵ contains a school-level error term μ_s and a child-level error term v_{ist}. β_1 captures trends in achievement between 2002 and 2009 common to children in both types of school. β_2 captures time-invariant differences between students in shift and reform schools.

If $\mathrm{Post}_t = 1$ and $\mathrm{Reform}_s = 1$, the child attended a reform school in 2009, after the reform was applied in that school. $\triangle$ is the difference-in-differences estimate of the impact being in a reform school compared to being in a shift school. If it is positive and significant, changing from shift to full-day schooling has improved achievement for children in reform schools. $\triangle$ estimates $\triangle^{ATT}$ consistently if $f(X_i)$ and $g(Y_s)$ are specified correctly and if the equal time trends assumption holds.

The dataset examines the same schools over time and captures information on two cohorts of children, one attending school in 2002 and one attending in 2009. It is thus a repeated cross-section of children but a panel of schools. Additionally, there is very limited data available on the quality of children's schools in 2002, as a school survey was not conducted.

I examine three different specifications. Model I does not include $f(X_i)$ or $g(Y_s)$, but is restricted to the sample of children for whom all control variables X_i and Y_s are available. Model II, which corresponds to Equation 5.4, includes child- and school-level controls and grade fixed effects ϕ_g. $f(X_i)$ and $g(Y_s)$ are assumed to be linear for simplicity. The controls reduce bias from any differences in the type of children in shift and reform schools in the pre- and post-reform cross-sections. Substantial changes in the student body in shift or reform schools may still be of concern if they indicate problematic selection, an issue discussed further below. Model III, the most complete, adds school-specific intercepts μ_s to control for time-invariant differences between schools not due to treatment and for the mean over time of any school-year shocks μ_{st}. $school * cohort$ or $school * year$ shocks μ_{st} cannot be included: they would be perfectly collinear with the treatment effect Post $*$ Reform. But there is a large number of schools, so it is plausible that any $school * year$ shocks average out to zero. The issue is then how one deals with clustering in the error term.

$$A_i = \beta_0 + \beta_1 \mathrm{Post}_t + \triangle \mathrm{Post}_t * \mathrm{Reform}_s + f(X_i) + \mu_s + \phi_g + \epsilon \tag{5.5}$$

The reading and writing variables are ordinal: categories can be ranked but the distances between adjacent categories are not known. I therefore present ordered logit models for the reading and writing variables and a logit model for the numeracy variable. To control for school fixed effects for the binary logit model, I use the Chamberlain (1980) conditional logit estimator. There is no similarly well-established fixed effects ordered logit estimator. Often, researchers collapse the ordinal variables into binary variables and estimate models with fixed effects using the conditional logit model. But here, in models without fixed effects, the threshold parameters are statistically different from each other, so the categories should not be collapsed. Instead, I use one of a range of new estimators: the so-called "Blow Up and Cluster" estimator, proposed by Staub, Winkelmann, and Baetschmann (2011).[4]

In all non-linear models, the coefficients displayed are odds ratios. A coefficient larger than one is equivalent to a positive coefficient in a linear regression, while a coefficient smaller than one is equivalent to a negative coefficient. It would be preferable to present marginal effects, but it is not possible to compute the marginal effect of an interaction term in the conditional logistic model (Karaca-Mandic, Norton, and Dowd, 2011, 267).

I use Liang and Zeger (1986) standard errors to account for clustering at the school level, rather than at the $school * year$ level.[5] Students at the same school, both at the same time and in different time periods, are likely to share unobservable characteristics, resulting in serial correlation in the error term (Moulton, 1986). In addition, while additive time effects are included in all models and additive school effects are included in Model III, $school * year$ shocks μ_{st} are not included in any regression. There are a number of schools and two time periods, so it is assumed that $E(\mu_{st}) = 0$,

[4]Instead of selecting one cut-off point to dichotomise the ordinal variable, they "blow up" the sample size by replacing every observation with $K - 1$ copies of itself and dichotomise every copy of the individual at a different cut-off point. They then estimate the conditional maximum likelihood logit using the whole sample. Some groups contribute to several terms in the log likelihood (e.g. a school where one child scores four and one scores one will contribute to the score if the cut-off is two or three), so there is dependence between these terms which is accounted for in the standard errors by using a cluster-robust variance estimator. The sample sizes displayed include the replications of each individual.

[5]Estimation techniques that account for clustering use a pseudo-likelihood, rather than the true likelihood, to estimate coefficients, so likelihood ratio tests cannot be used (Chambers and Skinner, 2003, 22). I therefore use an adjusted Wald test, which approximates a likelihood ratio test asymptotically and if the model is correctly specified. I base inference on a t distribution with $g - k$ degrees of freedom, where g is the number of groups, rather than on the standard normal distribution, because Liang and Zeger (1986) standard errors can be unreliable if there are fewer than about 100 clusters. Hansen (2007) finds when the t distribution is used, standard errors provide reasonably good corrections for serial correlation in panels with only ten clusters (across the different specifications used, the smallest number of schools is 72).

even if some shocks occur. However, some individual μ_{st} may not be equal to zero. Bertrand, Duflo, and Mullainathan (2004) argue adjustments must also be made for serial correlation in the μ_{st} themselves.

Finally, I examine treatment effects for different types of children. In Equation 5.6, for example, I examine if there are differences in the effect of being in a reform school for boys and girls. If $\triangle_2$ is significantly larger than zero, this suggests that the treatment effect on the treated of being in a reform school after the reform was implemented is larger for boys than girls. $\triangle_1$ gives the treatment effect for girls and $\triangle_1 + \triangle_2$ gives the treatment effect for boys.

$$A = \beta_0 + \beta_1\text{Post} + \triangle_1\text{Post} * \text{Reform}+\triangle_2\text{Post*Reform*Male} + f(X) + \phi_g + \epsilon \tag{5.6}$$

6. Plausibility of the equal time trends assumption

To obtain valid estimates from a differences-in-differences model, the equal trends assumption must be plausible. The assumption is that, in the absence of the reform, average unobserved characteristics of shift and reform schools and the children in them would have changed over time in the same way in shift and reform schools. In this section, I examine whether three violations of this assumption occur. This does not demonstrate that the equal trends assumption holds, but merely that three common violations of this assumption do not occur.

Firstly, if shift and reform schools or the children in those schools are very different on observables at baseline, this suggests that children's unobserved characteristics, or the unobserved characteristics of schools, are unlikely to change in the same way over time. Secondly, if the reform induced migration of children or teachers into or out of reform or shift schools, it might be these changes in composition of the children in shift and reform schools, rather than the reform, which cause differences in outcomes. Finally, there may be shocks or policy changes which are correlated with treatment.

Importantly, the equal time trends assumption does not require that treatment is randomly assigned, and random assignment of treatment is not sufficient for the equal trends assumption to hold. If the reform had been randomly assigned, shift and reform schools and the children in them would not differ in expectation on observables or unobservables at baseline and would change over time in similar ways. However, treatment-induced changes in composition of students in shift or reform schools and policy changes or shocks correlated with treatment could still jeopardise identification of the treatment effect.

6.1. **Incomparable treatment and control groups.** Table 8 presents descriptive statistics on the children in the sample. In Round 1 in 2002, very limited achievement measures were collected with the older cohort. These were repeated with the younger cohort in Round 3 in 2009. The numeracy item required children to solve a basic multiplication problem (2 x 4). The reading item required children to read three letters (T, A, H), one word ('hat'), and one sentence ('the sun is hot'). Children were coded into four categories: unable to read, able to read letters, able to read words or able to read the sentence. In the writing item, a fieldworker asked the children to write the sentence ‘I like dogs’. Children were coded into three categories: not being able to write, writing with difficulty or errors or writing without errors.

