HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Abstract

Faculty: Social Sciences

Degree programme: MSc Economics

Study track: General

Author: Dominic Savolainen

Title: What Are the Best Predictors of Learning Outcomes in Sub-Saharan Africa?

Level: Master's Degree

Month and year: August 2021

Number of pages: 44 pages (and 32 pages in the Appendix)

Keywords: learning outcomes, lasso, elastic net, regularisation, regularization, machine learning, classification, prediction, causality, education, teachers, schools, students, africa

Supervisor or supervisors: Matti Mitrunen

Where deposited: A Master's thesis at the Helsinki University Library

Additional information:

Abstract: This study attempts to discover the best predictors of mathematics and language learning outcomes across Kenya, Mozambique, Nigeria, Uganda, and Tanzania by analysing World Bank SDI data and using machine learning methods for variable selection purposes. Firstly, I use the SDI data to show the current fragilities in the quality of education service delivery, while also highlighting deficiencies in student learning outcomes. Then, I use CV Lasso, Adaptive Lasso, and Elastic Net regularisation methods to help discover the best predictors of learning outcomes. While the results from the regularisation methods show that private schools, teacher subject knowledge, and teacher pedagogical skills are good predictors of learning outcomes in a sample combining observations from Kenya, Mozambique, Nigeria, Uganda, and Tanzania, the results fail to infer causality by not distinguishing if unobservable factors are driving the results. To quantify the relationship of key predictors, and for statistical significance testing purposes, I then conduct subsequent OLS analysis. Despite not expecting the true partial derivative effects to be identical to the OLS coefficients presented in this study, this study highlights deficiencies in education service delivery and applies methods which help select key predictors of learning outcomes across the sampled schools in the SDI data.

What Are the Best Predictors of Learning Outcomes in Sub-Saharan Africa?

Dominic Savolainen

UNIVERSITY OF HELSINKI

A thesis presented for the degree of
Master's in Economics

Faculty of Social Sciences
University of Helsinki
Finland

Abstract

This study attempts to discover the best predictors of mathematics and language learning outcomes across Kenya, Mozambique, Nigeria, Uganda, and Tanzania by analysing World Bank SDI data and using machine learning methods for variable selection purposes. Firstly, I use the SDI data to show the current fragilities in the quality of education service delivery, while also highlighting deficiencies in student learning outcomes. Then, I use CV Lasso, Adaptive Lasso, and Elastic Net regularisation methods to help discover the best predictors of learning outcomes. While the results from the regularisation methods show that private schools, teacher subject knowledge, and teacher pedagogical skills are good predictors of learning outcomes in a sample combining observations from Kenya, Mozambique, Nigeria, Uganda, and Tanzania, the results fail to infer causality by not distinguishing if unobservable factors are driving the results. To quantify the relationship of key predictors, and for statistical significance testing purposes, I then conduct subsequent OLS analysis. Despite not expecting the true partial derivative effects to be identical to the OLS coefficients presented in this study, this study highlights deficiencies in education service delivery and applies methods which help select key predictors of learning outcomes across the sampled schools in the SDI data.

Contents

1 Introduction

Education has traditionally been, and remains, a prevalent area for economic research. Jacob Mincer (1958, 1974) pioneered research on the effect of schooling on future earnings. Additionally, neoclassical growth models build on Solow growth theory (1970) and often highlight human capital as a key factor in output growth. Also, the positive spillover effects from education are becoming increasingly recognised, highlighted by studying the effect of increased educational attainment on fertility (Kravdal, 2002), child health (Desai and Alva, 1998), crime (Lochner and Moretti, 2004) and political inclusion (Sondheimer and Green, 2010). Therefore, education provision is widely considered as a fundamental block in public policy and paramount for economic development.

Many economic models proxy 'years schooling' for 'education', but is it naïve to assume that schooling is directly synonymous with learning? School enrolment statistics have universally increased over the last couple of decades in low-income countries but many children in low-income countries who complete their primary school education still lack basic reading, writing and arithmetic skills. For example, Pritchett and Sandefur (2020) report that only 11% of girls in a sample of grade 4 primary school students in Nigeria could read.

Uniquely, the amount of extremely poor people in Sub-Saharan Africa (SSA) is rising, from 278,000,000 in 1990 to 413,000,000 in 2015 (World Bank). In 2015, SSA was home to 27 of the world's 28 poorest countries and had more extremely poor people than the rest of the world combined (World Bank). Also in 2015, the United Nations reviewed their initial 8 international development goals and agreed on 17 revised global development goals, whereby a preliminary goal of universal primary education was extended into a goal of quality universal primary education. Therefore, when considering schooling, particularly in low-income countries, we should not only be concerned about the accessibility of schooling, but also the quality of it.

With scarce resources and comparatively tighter credit constraints, discovering the best predictors of learning outcomes in SSA could prove extremely significant when aiming to improve the quality of education service delivery by optimising the allocation of resources. While learning outcomes in SSA remain at low levels, compared to developed countries, it seems plausible that the global dissonance between the rich and poor will remain.

The purpose of this thesis is to discover the best predictors of learning outcomes in SSA[1] by analysing the Service Delivery Indicators (SDI) data[2], launched by the World Bank in partnership with the African Economic Research Consortium (AERC).

Firstly, I use the SDI data to show the current fragilities in the quality of education service delivery, while also highlighting deficiencies in student learning outcomes. Then, I use CV Lasso, Adaptive Lasso, and Elastic Net regularisation

[1]When I reference SSA in the context of my personal analysis, I am referring to Kenya, Mozambique, Nigeria, Uganda, and Tanzania as the sample.

[2]Because to my knowledge, this is the first data set that attempts to measure the quality of education service delivery, in such detail, across primary schools in SSA.

methods to help discover the best predictors of learning outcomes. While the results from the regularisation methods show that private schools, teacher subject knowledge, and teacher pedagogical skills are good predictors of learning outcomes in a sample combining observations from Kenya, Mozambique, Nigeria, Uganda, and Tanzania, the results fail to infer causality by not distinguishing if unobservable factors are driving the results. To quantify the relationship of key predictors, and for statistical significance testing purposes, I then conduct subsequent OLS analysis. Despite not expecting the true partial derivative effects to be identical to the OLS coefficients presented in this study, this study highlights deficiencies in education service delivery and applies methods which help select key predictors of learning outcomes across the sampled schools in the SDI data.

2 Institutional Backgrounds

In this section I briefly introduce the primary education sector across the sampled countries to help contextualise the findings in this thesis.

Notably, the sampled countries have all experienced prior colonial ruling which has inevitably shaped the current state of institutions and influenced official national languages. In Kenya, Nigeria, Uganda, and Tanzania[3], the official language is English, whereas in Mozambique the official language is Portuguese. All sampled countries gained political independence in the 20th century and range from Eastern sub-Saharan Africa[4] to Western sub-Saharan Africa[5].

The following net enrolment[6], gross enrolment[7], and net completion[8] statistics are from the World Bank[9].

Kenya is split into 8 administrative provinces[10] and within each province, there are administrative districts which are split further into educational divisions. The government recognises 42 tribes and each has their own local dialect. Notably, in 2003, the government implemented a free primary education program, and later in 2008, public secondary schools were also made freely available. As a result, attendance statistics increased by almost 40% within four years, from 5,900,000 in 2003 to 8,200,000 in 2007[11]. In 2016, the gross

[3] Kiswahili is also an official language in Tanzania.

[4] Kenya, Mozambique, Tanzania, and Uganda.

[5] Nigeria.

[6] The total number of appropriately aged children for the given level of education that are enrolled, expressed as a percentage of the total population in the same given age group.

[7] The total number of enrolled students, irrespective of age, expressed as a percentage of the official school age population for the given level of education.

[8] The total number of appropriately aged children for the given level of education that completed school, expressed as a percentage of the total population in the same given age group.

[9] World Bank Metadata, UNESCO Institute for Statistics can be downloaded here: https://api.worldbank.org/v2/en/indicator/SE.PRM.CMPT.MA.ZS?downloadformat=excel. Also, the corresponding statistics used in this study represent the most recent openly available data from the World Bank at the time of writing.

[10] Central, Coast, Eastern, Nairobi, North Eastern, Nyanza, Rift Valley and Western.

[11] https://wenr.wes.org/2015/06/education-kenya.

enrolment rate into primary education was 103.2%[12], in 2019 the net enrolment rate into primary schools was 80%, and in 2019 the net primary school completion rate was 99.7%. Typically in Kenya, schooling begins for children at 6 years old and basic primary school education is split into lower, middle and upper primary. Grades 1 to 3 are in lower primary, grades 4 and 5 are in middle primary, and grades 6 to 8 are in upper primary. At the end of basic primary school, students are required to take the Kenya Certificate of Primary Education examination (KCPE) which is supervised by the Kenya National Examination Council (KNEC) and led by the Ministry of Education. Then, students are ranked and streamed into secondary and technical schools dependent on their exam results. Unlike primary schools, secondary schooling is not compulsory in Kenya.

Mozambique is divided into 11 administrative provinces[13]. The provinces are divided further into districts, and districts are divided further into municipalities. Despite the official language being declared as Portuguese, native languages such as Makuha, Sena and Kiswahili are also commonly used. In 1983 The National System of Education (SNE) was introduced to provide an official public provision of education. Primary education is free and compulsory in Mozambique and is split into lower primary and upper primary. Lower primary consists of grades 1 to 5, and upper primary includes grades 6 and 7. Typically, students start primary school at 6 years old and after 7 years of primary school, students can choose to enrol into secondary education or other alternatives. In 2019, Mozambique had a 116.4% gross primary school enrolment rate and a 93.9% net enrolment rate in 2018. Also in 2019, Mozambique had a net primary school completion rate of 54.7%.

Nigeria operates with a federal government across 36 states and the Federal Capital Territory of Abuja. States are then divided further into local governments. The education system is directed by the Federal Ministry of Education. Despite this, local authorities[14] have the responsibility for the implementation of public policy regarding the provision of public education. The education system progresses from kindergarten, through to primary school and secondary school education, and then further to tertiary education institutions. Nigeria's national policy on education promotes children learning in their indigenous language for the first 3 years of schooling. Primary school education is officially free and compulsory. Typically primary schooling begins when students are 5 years old and then students are awarded with a Primary School Leaving Certificate (PSLC) on completion of grade 6 which is based from a continuous assessment approach. Progression to junior secondary education is automatic and compulsory. In 2016, Nigeria had an 84.6% gross primary school enrolment rate and a 64.1% net enrolment rate in 2010. Also in 2010, Nigeria had a net primary school completion rate of 73.8%

[12] The gross enrolment rate can be over 100% because of the inclusion of over-aged and under-aged students due to early/late entrants, and grade repetition.

[13] Niassa, Cabo Delgado, Nampula, Tete, Zambezia, Manica, Sofala, Gaza, Inhambane, Maputo City, and Maputo.

[14] There are 774 local authorities in Nigeria.

Uganda is divided into 4 administrative regions and within those regions there are 135 districts, alongside the capital city of Kampala. Districts are then divided further into counties, sub counties, and municipalities. Since the introduction of Universal Primary Education (UPE) in 1997 primary school education has been free[15]. Children are typically 6 years old when they start primary school, and spend 7 years in primary school education, from grade 1 to grade 7. The National Curriculum Development Center (NCDC) designs the national curriculum for use at all UPE public primary schools. The curriculum for the 7 years of Ugandan elementary education is divided into 3 cycles: lower primary (grade 1 to grade 3), transition (grade 4), and upper primary (grade 5 to grade 7). In lower primary, classes are taught in the local language, if possible, and then English is typically introduced as the primary language of instruction in grade 4. Before leaving primary school, students must take the Primary Leaving Examination (PLE) to be awarded with their Primary School Leaving Certificate administered by the Uganda National Examination Board (UNEB). This exam is a requirement for students aiming to continue into secondary schools and vocational programs. Despite Uganda becoming the first sub-Saharan African country to introduce a Universal Secondary Education (USE) program, students must score high enough in their PLE in order to attain the government-funded enrolment status. In 2017, Uganda had an 102.7% gross primary school enrolment rate and a 95.5% net enrolment rate in 2013. Also in 2017, Uganda had a net primary school completion rate of 52.7%.

Tanzania is divided into 31 regions which are subdivided into districts, districts are then further subdivided into local wards. Within Tanzania there are believed to be around 120 different tribes with varying languages and dialects. Since 2001, the public provision of primary school education has been free for students and attending primary school is compulsory. Teaching in public schools is usually instructed in Kiswahili, whereas private schools tend to instruct in English. Before entering primary school, some students spend the first 2 years of their schooling career in a pre-primary educational institution. Primary school should last for 7 years and students should legally begin primary school at the age of 7. At the end of grade 7, students take the Primary School Leaving Examination (PSLE), which acts as selection examinations for entry into secondary school. Irrespective of their test result, all students will also receive a Primary School Leaving Certificate.

[15] For initially up to 4 children per family before the policy was revised to further improve enrolment and educational attain for families with more than 4 children.

3 Education Production Function Theory

In this section basic education production theory is presented to provide clarity on the different relationships that researchers can estimate when attempting to discover the best predictors of learning outcomes. This section also provides a framework to follow when reviewing the literature, and conducting analysis, aiding the process of discovering the best predictors of learning outcomes by classifying effects.

Glewwe and Muralidharan (2016) explain a standard theoretical framework for understanding how various inputs at the household, school and classroom level translate into learning outcomes.

$$A = a(S, Q, C, H, I) \tag{1}$$

$$S = f(Q, C, H, P) \tag{2}$$

$$I = g(Q, C, H, P) \tag{3}$$

$$A = h(Q, C, H, P) \tag{4}$$

Equation (1) represents the education production function, where A = achievement (learning outcomes), S = years of schooling, Q = a vector of schooling quality, C = a vector of child characteristics, H = a vector of household characteristics and I = a vector of school inputs that households control (like attendance, effort, homework). In equations (2), (3), and (4), P is price.

Equations (2) and (3) represent the household decision making process on both school attendance and the extent of investments in education, based on optimising household utility, subject to the education production function and a set of constraints. Assume that C and H are exogenous to households, and that S and I are endogenous and dependent on another important set of variables, like P. In a simple scenario, we could assume that there is only one school option and parents cannot change the characteristics of such school, making Q and P exogenous. Parents then choose S and I to maximise household utility. Inserting equations (2) and (3) into (1) creates (4). Equation (4) highlights a reduced form equation which expresses a causal relationship but is not a production function because it reflects household preferences and includes P.

Policy makers should be concerned with the impact of education policies on academic achievement A. Equation (1) is significant because it shows how Q affects A ceteris paribus, and therefore provides the partial derivative of A with respect to Q. Whereas equation (4) provides the total derivative of A with respect to Q as it allows for changes in S and I in response to the change in Q. This represents a policy parameter as parents may respond to better school quality by increasing their provision of education inputs if seen as a complement, or could reduce their provision of educational inputs if considered as a substitute. So, the partial and total derivative effects could be quite different in reality and researchers should be clear which relationship they are estimating. Knowing both the nominal production function impact of a policy change and its total real policy impact may also capture overall welfare effects.

Glewwe and Muralidharan highlighted two main challenges when attempting to estimate the relationship in either equation (1) or equation (4). First is that these equations represent the relationship between inputs and the total stock of human capital, therefore, to accurately estimate the production function in equation (1), the researcher should have data on all prior inputs of human capital which is extremely challenging and perhaps unfeasible. So, a standard approach when estimating the education production functions is to treat the lagged test score as a sufficient statistic to represent prior inputs into learning, and to use a value-added model to study the impact of changing contemporaneous inputs into education on test scores. Todd and Wolpin (2003) expand on this model and the assumptions needed for this approach to yield consistent estimates of production function parameters of interest.

$$A_{i,t} = \gamma A_{i,t-1} + \beta X_{i,t} + \epsilon_{i,t} \tag{5}$$

In equation (5), A is again the child's test score, but present and lagged. X represents a full vector of contemporaneous home and school inputs. While the production function above is linear in terms of X, the specification does not have to be as restrictive because X can include quadratic terms in individual inputs and also include interaction terms between specific sets of inputs. However, even with the value-added equation (5), the second challenge to consistently estimate β is that the variation in the independent variables are likely to be correlated with the error term. Essentially, variations in observed school, teacher and household characteristics are all likely to be correlated with unobserved/omitted school, teacher and household variables that directly determine learning outcomes, leading to biased estimates of β[16].

4 Literature Review

In this section I review causal literature that explores improving learning outcomes in lesser-developed countries. Notably, the literature examining learning outcomes in the developing world is growing and this is represented by multiple extended literature reviews ((Banerjee et al., 2013), (Kremer et al., 2013),(Krishnaratne et al., 2013), (McEwan, 2015), (Murnane and Ganimian, 2014), and (Glewwe and Muralidharan, 2016)).

While regression analysis is commonly used to analyse the impact of policy interventions on learning outcomes, a key problem is endogeneity. Often, true underlying factors driving learning outcomes are rarely captured in regressions which leads to omitted variable bias in the estimates of the effect of education policies on learning outcomes, and thus extends to misleading results. Like

[16]For example, communities and parents that care more about education may be more likely to successfully influence school inputs, and also more likely to provide unmeasured inputs into their child's education, which would lead to an upward bias on β when equation (5) is estimated using cross-sectional data. In other cases, governments may target inputs to disadvantaged areas, in which case areas with higher values of X may have lower values of ϵ, leading to a negative correlation between X and ϵ, and thus a downward bias estimate of β.

Glewwe et al. (2004) found in the context of estimating the effect of flipcharts on learning outcomes in Kenya, regression estimates predicted a rise in test scores of 0.2 standard deviations, whereas an Randomised Controlled Trial (RCT) follow up study found no impact on test scores. Now, this is not to say that regression coefficients are useless, but in this literature review, I will try to focus on analysis which follows quality causal frameworks. High quality descriptive work is also insightful because the findings can highlight key relationships which then present areas for conducting follow-up causal experimental designs. The literature review in this thesis suggests that there are multiple avenues in which the efficiency of education spending in developing countries can be improved by reallocating public expenditure from less cost-effective to more cost-effective solutions. But clearly, making the optimal decisions for learning outcomes is complicated and varies contextually.

4.1 Schooling Time

Schools, and the time that students are in schools, represents initial determinants of learning outcomes. Burde and Linden (2013) used an RCT to explore the impact of opening primary schools on children's enrolments rates and learning outcomes in rural areas of Afghanistan where schools were scarce geographically. The study was conducted in the Ghor province, where only 29% of families lived within 5km of a primary school in 2007 and they found that the program increased enrolment rates by 51.5 percentage-points for girls and 34.6 percentage-points for boys. Also, after 6 months, average test scores increased by 0.66 standard deviations for girls and 0.41 standard deviations for boys, and notably, both results were statistically significant and included children who were not enrolled in school. The heterogeneous enrollment effects between boys and girls could be reflected by the initial disparity between boy and girl enrolments rates, which could be affected by girls not being allowed to travel to neighboring villages to attend schools.

Using a regression-discontinuity-design (RDD) Kazianga et al. (2013) estimated the causal effects of constructing "girl-friendly" primary schools in rural Burkina Faso on enrolment rates and learning outcomes. These schools were attributed with higher qualities amenities for girls, compared to general schools prior to the intervention. Overall, the schools increased enrolment rates for boys and girls by 20 percentage-points, whereby the effect for girls was 5 percentage-points higher than boys. They also found that test scores increased by 0.41 standard deviations in villages that previously had no primary school[17].

These two studies suggest that building schools in areas with low school availability can lead to large increases in school enrolment and subsequent learning outcomes.

Building on school construction, instruction time is another factor affecting student's learning outcomes. Bellei (2009) used a differences-in-differences

[17]This result was an average between French and Mathematics exams for both boys and girls.