It is more reasonable to assume equal changes over time for two similar groups than two very different groups. A common check is whether treatment and control groups are similar at baseline on observable variables and whether trends in observable variables are the same over time. Observable differences between groups are accounted for by the difference-in-differences specification. In all three specifications used here, the Post term captures any trends over time in achievement that occur across all schools, and the controls capture the effect on outcomes of any differences in observable variables from the pre- to post-reform period in the type of children in shift and reform schools. In Model III, the school fixed effects also account for all unobservable differences betweens schools that occur at baseline and do not change over time, and for the mean over time of any school-year shocks μ_{st}. The concern, however, is that substantial differences in observables between shift and reform schools may suggest that unobservable trends in the groups are not similar.

The first three columns of Table 8 describe children in shift and reform schools in 2002, before the reform was implemented, on outcome variables and on control variables at child level which may affect achievement. Column 7 describes differences between reform and shift schools for the older cohort and Column 9 gives the second difference: the difference in changes in control and outcome variables between pre- and post-reform students in shift and reform.

Shift and reform schools were largely similar at baseline. There are significant differences at baseline in the proportion of children who can read a sentence: more children in schools which later applied the reform can read a sentence. However, there are no other differences in outcome variables at baseline. There are some differences in control variables at baseline. Children in reform schools are more likely to have mothers with some education and to be wealthier. There is only limited data available on schools at baseline. A higher proportion of children in reform than shift schools also attend an urban school at baseline, but there are no major pre-treatment differences between children in shift and reform schools in whether their schools cover eight grades. Schools which cover fewer than eight grades tend to have been more recently built and have fewer resources. More importantly, the difference-in- differences estimates in Column 9 are not significantly different from zero in all observable covariates, indicating that time trends in observable variables in shift and reform schools on these covariates were approximately equal.

I present an additional check in Table 10. I fit regression models only for children in 2002, before the reform was applied. If reform schools had systematically lower or higher achievement before the reform was implemented, this might suggest that shift and reform schools were likely to change in different and unobservable ways. Model I does not include controls; while Model II includes child controls, school controls and grade fixed effects. There are no significant pre-treatment differences between shift and reform schools in writing or numeracy. There are significant differences in reading in Model I, with no controls. This is to be expected from the descriptive statistics. However, these disappear when controls are added for child and school variables. However, results for reading should be treated with some caution as they may reflect pre-treatment differences between shift and reform schools.

The second reason to suggest that shift and reform schools are comparable is that interviews with regional education bureau officials indicate that there was no policy of targeting particular types of schools systematically for the full-day reform. Particular types of districts were not targeted to implement the reform and schools did not have to have certain numbers of classrooms or teachers available to implement the reform. In addition, the probability that a school implements the reform is not correlated with a wide range of observable variables, indicating that district-level bureaucrats were not unwittingly targeting particular types of schools, at least on observable variables. Table 9 shows a linear regression which examines associations between the probability that a school applied the full-day reform in Grade 1–4 between 2002 and 2009 and two sets of control variables. Model I includes control variables describing characteristics of the school at baseline in 2002. Model II includes a more extensive set of control variables from the school survey, but these are measured in 2009 after the reform was implemented, so reverse causality is possible. In Model II, implementing the reform is not correlated with any observable variables, including the average level of mothers' education, average wealth of children in the school, the age of the school, measures of school quality and the school's other contact with the district and obedience to other regulations. It remains possible that implementing the reform is correlated with unobservable variables.

6.2. **Changes in group composition.** One might expect parents to think more instructional time signals a better quality school and move children into schools which had applied the full-day reform, resulting in differences in unobservables (such as parental motivation or child ability) between children in shift and reform schools. Poorer students might also moved out of reform schools in response to the reform as they could not take time from work to attend a longer school day, also resulting in changes in the composition of the student body on unobservable variables. However, this issue is of limited concern in this context.

Firstly, in Ethiopia, children have fewer choices of school than in many other countries. In rural areas, there is usually one Grade 1–4 primary school per village, and one Grade 1–8 primary school for every three or four villages. In this sample, even children who had more than one choice of school rarely had the ability to choose between shift and reform schools. Table 7 shows that in 17 of 30 districts, either all the schools had applied the reform or no schools had.

Furthermore, even if parents had a choice of school, they often do not have the resources to take advantage of these choices. Distance to school strongly influences where parents send children to school, possibly because of the cost of transport, worries that girls will not be safe on the way to school, or the fact that even the closest school can be quite far away (Camfield, 2009). Parents are also unlikely to be aware of exactly when and in which grades a school was implementing the reform in order to select a school or to move children to a different school. Indeed, the difference-in-differences estimates in Column 9 of Table 8 are not significantly different from zero in all observable covariates, indicating that time trends in the composition of the student body in shift and reform schools on these covariates were approximately equal. There is no evidence that certain types of children moved into or out of shift or reform schools in response to the reform.

Similarly, there may be a worry that once changes in the school day have been made, better, more motivated teachers may sort into schools based on the length of the day. But in Ethiopian government schools, teachers have limited ability to move between schools. Teachers must teach in a government school in the region where they trained for the first two years after finishing training and are allocated to schools by lottery. After teachers have completed this two-year period, they can apply to transfer school, but more transfers are requested than are granted, and transfers may take years. Table 11 presents descriptive statistics for a range of school quality variables in 2009. The third and fourth panel of Table 11 show that there are no significant differences in the qualification levels, motivation levels or absenteeism of teachers in shift and reform schools after the reform is implemented.

6.3. **Correlated shocks or policy changes.** If one of the schools experiences a policy change or shock at the same time as the policy change of interest, this may confound results. This does not seem to have occurred here. Firstly, the reform was not attached to any other policy reforms, according to interviews with federal officials and the ESDP-III documents (Ministry of Education, 2005). Federal officials report that the curriculum was not changed for reform schools and the schools were not assessed differently in the National Learning Assessment. In other studies, interventions to change instructional time were part of school management or funding reforms which affect other school-level variables, making it difficult to identify effects due only to variation in instructional time. In Chile, full-day schools got a one-time investment in school facilities and a permanent increase in the amount of the monthly public per-student subvention (Bellei, 2009). In Uruguay, the full-day school programme included funding for construction of new classrooms, a reduction in class size, teacher training in a new pedagogical model, provision of teaching materials and nutritional and health care support (Cerdan-Infantes and Vermeersch, 2007).

Secondly, reform schools do not have substantially higher or lower levels of per-student funding or physical resources. If implementing the reform had resulted in other unobserved changes in reform schools, such as improvements in resources because of extra funding or, alternatively, an increase in overcrowding, the effect of the full-day reform may pick up the effect of these changes, not the effect of a change in instructional time. However, in this sample, as shown in Table 11, there are few significant differences between shift and reform schools in basic resources, on measures of overcrowding or on other measures of school quality after the implementation of the reform.[6] There are two exceptions. After the reform, reform schools have fewer brick classrooms, as might be expected if they built extra classrooms with local materials to cater for teaching all children at once. Reform schools also have significantly lower total enrolment than shift schools, which is expected as shift schools are designed to serve double the number of children with the same resources.

There are two possible explanations for the similarity in post-reform resource levels. Reform schools may have had similar resource levels to shift schools at baseline and been given enough resources to keep per-student resource levels the same as in shift schools. Alternatively, these schools may have had more classrooms, teachers and resources than shift schools before they implemented the reform. Given the scarcity of resources in Ethiopia in general, it is unlikely that schools had spare resources lying unused, so it is likely schools received small amounts of extra resources to implement the reform and maintain the same level of per-student resources.

[6]It is not possible to examine pre-treatment differences in school quality at baseline because there is no baseline data on school quality.

This pattern agrees with policy documents about how the reform was implemented. Reform schools continued to receive the same per-student grant of funding for running expenses (Dom, 2010). Schools often constructed extra classrooms or bathrooms to implement the shift system, but any new classrooms or toilet facilities required in schools should be built with unpaid labour and local materials donated by the community, while the district donated only more expensive building materials (Ministry of Education, 2005, 40), as was done for other school expansions in the period (Garcia and Rajkumar, 2008). If anything, reform schools may have been disadvantaged by implementing the reform and having to stretch the same physical resources over more students, so the treatment effect may be attenuated by reform schools having lower levels of resources, but is unlikely to be upward-biased.

Finally, it might be expected that lengthening the day increase teachers' teaching loads, so that teachers in reform schools may have been more tired or demotivated than those in shift schools. Federal officials noted that teachers were not paid more in schools which lengthened the day. However, the limited evidence suggests that before the reform, teachers often taught both shifts, so teachers' teaching loads in reform schools were actually lower than or relatively similar to those in shift schools.

There is no national data on teaching loads, but a teacher census in Oromia Region (the country's largest region) in 2002/3, shown in Table 12, found an average teaching load of 26.0 hours for Grade 1–4 teachers (39 lessons a week). Average loads were 21.5 hours (33 lessons) for Grade 1–4 teachers in urban areas and 28.6 hours (43 lessons) in rural areas (World Bank, 2005, 145).[7] Teachers in rural Oromia were thus on average teaching in more than one shift, as there are only 30 lessons in a week in shift schools. In rural schools which applied the full-day reform, teachers' teaching loads would actually have decreased from 43 to 35 lessons. In urban schools loads would have increased marginally from 33 to 35 lessons. If other regions were similar to Oromia, moving from shift to full-day would have caused little change in teaching loads for most teachers.

7. Results

In Table 13, I show difference-in-difference estimates of the effect of schools changing from shift to full-day on the achievement of children aged eight. Model I includes only variables reflecting whether the school implemented the reform. Model II adds child and school controls and grade fixed effects. Model III, which controls for child- and household-level factors and school fixed effects, provide the most complete estimates. School-level variables are not identified in Model III because they only change over time in a tiny fraction of schools, so they are not included.

The full-day reform improves writing and numeracy scores significantly. Ordered logit and logit estimates for Model III, which includes school fixed effects, show that children in schools which have implemented the full-day reform are 2.17 times more likely to be numerate than children in those same schools before the reform was applied. Model II, without school fixed effects, gives a similar estimate of 2.74. For writing, in Model III, children in schools which have implemented the full-day reform are 3.51 times more likely to be able to write with difficulty or write easily, versus not being able to write, than children in those same schools before the reform was applied (they

[7]This calculation assumes lessons are 40 minutes long.

are also 3.51 times more likely to be able to write easily versus not being able to write or being able to write with difficulty). The odds ratio from Model II, without school fixed effects, is similar at 4.18.[8] The full-day reform has a positive but not significant effect on reading, although these results should be treated with slight caution because of pre-treatment differences between reform and shift schools.

The control variables behave as expected. In all specifications for all achievement measures, boys perform better than girls, although the differences are not significantly different from zero. Children who are stunted perform less well on all achievement measures. Children whose households are wealthier perform better, although in regressions for writing and numeracy with school fixed effects, the coefficient on the wealth index is positive but not significant. This suggests household wealth is correlated with time-invariant school-level unobservables, which may capture variation in school quality. In Model II, which does not include school fixed effects, children in urban schools perform better on all measures. Children in schools with fewer than eight grades, which tend to have lower levels of resources, perform worse, but coefficients are not statistically significant.

In Table 14, I examine whether treatment effects differ for different types of children. Most data on children's achievement at school level is collected at schools and does not contain child- or household-level variables, but this data matches household data with data on schools. In equation 5.6, the treatment effect for girls is the coefficient $\triangle_1$. The treatment effect for boys is $\triangle_1 + \triangle_2$. The coefficient labelled 'Difference' in Table 14 is $\triangle_2$, the coefficient on $\text{Post} * \text{Reform} * \text{Male}$, or the difference between the treatment effect for boys and the treatment effect for girls. If $\triangle_2 \neq 0$, there is a significant difference in the treatment effects for boys and girls.

Lengthening the school day does not appear to affect reading scores, and this is true for all groups of children examined. For both writing and numeracy, however, a consistent pattern emerges. Girls who experience the full-day reform see a larger treatment effect on their writing and numeracy scores than boys do (differences are statistically significant for numeracy but not writing). This is possibly because Ethiopian girls are expected to do many hours of domestic chores when at home (Camfield, 2009; Poluha, 2007). Boys also work for the household, but often do activities such as herding cattle while they graze, which allow them to study Orkin (2012). When the school day is short, boys may be able to spend time on schoolwork outside school to learn literacy and numeracy skills independently. Girls do not have as much free time, so attending a full-day school where more of their learning occurs during the formal school day has larger benefits for girls than boys.

Lengthening the school day has a larger treatment effect for children whose households have above the median level of wealth for their district (again, differences are significant for numeracy at the ten per cent level, but not for writing). Children whose households have below the median wealth level also benefit, but benefits are larger for children from wealthier families. This suggests that wealthier children have better nutrition and access to school materials and thus learn more in the same amount of time than poorer children. Similarly, stunted children do not benefit from

[8]Estimates of the treatment effect are not sensitive in magnitude, sign or significance to the imposition of various sample restrictions. One such restriction examines only children in schools where there is one child in each cohort. Another such restriction examines only the sample of children examined in Model III, who are in schools where there is within-school variation in outcomes.

increase in time in instruction, while children who are not stunted derive a large benefit.[9] Stunted children are likely to learn less in a unit of time than non-stunted children as they may battle to concentrate. It is also likely that long-term stunting is correlated with poor current levels of nutrition, and children who are hungry are unlikely to benefit from more time at school.

Finally, there is suggestive evidence that higher scores among children who experience the full-day reform may be driven largely by substantial improvements by children in better-quality schools. Children in urban reform schools do substantially better than children in urban shift schools. In contrast, children in rural reform schools do not do significantly better than children in rural shift schools. The quality of schooling is usually higher in urban than rural areas. This is unsurprising: children derive more benefit from an increase in time in instruction when the quality of instruction is higher. Similarly, as noted above, being from a wealthier family appeared to be correlated with time-invariant school unobservables which might reflect school quality. The result that wealthier children experience higher treatment effects may also suggest that children in better schools benefit more from an increase in time than children in less good schools.

8. Conclusion

In qualitative interviews, national and regional officials and nearly all teachers and principals articulated why they believed teaching for a full day improves education quality: an increase in instructional time gives teachers more time to cover the curriculum, explain concepts in different ways and use different methods of instruction; and gives children more time-on-task in a classroom setting to absorb concepts and practise skills. This paper shows that lengthening the amount of instructional time children receive has large, positive effects on writing and maths scores, but not on reading.

It is unsurprising that effects are larger than others in the literature. In Ethiopia, most schools have only four hours of instructional time before the reform, so instructional time increases by roughly 30 per cent because of the reform. Other research has largely examined increases of roughly 10 per cent in instructional time in countries where the school days were longer to begin with. Other work finds that there are diminishing marginal returns to instructional time (Lavy, 2010). If there are diminishing marginal returns, increases in instructional time will have larger effects when there are low initial amounts of instruction than when the initial amounts of instruction were higher.

Theoretical work on instructional time also predicts that there will be diminishing marginal returns to instructional time. Levin and Tsang (1987) explicitly model student time and effort as inputs into the education production function.[10] They assume that there are trade-offs between the time spent in school and the effort students put into learning. Students have an optimum amount of learning activity that maximises their utility. If a student is at their optimum amount of learning and there is an externally imposed increase in time spent in learning, they will respond

[9]Differences are not statistically significant, but this may be because a small proportion of the sample is stunted so the test has limited power.

[10]General education production functions often include years spent in formal schooling as an input into the production of achievement (Hanushek, 1979). This captures both the completion of a number of years of the curriculum and time spent in school. However, it is difficult to capture what happens when children spend more or less time in school covering the same amount of material. This can be examined in the Levin and Tsang (1987) model.

by reducing effort so that their amount of learning is unchanged. If the increase in time is small, the model predicts students can reduce effort so there is no overall increase in learning activity. If the increase is large, students cannot reduce effort enough to achieve equilibrium. Achievement is predicted to increase by a small amount, because the increase in time in school is slightly larger than the decrease in student effort. This prediction has been accurate in most studies of increases in instructional time in developed countries. However, if children are receiving a small initial amount of instruction, they may not yet at the equilibrium point where they are engaging in the optimum amount of learning activity, and therefore that there is some room to increase time in instruction, overall learning activity and hence student scores without triggering a reduction in student effort.