(DiD) estimation framework and found a significant positive estimation on the effect of increasing the length of the school day on learning outcomes in Chile. Similarly, Orkin (2013) found a similar relationship using the same empirical strategy in Ethiopia, but many of the results on learning outcomes were insignificant. Assuming that student absenteeism does not increase, it seems plausible that increasing the instruction time will increase learning outcomes, if instruction times are not already set at excessive levels.

4.2 Teachers

Teachers are one of the key stakeholders in schools and represent a focused area in research when studying the optimisation of learning outcomes.

Exploring teacher absence rates, Chaudhury et al. (2006) present findings from a multi-country study where unannounced visits were made to public schools in order to measure teacher attendance and general activity. Across Bangladesh, Ecuador, India, Indonesia, Peru and Uganda, they report average absence rates amongst teachers of 19%. Kremer et al. (2005) report findings which support the idea that teacher absenteeism may be a problem in the developing world. They report that, on average, 25% of teachers in their sample of India were absent, and that another 25% of teachers were in school but not teaching. Therefore, they conclude that around half of the teachers were found to be engaged in teaching. Muralidharan et al. (2016) present results from a nationally representative panel survey which revisited the rural villages surveyed by Kremer et al. (2005), and found a small reduction in teacher absence rates, to 23.7%. They then calculated the fiscal cost of teacher absence in India which totalled $1.5billion, highlighting the economic cost of low teacher effort and potentially poor governance. Building on absence rates in India, Duflo et al. (2012) found that teacher absence rates were above 40% in schools run by an NGO in Rajasthan. Muralidharan and Sundararaman (2013) studied teacher absence with multiple unexpected visits to a representative sample of rural public primary schools in Andhra Pradesh. They discovered high civil service teacher absenteeism, estimated at 27%, whereas contract teachers were less likely to be absent from school with their absenteeism estimated at 18%. Also, their experimental evidence suggests that students in schools assigned an extra contract teacher scored higher in mathematics and language tests. Despite contract teachers earning roughly 1/5th of a civil service teacher's wage and typically attributed with less formal education, the results suggest that they are no less effective at raising pupil learning outcomes.

Contract teachers comprise around a third of public-school teachers across twelve countries in Africa (Bourdon et al. (2007)), therefore it seems relevant to review some of the literature which causally links this shift to learning outcomes, as contracts with periodic renewals could increase teachers' effort levels. Opposing this, some argue that using under-qualified and under-trained contract teachers may increase teacher presence but will not translate into learning outcomes, and that the use of contract teachers can reduce the professionalism of teaching as a practice, reducing the motivation of all teachers (Govmda and

Josephine (2005)).

Duflo et al. (2015) conducted an experimental evaluation of a program in Kenya which randomly allocated an extra contract teacher to selected schools, which on average reduced first grade class sizes from around 80 to 40. Half of the students were then randomly assigned to the contract teachers, and the other half to the existing civil service teacher. The results show that reducing class sizes in regards to the civil service teaching had no effect on test scores, whereas the students who experienced reduced class sizes but were then taught by the contract teacher scored significantly higher in test results of around 0.29 standard deviations. They found that, even when controlling class sizes, students taught by contract teachers scored significantly higher than those taught by civil service teachers, despite the contract teachers remuneration being substantially lower.

These results suggest that contract teachers may make up for their lack of conventional teaching training and education through increased effort. This is significant for policymakers aiming to increase learning outcomes given tight public budget constraints. But the management and administration of such programmes could be pivotal, as Duflo et al. (2015) found that when the contract teacher program was implemented by the government, the effect on learning outcomes was non-existent, compared to positive effects when implemented by a non-profit partner.

Therefore, the pay structure of teachers could be an interesting factor which could increase learning outcomes. Salary structures which motivate teachers to put in more effort should translate into increased learning outcomes for students if implemented correctly. However, there's a debate on how to measure teacher performance fairly and further concerns that link pay to performance on measureable outcomes which may lead to a diversion of effort way from other valuable tasks which measureable outcomes fail to capture (Baker (1992); Holmstrom and Milgrom (1991)). Despite this, low levels of teacher attendance, perceived as effort, in developing countries have led policymakers and researchers to analyse the possibilities of introducing performance pay linked contracts for teachers to boost learning outcomes.

Muralidharan and Sundararaman (2011) conducted a RCT in India to measure the impact of teacher related bonus payments based off student test scores, and the subsequent effect on learning outcomes. They found that after two years, students in the incentive schools performed significantly better than students in control schools by 0.27 standard deviations in mathematics and 0.17 standard deviations in language tests. Treated students also performed better in subjects where teachers' bonuses were unrelated, suggesting positive spill over effects. The treatment group was split between individual incentivised teachers and collective incentivised teachers, the research found no significant difference on learning outcomes. This incentive based intervention was implemented at the same time as other experimental evaluations with equal cost and this allowed the researchers to conclude that the teacher incentive program effects on learning outcomes were greater than other greater than the effects from other additional schooling inputs of the same value. Muralidharan (2012) presents further evidence after a long-term follow-up in the same context after the program was extended for 5 years and found that individual teacher incentive programmes effects on learning outcomes were larger, and in general, more significant than group incentive programmes. This study suggests that incentive programmes can be up to 20 times more cost effective than default policy which aims at reducing pupil-teacher ratios when analysing the effect on learning outcomes.

Glewwe et al. (2010) conducted a natural experiment in Kenya to analyse the effects of school level group incentives for high test scores. The prizes were distributed via tournament design and the authors conclude that the incentive programme led to teachers increasing effort for test-preparation but not in activities that could lead to increases in long term learning. Factors such as absence rates were not improved. They found that treated students performed better on high stake tests but not low stake tests, and that the gains evaporated after the incentive programme finished. They interpret these results by highlighting that teacher incentives may not be an effective strategy to increase long term learning outcomes. But importantly, it is clear that many interventions seem to have high levels of test score decay (Andrabi et al. (2011)), and that grouped incentive programmes can foster environments inducing free riding (Muralidharan (2012)).

Contreras and Rau (2012) used a DiD approach to evaluate a programme which provided teacher bonus payments in Chile based on student's test scores. Their estimates indicate that the programme led to significant increases of 0.29 standard deviations on mathematics scores. Incentives could be based off attendance rather than student's test scores which might be easier to monitor and manage yet provide positive results on learning outcomes. Like Duflo et al. (2012) found, paying teachers based on their attendance reduced teacher absence rates by 21% while significant increases in student test scores also occurred. Alternatively, De Ree et al. (2018) found that an unconditional increase in teacher salaries led to no significant improvements in student learning outcomes after a couple of years in Indonesia. The literature suggests that it is extremely important to design bonus schemes well and ensure that designs reflect insights from economic theory, so that modest changes to compensation schemes and generate substantial improvements in learning outcomes.

4.3 Class Sizes and Sorting

Education systems globally unanimously agree that student-teacher ratios are an important driver of education quality which influence learning outcomes. This is evident from many policy recommendations globally which impose a limit on how many students should be in each classroom, based on student-teacher ratios. Essentially, if a teacher has a larger class, they will be able to give less individual attention to students compared to the identical teacher with a smaller class and the same instruction time.

Urquiola (2001) used a RDD method and estimated that lower student-teacher ratios in Bolivia resulted in significantly higher language scores for students, but reported no highly significant effects on mathematics scores. Uroquiola and Verhoogen (2009) also used a RDD method to estimate the impact of class sizes on test scores, but this time in Chile. The authors reinforce the findings that smaller class sizes can have significantly positive effects on learning outcomes. In their study, there was a positive effect on languages and also mathematics[18].

Duflo et al. (2015) explored the effect of reducing class sizes on learning outcomes through analysing the impact of programme which randomly assigned contract teachers to school in Kenya. Despite a reduction in class size from 82 to 44 on average, students assigned to stay in existing classes taught by civil service teachers experienced no significant increases in test scores. However, test results increased for the students that were assigned to locally hired contact teachers. The authors believe that absence rates and effort may have varied between the civil service teachers and the locally hired contract teachers, with the later being further incentivised by their contractual structure. Hence, they precautiously conclude on the effect of reduced class sizes on learning outcomes.

While the evidence from developing countries is relatively small, assuming that student-teacher ratios affect learning outcomes seems highly plausible. However, further credible evidence on these effects in developing countries is needed as the cost of increasing the supply of teachers to reduce class sizes can be very expensive. Further analysis is needed to see if the cost of providing more teachers produces significant returns on learning outcomes, compared to cheaper interventions.

Another way to change the student experience in classrooms is by sorting, or streaming, because public and private schools often sort classrooms based on ability. Interestingly, Duflo et al. (2011) conducted an experimental evaluation which tracked and streamed pupils into relevant ability sorted classes in Kenya. They found highly significant positive effects on test scores in both the short and the long term, and these effects persisted after 1 year. But it raises the question if the top percentile of students' gains were disproportionately higher than the lower percentage of students (students sorted by test results before the intervention). This type of programme may raise average test scores, but it could increase inequality amongst learning outcomes.

4.4 Physical Schooling Inputs

The supply of school facilities, or physical schooling inputs, can also dictate the quality of schooling and analysing the effects of the provision of physical schooling inputs on learning outcomes seem to be well covered in the literature. A positive relationship between school facilities is expected as inputs such as books, pens, paper are considered complementary in the learning process.

Glewwe et al. (2009) conducted an RCT to study the effect of textbook provision on learning outcomes in Kenya. Surprisingly, they found no statistically significant effects. Subsequently, the authors explored the reasons behind these

[18] However around half of the sampled schools in Uroquilo and Verhoogen's study were private schools.

findings and they discovered that the textbooks, provided by the government, were too difficult for the average child to read in districts of Busia and Teso in Kenya. When the sample was restricted to the top 20% of students, and the top 40% in some samples, the textbooks did indeed improve student's learning outcomes.

In Sierra Leone, Sabarwal et al. (2014) also conducted an RCT to measure the impact of textbook provision on learning outcomes and they found no positive effects. However, their surprising results may have hinged on the fact that the schools failed to distribute the textbooks to the students. The programme was initiated by the Ministry of Education without encouraging the schools to distribute the textbooks to students. Resultantly, most textbooks were stored in the school and Sabarwal et al. provided some evidence which suggested that negative expectations of future resources from staff led to this failure in distribution.

Overall, it seems that textbooks, that are actually provided to students, can have positive impacts on learning outcomes if they are of the appropriate level to the students. If textbook complexity is misaligned with student knowledge levels, they can have minimal or zero effects on learning outcomes.

Borkum et al. (2012) used an RCT to study the effect of libraries on attendance and also learning outcomes. They study the effect of in-school and traveling library provisions in India. They found that in-school libraries had no effect on language scores and that travelling libraries had a negative effect of 0.22 standard deviations on language scores.

Also as previously mentioned, Glewwe et al. (2004) found no impact of flip charts on learning outcomes after following an RCT experimental design after the initial regression results pointed towards a positive relationship.

Tan et al. (1999) estimate the impact of providing three types of 'multilevel learning materials' alongside 'parent-teacher partnerships' by conducting an RCT in the Philippines. They found that two out of the three multilevel materials positively and significantly impacted test scores across language and mathematics. They also found that all three multilevel materials combined with parent-teacher partnerships had significantly positive results on learning outcomes across language and mathematics test scores. However, it's unclear to what extent each combination of the parent-teacher partnerships and multilevel learning materials had on learning outcomes because these differences were not clearly distinguished.

Technological improvements have typically played a pivotal role increasing productivity, hence some believe that technological investments in education could be extremely positive for improving learning outcomes.

Banerjee et al. (2007) found significant positive impacts of computer aided learning programme interventions in India. While Yang et al. (2013) and Mo et al. (2014) found significant positive effects of the computer aided learning programme on learning outcomes in China, but questions arose around the cost of such programmes and how efficiently they can be implemented. Alternatively, Barrera-Osorio and Linden (2009) conducted a RCT in Colombia and found no impact of the provision of computers on learning outcomes. One caveat is if the lack of impact was because poor implementation and the inability of teachers to optimally incorporate the technology in their teaching. Similarly, in Peru, Beuermann et al. (2013) studied the impact of the 'One Laptop Per Child' programme through a large scale RCT and found no impact on test scores across language and mathematics. Both results suggest that the magnitude of investment in education does not always translate directly into learning outcomes.

Opposing this, Linden (2008) conducted a RCT and found negative impacts of the Gyan Shala Computer Assisted Learning (CAL) project on learning outcomes in India when the technology was implemented in class and acted as a substitute for regular instruction. Linden believes implementation of the technology could explain negative impact of 0.57 standard deviations on test scores because it was treated as an 'in-class substitution'. When the technology was implemented as an 'out-of class complement', test scores rose (insignificantly). Malamund and Pop-Eleches (2011) conducted a RDD study to measure the impact of distributing computer vouchers on learning outcomes for middle-school students in Romania but their results were mixed.

These results suggest that wariness is required when policymakers opt for expensive large scale technological interventions like providing computers for all. From the outset, these policies seem extremely positive and ambitious. There are many reasons to be excited about technology significantly improving learning outcomes. However, there is also evidence that questions the effectiveness of such policies which can depend crucially on the implementation. Effective pedagogical techniques need to also be developed at the same time of technological roll-outs, otherwise teachers may not be equipped to optimise the potential returns that translate in increased learning outcomes. It's clear that more research is needed here, especially when public funds are highly constrained, the opportunity cost becomes more significant.

4.5 Student Health Related Provisions

There is substantial evidence that well-nourished, healthy children have better learning outcomes, yet child malnutrition remains a prevalent issue not just in SSA but across many developing countries. Providing nutrition in schools is one way to ensure students have access to good nutrition, while also increasing the demand for education. Glewwe and Miguel (2008) support this notion. Many developing countries now have opted to implement programmes which allocate meals to students and sometimes to their families too.

Kazianga et al. (2012) evaluated two programmes in Burkina Faso which provided meals to school children from low income households by conducting an RCT. They found that take-home rations increased mathematics scores by 0.08 standard variations, and this result was statistically significant.

Utilising a RDD method, McEwan (2013) found statistically insignificant impacts of a school feeding program on language and mathematics test scores amongst grade 4 children in Chile. Also, Adrogue and Orlicki (2013) executed a DiD model and found a small and statistically insignificant impact of the provision of school meals on mathematics scores of grade 3 students in Argentina.

However, Tan et al. (1999) found significantly positive impacts of school meal provision on Mathematics and Filipino test scores through their DiD estimates in the Philippines. Mathematics test scores increased by 0.25 standard deviations, while Filipino scores increased by 0.16 standard deviations.

After examining the provision of food affecting students' educational outcomes, studying whether the provision of healthcare could also drive improved learning outcomes seems reasonable. After all, it is widely accepted that many individuals in developing countries are more exposed to severe health problems than individuals living in developed countries.

Miguel and Kremer (2004) implemented an RCT in rural Kenya to estimate the impact of providing deworming medicine to primary school students. They discovered no effect on learning outcomes. Notably, an interesting follow-up paper by Ozier (2018) examined the spill-over effects stemming from the Kenyan deworming programme on younger siblings of treated children. Ozier found significant positive effects on the test scores of children who were younger than 1 when the deworming programme was implemented and also after 10 years. Another follow up study by Baird et al. (2016) found that after 10 years, women who were eligible for the programme as school girls were 25% more likely to have attended secondary school and that boys were more likely to have completed primary school. Baird et al. also found large positive labour market effects. Combined, these findings suggest that long term gains exist from the deworming programme, even if the short-term gains on learning outcomes were minimal.

In China, Luo et al. (2012) and Sylvia et al. (2013) provide some evidence that iron supplements can increase student learning outcomes. However, the external validity must be questioned. Also in China, Glewwe et al. (2016) conducted an RCT to estimate the impact of the provision of glasses on learning outcomes within rural Chinese primary schools. They found that the glasses significantly raised average test scores amongst by at least 0.16 standard deviations.

4.6 Finance

Increasing school's budgets seems a relatively simple measure to improve educational outcomes as schools are able to allocate more resources into beneficial programmes. Das et al. (2013) presented a dynamic household optimisation model relating test scores to school and household inputs. In India, they showed how unexpected school block grants of around \$3 per student, increased test scores by 0.09 standard deviations. However, they also showed that expected grants either had no effect or very little, as households responded by decreasing spending on education. This suggests that providing additional schooling finance in India could be seen as a substitute, rather than a complement for households' education investment decisions. Olken et al. (2014) conducted a field experiment in Indonesian villages to see if block grants improve health and education outcomes. Specifically, they found no positive effects on student test scores. Additionally, Reinikka and Svensson (2004) showed evidence of financial leakage in a school government transfer programme in Uganda, they also anticipated that this phenomena is not exclusively happening in Uganda. Therefore, policies which aim to provide financial injection to improve learning outcomes should also be monitored to limit corruption.

4.7 Monitoring and Community Engagement

Monitoring and collective community involvement in the provision of schooling could be a key factor when aiming to optimise learning outcomes.

Muralidharan et al. (2016) used a nationally representative village level panel data set from India on teacher absence to examine the correlations between changes in various school and management characteristics and changes in teacher absence. They found that increasing the probability of a school being inspected within three months was correlated with a 7 percentage-point reduction in teacher absence. This estimate was similar in cross-sectional and panel estimates, and also bivariate regressions, with and without district fixed effects. The researchers predicted that monitoring could be over 10 times more cost effective at increasing teacher led instruction time than hiring more teachers.

However, evidence on the impact of monitoring teachers time spent teaching on learning outcomes in developing countries is small. Banerjee et al. (2010) assessed the impact of bottom up monitoring through the community, but they found no positive effects on teacher effort or learning outcomes.

Other bottom up monitoring polices have been analysed and one study by Pandey et al. (2009) used an experimental evaluation of an information campaign to improve parental participation in village education committees. They found positive effects on learning outcomes, but many of the estimated impacts were statistically insignificant. Pradhan et al. (2011) also conducted an experimental evaluation of multiple interventions which aimed at increasing community participation in school management in Indonesia. They only found one significantly positive impact on tests scores form the "linkage" intervention, which facilitated meetings between school committees and village councils.

Beasley and Huillery (2017) conducted a randomised experiment to explore the effect of a parent-empowerment programme in Niger. The programme allocated grants to school committees to encourage parent participation in school management processes. However, the programme had no measured impact on learning outcomes. The authors suggest that this may because many parents do not have the knowledge and necessary information to make effective decisions to optimise educational quality. Glewwe and Maïga (2011) and Lassibille et al. (2010) both conducted experimental evaluations of the Amélioration de la Gestion de l'Education à Madagascar (AGEMAD) programme in Madagascar which tried to strengthen school management practices with the community at various levels. However, both studies also found limited measured significant impacts on student test scores.

Similarly, Duflo et al. (2015) conducted an experimental design to explore the impact of the School Based Management (SBM) programme in Kenya and found that training school management committees to evaluate the performance of contract teachers had significantly positive impacts on the performance of contract teachers and student test scores. This adds to an interesting branch of research suggesting that improving teachers' motivation could be a cost-effective method when trying to improve test scores. Perhaps this is an interesting intervention for schools with a high supply of contract teachers where monitoring isn't frequently and effectively occurring.

Gertler et al. (2012) conducted DiD analysis on the impact of a programme in Mexico which empowered parents to improve school quality. They found that the programme reduced student failure rates and grade repetition rates, but had no impact on dropout rates. They also found that the AGE programme had no impact in poorer communities, suggesting that community empowered governance programmes may increase inequality across schools. Santibanez et al. (2014) also used a DiD strategy to evaluate another school based management project in Mexico. They found that the programme had no general effect on student test scores.

It seems that the literature suggests that there may be problems with collective action and community involvement in school based management projects, especially in disadvantaged areas with low levels of parental education, which may make community monitoring less effective than top down monitoring from government administrations.