However, this study still finds larger effects than other studies examining increases in time of a similar magnitude from a similar short initial school day. Only two other studies examine increases in time of a similar magnitude from a similar short initial school day. Cerdan-Infantes and Vermeersch (2007) find that a 100 per cent increase in Grade 6 from a three-and-a-half to a seven hour day in Uruguay increased student test scores in third grade by 0.063 standard deviations per year in mathematics and 0.044 standard deviations in language. These effect sizes are roughly equivalent to odds ratios of 1.121 for maths and 1.095 for language.[11] Bellei (2009) finds that a 27 per cent increase from a four-and-a-half to six hour day in Grade 10 in Chile increased student test scores in third grade by 0.07 standard deviations per year in mathematics and 0.05 standard deviations in language. These effect sizes are roughly equivalent to odds ratios of 1.135 for maths and 1.082 for language.

This plausibly occurs because the study is conducted with younger children than others to date. Marcotte (2007) and Lavy (2010) find greater effects of changes in instructional time for younger children. Similar results are found in other studies of instructional time which do not focus exclusively on lengthening the school day. In a study of allocation of time between activities in Californian classrooms, Brown and Saks (1987) found total elasticities of achievement with respect to total time allocated to teaching were 0.24 and 0.13 for maths and reading scores respectively for Grade 2 children, but were no different from zero and 0.07 for Grade 5 children. They argue that learning more advanced material may require students to learn mainly outside class: for example, second grade reading requires decoding skills, which are learned well in intensive reading classes, but by fifth grade, comprehension and inference are required, which probably require self-study as well as more time in instruction. Similarly, Jacob and Lefgren (2004) find that low-achieving Chicago Grade 3 children forced to attend a summer school which revised the year's material had reading and maths scores 12 per cent of the average annual learning gain higher after two years. The effect was half as large for Grade 6 children.

This paper shows that more instructional time has greater effects on maths than on reading and writing. This also agrees with the literature. Of the natural experiment studies which examine subjects separately, all but one find instructional time has positive effects on maths, but much smaller or no effects on reading or language (Sims, 2008; Brown and Saks, 1987; Lavy, 2012; Bellei,

[11] The logistic and Normal distributions differ little, except in the tails of the distributions. The standard logistic distribution has variance $\frac{\pi^2}{3}$. A difference in normal equivalent deviate can be converted into the approximate difference in log odds by multiplying by $\frac{\pi}{\sqrt{3}}$, which is 1.81 to 2 decimal places. Log odds can then be transformed to an odds ratio (Chinn, 2000).

2009; DeCicca, 2007; Marcotte, 2007). Marcotte (2007) hypothesises that in maths students receive least guidance at home and may require more review, and that maths curricula tend to be more rigid and content-focussed than reading or language. In cross-country work, Lee and Barro (2001) find increased instructional time improves maths scores but actually decreases reading scores. They argue that children read in their spare time but do not practise maths and science, for which instruction and practice in classrooms is important.

Finally, the paper finds significant differences in the effect of the reform on different types of children. The reform has larger benefits for children in urban schools who are also likely to be in better schools. It also has larger benefits for children who may learn more in from a unit of time, such as children who are not stunted and children who are wealthier and may have better access to school materials and nutrition. Finally, it has larger benefits for girls, who may struggle to spend time on schoolwork outside of school because of household chores. Boys in shift schools may be more able to spend time on homework to compensate for a shorter school day.

This research has a number of policy implications. Firstly, increasing instructional time is more effective for younger children. Currently many primary schools in Ethiopia implement the reform only in the higher grades and many countries lengthen the school day as children get older, but this may not be an optimal use of resources. Secondly, the reform on its own may exacerbate inequalities between children, without other initiatives to improve the quality of instructional time or to address disadvantages such as poverty or poor nutrition that affect children's ability to learn during the instructional time that they receive. Thirdly, there are often concerns about children's ability to balance work and school when the school day is longer. However, this study finds that attendance is no lower in schools with a longer day. In addition, extra instructional time can be provided in other ways. The international research suggests that the effects of extra instructional time are similar if it is the year or the school day that is lengthened. In Senegal, for example, shift schools teach for an extra 10 days a year. In Hong Kong, shift schools teach every second Saturday (Bray, 2008, 26). This may be a plausible way of achieving the benefits of an increase in instructional time without the costs to children of less flexibility in their time allocation.

References

Abadzi, H. 2009. "Instructional time loss in developing countries: Concepts, measurement, and implications." *The World Bank Research Observer* 24:267–290.

Agüero, J.M., and T. Beleche. 2013. "Test-Mex: Estimating the effects of school year length on student performance in Mexico." *Journal of Development Economics* 103:353–361.

Bellei, C. 2009. "Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile." *Economics of Education Review* 28:629–640.

Bertrand, M., E. Duflo, and S. Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics* 119:249–275.

Bray, M. 2008. *Double-shift schooling: Design and operation for cost-effectiveness*. Paris: UNESCO: International Institute for Educational Planning.

Brown, B.W., and D.H. Saks. 1987. "The microeconomics of the allocation of teachers' time and student learning." *Economics of Education Review* 6:319–332.

Camfield, L. 2009. "'A girl never finishes her journey': Mixing methods to understand female experiences of education in contemporary Ethiopia." *Research Papers in Education* 26:1–11.

Cerdan-Infantes, P., and C. Vermeersch. 2007. "More time is better: An evaluation of the full-time school program in Uruguay." *World Bank Policy Research Working Paper* 4167:1–25.

Chamberlain, G. 1980. "Analysis of covariance with qualitative data." *Review of Economic Studies* 47:225–238.

Chambers, R.L., and C.J. Skinner. 2003. *Analysis of survey data*. Chichester, UK: John Wiley & Sons.

Chinn, S. 2000. "A simple method for converting an odds ratio to effect size for use in meta-analysis." *Statistics in Medicine* 19:3127–31.

DeCicca, P. 2007. "Does full-day kindergarten matter? Evidence from the first two years of schooling." *Economics of Education Review* 26:67–82.

Dom, C. 2010. *Ethiopia country desk study: Mid-term evaluation of the Education For All Fast Track Initiative*. Cambridge: Cambridge Education, Mokoro Consultants and Oxford Policy Management.

Eide, E., and M.H. Showalter. 1998. "The effect of school quality on student performance: A quantile regression approach." *Economics Letters* 58:345–350.

Engel, J. 2010. *Ethiopia's progress in education: A rapid and equitable expansion of access*. London: Overseas Development Institute.

Fitzpatrick, M.D., D. Grissmer, and S. Hastedt. 2011. "What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment." *Economics of Education Review* 30:269–279.

Garcia, M., and A.S. Rajkumar. 2008. "Achieving better service delivery through decentralization in Ethiopia." *World Bank Working Paper: African Human Development Series* 131:1–134.

Glick, P., and D.E. Sahn. 2009. "Cognitive skills among children in Senegal: Disentangling the roles of schooling and family background." *Economics of Education Review* 28:178–188.

Hansen, C.B. 2007. "Asymptotic properties of a robust variance matrix estimator for panel data when T is large." *Journal of Econometrics* 141:597–620.

Hanushek, E.A. 1979. "Educational production functions." *Journal of Human Resources* 14:351–388.

Jacob, B., and L. Lefgren. 2004. "Remedial education and student achievement: A regression-discontinuity analysis." *Review of Economics and Statistics* 86:226–244.

Karaca-Mandic, P., E.C. Norton, and B. Dowd. 2011. "Interaction terms in nonlinear models." *Health Services Research* 47:255–275.

Kebede, M. 2006. "The roots and fallouts of Haile Selassie's educational policy." *UNESCO Forum occasional paper series* 20:1–35.

Kremer, M., and A. Holla. 2009. "Improving education in the developing world: What have we learned from randomised evaluations?" *Annual Review of Economics* 1:513–542.

Lam, D., C. Ardington, and M. Leibbrandt. 2011. "Schooling as a lottery: Racial differences in school advancement in urban South Africa." *Journal of Development Economics* 95:121–136.

Lavy, V. 2010. "Do differences in school instruction time explain international achievement gaps in math, science and reading? Evidence from developed and developing countries." *National Bureau of Economic Research Working Paper* 16227:1–48.

—. 2012. "Expanding school resources and increasing time on task: Effects of a policy experiment in Israel on student academic achievement and behavior." *National Bureau of Economic Research Working Paper* 18369:1–45.

Lee, J.W., and R.J. Barro. 2001. "Schooling quality in a cross-section of countries." *Economica* 68:465–489.

Levin, H.M., and M. Tsang. 1987. "The economics of student time." *Economics of Education Review* 6:357–364.

Liang, K.Y., and S.L. Zeger. 1986. "Longitudinal data analysis using Generalized Linear Models." *Biometrika* 73:13–22.

Marcotte, D.E. 2007. "Schooling and test scores: A mother-natural experiment." *Economics of Education Review* 26:629–640.

Ministry of Education. 1998. *Education Sector Development Plan I (ESDP-I) 1997/1998-2001/2002: Programme action plan*. September, Addis Ababa: Government of Ethiopia.

—. 2005. *Education Sector Development Plan III (ESDP-III) 2005/2006-2010/2011: Programme action plan*, vol. 2011. Addis Ababa.

—. 2011a. *Education Sector Development Plan IV (ESDP-IV) 2010/2011-2014/2015: Programme action plan*. August 2010, Addis Ababa.

—. 2002a. *Education statistics annual abstract 1994 E.C./2001-2002 G.C.*. Addis Ababa: Government of Ethiopia.

—. 2003. *Education statistics annual abstract 1995 E.C./2002-2003 G.C.*. Addis Ababa: Government of Ethiopia.

—. 2011b. *Education statistics annual abstract 2003 E.C./2010-2011 G.C.*. Addis Ababa: Government of Ethiopia.

—. 2002b. *The Education and Training Policy and its implementation*, vol. 27. Addis Ababa: Government of Ethiopia.

Ministry of Education/USAID. 2008. *Ethiopian Third National Learning Assessment of Grade 4 Students*. Addis Ababa: General Education Quality Assurance and Examinations Agency, Federal Ministry of Education.

Moulton, B. 1986. "Random group effects and the precision of regression estimates." *Journal of Econometrics* 32:385–397.

Negash, T. 2006. "Education in Ethiopia: From crisis to the brink of collapse." *Nordiska Afrikainstitutet Discussion Paper* 33:1–56.

Orkin, K. 2012. "Are Work and Schooling Complementary or Competitive for Children in Rural Ethiopia?" In J. Boyden and M. Bourdillon, eds. *Childhood Poverty: Multidisciplinary Approaches*. Basingstoke: Palgrave Macmillan, pp. 298–315.

Patall, E.A., H. Cooper, and A.B. Allen. 2010. “Extending the school day or school year: A systematic review of research (1985-2009).” *Review of Educational Research* 80:401–436.

Poluha, E. 2007. *The world of boys and girls in rural Ethiopia*. Addis Ababa: Forum for Social Studies.

Sims, D.P. 2008. “Strategic responses to school accountability measures: It’s all in the timing.” *Economics of Education Review* 27:58–68.

Staub, K.E., R. Winkelmann, and G. Baetschmann. 2011. “Consistent estimation of the fixed effects ordered logit model.” *Institute for the Study of Labour (IZA) Discussion Paper* January:1–35.

Woldehanna, T., M. Alemu, and A. Tekie. 2008. “Young Lives Ethiopia Round Two survey report.” *Young Lives Survey Report* 2:1–80.

World Bank. 2005. *Education in Ethiopia: Strengthening the foundation for sustainable progress*. Washington, D.C.: World Bank.

Wößmann, L. 2003. “Schooling resources, educational institutions and student performance: The international evidence.” *Oxford Bulletin of Economics and Statistics* 65:117–170.

Table 1: Average instructional hours per year in primary schools in African countries

Country	Hours per day in full-day schools	Hours per day in shift schools	Percentage difference
Burkina Faso	603	858	42.3
Mali	645	888	37.7
Ethiopia	840	1,100	31.0
Ivory Coast	580	754	30.0
Guinea	585	747	27.7
Senegal	547	675	23.4
Ghana	772	903	17.0
Gambia	936	1,024	9.5
Zambia	804	804	0.0

Source: Abadzi (2009), World Bank (2005)

Table 2: Cost of implementing full-day schooling in selected countries

Country	Type of reform	Cost implications
Chile	Public schools required to teach a full day	Operational cost of public education increased 25 per cent
Jamaica	Move from single to double shift schooling	Saved 32 per cent of cost of facilities
Uruguay	New schools set up with an eight hour day	Schools cost 60 per cent more than normal schools
Zambia	Shortening of school day for rural Grade 5-7	Reduced costs of achieving UPE by 46 per cent
Vietnam	Estimate of cost to move from half to full day	Recurrent spending would increase by 40 per cent

Source: Bellei (2003), Leo-Rhynie (1981), Cerdan-Infantes and Vermeersch (2007), Bray, (2008), Carr-Hill (2010).

Table 3: The effect of increasing time in instruction by 10 per cent in various studies

Authors, Place	Unit of increase in time	Age/grade	Pooled	Reading	Language	Maths
Cross country						
Lee and Barro (2001), 58 developed countries	10 %, days in school year	Age 10 and 14, primary		1.8 percentage point decrease		2.6 percentage point increase
Wößmann (2003), 39 developed countries	10 %, minutes of instruction per year	Age 13, primary	0.015 SD increase (maths and science)			
Lavy (2010), 57 countries	10 %, hours of instruction per week	Age 15, secondary	Developed countries: 0.049 SD increase Developing countries: 0.024 SD increase			
Within student						
DeCicca (2007), US	10%, hours of instruction	Grade 1 White		0.028 SD increase		0.037 SD increase
		Grade 1 Black		0.011 SD increase		0.037 SD decrease
		Grade 1 Hispanic		0.039 SD increase		0.021 SD increase
Eide and Showalter (1998), US	10 %, days in school year	Grade 10, 12				25th quantile: no effect 50th: 0.1 SD increase 75th: 0.09 SD increase
Natural experiment						
Fitzpatrick, Grissmer, and Hastedt (2011), US	10 %, days in school year	Last year of kinder-garden		0.12 SD increase		0.09 SD increase
Sims (2008), Wisconsin, US	10 %, days in school year	Grade 3		No effect	No effect	No effect
		Grade 4		No effect	No effect	0.05 SD increase
Marcotte (2007), Maryland, US	Increase of one SD in snowfall for winter	Grade 3		0.78 % fewer got satisfactory	0.61 % fewer got satisfactory	1.2 % fewer got satisfactory
		Grade 5		No effect	0.56 % fewer got satisfactory	0.93 % fewer got satisfactory
		Grade 8		No effect	No effect	0.94 % fewer got satisfactory
Lavy (2012), Israel	10 %, hours of subject instruction per week	Grade 5	0.026 SD increase (English, maths and science)		0.023 SD increase	0.025 SD increase
Cerdan-Infantes and Vermeersch (2007), Uruguay	10 %, length of school day	Grade 6			0.009 SD increase	0.013 SD increase
?, Mexico	10 days in school year	Grade 4				0.04 SD increase
		Grade 5				0.07 SD increase
Bellei (2009), Chile	10 %, length of school day	Grade 10			0.023 SD increase	0.032 SD increase

Table 4: Young Lives survey waves used in this paper

Wave	Year	Younger cohort (born 2000)	Older cohort (born 1994)
Round 1	October–December 2002	6–18 months	7–8 years
Round 3	October 2009–March 2010	7–8 years	14–15 years
School survey	March–May 2010	7–8 years	14–15 years

Table 5: Number of full-day and shift schools in Ethiopia in 2002 and 2010

Region	2002/03				2010/11			
	Full-day	Shift	Total	Full-day (%)	Full-day	Shift	Total	Full-day (%)
Tigray	537	422	959	56.0	295	1 683	1 978	14.9
Amhara	1 562	1 432	2 994	52.2	2 449	4 735	7 184	34.1
Oromia	1 864	2 834	4 698	39.7	1 903	9 478	11 381	16.7
SNNP	1 111	1 279	2 390	46.5	1 237	3 906	5 143	24.1
Addis Ababa	78	201	279	28.0	2	726	728	0.3
Five regions	5 152	6 168	11 320	45.5	5 886	20 528	26 414	22.3
National	5 270	6 694	11 964	44.0	6 071	22 278	28 349	21.4

Source: Ministry of Education (2003, 88; 2011, 38).