4.8 Private Schools

Private schools now account for over 20% of total primary school enrolment in low-income countries (Baum et al. (2014)). This raises questions of the efficiency of public schooling provision compared to private school provision on learning outcomes, and whether policymakers should respond to their growth. A weakness of private schools is the lack of accessibility to the poor and many argue that policymakers should allocate support programmes enabling a larger share of the population from lower socio-economic backgrounds to have the opportunity to attend private schools. (Tooley et al. (2007); Muralidharan and Kremer (2006); Goyal and Pandey (2009)).

Angrist et al. (2002) and Angrist et al. (2006) study the randomised assignment of vouchers from the PACES programme in Colombia to estimate the effect on learning outcomes from private secondary school attendance. They found that treated students who received vouchers scored higher in mathematics and reading tests after three years and that they completed more schooling years with a lower chance of grade repetition. The later study discovered that treated students also had significantly higher graduation rates and scored higher on college entrance exams, even after controls for differential attrition. However it must be noted that the PACES programme allowed parents to supplement vouchers with their own income and that students had to maintain certain academic standards to continue with the programme. So the results do not distinguish the effectiveness of student incentives, additional education spending, and private school effects on learning outcomes.

Muralidharan and Sundararaman (2015) conducted a RCT to measure the effect of the school choice programme in Andhra Pradesh, India, on learning outcomes. This intervention allowed a randomisation of selection into private schools and the researchers created a set of control villages allowing a robust presentation of individual and aggregate returns to the treatment. They explain that differences in the measured test outcomes[19] were mainly driven by omitted variables, but they did find that private schools spent significantly less time on teaching mathematics and the native language Telugu, compared to public schools. As a result, private school students spent more time on other subjects like languages, sciences, and social studies, which were not taught in the public schools. Assuming equal weights across all subjects, the researchers found that the treated students who received the vouchers and were schooled privately had higher average test scores of 0.26 standard deviations which were significantly significant, but this was mainly driven by Hindi. Even without assuming the equal weights across subjects, the results point to private schools being more efficient and being able to attain equal levels of learning outcomes with less instruction time. Also, in the sample, the annual cost of a student in a public school is believed to be over three times the mean cost per student in private schools.

Lucas and Mbiti (2014) support the notion that elite private schools may not add much added value to the marginal, or less academically gifted students in terms of learning outcomes, but that private schools tend to select the most intellectual students which can skew the perceived effects of attending private schools. While Hsieh and Urquiola (2006) found evidence of a private school voucher programme in Chile resulting in 'sorting', where the best public school students left for private school, they found no effects on educational outcomes from students who switched from public to private schools. The challenge empirically is to compare the counterfactual outcomes of attending private school or public school for a given individual to account for unobserved individual characteristics. Addressing this challenge, Lara et al. (2011) employed propensity score econometric techniques and a changes-in-changes estimation method and used data which contained the previous academic achievement levels of students in Chile, this identification strategy allowed the authors to identify differences in students' unobservable characteristics and they found that the effect of the private school voucher education amounted to a small increase in test scores.

In general, it seems that private schools could be more productive at achieving learning outcomes. Much of the literature suggests that the allocation of resources in the public sector is inefficient in comparison. More evidence is needed from a wider range of countries to understand whether this trend extends across much of the developing world.

[19]Mathematics and Telugu.

5 SDI Data and Samples

In this section, I present the SDI data[20] which I use to help discover what the best predictors of learning outcomes might be. I also show some initial descriptive findings on the current state of the quality of education service delivery and resultant learning outcomes, before finally outlining some concerns with the data. The SDI data reports information ranging across physical schooling inputs, teacher characteristics, school finances, classroom observations, pupil assessments, and teacher assessments. Table 10 in A.1 highlights the data collection methodology followed by the SDI team in more detail. Also please refer to Table 11 in A.1 for more information regarding SDI data modules.

I selected the samples surveys from Kenya (2012), Nigeria (2013), Uganda (2013), Tanzania (2014), and Mozambique (2014) for the analysis in this thesis due to the heightened comparability of the survey designs[21] and due to the scarce availability of this type of data in SSA. Within each country, the sampled schools originate from all different states and regions[22]. The SDI data was intended to provide a representative snapshot of education service delivery by using a multistage, cluster-sampling design. Survey weights are available with the data but I did not use these due to a lack of information available regarding the consideration of missing observations on the design of the weights. Therefore, the results I present represent a snapshot of the selected schools in the sample, rather than the whole population of interest in the selected countries[23]. All sampled schools vary categorically as either urban or rural, and whether they are publicly owned or privately owned[24]. Because of the relatively small sample sizes, I also analysed a sample of 'All Countries', combining all school observations from the sampled countries with equal weights for each school.

[20] To my knowledge, there is no other set of indicators that exist which measures the quality of service delivery in SSA in the same detail as the SDI indicators.

[21] Pilot surveys administered in Senegal and Tanzania record slightly different data to the samples I selected, and the data for Togo doesn't distinguish teachers in the same level of detail as the other selected samples.

[22] Apart from Nigeria, where surveys represent Anambra, Bauchi, Ekiti and Niger states.

[23] In Kenya 306 schools were sampled, in Tanzania 400 schools were sampled, in Uganda 400 schools were sampled, in Mozambique 203 schools were sampled, and in Nigeria 706 schools were sampled.

[24] Apart from Mozambique where all sampled schools are public.

5.1 Descriptive Findings

In this subsection, I present the descriptive findings from the SDI data to highlight the current state of the provision of schooling across the sampled countries.

To discover the best predictors of learning outcomes, learning outcomes are proxied by student test results. Table 1 shows a comparison of literacy/language and mathematics student test results across the sampled countries. Staggeringly, we can see that across the whole sample, on average, 4th grade students in SSA could only read 43% of a selected paragraph in their chosen language correctly. Also across the sample, we can see that, on average, 55% students could successfully add double digits numbers. The measures selected in Panel A and Panel B represent skills that 4th graders are expected to perform relatively well in. Consequently, the results heighten the significance discovering the best predictors of learning outcomes in SSA.

Student Learning Outcomes

	Kenya	**Mozambique**	**Nigeria**	**Tanzania**	**Uganda**	**All**
Panel A: Average Student Literacy/Language Levels						
Pupil literacy test score (%)	80	22	46	50	53	50
Identifies letters (%)	97	63	83	80	87	82
Identifies words (%)	93	40	57	80	80	70
Can read a sentence (%)	91	17	52	26	58	49
Can read paragraph (%)	83	17	41	26	48	43
Total students sampled for testing	*2952*	*1758*	*6735*	*3999*	*3957*	*19401*
Panel B: Average Student Mathematics Levels						
Pupil mathematics test score (%)	60	25	40	49	46	44
Recognises numbers (%)	100	90	83	97	97	93
Can order numbers (%)	73	21	40	48	51	47
Can add single digits (%)	92	48	71	81	84	75
Can add double digits (%)	85	18	50	64	60	55
Can add triple digits (%)	88	9	41	64	60	52
Can subtract single digits (%}	89	30	65	76	79	68
Can subtract double digits (%)	65	5	31	42	31	35
Can multiply single digits (%)	52	4	33	40	27	31
Can multiply double digits (%)	10	0	8	14	2	7
Can multiply triple digits (%)	4	0	6	11	1	4
Can divide single digits (%)	63	8	29	40	41	36
Can divide double digits (%)	38	2	19	20	17	19
Can complete a sequence (%)	26	4	20	16	13	16
Total students sampled for testing	*2952*	*1758*	*6735*	*3999*	*3957*	*19401*

Table 1: Reports the average of competencies of 4th grade students sampled across both government and private primary schools. In Kenya students were tested in English, in Mozambique students were tested in Portuguese, in Nigeria students were tested in English, in Tanzania students were tested in either English or Swahili, in Togo students were tested in French, and in Uganda students were tested in English. These results are not weighted, however, the average of 'All' countries is calculated by averaging all countries' totals, whereby each country has an equal weight.

Similar to Bold (2017), combining observational classroom data, absenteeism data, and administrative opening times data, Table 2 compares predicted measures that students experience regarding learning times, unsupervised classrooms, and teacher absenteeism. Across the sample, we can see considerable gaps between scheduled teaching times and estimated effective teaching times.

Effective Teaching Time

	Kenya	Mozambique	Nigeria	Tanzania	Uganda	All
Panel A: Daily Learning Time						
Scheduled teaching time	5h 36m	4h 21m	5h 1m	5h 56m	7h 21m	5h 39m
Estimated effective teaching time	2h 44m	1h 41m	3h 42m	2h 55m	3h 24m	2h 53m
Total schools sampled	*306*	*203*	*760*	*400*	*400*	*2069*
Total classes observed	*306*	*203*	*729*	*399*	*398*	*2035*
Panel B: Orphaned Classrooms						
Unsupervised classrooms (%)	36	29	19	36	36	31
Total classrooms observed	*300*	*153*	*721*	*396*	*393*	*1963*
Panel C: Teacher Absenteeism						
Teachers absent from class (%)	45	56	16	48	52	43
Teachers absent from school (%)	16	43	10	15	23	21
Total teachers sampled	*2930*	*1006*	*5357*	*3664*	*3764*	*16721*

Table 2: Reports the scheduled instruction times, estimated effective teaching times, unsupervised classrooms, and teacher absenteeism from the classroom and school. The sampled schools include both government and private schools and the sample is not weighted. Teachers' absenteeism data is collected from the second unannounced visit, whereby they are marked as 'absent from school' if they are missing from the school premises, and they are marked as 'absent from the classroom' if they are not found in the class. The scheduled teaching time is calculated by deducting break times from the length of the school day. The estimated effective teaching time is calculated by adjusting the length of the scheduled teaching time by the share of teachers who are present in the classroom and by the average time that observed teachers spend teaching while in the classroom. The unsupervised classrooms label represents a percentage of classrooms with students inside without the presence of a teacher compared to the total number of classrooms with students inside, with or without a teacher. The 'All' column represents an average of all countries' totals, whereby each country is given an equal weight.

The results of the tests administered to teachers across the sample can be found in Table 3. These results will be used in this thesis as a proxy for teacher subject knowledge. Following Bold et al. (2017), teachers were deemed to have the minimum subject knowledge to teach if they corrected over 80% of the students mistakes accurately. Again following Bold et al.'s categorisation, on average, 11% of teachers across the sampled countries have the minimum literacy knowledge to teach 4th grade students in literacy/language, and 71% of teachers have the minimum mathematics knowledge to teach 4th grade students in mathematics.

Teacher Subject Knowledge

	Kenya	Mozambique	Nigeria	Tanzania	Uganda	All
Panel A: Average Teacher Literacy/Language Levels						
Have the minimum literacy knowledge to teach in the 4th grade (%)	37	0	0	2	17	11
Can use correct grammar (%)	92	82	66	71	87	80
Can mark student grammar in sentence (%)	1	0	0	1	1	1
Can correct student spelling, grammar, syntax, and punctuation in a paragraph (%)	52	10	24	22	41	30
Total teachers sampled for testing (%)	*771*	*320*	*1175*	*747*	*702*	*3715*
Panel B: Average Teacher Mathematics Levels						
Have the minimum mathematics knowledge to teach in the 4th grade (%)	91	52	62	79	69	71
Can add double digits (%)	98	83	89	97	98	93
Can subtract double digits (%)	87	62	70	84	73	75
Can multiply double digits (%)	87	46	62	64	67	65
Can understand a Venn diagram (%)	55	14	26	33	53	36
Can interpret data in a graph (%)	40	2	7	3	8	12
Can solve algebra (%)	76	4	18	46	45	38
Total teachers sampled for testing	*791*	*317*	*1174*	*916*	*738*	*3936*

Table 3: Presents language and mathematics scores of sampled teachers who currently teach a grade 4 class, or who taught a grade 3 class in the previous year. This table only reports teachers who teach or taught the subject of interest. Therefore, if a teacher did not teach mathematics or language, their test results were not included in this table. These estimates are unweighted and represent the sample, rather than the wider population of interest. In Panel A, a teacher is defined the have the 'minimum language knowledge to teach 4th graders' if the teacher scored at least 80% on the 3 tasks in their test (1: using grammar, 2: marking student grammar responses, 3: correcting spelling, grammar, syntax, and punctuation mistakes in a student letter). In Panel B, a teacher is defined to have the 'minimum mathematics knowledge to teach 4th graders' if the teacher scored at least 80% on the tasks covered in the 4th grade curriculum. Note that a teacher was required to answer both questions correctly related on the Venn diagram and graphical data section to be deemed able to understand and interpret respectively. The 'All' column represents an average of all countries' totals, whereby each country is given an equal weight.

Pedagogy represents the method and practice of teaching. Essentially, pedagogy levels should explain how well teachers can teach. The findings in Table 4 suggest deficiencies in teachers' abilities to effectively teach and assess students.

Teacher Pedagogical Skills

	Kenya	Mozambique	Nigeria	Tanzania	Uganda	All
Panel A: General Pedagogical Knowledge						
Percentage of teachers with the minimum pedagogical knowledge to teach (%)	19	2	2	36	5	13
Average teacher score when comprehending a factual piece of text (%)	65	25	32	81	50	51
Average teacher score when formulating aims and learning outcomes (%)	41	12	16	40	19	30
Total teachers sampled for testing	*1118*	*329*	*1345*	*1357*	*1194*	*5343*
Panel B: Assessing Students						
Percentage of teachers with the minimum knowledge to assess students effectively (%)	1	0	0	0	0	0
Average teacher score when formulating questions to check students' understanding (%)	55	13	6	53	5	26
Average teacher score when assessing and comparing students' abilities (%)	41	10	12	18	24	21
Average teacher score when evaluating students' progress (%)	29	6	8	22	10	15
Total teachers sampled for testing	*1118*	*329*	*1345*	*1357*	*1194*	*5343*
Panel C: Average Observed Teacher Skills & Application Of Practices						
Lesson appeared planned (%)	77	71	70	79	47	69
Lesson appeared structured (%)	63	19	60	42	30	43
Asked questions to students with varied difficulty (%)	34	14	44	42	45	36
Gave positive feedback, praise, and corrected mistakes (%)	73	28	47	50	76	55
Engaged in all of the above practices (%)	20	1	14	15	5	11
Total observed classes sampled	*233*	*200*	*721*	*392*	*335*	*1881*

Table 4: Presents information on pedagogical skills and classroom observations of sampled teachers. These estimates are unweighted and represent the sample, rather than the wider population of interest. The 'All' column represents an average of all countries' totals, whereby each country is given an equal weight. Panel A reports on the minimum pedagogical knowledge to teach and percentage scores on specific pedagogical tasks for teachers who currently teach grade 4 or who taught grade 3 in the previous year. To be deemed to have the minimum pedagogical knowledge to teach, teachers had to answer at least 80% on the tasks relating to general pedagogy correctly, which was comprised of the factual text comprehension and subsequently being able to formulate relevant aims and learning outcomes on the given topic. Panel B reports on the minimum knowledge to effectively assess students alongside scores on related tasks for teachers who currently teach grade 4 or who taught grade 3 in the previous year. To be deemed to have the minimum knowledge to assess students effectively, teachers had to answer at least 80% on the tasks relating to assessment, which was comprised of comparing and monitoring the progress of students. Panel C presents observed teacher skills and practices in grade 4 classrooms.

5.2 Data Limitations

In this subsection, I present some concerns regarding the SDI data which may limit the validity when finding the best predictors of learning outcomes across the sampled countries.

The SDI data collection method offers a unique data resource in the SSA region, and hence, the ambitious nature behind the creation of the SDI could lead to data salience[25]. Also, the data regarding the quality and availability of inputs is conventional, but accurately assessing teachers' ability and teachers' effort can be more challenging. Hence, the data collected for teachers' ability and teachers' effort should be interpreted as a proxy, and reinforce the assumption that the SDI data does not perfectly represent true levels of providers' ability and effort.

Some elements of the data collection was subjective based on enumerators' perception, therefore there may be some measurement inaccuracies in the data[26]. Also, some designs of the SDI data collection methods were sub-optimal when creating an environment to effectively proxy variables such as teacher knowledge[27].

Furthermore, the data collection and formatting followed anonymisation methods to mitigate the risk of breaching the privacy of the respondents in line with Statistical Disclosure Controls processes[28]. While this protects the identity of respondents and follows ethical guidelines, this limits the detail of analysis which could have been conducted[29].

Another issue with the data is the amount of missing observations, which reduced already small sample sizes, and the lack of information on if these missing observations reduces the accuracy of the available survey weights. If the missing observations are randomly distributed, and the sampled schools were randomly selected, then selection bias will be limited despite the results not perfectly representing the countries of interest. Also, like most observational studies, the Hawthorne effect could also prove to be a detrimental feature of this data collection method[30].

[25]The issue of data salience could exist when political bias may influence the initial variable selection in the anticipation that it will capture the attention of policy makers, rather than collecting data on optimal education production function inputs.

[26]For example, Panel C in Table 4 reports findings on whether a lesson *appeared* planned and structured. Also, because data was recorded manually by enumerators, there could be some inaccuracies due to human error.

[27]For example, the teacher tests to measure subject knowledge were administered during their lunch breaks. Resultantly, some teachers may have exerted less effort and less care in an attempt to complete the tasks as quick as possible in order to maxmimise their effective break time, leading to measurement errors.

[28]For an example, Kenya's SDC process can be downloaded here `https://microdata.worldbank.org/index.php/catalog/2755/download/39311`.

[29]For example, if the data showed which teachers directly taught individual students, then stronger causal claims of teacher characteristics on learning outcomes could be explored from microeconometric analysis (if the allocation of teachers to teachers was deemed random and other factors influencing learning outcomes were equally distributed).

[30]Whereby observed individuals alter their behavior because of their awareness of being observed and potential implications or benefits that arise from their behaviour. For example,

The Education Field Manual[31] was made to provide guidance to field workers, containing detailed information on procedures for carrying out field data-collection work to ensure homogeneity in data collection methods. However, the SDI team recognise that the manual is sub-optimal as it states that versions will be "updated periodically and supervisors and enumerators are encouraged to give feedback and recommendations that will enhance the quality and utility of the manual"[32]. Homogeneity across the data collection is a key concern and could even extend to student language test difficulty. For effective descriptive comparisons of student learning outcomes, we must assume that there is no variation in difficulty dependent on the language of the test that was administered. But, a direct comparison of the countries is not recommended due to heterogeneities across the schools selected for sampling and heterogeneities across recorded variables in each country. Some of the countries have data collected on areas of service delivery where other countries do not[33]. Furthermore, I anticipate that categorical data entries for some covariates used in the SDI data increase measurement errors in the findings in this thesis[34][35].

6 Empirical Strategy

In this section I present the empirical methods used to help select, quantify, and statistically test the effect of key predictors of learning outcomes.

To classify the best predictors of learning outcomes, I used machine learning regularisation methods to initially produce sparse models and help synthesise the SDI data set by selecting a smaller sub-set of key variables from a large potential covariate list. These techniques were used to help uncover key associations. The use of machine learning techniques in economics for classification is becoming increasingly popular, potentially due to the rapid increase in data-driven collection initiatives which sometimes develops at a faster pace than the economic theories to simplify them. It is important to note that machine learning methods can utilise infinitely many possible restrictions which can each lead to a unique solution, so the ambiguous nature of the results are never removed

despite the second visit to schools being unannounced, schools were given a broader time-period as to when they will receive the second visit, hence some teachers may exert more effort to attend during this broader period in an attempt to improve the perception of their effort.

[31]For an example, Kenya's Education Field Manual can be downloaded here https://microdata.worldbank.org/index.php/catalog/2755/download/39313.

[32]Some tasks require emotional intelligence, like interviewing. When interviewing, the ability to build rapport, instil trust, and potentially calm the interviewee can affect the responses from the interviewee. Hence, a reasonable assumption is that the field manual will not be able to create perfect homogeneity across enumerators interviewing skills because of pre-existing heterogeneous levels of emotional intelligence.