Table 6: Children and schools in difference-in-differences sample

	(1)	(2)	(3)	(4)
	School survey sample		No missing values	
	Children	Schools	Children	Schools
Shift				
2002 cohort	566	46	551	46
2009 cohort	1007	77	934	74
Total	1573		1485	
Reform				
2002 cohort	84	17	83	17
2009 cohort	254	31	226	30
Total	338		309	
full-day				
2002 cohort	2	2		
2009 cohort	9	4		
Total	11			
Total				
2002 cohort	652	65	634	63
2009 cohort	1270	112	1160	104
Total	1922		1794	

Column 1 shows all the children attending schools that were covered by the school survey in 2010. Column 2 shows the number of schools they attend. 57 of the 65 schools attended by older cohort children are also attended by younger cohort children. Columns 3 and 4 show the children who do not have any missing values on variables used in the regressions in this paper and the schools they attend.

Table 7: Application of the full-day reform and teaching time in Young Lives districts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Whether school reports implementing reform				Teaching time: Grade 1–4	
Implementation within the district	Districts	Schools	Shift school, directive not received	Shift school, directive received but not implemented	School with reform applied in Grades 1-4	School with reform applied only in higher grades	Less than 4 hours	More than 4 hours
All shift	11	27	14	13	0	0	27	0
Majority shift	5	42	11	15	11	5	33	9
Equal numbers	3	8	3	1	4	0	4	4
Majority reform	5	21	1	4	8	8	14	7
All reform	6	16	0	0	9	7	8	8
Total	30	114	29	33	32	20	86	28

Column 1 categorises the 30 Young Lives districts according to the proportion of schools that have implemented the full-day reform. Column 2 shows, for example, that there are 27 schools in the 11 districts where all schools were still teaching in shift. Columns 3-6 give the number of schools in each type of district that report implementing the full-day reform in Grades 1 to 4. Column 3 shows shift schools that did not receive a directive from the district. Column 4 shows shift schools that received a directive but had not yet implemented it. Column 5 shows schools which implemented the reform in Grades 1 to 4 between 2002 and 2009 (they could also have implemented the reform in other grades). Column 6 shows schools which implemented the reform, but only in grades above Grade 4. In Columns 7 and 8, schools are classified according to the total amount of teaching time children receive in Grade 1 to 4. 32 schools state that they implemented the full-day reform in Grades 1 to 4, but four of these schools did not increase the amount of teaching time to more than four hours.

Table 8: Descriptive statistics at child level for eight-year-olds in 2002 and 2009

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Older cohort			Younger cohort			Differences		Second diff.
	Shift	Reform	Total	Shift	Reform	Total	(2) - (1)	(5) - (4)	ΔRef - ΔShi
Outcome variables									
Can't read anything	0.443	0.253	0.418	0.450	0.327	0.426	-0.190	-0.094	0.095
	(0.048)	(0.095)	(0.045)	(0.043)	(0.079)	(0.038)	(0.106)	(0.095)	(0.119)
Reads letters	0.270	0.157	0.256	0.202	0.146	0.191	-0.114	-0.060	0.054
	(0.039)	(0.048)	(0.036)	(0.019)	(0.033)	(0.016)	(0.062)	(0.041)	(0.077)
Reads words	0.073	0.072	0.073	0.086	0.080	0.084	-0.000	-0.013	-0.013
	(0.010)	(0.026)	(0.010)	(0.014)	(0.019)	(0.012)	(0.028)	(0.022)	(0.035)
Reads sentence	0.214	0.518	0.254	0.262	0.447	0.298	0.304**	0.167	-0.136
	(0.048)	(0.092)	(0.048)	(0.042)	(0.084)	(0.038)	(0.104)	(0.100)	(0.099)
Can't write	0.454	0.410	0.448	0.552	0.434	0.529	-0.044	-0.107	-0.063
	(0.054)	(0.090)	(0.048)	(0.057)	(0.084)	(0.049)	(0.105)	(0.108)	(0.124)
Writes with	0.241	0.313	0.251	0.297	0.261	0.290	0.072	-0.048	-0.120
difficulty	(0.027)	(0.057)	(0.025)	(0.037)	(0.042)	(0.031)	(0.063)	(0.058)	(0.083)
Writes without	0.305	0.277	0.301	0.151	0.305	0.181	-0.028	0.155*	0.183
difficulty	(0.042)	(0.094)	(0.039)	(0.028)	(0.062)	(0.026)	(0.103)	(0.073)	(0.095)
Multiplication	0.584	0.506	0.574	0.439	0.535	0.458	-0.029	0.069	0.098
correct	(0.040)	(0.042)	(0.035)	(0.036)	(0.060)	(0.031)	(0.072)	(0.070)	(0.086)
Child controls									
Male	0.499	0.506	0.500	0.528	0.509	0.524	0.007	-0.011	-0.018
	(0.015)	(0.070)	(0.016)	(0.019)	(0.033)	(0.017)	(0.071)	(0.040)	(0.088)
Stunted	0.261	0.253	0.260	0.161	0.133	0.155	-0.008	-0.016	-0.007
	(0.030)	(0.080)	(0.028)	(0.013)	(0.025)	(0.012)	(0.085)	(0.029)	(0.091)
Mother has	0.441	0.711	0.476	0.385	0.624	0.432	0.270**	0.213***	-0.056
education	(0.052)	(0.065)	(0.048)	(0.043)	(0.040)	(0.038)	(0.083)	(0.059)	(0.081)
Wealth index	0.225	0.357	0.243	0.332	0.376	0.340	0.131**	0.030	-0.102*
	(0.026)	(0.035)	(0.024)	(0.019)	(0.032)	(0.017)	(0.044)	(0.039)	(0.039)
Log household	1.795	1.785	1.794	1.770	1.744	1.765	-0.011	-0.026	-0.015
size	(0.031)	(0.047)	(0.028)	(0.024)	(0.034)	(0.020)	(0.057)	(0.044)	(0.060)
Grade child is	1.510	1.542	1.514	1.833	1.695	1.806	0.032	-0.122	-0.154
in	(0.049)	(0.120)	(0.046)	(0.077)	(0.095)	(0.066)	(0.129)	(0.129)	(0.147)
Hours/day on	35.082	30.422	34.472	1.540	1.518	1.535	-4.660	-0.096	4.564
chores	(4.935)	(3.777)	(4.355)	(0.097)	(0.147)	(0.083)	(6.214)	(0.183)	(6.223)
Age of child in	96.343	96.587	96.375	98.859	99.041	98.895	0.243	0.077	-0.166
months	(0.305)	(0.406)	(0.270)	(0.181)	(0.267)	(0.155)	(0.508)	(0.329)	(0.575)
School controls									
Urban school	0.417	0.795	0.467	0.316	0.553	0.362	0.378*	0.173	-0.204
	(0.108)	(0.128)	(0.098)	(0.090)	(0.142)	(0.077)	(0.168)	(0.172)	(0.121)
Fewer than 8 grades	0.114	0.145	0.118	0.439	0.460	0.443	0.030	0.039	0.009
	(0.061)	(0.079)	(0.054)	(0.094)	(0.134)	(0.080)	(0.100)	(0.173)	(0.146)
Children (Schools)	634 (63)			1160 (104)					

Column 7 is the difference between Columns 1 and 2. Column 8 is the difference between Columns 4 and 5. Column 9 is (Column 5 - Column 2) - (Column 4 - Column 1). If the p-value of the adjusted Wald test of the difference is significant, the difference is starred. *p<0.01 **p<0.05 ***p<0.01. T stats, adjusted Wald tests and cluster-robust standard errors are used, where the cluster is the school.

Table 9: Correlates of whether schools applied the full day reform between 2002 and 2009

	(I)	(II)
Demographics of children's parents		
Mean consumer durables index	0.22	0.61
	(0.36)	(0.46)
Mean housing quality index	−0.37	−0.49
	(0.32)	(0.36)
Mean mother's education	0.059**	0.044
	(0.024)	(0.028)
Mean stunting	0.15	−0.0063
	(0.20)	(0.22)
Period when school was established (Pre-Derg=0)		
Derg and Civil War	−0.29**	−0.23
	(0.13)	(0.14)
EPRDF government pre-1997	−0.063	−0.12
	(0.17)	(0.19)
ESDP-I onwards	−0.032	−0.015
	(0.15)	(0.17)
Grades covered (First cycle=0)		
Incomplete full cycle	0.089	0.13
	(0.17)	(0.18)
Full cycle or junior sec.	0.37***	0.45***
	(0.13)	(0.16)
Secondary	0.18	0.073
	(0.15)	(0.21)
Small town school (Urban school=0)	−0.24*	−0.26
	(0.15)	(0.17)
Rural school	−0.015	0.079
	(0.14)	(0.16)
Contact with region		
Directives not received out of 3		0.029
		(0.060)
Directives received but disobeyed out of 3		−0.17
		(0.11)
District officials visit school more than once/fortnight		−0.040
		(0.11)
General school quality		
% teachers with diplomas		−0.18
		(0.21)
Has piped drinking water		−0.057
		(0.12)
Constant	0.24	0.35
	(0.23)	(0.30)
Schools	112	97
F stat	2.87	2.12
Adj. R squared	0.17	0.17

Coefficients are from a standard linear regression. *p<0.10 **p<0.05 ***p<0.01. The regions issued three other national directives in the period, on kindergarten provision, automatic promotion and mother tongue education. Directives not received is the number of these directives the school did not receive. Directives disobeyed is the number disobeyed.

Table 10: Average fitted pre-treatment differences in scores for eight-year-olds in 2002

	Reading		Writing		Numeracy	
	(I)	(II)	(I)	(II)	(I)	(II)
Reform	3.30**	1.97	1.05	0.54	0.73	0.71
	(1.62)	(0.90)	(0.43)	(0.22)	(0.17)	(0.19)
Controls	No	Yes	No	Yes	No	Yes
School fixed effects	No	No	No	No	No	No
Grade fixed effects	No	Yes	No	Yes	No	Yes
Children	634	634	634	634	634	634
Schools	63	63	63	63	63	63
Adj. Wald test: F stat	5.93	14.28	0.02	7.59	1.81	2.29
Adj. Wald test: Prob>F	0.02	0.00	0.90	0.00	0.18	0.03

Coefficients are odds ratios, so a coefficient less than one denotes a negative effect. T stats, adjusted Wald tests and cluster-robust standard errors are used, where the cluster is the school. *p<0.10 **p<0.05 ***p<0.01.

Table 11: Descriptive statistics at school level in 2009

	(1) Shift	(2) Reform	(3) Total	(4) (2) - (1)
Basic school quality				
Fewer than 8 grades	0.507	0.500	0.505	-0.007
	(0.060)	(0.095)	(0.051)	(0.118)
Average classroom is brick	0.576	0.467	0.556	-0.327*
	(0.061)	(0.130)	(0.056)	(0.147)
Students have pit latrines	0.712	0.600	0.691	-0.274
	(0.056)	(0.127)	(0.052)	(0.160)
Students have flush toilets	0.091	0.267	0.123	0.278
	(0.036)	(0.115)	(0.037)	(0.150)
Has piped drinking water	0.545	0.600	0.556	-0.117
	(0.062)	(0.127)	(0.056)	(0.162)
Has electricity	0.431	0.467	0.438	-0.086
	(0.062)	(0.130)	(0.056)	(0.158)
Has library	0.591	0.533	0.580	-0.145
	(0.061)	(0.130)	(0.055)	(0.162)
Cluster resource centre	0.516	0.533	0.519	0.030
	(0.064)	(0.130)	(0.057)	(0.163)
Overcrowding				
Total enrolment in all grades	1011.328	719.632	944.554	-497.619***
	(91.373)	(144.758)	(79.027)	(136.909)
Enrolment in first shift/whole day	522.856	473.500	513.161	-103.073
	(60.973)	(98.587)	(52.751)	(108.649)
Pupil-teacher ratio	40.318	38.052	39.876	-2.452
	(1.363)	(2.807)	(1.230)	(3.949)
Pupil-classroom ratio	48.105	49.591	48.375	0.088
	(2.361)	(5.768)	(2.200)	(7.083)
Pupil-latrine ratio	91.963	90.912	91.739	-9.648
	(12.178)	(21.776)	(10.648)	(28.196)
Children usually sit on floor	0.338	0.400	0.350	0.016
	(0.059)	(0.127)	(0.054)	(0.157)
Number of classrooms	11.687	9.278	11.176	-2.643
	(1.154)	(1.424)	(0.964)	(1.875)
Staff qualifications, motivation, salary				
Percent of teachers with degrees	0.027	0.022	0.026	-0.016
	(0.011)	(0.014)	(0.009)	(0.015)
Percent of teachers with diplomas	0.536	0.465	0.522	-0.084
	(0.035)	(0.065)	(0.031)	(0.086)
Percent of teachers with Grade 10	0.131	0.057	0.116	-0.066
	(0.028)	(0.032)	(0.023)	(0.048)
Average annual salary, Grd 1-4 full time teacher	10037.111	10629.883	10146.306	-855.426
	(335.499)	(1032.362)	(334.334)	(1030.681)
Average teacher job satisfaction out of 5	1.826	1.857	1.832	-0.009
	(0.033)	(0.073)	(0.030)	(0.088)
Principal motivation index out of 5	4.645	4.667	4.649	-0.121
	(0.080)	(0.155)	(0.071)	(0.213
Yearly salary and benefits	17589.677	14777.067	17041.766	-3926.727
	(1174.620)	(1412.976)	(993.290)	(2038.095)
Absenteeism				
Mean days teacher absence last week	0.106	0.171	0.118	0.108
	(0.029)	(0.067)	(0.027)	(0.090)
Mean days teacher absence in year	3.671	3.672	3.671	0.389
	(0.368)	(0.826)	(0.337)	(1.142)
Mean absenteeism this year, Grd 1-4	10.706	4.207	9.488	-6.396***
	(1.578)	(0.766)	(1.321)	(1.797)
Possible teacher rewards and censures				
Bonus or promotion	0.226	0.333	0.247	0.136
	(0.053)	(0.123)	(0.049)	(0.155)
Pay deduction	0.597	0.267	0.532	-0.197
	(0.063)	(0.115)	(0.057)	(0.158)
Demoted or transferred	0.226	0.267	0.234	0.045
	(0.053)	(0.115)	(0.049)	(0.145)
Schools	77			

Column 4 gives the difference between Column 2 and Column 1. If the p-value of the adjusted Wald test of the difference is significant, the difference is starred. *p<0.01 **p<0.05 ***p<0.01.

Table 12: Average weekly teaching loads in hours by qualification certification and the grade the teacher teachers, Oromia Region, 2002/03

	Teacher qualification				
Locality & grade taught	10+1; 10+2	12; 12+1	1 year certificate	2 year diploma	All teachers
Oromia	17.1	21.0	24.2	18.1	23.8
Grades 1-4	18.0	27.4	26.0	22.2	26.0
Grades 5-8	17.5	16.2	21.2	18.0	20.6
Urban schools	17.5	16.6	19.9	17.0	17.6
Grades 1-4	18.1	21.6	21.6	18.1	21.5
Grades 5-8	17.1	15.8	17.9	17.0	17.6
Rural schools	23.0	27.7	27.3	22.3	27.2
Grades 1-4	15.0	28.9	28.6	28.9	28.6
Grades 5-8	27.0	20.3	24.5	21.9	24.3

Source: World Bank (2005, 176).
10+1 is Grade 10 plus a one year certificate. 10+2 is Grade 10 plus a two year diploma.