[33]The Mozambique sample has no private schools and the SDI data doesn't record *if schools have Parent-Teacher Associations* in Mozambique and Tanzania.

[34]For example, variables like *teacher age* and *the total number of teachers per school* were entered into categories with 10 year spans.

[35]Also, further measurement errors are likely if assuming that one classroom observation per school from Module 4 is representative of a school average.

entirely, but rather transferred to the choice of the constraint (Friedman et al., 2001).

Supervised machine learning regression methods estimate coefficients by observing inputs and outputs in a training set of observations, whereby inputs are fed into a learning algorithm which produces given outputs in response to the artificial learning system. The learning algorithm can impose restrictions on the function with the motivation to improve the performance of a model with unseen data. To mimic this setting, a data set can be split into a training and validation sample. The training sample provides the environment for the machine learning model to impose restrictions on a given function, and the validation sample provides an environment to test the performance of the restrictions imposed by learning from the training sample. The concept of regularisation was first defined by Ridge regression methods. Hoerl and Kennard (1970) show that the properties of Ridge regression estimates can improve the mean squared error of estimation by attempting to reduce the variance of coefficients by imposing a penalty term to the size of the coefficients. This is also known as shrinkage, or regularisation. Before conducting any shrinkage, coefficients are typically standardised. Standardised variables each have mean of 0 and standard deviation of 1 which ensures that the shrinkage doesn't effect coefficients disproportionately. The Ridge coefficients minimise a penalised residual sum of squares

$$\hat{\beta}^{ridge} = \arg\min_{\beta}(\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2). \tag{6}$$

In the Ridge minimisation equation 6 (Lagrangian form), $\lambda \geq 0$. The λ parameter controls the amount of regularisation in the model, therefore λ is known as a shrinkage parameter because a larger λ value increases the penalty imposed on coefficients, subsequently leading to more shrinkage of the coefficients. Another way to write the Ridge problem to clearly define the size constraint imposed on coefficients is

$$\begin{aligned} \hat{\beta}^{ridge} = \arg\min_{\beta} \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2, \\ \text{subject to} \sum_{j=1}^{p} \beta_j^2 \leq t. \end{aligned} \tag{7}$$

Whereby there is a one-to-one correspondence between λ in equation 6 and t in equation 7.

However this thesis uses regularisation methods for classification, rather than prediction and the Ridge penalty cannot directly perform variable selection by shrinking variable coefficients exactly to 0, unlike the Lasso and Elastic Net methods. Nevertheless, the fundamental shrinkage ideology stemming from Ridge regressions help explain Lasso and Elastic Net models. The Lasso model (Tibshirani, 1996) performs regularisation with the ability of shrinking coefficient values to 0. Therefore, the Lasso is able to perform variable selection which can enhance the interpretability of models.

The Lasso estimate is defined by

$$\begin{aligned} \hat{\beta}^{lasso} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \\ \text{subject to} \sum_{j=1}^{p} |\beta_j| \leq t. \end{aligned} \tag{8}$$

In equivalent Lagrangian form, the Lasso problem can be written as

$$\hat{\beta}^{lasso} = \arg\min_{\beta} (\frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|). \tag{9}$$

The key difference between the Ridge and Lasso model are the constraints imposed on the coefficients. The Lasso l_1 penalty $\sum_{j=1}^{p} |\beta_j|$ will make solutions non linear for y_i, and there is no closed form expression. Hence, computing the Lasso is a quadratic programming problem and because of the constraint imposed on the model, coefficients can be shrunk to exactly 0, if t is small enough. This feature selection element is the key difference between the Lasso and Ridge method. The Ridge l_2 penalty $\sum_{j=1}^{p} \beta_j^2$ has a closed form solution but cannot shrink coefficients exactly to 0.

Figure 1 (Friedman et al., 2001) shows a common visualisation which compares the l_2 Ridge regularisation against the l_1 Lasso. This picture depicts that for the same regularisation cost (the blue area) that the two different methods can result in different solutions.

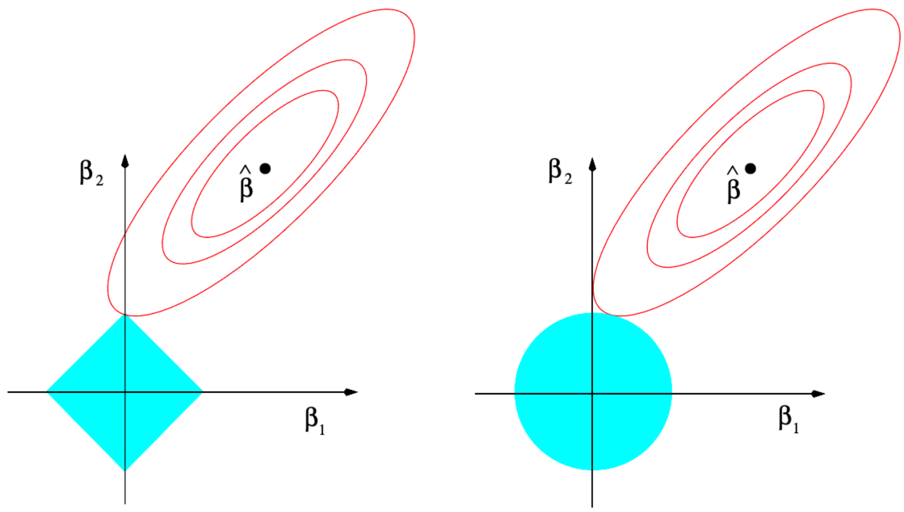

Figure 1: Comparing l_1 Lasso (left) and l_2 Ridge (right) estimations. Shown are contours of the error and constraint functions. The solid blue areas are the constraint regions $|\beta_1| + |\beta_2| \leq t$ and $\beta_1^2 + \beta_2^2 \leq t^2$, respectively, while the red contours are the contours of the least squares error function.

In Figure 1, the $\hat{\beta}$ point shows the least squares fit and the contours represent different lambda values with the ones furthest away from $\hat{\beta}$ representing a larger lambda. Visibly in Figure 1, the Lasso is able to shrink one of the parameters, β_1 in this case, to 0 because of the properties of the constraint, highlighted by the blue region. However, in cases where $p > N$, the Lasso model selects at most N variables before the model saturates. Also, in the presence of multiple highly correlated variables, the Lasso tends to select one variable from the group and ignore the others. The later limitation is particularly inefficient when using the Lasso method for classification.

But methods to select the optimal amount of shrinkage in regularisation methods vary. A common way to select the size of the penalty (tuning parameter) when using Lasso by Cross Validation (CV), a method used in this thesis.

Briefly to explain, given a training data set $(x_i, y_i), i = 1, ...n$ and an estimator $\hat{f}_\theta$, depending on a tuning parameter θ of some discrete set $(\theta_1, ...\theta m)$, CV estimates the prediction error. In CV, the training data set $(1, ...n)$ is divided into K folds of roughly equal size (in this study $K = 10$), $F_1, ...F_K$. For $k = 1, ...K$, consider training on $(x_i, y_i), i \notin F_k$, and validating on $(x_i, y_i), i \in F_k$. Then for each tuning parameter value $\theta \in (\theta_1, ...\theta_m)$, compute the estimate of $\hat{f}_\theta^{-k}$ on the training set, and record the total error on the validation set $e_k(\theta) = \sum_{i \in F_k} (y_i - \hat{f}_\theta^{-k}(x_i))^2$. Then the average error across all folds for each tuning parameter θ is computed,

$$CV(\theta) = \frac{1}{n}\sum_{k=1}^{K} e_k(\theta) = \frac{1}{n}\sum_{k=1}^{K}\sum_{i\in F_k}(y_i - \hat{f}_\theta^{-k}(x_i))^2. \tag{10}$$

Then the value of the tuning parameter which minimises this $CV(\theta)$ curve is chosen

$$\hat{\theta} = \underset{\theta\in(\theta_1,\dots\theta_m)}{\arg\min}\ CV(\theta). \tag{11}$$

However, in this thesis the Adaptive Lasso is also used. The Adaptive Lasso essentially performs a two-stage CV selection. Both Zou (2006) and Bühlmann and Van de Geer (2011) explore how the Adaptive Lasso can lead to more parsimonious models compared to the CV Lasso when over-selection is an issue. The Adaptive Lasso selection first starts with an initial CV, before then conducting another CV amongst the selected covariates from the first round. Interestingly in the second step, weights are applied to the penalised coefficients from the first step. Therefore, covariates with smaller coefficients are more likely to be shrunk to 0 in the second step. The Adaptive Lasso penalty can be defined by:

$$\lambda\sum_{j=1}^{p}\hat{w}_j|\beta_j|. \tag{12}$$

Where the weight vector can be defined by $\hat{w} = 1/|\hat{\beta}_j|^\gamma$ (Assuming that $\hat{\beta}$ is a root-n-consistent estimator; for example, $\hat{\beta}^{ols}$. Also a positive γ is selected, such that $\gamma > 0$ (Zou, 2006)). Zou (2006) show that the Adaptive Lasso enjoys the oracle properties[36] by using the adaptively weighted l_1 penalty. Resultantly, the Adaptive Lasso tends to lead to more parsimonious models than the CV Lasso.

Specifically in this study, I first followed used the CV Lasso approach with the aim of selecting a subset of key covariates, smaller than the original covariate list[37]. I then used the Adaptive Lasso method to create more parsimonious models compared to the CV Lasso method, with the hope of increasing the interpretability of the output.

Zou and Hastie (2005) introduced the Elastic Net penalty in an attempt to outperform the Lasso model in terms of prediction and interpretability during classification, while still being attributed with a similar sparsity of representation. The penalisation penalty introduced by Zou and Hastie incorporates elements from both the l_1 Lasso and l_2 Ridge norms, allowing the Elastic Net to select variables like the Lasso, but collectively shrink correlated predictors like in the Ridge method. The Elastic Net penalty is be defined by

[36]The oracle property known when the asymptotic distribution of an estimator is the same as the asymptotic distribution of the MLE (maximum likelihood estimation) on only the true support. Therefore, in the context of asymptotic distribution, the estimator adapts to knowing the true support without paying a price.

[37]Section A.2 Figure 2 in the Appendix.

$$\lambda \sum_{j=1}^{p} (\alpha\beta_j^2 + (1-\alpha)|\beta_j|). \tag{13}$$

Equation (13) shows the combination of the l_1 and l_2 norm in the Elastic Net's penalisation term.

The Elastic Net method encourages a grouping effect, which means that groups of highly correlated predictors will collectively tend to be in or out of the model after selection. This is different to the Lasso where certain covariates from a highly correlated group of covariates can be selected. Hence, in this thesis I use the Elastic Net method as a type of selection validity check to mitigate against potential selection errors from the covariates that the Lasso models may exclude[38].

Combining the output from the regularisation methods with the findings from the causal literature reviewed around improving learning outcomes, I created an OLS regression model to uncover coefficients of key predictors closer to partial derivative effects relating to learning outcomes. The following model[39] was estimated for both mathematics and language learning outcomes across Kenya, Mozambique, Nigeria, Uganda, Tanzania, and 'All Countries'[40] samples:

$$\begin{split} A_i = \beta_0 + \beta_1 TSK_i + \beta_2 TPS_i + \beta_3 TE_i + \beta_4 TA_i + \beta_5 TB_i + \beta_6 STR_i + \beta_7 TST_i \\ + \beta_8 F_i + \beta_9 PTA_i + \beta_{10} SC_i + \beta_{11} SPS_i + \beta_{12} PS_i + \beta_{13} UR_i + \beta_{14} SB_i + \beta_{15} SA_i + \epsilon_i. \end{split} \tag{14}$$

Each observation $_i$ represents an average in each selected school from the SDI sample. Notably, $\beta_1, ..., \beta_{15}$ will attempt to uncover the partial derivative effects rather than highlighting the total derivative effects. Table 5 briefly presents a list of the covariates used in the OLS model (14). For full descriptions, please refer to Table 12 in the Appendix.

[38] With the hope that the Elastic Net models will initially select more covariates compared to the Lasso models which potentially include key, but correlated, predictors of learning outcomes.

[39] Note that no private schools (P) were sampled in Mozambique, and that no data regarding Parent-Teacher Associations (PTA) was collected in Mozambique or Tanzania.

[40] Because of the small sample sizes, and missing observations further crippling the size of the samples, I created the 'All Countries' sample which combined all observations, irrespective of country.

OLS Variable List	
A_i	Learning Outcomes
TSK_i	Teacher Subject Knowledge
TPS_i	Teacher Pedagogical Skill
TE_i	Teacher Education that Surpasses Secondary School
TA_i	Teacher Absenteeism
TB_i	Positive Teacher Behaviour Index
STR_i	Student-Teacher Ratio
TST_i	Share of Instruction Time Spent Teaching
F_i	Facilities Index
PTA_i	Parent Teacher Association Dummy
SC_i	School Committee Dummy
SPS_i	Students Per Classroom Stream
P_i	Private School Dummy
U_i	Urban Dummy
SB_i	Students Breakfast
SA_i	Student Absence Rate

Table 5: Presents the OLS variable list and short descriptions on the right of the corresponding selected variables.

The OLS models were then tested for heteroskedasticity[41] and multicollinearity[42] in the form of White (1980), Breusch–Pagan (BP) (1979) and Variance Inflation Factor (VIF) tests. The VIF process creates auxiliary regressions for each covariate, to see how much each covariate is being explained by the other covariates. To avoid initial excess collinearity, where necessary, observations were averaged at the school level.

[41]Heteroskedasticity occurs when $V(Y|X)$ is not constant which means that the residuals do not have a constant variance.

[42]Multicollinearity occurs when the X variables are correlated strongly with each other, which leads to obscured coefficients. The regression estimates will then struggle to distinguish which covariate might be the best predictor of learning outcomes. If present, multicollinearity will inflate the variance of the affected variables. Therefore, standard errors would be high and perhaps there would be a lack of statistically significant predictor variables.

7 Results

In this section, I present the results from the regularisation techniques, OLS regressions, and tests for multicollinearity and heteroskedasticity. The Lasso and Elastic Net methods should help discover what the key predictors of learning outcomes might be in the SDI data, whereas the OLS method will produce more quantifiable coefficients which aim to mimic partial derivative effects of the selected key covariates. Tests for multicollinearity and heteroskedasticity will uncover aspects of how valid the OLS results are.

Firstly across all samples, the CV Lasso method successfully performed variable selection by selecting a group of covariates with non-zero coefficients smaller than the original list of potential covariates inputted into the models.

Secondly and also across all samples, the Adaptive Lasso method successfully produced more parsimonious models by performing variable selection which selected a smaller group of covariates with non-zero coefficients compared to the CV Lasso.

Thirdly, the Elastic Net method successfully produced less parsimonious models, for the purpose of variable selection validation, by performing variable selection which selected a larger group of covariates with non-zero coefficients compared to the CV Lasso and Adaptive Lasso models[43].

In terms of prediction, the Adaptive Lasso outperformed the CV Lasso and Elastic Net methods across all samples[44]. For statistical summaries of the regularisation methods, please see section A.5 in the Appendix.

Table 6 reports the largest 3 non-zero post shrinkage coefficients from the CV Lasso, Adaptive Lasso, and Elastic Net methods across all samples.

For full lists of the selected covariates across all samples, please refer to section A.4 in the Appendix, where the tables also report the selected post-shrinkage standardised coefficients.

Table 7 and 8 show the OLS regression results for the selected samples across mathematics and language learning outcomes respectively, following the estimation equation (14).

$$A_i = \beta_0 + \beta_1 TSK_i + \beta_2 TPS_i + \beta_3 TE_i + \beta_4 TA_i + \beta_5 TB_i + \beta_6 STR_i + \beta_7 TST_i$$
$$+\beta_8 F_i + \beta_9 PTA_i + \beta_{10} SC_i + \beta_{11} SPS_i + \beta_{12} PS_i + \beta_{13} UR_i + \beta_{14} SB_i + \beta_{15} SA_i + \epsilon_i{}^{45}.$$

The resultant descriptive statistics for the OLS models can be found in section A.6, in the Appendix.

[43]Apart from in Mozambique, where the Elastic Net method selected the same number of non-zero coefficients as the CV Lasso.

[44]In terms of having the highest out of sample R^2 and lowest CV mean prediction error.

Top 3 Non-Zero Coefficients From Machine Learning Methods						
Panel A: Mathematics	**Kenya**	**Mozambique**	**Nigeria**	**Uganda**	**Tanzania**	**All**
CV Lasso	P+ G3 SPS- SB+	TB+ Student Textbook %+ SA-	P+ % Male Teachers- TSK+	G4 SPS- P+ Occupied Classrooms+	Total Teachers+ Students Teacher Visited- Toilets Clean+	TSK+ P+ TPS+
Adaptive Lasso	P+ G3 SPS- SB+	Student Textbook %+ SA- School Days+	P+ % Male Teachers- Teacher Age-	G4 SPS- P+ Total Teachers+	Total Teachers+ Students Teacher Visited- Chalkboard Lux-	P+ TSK+ Girls' Toilets+
Elastic Net	P+ SB+ STR-	TB+ Student Textbook %+ SA-	P+ % Male Teachers- Teacher Age-	P+ G4 SPS- Occupied Classrooms+	Students Teacher Visited- STR- Total Teachers+	TSK+ P+ STR-
Panel B: Language						
CV Lasso	G3 SPS- P+ SB+	SA- TB+ SC Meetings+	P+ Student Textbook %+ Local Language-	G3G4 SPS- P+ Material on Wall+	Total Teachers+ Toilets Clean+ U+	P+ TSK+ TPS+
Adaptive Lasso	Total G4 Boys- Total Classrooms+ Total G3 Students-	SA- TB+ SC Meetings+	P+ Local Language- Student Textbook %+	G3G4 SPS- P+ School Opening Year-	Total Teachers+ Toilets Clean+ Total G4 Boys-	P+ TSK+ Total Teachers+
Elastic Net	G3 SPS- P+ SB+	SA- TB+ SC Meetings+	P+ Local Language- Student Textbook %+	P+ Material on Wall+ School Opening Year-	Total Teachers+ Toilets Clean+ U+	P+ TSK+ TPS+

Table 6: Shows the top 3 largest post shrinkage coefficients across each sample and their corresponding coefficient sign. Across the individual models and samples, the coefficients are ordered so that the largest is top. Note that P = private school dummy, $G3$ SPS = students per stream in grade 3, SB = percentage of children who ate breakfast, STR = student-teacher ratio, Total G4 Boys = total male students in grade 4, Total Classrooms = total number of classrooms, Total G3 Students = total number of students in grade 3, TB = teacher behaviour categorical variable, Student Textbook % = percentage of children in the classroom who had a textbook, SA = average student absenteeism rate, School Days = total number of school days in operation, SC Meetings = number of meetings from the school committee in the year, % Male Teachers = percentage of male teachers sampled, TSK = proxy for teacher subject knowledge, Teacher Age = average teacher age, Local Language = if the teacher used a local language for classroom instruction dummy, $G4$ SPS = students per stream in grade 4, Occupied Classrooms = total number of classrooms with students inside, Total Teachers = total number of teachers registered, $G3G4$ SPS = average of students per stream across grade 3 and grade 4, Material on Wall = dummy if material other than students work was displayed on the observed classroom wall, School Opening Year = year that the school opened, Students Teacher Visited = total number of students that the teacher individually visited, Toilets Clean = dummy for if the student toilets were clean, Chalkboard Lux = the chalkboard lux measure, STR = the school's student teacher ratio, U = urban dummy, TPS = teacher pedagogical skill proxy, and Girls' Toilets = total number of girls' toilets. For full variable descriptions and selection lists, please refer to Appendix sections A.2 and A.4 respectively.