Table 13: Difference-in-difference estimates for eight-year-olds in 2002 and 2009

	Reading			Writing			Numeracy		
	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)
Post	1.14	1.22	1.51	0.58**	0.26***	0.29***	0.57***	0.54**	0.62
	(0.21)	(0.32)	(0.45)	(0.13)	(0.06)	(0.07)	(0.09)	(0.16)	(0.20)
Reform	3.09**	1.84*		1.06	0.45*		0.73	0.47**	
	(1.40)	(0.66)		(0.46)	(0.21)		(0.17)	(0.14)	
Post*Reform	0.59	0.96	1.12	1.69	4.18***	3.51**	1.81*	2.74***	2.17**
	(0.27)	(0.35)	(0.47)	(0.80)	(1.96)	(1.83)	(0.58)	(0.87)	(0.68)
Child controls									
Male		1.03	1.04		1.11	1.09		1.11	1.08
		(0.11)	(0.12)		(0.11)	(0.12)		(0.13)	(0.13)
Stunted		0.53***	0.64***		0.55***	0.57***		0.92	1.08
		(0.08)	(0.10)		(0.09)	(0.09)		(0.13)	(0.16)
Mother's education		1.15	1.02		1.14	1.18		1.14	1.06
		(0.10)	(0.14)		(0.13)	(0.16)		(0.14)	(0.13)
Wealth index		9.54***	3.49**		9.51***	1.75		3.17**	1.84
		(4.60)	(1.84)		(5.95)	(1.04)		(1.61)	(0.82)
Log household size		0.87	1.02		1.06	1.13		1.17	1.01
		(0.18)	(0.21)		(0.24)	(0.26)		(0.16)	(0.15)
Not biological child		0.58**	0.58*		1.02	1.07		0.87	0.96
		(0.13)	(0.18)		(0.24)	(0.29)		(0.26)	(0.29)
School controls									
Urban school		3.10***			2.89***			1.64***	
		(0.79)			(0.73)			(0.31)	
School $<$ 8 grades		0.81			0.81			0.66*	
		(0.18)			(0.22)			(0.16)	
Constant							1.41**	0.49*	
							(0.23)	(0.18)	
Threshold 1	0.88	2.73**		0.75	3.59**				
	(0.16)	(1.31)		(0.17)	(2.06)				
Threshold 2	2.15***	9.14***		2.66***	20.13***				
	(0.44)	(4.36)		(0.57)	(11.92)				
Threshold 3	3.14***	15.29***							
	(0.67)	(7.48)							
School fixed effects	No	No	Yes	No	No	Yes	No	No	Yes
Grade fixed effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Children	1794	1794	4711	1794	1794	3076	1794	1794	1742
Schools	112	112	78	112	112	72	112	112	74
Children dropped			671			512			52
Schools dropped			144			101			38
Adj. Wald F stat	2.09	35.60	289.15	2.27	39.23	344.26	3.84	10.57	7.06
Adj. Wald Prob>F	0.11	0.00	0.00	0.08	0.00	0.00	0.01	0.00	0.00

Coefficients are odds ratios, so a coefficient less than one denotes a negative effect. t stats, adjusted Wald tests and cluster-robust standard errors are used, where the cluster is the school. *p<0.10 **p<0.05 ***p<0.01. The sample sizes for Model III for reading and writing reflect the replications of each individual generated for the Blow Up and Cluster estimator used to account for school fixed effects. Schools with no within school variation in outcomes are dropped from Model III. The number of children and schools dropped are shown in the rows 'Children dropped' and 'Schools dropped'.

Table 14: Heterogeneous treatment effects for eight-year-olds in 2002 and 2009

Average treatment effect on treated	All	Girls	Boys	Median wealth		Stunted		Urban	Rural
				Above	Below	Yes	No		
Reading	0.96	1.01	0.79	0.93	0.87	0.54	0.96	1.35	0.60
	(0.35)	(0.46)	(0.27)	(0.36)	(0.30)	(0.39)	(0.34)	(0.52)	(0.36)
Difference		0.78		0.87		0.57		2.27	
		(0.21)		(0.30)		(0.28)		(1.32)	
Writing	4.18***	3.64**	2.53**	3.51**	2.70*	2.45	3.06**	3.85***	2.03
	(1.96)	(1.91)	(1.09)	(1.70)	(1.39)	(1.63)	(1.38)	(1.57)	(1.47)
Difference		0.69		1.30		0.80		1.90	
		(0.16)		(0.45)		(0.61)		(0.29)	
Numeracy	2.74***	3.12***	1.96**	3.15***	1.82*	1.35	2.67***	3.53***	1.58
	(0.87)	(1.04)	(0.61)	(1.09)	(0.60)	(0.65)	(0.80)	(1.07)	(0.48)
Difference		0.63**		1.73*		0.51		2.23***	
		(0.14)		(0.49)		(0.22)		(0.56)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	No	No	No	No	No	No	No	No	No
Grade fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Children	1794	1794		1726		1794		1794	
Schools	112	112		100		112		112	

The median for wealth is in the sample of Young Lives households in the district where the Young Lives child lives. Coefficients are odds ratios, so a coefficient less than one denotes a negative effect. T stats, adjusted Wald tests and cluster-robust standard errors are used, where the cluster is the school. *p<0.10 **p<0.05 ***p<0.01. All regressions use the specification for Model II, which includes child and school controls and grade fixed effects but not school fixed effects. The difference between the coefficients for boys and girls, for example, is the coefficient on the term Post $*$ Reform $*$ Male.

The Effect of Lengthening the School Day on Children's Achievement in Ethiopia

Many schools in developing countries have four-hour school days and teach two groups of children each day. Governments are considering lengthening the school day, at great expense, to improve school quality. Advocates of the shift system argue the reform is unnecessary, as evidence from developed countries suggests increasing instructional time only improves achievement scores by small amounts. This paper is the first study of the effect of a large increase in instructional time in a low-income country. In 2005, the Ethiopian federal government directed school districts to abolish teaching in shifts and to lengthen the school day from four to six hours. Districts implemented the reform at different times, creating exogenous variation in instructional time. I use a difference-in-difference specification controlling for time-invariant unobservables at school level on a unique longitudinal dataset. For 8-year-old children, a longer school day improved writing and mathematics scores, but had no significant effect on reading. However, effects are larger among better-off children: children who are not stunted, children from richer households, and children in urban schools. The exception is that the reform has larger positive effects on girls than boys. The reform thus improves achievement on average, but may exacerbate gaps between wealthier and poorer children.

Young Lives
An International Study of Childhood Poverty

Young Lives
An International Study of Childhood Poverty

About Young Lives

Young Lives is an international study of childhood poverty, involving 12,000 children in 4 countries over 15 years. It is led by a team in the Department of International Development at the University of Oxford in association with research and policy partners in the 4 study countries: Ethiopia, India, Peru and Vietnam.

Through researching different aspects of children's lives, we seek to improve policies and programmes for children.

Young Lives Partners

Young Lives is coordinated by a small team based at the University of Oxford, led by Professor Jo Boyden.

- *Ethiopian Development Research Institute, Ethiopia*
- *Pankhurst Development Research and Consulting plc*
- *Save the Children (Ethiopia programme)*
- *Centre for Economic and Social Sciences, Andhra Pradesh, India*
- *Save the Children India*
- *Sri Padmavathi Mahila Visvavidyalayam (Women's University), Andhra Pradesh, India*
- *Grupo de Análisis para el Desarollo (GRADE), Peru*
- *Instituto de Investigación Nutricional, Peru*
- *Centre for Analysis and Forecasting, Vietnamese Academy of Social Sciences, Vietnam*
- *General Statistics Office, Vietnam*
- *University of Oxford, UK*

Contact:
Young Lives
Oxford Department of International Development,
University of Oxford,
3 Mansfield Road,
Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Email: younglives@younglives.org.uk
Website: www.younglives.org.uk