OLS: Mathematics Learning Outcomes

	Kenya	Mozambique	Nigeria	Tanzania	Uganda	All
TSK_i	.0456	.0623**	.138***	.0983***	.0868***	.208***
	(.0399)	(.0311)	(.0294)	(.0345)	(.0278)	(.0146)
TPS_i	-.0128	-.0086	.0853	.0087	.086**	.1754***
	(.0369)	(.0418)	(.0573)	(.0505)	(.0418)	(0.227)
TE_i	.0472***	-.0017	.033	.0121	.0349	-.0017
	(.0156)	(.0116)	(.0416)	(.0241)	(.0222)	(.0078)
TA_i	-.0304**	-.0131	-.0557**	-.0358**	-.0381***	-.0163*
	(.0151)	(.0116)	(.0256)	(.0164)	(.0124)	(.0086)
TB_i	.0006	.0052**	.0028	.006**	.0015	.0037***
	(.0019)	(.0021)	(.0021)	(.0024)	(.0022)	(.0011)
STR_i	-.0001*	.00002	-.00003	-.0001***	-.0001***	-.0001***
	(.0001)	(.00003)	(.00004)	(.00003)	(.00002)	(.00001)
TST_i	.0473	.0282	-.0344	.0298	.0328	-.0308*
	(.0342)	(.0584)	(.0258)	(.0355)	(.0398)	(.0158)
F_i	.0035	.0039	.0121***	.014***	.0149***	.0127***
	(.0045)	(.0029)	(.0022)	(.0046)	(.0038)	(.0015)
PTA_i	-.0198*	N/A	-.0259	N/A	.0535***	N/A
	(.0107)		(.0283)		(.0157)	
SC_i	-.0004	-.0001	.008	-.0065	-.0173	.0171*
	(.0231)	(.0115)	(.0131)	(.074)	(.0284)	(.009)
SPS_i	-.0016***	.0001	-.0012***	-.0001	-.0005***	-.0002*
	(.0004)	(.0003)	(.0003)	(.0001)	(.0001)	(.0001)
P_i	.0747***	N/A	.118***	-.0135	.0683***	.0976***
	(.0161)		(.0115)	(.076)	(.0139)	(.0077)
U_i	.0082	-.0079	.0214	.0919***	.0269**	.0295***
	(.0117)	(.0139)	(.0135)	(.0161)	(.0128)	(.0072)
SB_i	.0613*	-.0288	.0573**	-.0714***	.0243	-.0282**
	(.0330)	(.018)	(.0287)	(.0209)	(.0196)	(.0109)
SA_i	-.0266	-.0586***	-.0083	-.0309	-.0364	-.0508***
	(.0365)	(.0216)	(.0228)	(.0292)	(.0312)	(0.013)
Constant	.4762***	.1972***	.1861***	.2962***	.2284***	.2111***
	(.0657)	(.065)	(.0533)	(.0924)	(.0659)	(.0226)
N	*216*	*176*	*647*	*321*	*321*	*1,697*
R^2	0.4686	0.176	0.391	0.3306	0.5230	0.4914
Adjusted R^2	0.4288	0.1099	0.3766	0.3	0.4995	0.4872
F Statistic	11.76	2.66	27.01	10.79	22.29	116.1
	(0)	(0.002)	(0)	(0)	(0)	(0)

Table 7: Shows the OLS regression results for mathematics outcomes across the samples. Regression coefficients are reported first, with standard errors reported in brackets underneath, however the brackets in the F statistic column report the corresponding P-values. For all coefficients, significance is reported, * = p < 0.1, ** = p < 0.05, *** = p < 0.01. Reported coefficients were rounded to 4 decimal places, or to 5 decimal places to capture the first non-zero integer.

OLS: Language Learning Outcomes

	Kenya	Mozambique	Nigeria	Tanzania	Uganda	All
TSK_i	.0469	.2349*	.2574***	.1299*	.0329	.336***
	(.1034)	(.1216)	(.0701)	(.0773)	(.0865)	(.0334)
TPS_i	.0409	-.0144	.3743***	.0015	.2031**	.3752***
	(.0665)	(.1086)	(.0844)	(.0718)	(.0926)	(.0348)
TE_i	.0677**	-.0647	.0289	.0079	.0346	.0171
	(.0285)	(.0671)	(.0271)	(.0298)	(.0417)	(.0122)
TA_i	-.0076	-.0437	-.0514	-.0149	-.051*	-.0249*
	(.02611)	(.0306)	(.0314)	(.0229)	(.0266)	(.0132)
TB_i	-.0008	.0182***	.0099***	.0072**	.0094*	.0097***
	(.0036)	(.0058)	(.0033)	(.0034)	(.0049)	(.0019)
STR_i	-.0002**	.00003	-.0001	-.0002***	-.0003***	-.0003***
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.00003)
TST_i	.0309	-.0286	-.0449	-.0137	.1055	-.0571**
	(.0620)	(.1573)	(.0422)	(.0509)	(.0911)	(.0272)
F_i	.0208**	.0125	.0261***	.0303***	.0433***	.0301***
	(.0080)	(.0077)	(.0038)	(.0067)	(.0082)	(.0026)
PTA_i	-.0137	N/A	.0326	N/A	.0233	N/A
	(.0195)		(.0466)		(.0341)	
SC_i	-.0028	-.02	.0027	-.2019*	-.0169	.0078
	(.0433)	(.0306)	(.0214)	(.1023)	(.0642)	(.0154)
SPS_i	-.0027***	-.0005	-.0024***	.0003	-.0012***	-.0005***
	(.0006)	(.0009)	(.0005)	(.0002)	(.0003)	(.0001)
P_i	.0486*	N/A	.2241***	.1418	.1086***	.1766***
	(.0285)		(.0188)	(.1063)	(.0295)	(.0129)
U_i	.0398*	-.0164	.09***	.1138***	.05*	.0685***
	(.0213)	(.038)	(.0217)	(.0235)	(.0279)	(.012)
SB_i	.1936***	.0027	.1231**	-.0538*	.18***	.0382**
	(.0581)	(.0492)	(.0478)	(.0303)	(.0427)	(.0184)
SA_i	-.1593**	-.317***	-.0642*	-.0674	-.2017***	-.1502***
	(.0666)	(.0589)	(.0379)	(.0412)	(.0656)	(.0227)
Constant	.457***	.1741	-.0761	.3969***	-.0259	.0206
	(.1286)	(.176)	(.0762)	(.131)	(.1444)	(.0383)
N	*219*	*176*	*571*	*306*	*314*	*1,603*
R^2	0.4287	0.3431	0.5575	0.3853	0.6118	0.5758
Adjusted R^2	0.3864	0.2904	0.5455	0.3557	0.5922	0.5721
F Statistic	10.15	6.51	46.62	13.03	31.3	153.99
	(0)	(0)	(0)	(0)	(0)	(0)

Table 8: Shows the OLS regression results for language outcomes across the samples. Regression coefficients are reported first, with standard errors reported in brackets underneath, however the brackets in the F statistic column report the corresponding P-values. For all coefficients, significance is reported, * = p < 0.1, ** = p < 0.05, *** = p < 0.01. Reported coefficients were rounded to 4 decimal places, or to 5 decimal places to capture the first non-zero integer.

Table 9 shows BP, White, and VIF test results for all sampled OLS models. Notably, the 'All Countries' Mathematics, 'All Countries' Language, and Nigeria Mathematics models are the only models which show no evidence of heteroskedasticity or excess multicollinearity[46].

OLS Robustness Checks

	BP Test	White Test	Mean VIF
'All Countries' Mathematics	39.35 (0)	226.94 (0)	1.35
'All Countries' Language	6.68 (0.0097)	252.83 (0)	1.33
Kenya Mathematics	1.34 (0.2467)	149.44 (0.1168)	1.30
Kenya Language	38.06 (0)	124.10 (0.6053)	1.27
Mozambique Mathematics	2.13 (0.1444)	96.63 (0.6314)	1.17
Mozambique Language	27.46 (0)	109.76 (0.2822)	1.15
Nigeria Mathematics	31.57 (0)	190.99 (0.0005)	1.27
Nigeria Language	0.04 (0.8358)	162.97 (0.0305)	1.26
Uganda Mathematics	0.02 (0.9009)	101.22 (0.9416)	1.25
Uganda Language	0.47 (0.4907)	133.42 (0.2658)	1.25
Tanzania Mathematics	0.52 (0.4727)	99.22 (0.3103)	1.14
Tanzania Language	0.42 (0.5151)	98.56 (0.3270)	1.16

Table 9: Shows the results of the robustness checks for each OLS model. In the BP Test and White Test columns, the Chi-squared is initially reported, and then in brackets the probabilities of the null hypothesis, H0, are reported.

[46]For lists of the highest and lowest VIF scores for each country, please see section A.7 the Appendix.

8 Discussion

The regularisation results from Table 6 suggest that private schools, teacher subject knowledge, and teacher pedagogical skills could be the best predictors of learning outcomes captured by the SDI data in the sample of 'All Countries'[47]. The complementary OLS results from Tables 7 and 8 show that the private school, teacher subject knowledge, and teacher pedagogical skills covariates also had the largest OLS regression coefficients with mathematics and language learning outcomes in the 'All Countries' sample[48]. Specifically, the teacher subject knowledge and teacher pedagogical skills covariates had the largest OLS coefficients in mathematics and language learning outcomes models respectively, implying that a 1% increase in teachers' mathematics test scores were attributed with higher student mathematics test scores of 20.8%, and that a 1% increase in teachers' pedagogical test scores were attributed with higher student language test scores of 37.52%. Both results were statistically significant at 1%.

Despite the largest coefficients from the 'All Countries' OLS models in Tables 7 and 8 mimicking the largest post-shrinkage coefficients in Tables 6, ranking the best predictors of learning outcomes by size of the OLS coefficients is not optimal because, unlike the regularisation methods, the coefficients from the OLS analysis are not standardised[49].

Across the regularisation results for the individual country samples, the results vary, but the private school covariate remained prevalent[50]. In the OLS analysis, the private school variable was positive and statistically significant at 1% in the models for mathematics and language learning outcomes in the 'All Countries' sample, implying that private schools were associated with higher test scores of 9.76% in mathematics student test scores and 17.66% in language student test scores. Across all individual country samples for mathematics and language learning outcomes which included private schools, the OLS coefficient for private schools was positive and statistically significant, apart from in Tanzania where the coefficients were statistically insignificant. These results point towards a strong association between private schools and higher learning outcomes for students, but similar to the literature, the positive association between private schools and learning outcomes fail to distinguish if a lack of random assignment and un-observable factors are key causal influences driving this effect ((Angrist, 2002), (Angrist et al., 2006), (Muralidharan and Sundararaman,

[47]With the 3 covariates most frequently appearing in the top 3 largest post-shrinkage coefficient lists across CV Lasso, Adaptive Lasso, and Elastic Net methods.

[48]Interestingly, these covariates had larger OLS coefficients when explaining language learning outcomes compared to mathematics learning outcomes.

[49]Thus, the effect implied from a regression coefficient from one standard deviation in one independent variable is not directly comparable to another standard deviation in an alternative independent variable.

[50]Apart from in Mozambique where no private schools were sampled. Tanzania represented the only country in the sample that included private schools where the private school covariate did not feature in the top 3 largest post shrinkage coefficients across mathematics and language learning outcomes using the CV Lasso, Adaptive Lasso, and Elastic Net methods.

2015), (Hsieh and Urquiola, 2006), and (Lucas and Mbiti, 2014)).

Again from Table 6 and the 'All Countries' sample, teacher subject knowledge was highlighted as a key predictor of learning outcomes, which may seem obvious, but prior literature around the association between teacher subject knowledge and learning outcomes in SSA is limited. Tables 13 and 14, show that all regularisation methods selected a positive teacher subject knowledge coefficient as a key predictor of learning outcomes, with the post-shrinkage covariate coefficient being the largest, or second largest, across the CV Lasso, Adaptive Lasso, and Elastic Net methods for both mathematics and language learning outcomes in the 'All Countries' sample. Furthermore, the findings presented in Tables 7 and 8 suggested that a 1% increase in teacher subject knowledge was associated with higher student mathematics scores of 20.8% and higher language scores of 33.6%[51]. Notably, Table 3 suggests that 71% of the sampled mathematics teachers had the minimum knowledge to teach mathematics in the 4th grade, and only 11% of the sampled language teachers had the minimum literacy knowledge to teach in the 4th grade[52], highlighting deficiencies in the knowledge of teachers, which could prove to be beneficial for student learning outcomes if addressed.

Similarly in Table 6, the 'All Countries' sample highlighted teacher pedagogical skills as a key predictor of learning outcomes and the OLS results in Tables 7 and 8 show that teacher pedagogical skills had a positive and significant association with both mathematics and language learning outcomes at 1% significance, suggesting that a 1% increase in teachers' pedagogical skills was associated with a 17.54% increase in mathematics scores, and a 37.52% increase in language scores. Despite this, it is estimated that only 13% of the sampled teachers have the minimum pedagogical knowledge to teach[53].

While the regularisation methods used in this thesis were able to perform variable selection and select key covariates which can be used to predict learning outcomes, they fail to infer causality and may frequently select covariates that are highly correlated to learning outcomes with low causal impacts while failing to select covariates with higher causal impacts. Hence, using Lasso and Elastic Net methods for classification presents challenges. Arguably, these methods are more suited for the purpose of prediction, compared to classification problems[54]. Nevertheless, the CV Lasso method accomplished selecting a group of covariates, smaller than the original group of covariates which did shed some light into certain relationships between the selected variables and learning outcomes. Also, the Adaptive Lasso model succeeded by selecting a smaller subset of covariates compared to the CV Lasso model. But, creating sparse explanatory models for phenomena which may be more complex and diverse in reality, may not provide much value when the main purpose of the analysis is for classification. Considering the tendency of the Lasso models to select one covariate

[51] Both results were statistically significant at 1%.

[52] Estimated in Panel's A and Panel B, and following Bold et al.'s (2017) classification.

[53] Estimated in Table 4 Panel A, and again following Bold et al.'s (2017) classification.

[54] Because the CV Lasso, Adaptive Lasso, and Elastic Net choose the optimal level of shrinkage with the goal of minimising the prediction error.

from a highly correlated group of covariates provides further issues for classification and the Elastic Net method proved valuable by selecting expected key, and intercorrelated, predictors of learning outcomes which the Lasso models omitted[55]. Therefore, the Elastic Net method was useful when performing a type of validity check for the Lasso variable selection methods.

Also the regularisation methods fail to convey interpretable coefficients mimicking partial derivative effects of the education production function. Therefore, I conducted subsequent OLS[56] analysis. I also then tested the statistical significance of the covariates, and checked for evidence of heteroskedasticity and excess multicollinearity. To detect heteroskedasticity, I conducted White and BP tests. A problem with the BP test is that it assumes that the heteroskedasticity is a linear function of the selected independent variables, whereas the White test can accommodate a non-linear relationship between the independent variables and the error variance. This can explain why the BP test in Table 9 failed to recognise heteroskedasticity in the models for Kenya and Mozambique language learning outcomes, despite the White test rejecting the null hypothesis of homoskedasticity. Heteroskedasticity could be present for many reasons; outliers in the data, and omitted variables are two reasons which I anticipate. To test for multicollinearity, I conducted VIF analysis, whereby examination of the latent roots and latent vectors of a correlation matrix can provide a sufficient procedure for detecting multicollinearity (Mansfield and Helms, 1982). Despite different researchers debating how much multicollinearity is unreasonable and excessive, no individual VIF scores looked excessively inflated from Table 9 so I deemed multicollinearity to not be a serious issue with the estimated OLS results[57].

The OLS results, which regressed the selected covariates[58] on mathematics and language learning outcomes in Kenya, Mozambique, Nigeria, Uganda, Tanzania, and the 'All Countries' sample, produced an array of results which could be subject to selection and measurement issues. Also, many variables were not statistically significant and robustness checks showed signs of heteroskedasticity in all models apart from Nigeria mathematics, 'All Countries' mathematics, and 'All Countries' language[59].

Noticeably, the coefficient for the average student breakfast covariate varied in the mathematics and language 'All Countries' OLS models. The coefficient was negative and significant at 5% in the mathematics OLS model, implying

[55] For example, in the 'All Countries' mathematics sample (Table 13), and the Kenya mathematics sample (Table A.4) the Elastic Net model selected expected key covariates like teacher absenteeism rates, and teacher subject knowledge, respectively, which the Lasso models omitted.

[56] The selected covariates forming the OLS models were influenced by the prior machine learning methods, education production function theory, and the literature reviewed.

[57] But multicollinearity did clearly affect the varying selection results from the CV Lasso, Adaptive Lasso, and Elastic Net methods.

[58] See Table 12.

[59] However, by combining all countries into one large sample, countries with comparatively more observations, like Nigeria, had a greater impact on the results compared to other countries will less observations.

that an 1% increase in children who ate breakfast before attending school was associated with lower student mathematics test scores by 2.28%. But, the coefficient in the language model was positive and also significant at 5%, implying that an 1% increase in children who ate breakfast before attending school was associated with higher student language test scores by 3.82%. The differing coefficient signs for the average student breakfast covariate highlight an interesting focus for further research as this variable could even proxy elements related to household's demand for education.

Perhaps unsurprisingly, the OLS results from Tables 7 and 8 show that teacher absenteeism rates and learning outcomes had a negative relationship for the 'All Countries' mathematics and language models, this relationship was significant at 10% in both models. Similarly, the teacher absenteeism post shrinkage coefficient was negative when selected in the 'All Countries' mathematics and language regularisation models. But notably, for mathematics, only the Elastic Net model selected the teacher absenteeism covariate, whereas in the alternate language model, the CV Lasso and Elastic Net methods did choose the covariate as a key predictor. The descriptive findings from Table 2 estimate that teachers are absent from class 43% of the time, as an average of all sampled countries, and again this highlights a striking deficiency in the provision of schooling across the sampled schools. Perhaps changing the contract structure for teachers which incentivises attendance could prove beneficial for student learning outcomes ((Muralidharan and Sundararaman, 2013), (Duflo et al., 2012)).

Also in the 'All Countries' OLS models, the coefficient for the facilities index was positive and significant at 1% for both mathematics and language learning outcomes. Also not only for the 'All Countries' regularisation models, but also in the individual countries samples, an array of physical schooling inputs/facilities were selected across CV Lasso, Adaptive Lasso, and Elastic Net frameworks. But, quantifying the effect of specific facilities and physical schooling inputs remains challenging ((Glewwe et al., 2009), (Sabarwal et al., 2014), (Borkum et al., 2012), (Glewwe et al., 2004)).

Nevertheless, I would not expect the true partial derivative effects to be identical to the OLS coefficients presented in this thesis[60]. A key omission when exploring the best predictors of learning outcomes in this study is a variable for innate ability. Ability is notoriously hard to measure, and perhaps the SDI data provides a proxy for ability if data regarding students' non-verbal test scores were used. However, I opted not to use this data as a potential control, because I believed that these results would not have been independent from effects of school related covariates. Hence, the Conditional Independence Assumption (CIA) would not hold and this would have been a poor control variable when trying to infer causality. Perhaps future studies could follow Lara et al.'s (2011) propensity score econometric techniques and changes-in-changes estimation methods with data which contains the previous academic

[60] Due to omitted variables, selection issues, and also measurement issues with some covariates (Especially covariates derived from Module 4 of the SDI data where one observation formed a school's average).

achievement levels of students to address concerns for differences in students' unobservable characteristics. This approach would also complement Todd and Wolpin's (2003) recommended approach.

While the results presented in this study do not imply causality, and could suffer from selection and measurement problems, the regularisation methods enabled feature selection, and ranking, of covariates, which provided some insight when exploring the best predictors of learning outcomes. The subsequent OLS results presented more interpretable coefficients between key covariates and learning outcomes which most likely represent biased partial derivative effects. The OLS results also highlighted contradictory relationships between the percentage of children who ate breakfast before attending school and the association with mathematics and language learning outcomes in the 'All Countries' model, perhaps a future study could explore these findings further[61].

To discover what the best predictors of learning outcomes are for progressive education policy decisions in SSA, more data is needed to estimate truer partial derivative effects, while experimental designs could aid the discovery of what the total derivative effects from the resultant policy actions might be.

9 Conclusion

This thesis further highlights deficiencies in the provision of education and learning outcomes across the sampled schools in SSA. The regularisation methods were able to standardise variables and perform feature selection to uncover what the best predictors of learning outcomes could be in the SDI data. The selection methods produced an array of results across the sample, but the private school covariate was often selected as a key predictor of learning outcomes. Other prominent covariates were related to teacher subject knowledge and teacher pedagogical skills[62]. Firstly, I used the CV Lasso method to perform variable selection, before using the Adaptive Lasso method to produce even more parsimonious models[63]. Then, I used the Elastic Net method as a selection validity check because of the tendency for Lasso models to exclude highly correlated covariates. The Elastic Net method proved that the Lasso methods omitted expected key predictors of learning outcomes[64].

I then conducted OLS analysis to discover quantifiable associations between key covariates and learning outcomes. The OLS results suggested that many of the samples suffered from heteroskedastic errors and it is anticipated that the coefficients do not mimic true partial derivative effects. The omission of key unobservable factors which influence student test results must also be noted. Nevertheless, the OLS results uncovered some peculiar results[65] and showed

[61] Assuming that the student breakfast covariate is a proxy for unobserved household characteristics like income and parental education, a negative relationship with mathematics learning outcomes is baffling and remains unexplained.

[62] Specifically in the 'All Countries' sample.

[63] In the hope of enhancing the interpretability of a smaller group of selected covariates.

[64] Like teacher subject knowledge and teacher absenteeism rates.

[65] relating to the varying associations between the average student breakfast covariate and

that the teacher subject knowledge and teacher pedagogical skills covariates had the largest OLS coefficients in mathematics and language learning outcomes models, respectively, in the 'All Countries' sample.

While the results from the CV Lasso, Adaptive Lasso, Elastic Net, and OLS analysis did not imply causality, and may have suffered from selection and measurement problems, the regularisation methods enabled feature selection and ranking of covariates which provided insight into further areas for research and data collection initiatives when exploring the best predictors of learning outcomes.

Future research efforts could extend the data collection methods and conduct experimental analysis to explore the total derivative effects of education production function inputs which can uncover household behaviour in response to policy implementation. If extending the data collection initiative further to households proves too challenging[66], then future studies could follow Lara et al.'s (2011) propensity score econometric techniques and changes-in-changes estimation methods with data containing the previous academic achievement levels of students which complements Todd and Wolpin's (2003) recommended approach.

learning outcomes.

[66]Due to research ethics and processes which must be followed to maintain the confidentiality of agents, open source data with the ability of individual tracking may be unfeasible.

References

Adrogue, C. and Orlicki, M. E. (2013). Do in-school feeding programs have an impact on academic performance and dropouts? the case of public schools in argentina. *education policy analysis archives*, 21:50.

Andrabi, T., Das, J., Khwaja, A. I., and Zajonc, T. (2011). Do value-added estimates add value? accounting for learning dynamics. *American Economic Journal: Applied Economics*, 3(3):29–54.

Angrist, J. (2002). How do sex ratios affect marriage and labor markets? evidence from america's second generation. *The Quarterly Journal of Economics*, 117(3):997–1038.

Angrist, J., Bettinger, E., and Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in colombia. *American economic review*, 96(3):847–862.

Baird, S., Hicks, J. H., Kremer, M., and Miguel, E. (2016). Worms at work: Long-run impacts of a child health investment. *The quarterly journal of economics*, 131(4):1637–1680.

Baker, G. P. (1992). Incentive contracts and performance measurement. *Journal of political Economy*, 100(3):598–614.

Banerjee, A., Glewwe, P., Powers, S., and Wasserman, M. (2013). *Expanding access and increasing student learning in post-primary education in developing countries: A review of the evidence*. Citeseer.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in india. *American Economic Journal: Economic Policy*, 2(1):1–30.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, 122(3):1235–1264.

Barrera-Osorio, F. and Linden, L. L. (2009). *The use and misuse of computers in education: evidence from a randomized experiment in Colombia*. The World Bank.

Baum, D., Lewis, L., Lusk-Stover, O., and Patrinos, H. (2014). What matters most for engaging the private sector in education.

Beasley, E. and Huillery, E. (2017). Willing but unable? short-term experimental evidence on parent empowerment and school quality. *The World Bank Economic Review*, 31(2):531–552.

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? results from a natural experiment in chile. *Economics of Education Review*, 28(5):629–640.

Beuermann, D. W., Cristia, J. P., Cruz-Aguayo, Y., Cueto, S., and Malamud, O. (2013). Home computers and child outcomes: Short-term impacts from a randomized experiment in peru. Technical report, National Bureau of Economic Research.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J., and Wane, W. (2017). Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in africa. *Journal of Economic Perspectives*, 31(4):185–204.

Borkum, E., He, F., and Linden, L. L. (2012). The effects of school libraries on language skills: Evidence from a randomized controlled trial in india. Technical report, National Bureau of Economic Research.

Bourdon, J., Frölich, M., and Michaelowa, K. (2007). Teacher shortages, teacher contracts and their impact on education in africa. *University of St. Gallen Economics Discussion Paper*, (2007-20).

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, pages 1287–1294.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

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

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

Contreras, D. and Rau, T. (2012). Tournament incentives for teachers: evidence from a scaled-up intervention in chile. *Economic development and cultural change*, 61(1):219–246.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics*, 5(2):29–57.

De Ree, J., Muralidharan, K., Pradhan, M., and Rogers, H. (2018). Double for nothing? experimental evidence on an unconditional teacher salary increase in indonesia. *The Quarterly Journal of Economics*, 133(2):993–1039.

Desai, S. and Alva, S. (1998). Maternal education and child health: Is there a strong causal relationship? *Demography*, 35(1):71–81.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74.

Duflo, E., Dupas, P., and Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from kenyan primary schools. *Journal of public Economics*, 123:92–110.

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

Friedman, J., Hastie, T., Tibshirani, R., et al. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Gertler, P. J., Patrinos, H. A., and Rubio-Codina, M. (2012). Empowering parents to improve education: Evidence from rural mexico. *Journal of Development Economics*, 99(1):68–79.

Glewwe, P., Ilias, N., and Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 2(3):205–27.

Glewwe, P., Kremer, M., and Moulin, S. (2009). Many children left behind? textbooks and test scores in kenya. *American Economic Journal: Applied Economics*, 1(1):112–35.

Glewwe, P., Kremer, M., Moulin, S., and Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: the case of flip charts in kenya. *Journal of development Economics*, 74(1):251–268.

Glewwe, P. and Maïga, E. W. (2011). The impacts of school management reforms in madagascar: do the impacts vary by teacher type? *Journal of development effectiveness*, 3(4):435–469.

Glewwe, P. and Miguel, E. (2008). The impact of child health and nutrition in the handbook of development economics.

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

Glewwe, P., Park, A., and Zhao, M. (2016). A better vision for development: Eyeglasses and academic performance in rural primary schools in china. *Journal of Development Economics*, 122:170–182.

Govmda, R. and Josephine, Y. (2005). Para-teachers in india: A review. *Contemporary Education Dialogue*, 2(2):193–224.

Goyal, S. and Pandey, P. (2009). *How do government and private schools differ? Findings from two large Indian states*, volume 30. International Bank for Reconstruction and Development/The World Bank.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, 7:24.

Hsieh, C.-T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from chile's voucher program. *Journal of public Economics*, 90(8-9):1477–1503.

Kazianga, H., De Walque, D., and Alderman, H. (2012). Educational and child labour impacts of two food-for-education schemes: Evidence from a randomised trial in rural burkina faso. *Journal of African Economies*, 21(5):723–760.

Kazianga, H., Levy, D., Linden, L. L., and Sloan, M. (2013). The effects of" girl-friendly" schools: Evidence from the bright school construction program in burkina faso. *American Economic Journal: Applied Economics*, 5(3):41–62.

Kravdal, Ø. (2002). Education and fertility in sub-saharan africa: Individual and community effects. *Demography*, 39(2):233–250.

Kremer, M., Brannen, C., and Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340(6130):297–300.

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

Krishnaratne, S., White, H., et al. (2013). Quality education for all children? what works in education in developing countries. Technical report, International Initiative for Impact Evaluation (3ie).

Lara, B., Mizala, A., and Repetto, A. (2011). The effectiveness of private voucher education: Evidence from structural school switches. *Educational Evaluation and Policy Analysis*, 33(2):119–137.

Lassibille, G., Tan, J.-P., Jesse, C., and Van Nguyen, T. (2010). Managing for results in primary education in madagascar: Evaluating the impact of selected workflow interventions. *The World Bank Economic Review*, 24(2):303–329.

Linden, L. L. (2008). *Complement or substitute?: The effect of technology on student achievement in India*. InfoDev Working Paper, Columbia University.

Lochner, L. and Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American economic review*, 94(1):155–189.

Lucas, A. M. and Mbiti, I. M. (2014). Effects of school quality on student achievement: Discontinuity evidence from kenya. *American Economic Journal: Applied Economics*, 6(3):234–63.

Luo, R., Shi, Y., Zhang, L., Liu, C., Rozelle, S., Sharbono, B., Yue, A., Zhao, Q., and Martorell, R. (2012). Nutrition and educational performance in rural china's elementary schools: Results of a randomized control trial in shaanxi province. *Economic development and cultural change*, 60(4):735–772.

Malamud, O. and Pop-Eleches, C. (2011). Home computer use and the development of human capital. *The Quarterly journal of economics*, 126(2):987–1027.

Mansfield, E. R. and Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician*, 36(3a):158–160.

McEwan, P. J. (2013). The impact of chile's school feeding program on education outcomes. *Economics of Education Review*, 32:122–139.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, 85(3):353–394.

Miguel, E. and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of political economy*, 66(4):281–302.

Mincer, J. (1974). Schooling, experience, and earnings. human behavior & social institutions no. 2.

Mo, D., Zhang, L., Wang, J., Huang, W., Shi, Y., Boswell, M., and Rozelle, S. (2014). The persistence of gains in learning from computer assisted learning (cal): Evidence from a randomized experiment in rural schools in shaanxi province in china. *Unpublished manuscript. Stanford, CA: Rural Education Action Program (REAP).*

Muralidharan, K. (2012). Long-term effects of teacher performance pay: Experimental evidence from india. *Society for Research on Educational Effectiveness.*

Muralidharan, K., Das, J., Holla, A., and Mohpal, A. (2016). *The fiscal cost of weak governance: Evidence from teacher absence in India*. The World Bank.

Muralidharan, K. and Kremer, M. (2006). Public and private schools in rural india. *Harvard University, Department of Economics, Cambridge, MA.*

Muralidharan, K. and Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from india. *Journal of political Economy*, 119(1):39–77.

Muralidharan, K. and Sundararaman, V. (2013). Contract teachers: Experimental evidence from india. Technical report, National Bureau of Economic Research.

Muralidharan, K. and Sundararaman, V. (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in india. *The Quarterly Journal of Economics*, 130(3):1011–1066.

Murnane, R. J. and Ganimian, A. (2014). Improving educational outcomes in developing countries: Lessons from rigorous impact evaluations. *NBER Working Paper*, (w20284).

Olken, B. A., Onishi, J., and Wong, S. (2014). Should aid reward performance? evidence from a field experiment on health and education in indonesia. *American Economic Journal: Applied Economics*, 6(4):1–34.

Orkin, K. (2013). *The effect of lengthening the school day on children's achievement in Ethiopia*. Young Lives.

Ozier, O. (2018). Exploiting externalities to estimate the long-term effects of early childhood deworming. *American Economic Journal: Applied Economics*, 10(3):235–62.

Pandey, P., Goyal, S., and Sundararaman, V. (2009). Community participation in public schools: impact of information campaigns in three indian states. *Education economics*, 17(3):355–375.

Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Alishjabana, A., Gaduh, A., and Artha, R. P. (2011). *Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia*. The World Bank.

Pritchett, L. and Sandefur, J. (2020). Girls' schooling and women's literacy: schooling targets alone won't reach learning goals. *International Journal of Educational Development*, 78:102242.

Reinikka, R. and Svensson, J. (2004). Local capture: evidence from a central government transfer program in uganda. *The quarterly journal of economics*, 119(2):679–705.

Sabarwal, S., Evans, D. K., and Marshak, A. (2014). *The permanent input hypothesis: the case of textbooks and (no) student learning in Sierra Leone*. The World Bank.

Santibanez, L., Abreu-Lastra, R., and O'Donoghue, J. L. (2014). School based management effects: Resources or governance change? evidence from mexico. *Economics of Education Review*, 39:97–109.

Solow, R. M. et al. (1970). Growth theory. an exposition. In *Growth theory. An exposition.* Oxford: Clarendon Press.

Sondheimer, R. M. and Green, D. P. (2010). Using experiments to estimate the effects of education on voter turnout. *American Journal of Political Science*, 54(1):174–189.

Sylvia, S., Luo, R., Zhang, L., Shi, Y., Medina, A., and Rozelle, S. (2013). Do you get what you pay for with school-based health programs? evidence from a child nutrition experiment in rural china. *Economics of Education Review*, 37:1–12.

Tan, J.-P., Lane, J., and Lassibille, G. (1999). Student outcomes in philippine elementary schools: An evaluation of four experiments. *The World Bank Economic Review*, 13(3):493–508.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

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

Tooley, J., Dixon, P., and Gomathi, S. (2007). Private schools and the millennium development goal of universal primary education: A census and comparative survey in hyderabad, india. *Oxford Review of Education*, 33(5):539–560.

Urquiola, M. (2001). *Identifying class size effects in developing countries: Evidence from rural schools in Bolivia*. The World Bank.

Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, 99(1):179–215.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, pages 817–838.

Yang, Y., Zhang, L., Zeng, J., Pang, X., Lai, F., and Rozelle, S. (2013). Computers and the academic performance of elementary school-aged girls in china's poor communities. *Computers & Education*, 60(1):335–346.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

A Appendix

A.1 SDI Overview

To track the quality of service delivery in primary schools, SDI researchers embarked on nationwide data collection field trips whereby enumerators visited sampled schools (which contained at least one 4th-grade class) to complete surveys, distribute questionnaires, and administer assessments[67]. The SDI research team created a Education Field Manual to help with operational deployment of data collection agents. Each sampled school was visited twice, the first visit was announced and the second visit was unannounced, with the later aiming to capture teacher absenteeism data. Refer to Table 10 for enumerators' data collection schedule during observational visits.

<table>
<tr><th colspan="2">SDI Enumerator Data Collection</th></tr>
<tr><td colspan="2">1st Visit (Announced)</td></tr>
<tr><td>Enumerator 1</td><td>Enumerator 2</td></tr>
<tr><td colspan="2">- Arrive at school
- Meet with principle/head teacher, or the most senior teacher present
- Make an introduction, explain the purpose of the visit, and confirm their permission to collate data</td></tr>
<tr><td>- Complete Module 1 (school information)
- Complete Module 2A (list of teachers)
- Select a sample of 10 teachers
- Complete Module 2B (accompanied by a staff member from the school, randomly select and interview 10 to collect personal and absence information)
- Complete Module 3 (school finances, either with the principle/head teacher or school accountant/ treasurer)</td><td>- Complete Module 4 (observe a fourth-grade mathematics or language lesson)
- Complete Module 5 (randomly select 10 pupils from the fourth-grade class and administer the pupil test)
- Complete Module 6 (during the lunch break, administer the teacher test to all fourth-grade teachers, the previous year's third-grade teacher, and 3-5 teachers who teach the fifth-grade and above</td></tr>
<tr><td colspan="2">2nd Visit (Unannounced)</td></tr>
<tr><td colspan="2">- Arrive at school
- Meet with the principle/head teacher again to ask for permission again to collate data
- Accompanied with a member of staff, collect attendance data of the 10 pre-selected teachers from the 1st visit
- Conduct interviews with the teachers that were not present during the 1st visit to complete Module 2B (personal and absence information)
- Count the number of unstaffed classrooms
- Collect any outstanding information to complete all Modules</td></tr>
</table>

Table 10: Highlights the structure that enumerators followed when visiting sampled schools and collecting data.

[67] For an example, Kenya's Service Delivery Indicators Education Survey Questionnaire can be downloaded here: https://microdata.worldbank.org/index.php/catalog/2755/download/39308.

Table 2

SDI Module Descriptions		
Modules	**Data Sources**	**Descriptions**
Module 1 - School Information	Head Teacher	Collects information about school type, facilities, governance, student numbers and schools hours.
Module 2 - Teacher Roster	Head Teacher and Sampled Teachers	Collects a list of school teachers, while measuring absences rates and collecting information regarding teacher characteristics.
Module 3 - School Finances	Head Teacher	Collects information regarding school finances.
Module 4 - Classroom Observations	Classroom Observations	An observations module to monitor teacher behaviour and classroom conditions.
Module 5 - Pupil Assessment	Sampled Pupils	Organises and marks assessments to collects grade 4 test results in language/literature and mathematics.
Module 6 - Teacher Assessment	Sampled Teachers	Organises and marks teacher's assessments as for proxies for their language and mathematics subject knowledge, while also measuring pedagogical skill.

Table 11: Presents SDI module descriptions.

A.2 Full Variable Lists (CV Lasso, Adaptive Lasso, and Elastic Net Models)

Variable Name	Description	Module
av_student_math_p	School Average Creation: Percentage of correct answers from mathematics students assessments from Grade 4 students	5
av_student_lang_p	School Average Creation: Percentage of correct answers from language students assessments from Grade 4 students	5
average_student _breakfast	School Average Creation: Percentage of students who ate breakfast on the SDI first data collection day, taken from child level SDI data	1
urban	Urban school dummy variable: 1 = Urban, 0 = Rural	1
private	Private school dummy variable: 1 = Private, 0 = Public	1
urb_rur_semi	Urban, Rural, Semi-Urban school categorical variable	1
m1_school_type	Type of school categorical variable (day school, boarding school, both, special needs education school, other)	1
m1_school_cat	School Catergory (boys' school, girls' school, co-education)	1
m1_school_year	When the school began operating (categorised in increments of 10 years)	1
m1_sc	School Committee (SC) or a Board of Directors (BoD) dummy variable: 1 = Yes, 0 = No	1
m1_sc_meet_lastyr	How many times the SC/BoD met in the last year	1
m1_sc_minutes	If the enumerator see the minutes of the SC/BoD meetings: 1 = Yes, 0 = No	1
m1_pta	School Parent Teacher Association (PTA dummy: 1 = Yes, 0 = No	1
m1_pta_meet_lastyr	How many times the Parent Teacher Association met last year	1
m1_toilets	If the school has toilets for pupils dummy: 1 = Yes, 0 = No	1
m1_gender_toilets	If the school toilets are gender defined: 1 = Yes, 0 = No	1
m1_boys_toilets	Number of boys toilets the school has	1
m1_girls_n_toilets	Number of girls toilets the school has	1
m1_toilets_clean	Clean toilets dummy: 1 = Yes, 0 = No	1
m1_toilets_private	Private toilets dummy: 1 = Yes, 0 = No	1
m1_toilets_accessible	Accessible toilets dummy (unlocked, not overflowing): 1 = Yes, 0 = No	1
m1_streams_g3	Number of streams in Grade 3	1
m1_streams_g4	Number of streams in Grade 3	1
m1_n_boys_g4	Number of boys in Grade 4	1
m1_n_girls_g3	Number of girls in Grade 3	1
m1_n_girls_g4	Number of girls in Grade 4	1
m1_n_total_g3	Number of students in Grade 3	1
m1_n_total_g4	Number of students in Grade 4	1
m1_days_in_session	Actual number of days during which school was in session in the last year	1
m2_n_teachers	Number of teachers who work in the school	2
m2_n_classrooms	Number of classrooms in the school	2
m2_n_cl_wchild	Number of classrooms in the school which contained pupils	2
m2_n_cl_wchild_noteac	Number of classrooms in the school which contained pupils but no teacher present	2
m4_tid	Unique Teacher ID	4
m4b_room_total	How many pupils were in the classroom	4
m4b_room_boys	How many pupils in the classroom were boys	4
m4b_room_girls	How many pupils in the classroom were girls	4
m4b_cornerlibrary	Was there a corner library, or additional books, in the classroom dummy : 1 = Yes, 0 = No	4

m4b_board	Available blackboard or whiteboard in the classroom dummy: 1 = Yes, 0 = No	4
m4b_chalk	Available chalk or marker to write on the board in the classroom dummy: 1 = Yes, 0 = No	4
m4b_electricity	Working electricity in the classroom dummy: 1 = Yes, 0 = No	4
m4b_work_displayed	Children's work displayed on the classroom walls dummy: 1 = Yes, 0 = No	4
m4b_material	Other than children's work, was there other material displayed on the classroom walls dummy: 1 = Yes, 0 = No	4
m4b_classroom_hyg	Classroom hygiene categorical variable (clean, not clean, semi clean)	4
m4b_board_contrast	Sufficient blackboard or whiteboard contrast in the classroom dummy: 1 = Yes, 0 = No	4
m4_light_front	Sufficient light in the front of the classroom dummy: 1 = Yes, 0 = No	4
m4_light_back	Sufficient light in the back of the classroom dummy: 1 = Yes, 0 = No	4
m4_lux_measure	Lux measure at the chalkboard	4
m4b_pencilpen	Number of pupils with pencil or pen	4
m4b_exbook	Number of pupils with textbook	4
m4c_txtbook_teacher	Textbook used by teacher dummy: 1 = Yes, 0 = No	4
m4c_txtbook_pup	How many pupils used a textbook	4
m4c_board_teacher	Teacher write on board dummy: 1 = Yes, 0 = No	4
m4c_board_pup	Pupils write on board dummy: 1 = Yes, 0 = No	4
m4c_bookpen_n_pup	How many puoiles used paper/exercise book and pencil/pen	4
m4c_teach_localinfo	Teacher use local information to make learning relevant dummy: 1 = Yes, 0 = No	4
m4c_teach_sitting	Teacher sitting or standing infront of class at anytime dummy: 1 = Yes, 0 = No	4
m4c_teach_indch	Teacher visit individual children dummy: 1 = Yes, 0 = No	4
m4c_teach_indch_n	How many pupils did the teacher visit individually	4
m4c_teach_smile	Teacher laugh/smile/joke with pupils dummy: 1 = Yes, 0 = No	4
m4c_teach_hit	Teacher hit/pinch/slap with pupils dummy: 1 = Yes, 0 = No	4
m4c_teach_recall	Teacher ask questions that required learners to recall information dummy: 1 = Yes, 0 = No	4
m4c_teach_task	Teacher ask pupils to carry out a task dummy: 1 = Yes, 0 = No	4
m4c_teach_apply	Teacher ask pupils questions that required learners to apply information dummy: 1 = Yes, 0 = No	4
m4c_teach_creativity	Teacher ask pupils questions that required learners to use creativity dummy: 1 = Yes, 0 = No	4
m4c_teach_feedback	Teacher give feedback/praise/encouragement/support to students dummy: 1 = Yes, 0 = No	4
m4c_teach_feed_cor	Teacher give feedback correct students' mistake dummy: 1 = Yes, 0 = No	4
m4c_teach_feed_scold	Teacher give feedback that scolded students' mistake dummy: 1 = Yes, 0 = No	4
m4c_teach_intro	Teacher introduce lesson dummy: 1 = Yes, 0 = No	4
m4c_teach_summary	Teacher summarise lesson dummy: 1 = Yes, 0 = No	4
m4c_teach_hw	Teacher assign homework dummy: 1 = Yes, 0 = No	4
m4c_teach_hw_corr	Teacher collect homework dummy: 1 = Yes, 0 = No	4
m4c_teach_local_lang	Teacher use local language for instruction dummy: 1 = Yes, 0 = No	4

m4d_offic_records	Teacher keep official attendance records dummy: 1 = Yes, 0 = No	4
m4d_n_reg_pup	How many students are registered in the class currently	4
m4d_n_abs_pup	How many students are absent from the class currently	4
m4d_workscheme_mt	If the teacher has schemes of work for the month or term dummy: 1 = Yes, 0 = No	4
m4d_lessonplan	If the teacher has lesson plan dummy: 1 = Yes, 0 = No	4
m4d_pup_grade_rec	If the teacher has record of pupils' grades dummy: 1 = Yes, 0 = No	4
average_math_teacher_score	School Average Creation: Percentage of correct answers from mathematics teacher assessments from teachers (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	6
average_MT3_teacher	School Average Creation: Percentage of correct answers from Task 3 'Preparing to Teach' teacher assessments from teachers (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	6
average_MT4_teacher	School Average Creation: Percentage of correct answers from Task 4 'Assessing differences in children's abilities as learners' teacher assessments from teachers (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	6
average_MT5_teacher	School Average Creation: Percentage of correct answers from Task 5 'Evaluating the learning achievements and progress of students' teacher assessments from teachers (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	6
average_mathteacher_pedagogy	School Average Creation: Percentage of correct answers from Task 3, Task 4, and Task 5 teacher assessments from teachers who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year	6
av_teacher_edu_primary_math	School Average Creation: Percentage of teachers who finished formal education at primary school level (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	2
av_teacher_edu_secondary_math	School Average Creation: Percentage of teachers who finished formal education at secondary school level (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	2
av_teacher_edu_diploma_math	School Average Creation: Percentage of teachers who finished formal education at diploma/college level (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	2
av_teacher_edu_bachelors_math	School Average Creation: Percentage of teachers who finished formal education at bachelors/university school level (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	2
av_mathteacheredu_gtss	School Average Creation: Percentage of teachers who finished formal education that surpasses secondary school level (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	2
average_male_math_teacher	School Average Creation: Percentage of teachers who are male (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	2

average_math_ teacher_age	School Average Creation: Average teacher age (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	2
av_mathteacher_ absenteeism	School Average Creation: Average teacher absenteeism (who taught mathematics in Grade 4 in this current year, or Grade 3 the previous year)	2
average_lang_ teacher_score	School Average Creation: Percentage of correct answers from language teacher assessments from teachers (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	6
average_LT3 _teacher	School Average Creation: Percentage of correct answers from Task 3 'Preparing to Teach' teacher assessments from teachers (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	6
average_LT4 _teacher	School Average Creation: Percentage of correct answers from Task 4 'Assessing differences in children's abilities as l earners' teacher assessments from teachers (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	6
average_LT5 _teacher	School Average Creation: Percentage of correct answers from Task 5 'Evaluating the learning achievements and progress of students' teacher assessments from teachers (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	6
average_langteacher _pedagogy	School Average Creation: Percentage of correct answers from Task 3, Task 4, and Task 5 teacher assessments (from teachers who taught language in Grade 4 in this current year, or Grade 3 the previous year)	6
av_teacher_edu _primary_lang	School Average Creation: Percentage of teachers who finished formal education at primary school level (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	2
av_teacher_edu _secondary_lang	School Average Creation: Percentage of teachers who finished formal education at secondary school level (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	2
av_teacher_edu _diploma_lang	School Average Creation: Percentage of teachers who finished formal education at diploma/college level (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	2
av_teacher_edu _bachelors_lang	School Average Creation: Percentage of teachers who finished formal education at bachelors/university school level (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	2
av_langteacheredu _gtss	School Average Creation: Percentage of teachers who finished formal education that surpasses secondary school level (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	2
average_male_ lang_teacher	School Average Creation: Percentage of teachers who are male (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	2
average_lang _teacher_age	School Average Creation: Average teacher age (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	2

av_langteacher _absenteeism	School Average Creation: Average teacher absenteeism (who taught language in Grade 4 in this current year, or Grade 3 the previous year)	2
share_teaching	The share of instruction time spent teaching	4
m1_streams_g3_pps	Students per Stream in Grade 3	1
m1_streams_g4_pps	Students per Stream in Grade 4	1
oprhaned_classrooms	Percentage of classrooms which contained pupils but no teacher	2
teacher_student_ratio	Actually student-teacher ratio, how many students there are per teacher	1,2
m4b_txtbook_class	How many pupils had a language or mathematics book	4
perc_students_ withtextbook	Percentage of students who had a textbook	4
perc_students_ withpen	Percentage of students who had a pencil/pen	4
perc_students_ withexbook	Percentage of students who had a exercise book	4
perc_students_ usedtextbook	Percentage of students who had a textbook	4
perc_students_ usedpaperpen	Percentage of students who used paper/pen	4
student_absence	Percentage of students absent from class	4
teacher_behaviour	Categorical variable, a combination of: (m4c_teach_localinfo + m4c_teach_sitting + m4c_teach_indch + m4c_teach_recall + m4c_teach_task + m4c_teach_apply + m4c_teach_creativity + m4c_teach_feedback + m4c_teach_intro + m4c_teach_summary + m4c_teach_hw + m4c_teach_hw_corr)	4
pencilpen_perc	Percentage of students who had a pencil/pen	4
exbook_perc	Percentage of students who had a exercise book	4
pencilpen80	If 80% or more of students had a pencil/pen dummy: 1 = Yes, 0 = No	4
exbook80	If 80% or more of students had a exercise book dummy: 1 = Yes, 0 = No	4
g3g4_pps	Pupils per Stream average across Grade 3 and Grade 4	1

Figure 2: Shows the full variable list used by the 'All Countries', Kenya, Mozambique, Nigeria, Tanzania and Uganda samples. Note that, 'av_student_math_p' and 'av_student_lang_p' are the proxies for learning outcomes. The 'All Countries', Mozambique, and Tanzania samples are all missing the 'm1_pta' and 'm1_pta_meet_lastyr' variables, hence they have 106 potential variables, compared to the other samples with 108.

A.3 OLS Variable Descriptions

OLS Variable Descriptions	
A_i	The dependant variable, represents school-averaged mathematics or language test results and proxies learning outcomes. This variable represents an average test result for the 10 randomly selected grade 4 students in each school. The average scores were compiled with equal weights for each pupil test score. This data comes from Module 5 in the SDI data.
TSK_i	An independent continuous variable, represents school-averaged teacher subject (either mathematics or language) test results as a percentage and proxies teacher subject knowledge. The average scores were compiled with equal weights for each teacher test score. This data comes from Module 6 in the SDI data. Here, I employed a two-stage sample restriction on the teachers included where firstly, teachers were only included if they taught in grade 4 in the year of the data collection, or if they taught in grade 3 in the previous year. Secondly, the teachers who taught the dependent variable subject were included (So for example, only teachers who taught either grade 4 or grade 3 mathematics were included in the regression analysing mathematics learning outcomes).
TPS_i	An independent continuous variable, represents school-averaged teacher pedagogy test results as a percentage and proxies teacher pedagogical skill. The school-averaged scores were compiled with equal weights for each sampled teacher test score. This data comes from Module 6 in the SDI data and averages teacher test results from Task 3 (Preparing to teach) Task 4 (Assessing differences in children's abilities as learners) and Task 5 (Evaluating the learning achievements and progress of students), where percentage averages from each task were given an equal weight in the creation of this variable. I employed the same two-stage sample restriction on the teachers where firstly, teachers were only included if they taught in grade 4 in the year of the data collection, or if they taught in grade 3 in the previous year, and within that sample, the teachers who taught the dependent variable subject of interest were then included.
TE_i	An independent continuous variable, represents the percentage of the sampled teachers tested in Module 6 of the SDI data that have attained an official education that surpasses secondary school. This variable represents a limited proxy for teacher education. Each sampled teacher had an equal weight when calculating this variable and converting the individual categorical information into a school-average continuous variable. Again, I employed the same two-stage sample restriction on the teachers where firstly, teachers were only included if they taught in grade 4 in the year of the data collection, or if they taught in grade 3 in the previous year, and within that sample, the teachers who taught the dependent variable subject of interest were then included.
TA_i	An independent continuous variable, represents school-averaged percentages of teachers who were absent from the classroom, or the school, during the second unannounced SDI visit. This variable is intended to act as a proxy for teacher absenteeism. Each sampled teacher had an equal weight when calculating this variable and converting the individual categorical information into a school-average continuous variable. Again, I employed the same two-stage sample restriction on the teachers where firstly, teachers were only included if they taught in grade 4 in the year of the data collection, or if they taught in grade 3 in the previous year, and within that sample, the teachers who taught the dependent variable subject of interest were then included.
TB_i	An independent discrete variable, represents a total of observed teacher behaviour and practices from Module 4 (classroom observation) in the SDI data. Note that in each school, there was a maximum of 1 classroom observation. This variable totals binomial entries (1 = observed, 0 = not observed) for the following practices: If the teacher used local information to make learning relevant, if the teacher either sat or stood in front of the class at any time, if the teacher visited individual children, if the teacher asked questions that required students to recall information, if the teacher asked students to carry out a task, if the teacher asked questions that required students to apply information, if the teacher asked questions which required students to use their creativity, if the teacher gave feedback or praise/moral strengthening/encouragement to students, if the teacher introduced the lesson at the start of the class, if the teacher summarised the lesson at the end of the class, if the teacher assigned homework to the class, and if the teacher reviewed or collected homework from the class. This variable is intended to act as a proxy for positive teaching behaviour. Note that the classroom observation was either for a mathematics or language 4th grade lesson.
STR_i	An independent continuous variable, represents an ambiguous student-teacher ratio score for each school. This variable divides information from Module 1 (total number of students) by information from Module 2 (total number of teachers) and is ambiguous because the total number of teachers for each school was rounded to the nearest multiple of 10. Therefore, entries where the total number of teacher were rounded to 0, I changed these entries to 1. Hence, I reluctantly include this variable as a proxy for student-teacher ratios, with expectations of measurement issues stemming from the SDI data collection and input methods.
TST_i	An independent continuous variable, represents the estimated share of instruction time that is spent teaching for each school. This variable is the estimated share of time spent teaching during instruction time in each school. This indicator was compiled from observing 1 fourth-grade language or mathematics class for 1 hour in each school. Hence this variable is also subject to ambiguity.
F_i	An index of facilities score and is a created independent discrete variable representing a binomial total of facilities in each school. This index was calculated by binomially totalling (1 = observed, 0 = not observed) if each school had designated toilets for pupils, gender specific toilets, if the toilets were clean, if the toilets were private, if the 1 observed language or mathematics class per school had a corner library, if the same observed class had a board for teachers to write on, if the same observed class had chalk or a pen to use the board, and if the same observed class had working electricity. Therefore the maximum total here was 8.
PTA_i	A dummy variable for if the school has a parent teacher association.
SC_i	A dummy variable for if the school has a school committee.
SPS_i	An independent continuous variable, represents the average number of pupils per stream across grade 3 and grade 4 in each school. This variable builds on the literature of classroom sorting by also incorporating the size of the ability-based groups.
PS_i	A private school dummy variable for if the school is private.
U_i	A dummy variable for if the school is located in an urban area.
SB_i	An independent continuous variable, represents the percentage of students who ate breakfast before starting school, for each school. This variable was compiled by creating a percentage average of the 10 randomly selected grade 4 students (who were initially sampled for testing) and were asked if they had eaten breakfast that day. In this creation of this school average, each student was given an equal weight.
SA_i	An independent continuous variable, represents the average student absence rate in each school. This variable was calculated by dividing the total number of absent students by the total number of registered students in the one observed 4th grade language or mathematics class in each school.

Table 12: OLS variable descriptions. Note that Mozambique and Tanzania are the two countries in the sample where the PTA variable is not included in the baseline regression, due to omission in the SDI data. Also in the baseline regression model for Mozambique, the P variable is also excluded due to no private schools being sampled. Furthermore, Due to some data entry irregularities, 2 variable entries were dropped for SA in Kenya, 1 in Mozambique, 6 in Nigeria, and 5 in Tanzania.

A.4 Selected Covariates: Standardised Post-Shrinkage Coefficients

All Countries Mathematics Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
average_math_teacher_score	.0381214	.0380981	.0280888
private	.0369675	.039729	.0253732
average_mathteacher_pedagogy	.0134718	.0119342	.0075756
teacher_student_ratio	-.0127152	-.0129189	-.0105847
m1_girls_n_toilets	.0116338	.0145934	.0097356
m4d_pup_grade_rec	-.0093585	-.0080716	-.0082904
m1_sc_meet_lastyr	.0089891	.0128042	.0009434
student_absence	-.0080609	-.0085059	-.0084376
m1_toilets_clean	.0078735	.0087575	.0072247
m4b_material	.0073842	.0086342	.007437
pencilpen80	.0059103	.0082883	.006229
m4c_teach_hw	.0052947	.0071029	.0050436
average_male_math_teacher	-.0050867	-.0082176	-.0055769
m2_n_cl_wchild	.0049942	.001414	.005984
m1_toilets_accessible	.0049438	.0073847	.0060632
m2_n_classrooms	.0049019	.0067869	.0069672
m4c_txtbook_teacher	.0049002	.0059743	.0059077
g3g4_pps	-.0046381	-.0085187	-.0033483
m4c_teach_feedback	.0042593	.0073178	.0052123
m4c_teach_local_lang	-.0042295	-.0055735	-.0043812
m4d_n_reg_pup	-.0041434	-.0083024	-.0043124
perc_students_usedtextbook	.0039102	.0029577	.0036046
average_MT3_teacher	.0037755	.0045836	.00755
av_teacher_edu_diploma_math	.0037414	.008828	.0059319
m1_boys_toilets	.0036053	.0043315	.0067536
m4d_workscheme_mt	-.0027425	-.0028621	-.0040626
m1_gender_toilets	.002505	.003319	.0034799
urban	.0023211	.0028316	.0036736
m1_school_type	.0021361	.004456	.0046474
perc_students_usedpaperpen	.0020574	.0044385	.003444
m4c_teach_sitting	-.0020546	-.0081237	-.0042063
m4c_teach_feed_scold	-.0016854	-.0055127	-.003723
m4_light_front	.0015653	.0048587	.003025
m1_days_in_session	.0014257	.0045089	.0037553
m4c_teach_intro	.0013946	.0040008	.0027223
m4_tid	.0013942	.0044169	.0032801
m4_lux_measure	-.0013829	-.0024256	-.003141
m4c_teach_summary	.0010892	.0004026	.0031074
m4b_classroom_hyg	-.0004873		-.0024047
m4c_teach_hit	-.0003736		-.0016686
m1_n_boys_g4	-.0003076		-.0029655
m1_toilets	.0002998		.0016902
m4c_teach_smile	-.0001965		-.0034556
average_math_teacher_age	-.0001625		-.0042904
m4_light_back	.0001599		.0028111
m4c_teach_recall	.0000987		.0037607
m1_sc_meet_lastyr			.0086327
m1_streams_g4			-.0043102
m1_school_year			.0037549
m2_n_teachers			.0036586
av_mathteacher_absenteeism			-.0033975
m1_streams_g3_pps			-.0032232
average_MT5_teacher			.0030679
m4b_board_contrast			.0028934
m1_streams_g4_pps			-.0028602
average_MT4_teacher			.0026925
m1_streams_g3			-.002562
m4c_teach_indch			-.0019658
m1_sc_minutes			.0019175
m4c_teach_apply			-.0018942
share_teaching			-.001735
m2_n_cl_wchild_noteach			.0016623
m4c_teach_hw_corr			.0016438
m4b_electricity			.0014968
m4c_teach_localinfo			-.0014863
exbook80			.0012781
m4d_offic_records			-.0011899
m4c_teach_feed_cor			.001081
av_teacher_edu_secondary_math			-.0007847
m4b_cornerlibrary			.0006718
teacher_behaviour			.0006226
m1_toilets_private			.0005774
m4c_teach_indch_n			-.0004896
m1_school_cat			.0004556
m1_n_total_g4			-.0004385
perc_students_withtextbook			.0002947
m4b_work_displayed			.0002649
av_teacher_edu_bachelors_math			-.0002217
m4c_board_teacher			-.0001536
m4b_board			-.0000969
m4c_teach_task			.0000521

Table 13: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the 'All Countries' sample when predicting mathematics learning outcomes (106 potential covariates). Rounded to 7 decimal places.

All Countries Language Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
private	.0640232	.067441	.0625998
average_lang_teacher_score	.0480749	.0506121	.0468938
average_langteacher_pedagogy	.0268615	.0260523	.0257647
m1_n_boys_g4	-.0261279	-.0293591	-.0245819
m4b_material	.0247859	.0262271	.0245336
student_absence	-.0240224	-.0245724	-.0239529
m2_n_teachers	.0218717	.0319937	.020747
m4d_n_reg_pup	-.0216731	-.0258219	-.0208423
teacher_student_ratio	-.0202674	-.01673	-.0207382
teacher_behaviour	.0186675	.0253504	.0180045
average_student_breakfast	.0175088	.0215265	.0173248
urban	.017267	.0182416	.0173034
m1_boys_toilets	.0171977	.0226752	.0165956
m1_sc_meet_lastyr	.0168654	.0198304	.0166634
m2_n_cl_wchild	.0165118	.0151767	.0167811
m1_streams_g4	-.0151943	-.0225411	-.0157387
m4c_teach_local_lang	-.0148626	-.0170563	-.0148468
m1_streams_g3_pps	-.0130326	-.0170819	-.0109089
m4b_txtbook_class	.0123772	.0165135	.0119769
average_male_lang_teacher	-.0109876	-.0135311	-.0108303
m1_toilets_clean	.0108181	.0117796	.0107859
average_LT3_teacher	.0106155	.0102004	.0113978
m4c_txtbook_teacher	.0105384	.0108297	.0105692
m4c_teach_sitting	-.0099562	-.0169755	-.009299
m1_girls_n_toilets	.0097183	.0073611	.0101033
m4_tid	.009597	.0111685	.0096983
m4d_pup_grade_rec	-.0089123	-.0062666	-.0092449
m2_n_classrooms	.0088225	.0089481	.0090183
m4d_workscheme_mt	-.0082955	-.0082935	-.008381
m4b_board	.0072477	.0095632	.0071697
m1_school_type	.0071326	.0070405	.0073939
m1_toilets	.0063929	.0058181	.0064217
perc_students_usedtextbook	.0062114	.006908	.0062448
m1_toilets_accessible	.0059264	.0049286	.0061348
m4_light_front	.0059134	.0071146	.0059086
m4b_cornerlibrary	.0057738	.0058724	.0057274
m4b_classroom_hyg	-.0057551	-.0050482	-.0059143
pencilpen80	.0055696	.0055665	.0057229
m4b_electricity	.0053785	.0046438	.0056698
m4c_teach_creativity	.0046062	.0002644	.0050025
m4b_board_contrast	.0041202	.0040619	.0041301
perc_students_withtextbook	.0033065		.0034206
m4c_teach_localinfo	-.0032858	-.0030568	-.0032142
m4b_chalk	.0026592		.002798
m4c_teach_summary	.0020726		.0023836
m4c_teach_feed_scold	-.0020388		-.0021554
av_langteacher_absenteeism	-.0017056		-.00175
av_teacher_edu_secondary_lang	-.0012746		-.0014653
av_teacher_edu_primary_lang	.0005197		.0005428
m1_toilets_private	.0003501		.0006824
g3g4_pps			-.0033094
average_LT4_teacher			.0008012
m4c_teach_hw_corr			.0001682

Table 14: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the 'All Countries' sample when predicting language learning outcomes (106 potential covariates). Rounded to 7 decimal places.

Kenya Mathematics Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
private	.0287714	.0334464	.0171185
m1_streams_g3_pps	-.0176629	-.0236639	-.0077051
average_student_breakfast	.0141288	.0164755	.0103656
av_mathteacheredu_gtss	.0092913	.0123183	.0061724
teacher_student_ratio	-.0091916	-.0114057	-.0096893
m4_tid	.0085197	.0145168	.0068968
m1_pta	-.0079434	-.0112649	-.0073183
m4c_teach_hw_corr	.007817	.0143029	.0054115
m4c_teach_indch	-.0076075	-.0109652	-.0061245
m1_days_in_session	.0073123	.0116855	.0060491
m1_sc_meet_lastyr	.0064161	.0150263	-.0017654
av_teacher_edu_bachelors_math	-.0056787	-.0121947	-.004737
m4_light_front	.0056207	.012319	.004828
m4c_board_pup	.0047373	.0086327	.0035622
m2_n_cl_wchild_noteach	-.004719	-.0074333	-.0036115
m4b_material	.0047183	.0052091	.0054116
m1_school_year	.0044285	.0058927	.0085221
m1_n_boys_g4	-.0036445	-.0058778	-.0027478
m4b_board_contrast	.0034859	.0100848	.0029564
average_math_teacher_age	.003485	.0104452	.0022691
urban	.0034592	.0066322	.0024889
m4b_electricity	.0033832	.0043561	.0040594
perc_students_withtextbook	.0032021		.0043267
m4d_offic_records	.0031571	.0096034	.0010566
m4d_n_abs_pup	-.0031019	-.0034392	-.0036354
share_teaching	.0030502	.0040577	.0030668
m4c_teach_summary	.0029598	.0022061	.0040955
perc_students_usedtextbook	.0018851	.0048249	.0016896
m4d_pup_grade_rec	.0014843	.0009221	.0020964
m4c_board_teacher	-.0013456	-.0042178	-.0013006
m1_school_cat	-.001073	-.0007773	-.0006239
student_absence	-.0008618		-.0014325
m4c_txtbook_teacher	-.0003636		-.0007052
av_mathteacher_absenteeism	-.0000872		-.0030317
av_teacher_edu_diploma_math	.0000843		.0017664
m4c_teach_creativity	.0000195		.001177
urb_rur_semi	-.0000125		-.0024888
g3g4_pps			-.005419
m1_sc_meet_lastyr			.0041669
m1_streams_g4_pps			-.0021904
m1_n_total_g4			-.0017645
average_math_teacher_score			.0015313
m4c_txtbook_pup			-.0013776
m1_n_total_g3			-.00111
average_male_math_teacher			-.0009463
pencilpen80			-.0008947
m1_toilets_clean			.0008643
average_MT3_teacher			.0007139
av_teacher_edu_primary_math			.0003213
m4c_teach_feedback			-.0001804
oprhaned_classrooms			-.0001707
m4c_teach_recall			-.0001373
exbook80			-.0001292

Table 15: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Kenya sample when predicting mathematics learning outcomes (108 potential covariates). Rounded to 7 decimal places.

Kenya Language Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
m1_streams_g3_pps	-.0349797	-.02667	-.0329575
private	.0268946	.0292771	.0256272
average_student_breakfast	.0220944	.0170683	.0215739
m4_tid	.015099	.0187283	.0145934
m1_n_total_g3	-.0146787	-.0401386	-.0156737
m2_n_classrooms	.0135627	.0441026	.0136216
teacher_student_ratio	-.0128955	-.0029445	-.012589
student_absence	-.0124371	-.0147612	-.0114912
m4b_board	.0114721	.0142332	.0115016
m4b_cornerlibrary	.0112362	.0214715	.0108078
m1_n_boys_g4	-.0109477	-.0446851	-.0115702
av_teacher_edu_secondary_lang	-.0099945	-.016983	-.0097563
m1_days_in_session	.0083448	.0107453	.0081883
m4c_teach_hw	.0076963	.0121688	.0077765
m4b_material	.0076807	.0112576	.0079802
urban	.0070614	.0014728	.0039175
m4d_n_abs_pup	-.0065993	-.0048988	-.0078057
pencilpen_perc	.0062719		
average_LT5_teacher	.0061739	.0101074	.0064552
av_langteacheredu_gtss	.005888	.0079782	.0056876
m1_pta_meet_lastyr	-.0047464	-.0112245	-.0052946
m4b_txtbook_class	.0045387	.0113637	.0047966
m4d_pup_grade_rec	.0036283	.0189635	.0042513
m4_light_front	.0034284	.0110532	.0037951
m1_sc	-.0033955	-.0046164	-.0038431
m4b_work_displayed	.0033266	.0112651	.0038148
perc_students_usedpaperpen	-.0017481	-.0156762	-.0023481
m4c_teach_summary	.0017014	.0050135	.0021157
m2_n_teachers	.0011085	.030038	.0022021
m4c_teach_sitting	-.000968	-.0123485	-.0013234
m4d_workscheme_mt	-.0001903	-.0154343	-.0011887
urb_rur_semi	-.0000662		-.0039196
perc_students_withpen		.0184365	.0068067
m4c_teach_feed_cor			.0000994

Table 16: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Kenya sample when predicting language learning outcomes (108 potential covariates). Rounded to 7 decimal places.

Mozambique Mathematics Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
teacher_behaviour	.0107709	.0121503	.0098451
perc_students_withtextbook	.0091012	.0145021	.00849
student_absence	-.0085116	-.0137211	-.0080166
perc_students_usedtextbook	.0043093	.0087225	.0043062
m1_days_in_session	.0035953	.0126837	.0031921
pencilpen80	.0035493	.0108729	.0031968
m4d_lessonplan	.0033103	.0077637	.0032734
m4c_teach_hw	.0029938	.0086765	.0032567
m4c_teach_sitting	.0029868	.0084131	.0029657
pencilpen_perc	.0022802		.0024196
average_math_teacher_score	.0009527	.0032604	.0010944

Table 17: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Mozambique sample when predicting mathematics learning outcomes (106 potential covariates). Rounded to 7 decimal places.

Mozambique Language Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
student_absence	-.0529525	-.0816415	-.053211
teacher_behaviour	.036136	.0526999	.0365984
m1_sc_meet_lastyr	.0186532	.0342016	.0193933
perc_students_withtextbook	.0183699	.0324013	.0188603
m4c_teach_hit	.0089493	.0010762	.0093028
m4b_txtbook_class	.0075354		.0076445
perc_students_usedtextbook	.0061516		.0066241
m4b_material	.001371		.0019194

Table 18: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Mozambique sample when predicting language learning outcomes (106 potential covariates). Rounded to 7 decimal places.

Nigeria Mathematics Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
private	.032262	.0320647	.0236345
average_male_math_teacher	-.022444	-.0239897	-.0164161
average_math_teacher_score	.0186822	.0211053	.0142373
average_math_teacher_age	-.0177715	-.0236449	-.0152453
perc_students_withtextbook	.016833	.0192886	.0121204
perc_students_withpen	.0151326	.0184251	
m1_days_in_session	.011873	.0149321	.0100871
av_mathteacher_absenteeism	-.010173	-.0144437	-.0093512
m1_school_year	.0097624	.0128444	.0105826
m4b_exbook	-.0085582	-.0188927	-.0052074
m2_n_classrooms	.008245	.0127434	.0061878
m4c_teach_local_lang	-.0081519	-.0101347	-.0086139
m1_girls_n_toilets	.0081419	.0116388	.0054069
perc_students_usedpaperpen	.0076158	.0098893	.0079431
av_teacher_edu_diploma_math	.007144	.0092479	.006413
m4d_n_reg_pup	-.0071345	-.0050569	-.007152
m4b_material	.0071296	.0097299	.0065339
m4b_classroom_hyg	-.0063531	-.0071342	-.0076105
m4c_teach_task	.0063222	.0105651	.0067723
m4d_lessonplan	.0062915	.0067838	.0061646
m4c_teach_apply	-.0061097	-.0104215	-.0067069
m1_sc	.0054832	.0110208	.0042135
average_MT3_teacher	.0047713	.0045573	.0072074
m4c_teach_localinfo	-.0045698	-.0033281	-.0064495
m4_light_front	.0042002	.0077825	.0038321
m4b_room_boys	-.0037325		-.0043893
av_teacher_edu_secondary_math	-.0036026	-.0029277	-.0032302
av_mathteacheredu_gtss	.0032125	.0038703	.0038707
share_teaching	-.0028254	-.0055275	-.0027629
m4b_electricity	.0025787	.0006703	.0036627
m4b_board	-.0025025	-.0021795	-.0041168
m4c_board_teacher	-.0023322	-.0039661	-.0026917
m4c_teach_summary	.0022017		.0044863
m4_lux_measure	-.0021248		-.0039815
m2_n_cl_wchild	.0017115		.0057345
average_MT5_teacher	.0015547		.0034587
m4c_teach_sitting	-.0015264		-.0024307
m4c_teach_feed_cor	-.0012933		-.0031477
perc_students_usedtextbook	.0008187		.004569
m1_sc_meet_lastyr	.0007747		.0029008
av_teacher_edu_primary_math	-.0007011		-.0022589
m4c_teach_smile	-.0006512		-.0031587
m2_n_teachers	.0005914		.0020352
m1_school_cat	.0005549		.0013516
m4d_pup_grade_rec	-.0002315		-.0031297
m1_school_type	-.0001196		-.0015251
m4c_teach_recall	.0000152		.0027089
pencilpen_perc			.0121404
m4b_room_total			-.002829
m1_boys_toilets			.002776
m1_n_total_g3			-.0022696
m4c_board_pup			.0021175
m1_sc_minutes			-.0021117
m4_light_back			.0019432
oprhaned_classrooms			-.0017906
m1_toilets_accessible			.0017387
exbook_perc			.0016138
m4c_teach_feed_scold			.0015445
m1_toilets_clean			.0014993
m4c_teach_creativity			-.0011509
m4b_pencilpen			-.001118
m1_pta_meet_lastyr			.0010739
m4c_teach_intro			.0010627
average_MT4_teacher			-.0009275
m4_tid			.0009271
student_absence			-.0009264
m4b_work_displayed			-.0009178
average_mathteacher_pedagogy			.0005523
m1_toilets_private			.0004781
m4c_teach_indch			-.0002734
m1_streams_g3_pps			-.0002663
teacher_student_ratio			-.0002451
g3g4_pps			-.0002192
urb_rur_semi			-.0001797

Table 19: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Nigeria sample when predicting mathematics learning outcomes (108 potential covariates). Rounded to 7 decimal places.

Nigeria Language Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
private	.0791889	.0807118	.0683292
perc_students_withtextbook	.0384576	.0409056	.0334462
m4c_teach_local_lang	-.0357742	-.0428589	-.0361086
g3g4_pps	-.0317633	-.0393919	-.0143628
m4b_material	.0247524	.0307731	.0226655
m2_n_cl_wchild	.023303	.0254459	.0232876
m4b_classroom_hyg	-.023194	-.0266701	-.0216679
average_LT3_teacher	.0227227	.026665	.0220442
m1_toilets_clean	.0158711	.0168724	.0012751
average_langteacher_pedagogy	.0155081	.0161019	.0156767
urban	.0119721	.0134702	.0128162
av_langteacher_absenteeism	-.0118805	-.0198301	-.0122374
m2_n_teachers	.0117916	.0185129	.0114728
m4_light_front	.0101001	.0205055	.0105535
m1_school_year	.0096773	.0192022	.0155609
average_male_lang_teacher	-.009422	-.0087256	-.0104162
m4c_teach_intro	.0093971	.0150409	.0080994
oprhaned_classrooms	-.0088379	-.0092134	-.009669
perc_students_usedtextbook	.0087585	.0104456	.0106895
m4c_teach_task	.0077804	.0114408	.0086604
m4c_teach_creativity	.0073777	.0078922	.0077391
average_lang_teacher_score	.0072491	.0077697	.008529
m4d_n_abs_pup	-.0069817	-.003203	-.009172
average_student_breakfast	.0066945	.009306	.0079428
pencilpen_perc	.0065469		.0078107
av_langteacheredu_gtss	.0052206	.0110171	.0057932
m4c_teach_summary	.0038363	.0014632	.0049739
m4_tid	.0036198		.004925
perc_students_usedpaperpen	.002555		.0018692
m4d_lessonplan	.0017238		.0027151
m4c_txtbook_teacher	.001087		.0017469
m1_pta_meet_lastyr	.0005562		.0018365
perc_students_withpen		.0071941	
m1_toilets_clean			.0156608
m1_streams_g4_pps			-.0105264
m1_streams_g3_pps			-.0084582
m4b_electricity			.0023402
m4c_bookpen_n_pup			.0020373
m4c_board_pup			.0019116
m1_sc_minutes			-.0016564
m4b_chalk			.0010227
m1_toilets_private			.0010187
av_teacher_edu_primary_lang			.0006925
m4_lux_measure			-.0006066
m1_days_in_session			.0005378
m4c_teach_feed_scold			.0001575

Table 20: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Nigeria sample when predicting language learning outcomes (108 potential covariates). Rounded to 7 decimal places.

Uganda Mathematics Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
m1_streams_g4_pps	-.0212531	-.0285808	.0019555
private	.0167012	.0244722	.0129105
m2_n_cl_wchild	.0134255	.0090598	.009706
m2_n_teachers	.0122508	.0181337	.0068231
av_mathteacher_absenteeism	-.0098141	-.0115623	-.0082462
average_MT5_teacher	.0090387	.013488	.0070732
m4b_material	.0081017	.0074317	.0070437
m1_girls_n_toilets	.0075114	.0119505	.0052071
average_math_teacher_age	-.007456	-.0099261	-.0071418
m4b_work_displayed	.0069955	.0084015	.0064775
g3g4_pps	-.0060582	-.0022665	-.0087438
average_math_teacher_score	.0055513	.010465	.0049273
m1_pta	.0049317	.0119525	.0037899
teacher_student_ratio	-.0048304	-.0016934	-.0066925
m1_pta_meet_lastyr	.0044753	.0063852	.0046053
average_male_math_teacher	-.0040189	-.0089656	-.0039924
average_student_breakfast	.003813	.0044056	.004456
m1_school_type	.0022964	.0038287	.0032645
m4c_teach_feed_scold	-.0021753	-.0032692	-.0030725
m4b_board_contrast	.0019201	.0034509	.0021879
m1_sc_minutes	.0018939	.0107487	.0013393
m4c_teach_feed_cor	.0012676	.0064881	.0017936
m1_sc	-.0011924	-.0058247	-.0021849
average_mathteacher_pedagogy	.0007798		.0033545
share_teaching	.0004858		.0014477
m4c_teach_summary	-.0003006		-.0006744
m4b_electricity	.000262		.002924
m4c_teach_creativity	.0001032		.0011786
m1_streams_g4_pps			-.0112683
m1_streams_g3_pps			-.0043855
student_absence			-.0019049
m1_streams_g3			.0015385
m1_school_cat			.0009996
m4_tid			.0009336
av_teacher_edu_secondary_math			-.0009003
m2_n_classrooms			.0008139
m4c_teach_localinfo			.0006183
m4c_board_pup			-.0005044
m1_toilets_accessible			-.000425
m4c_teach_hw			-.0003663
m1_toilets_clean			.0003062
m4c_teach_intro			.000056

Table 21: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Uganda sample when predicting mathematics learning outcomes (108 potential covariates). Rounded to 7 decimal places.

Uganda Language Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
g3g4_pps	-.0677861	-.0864954	-.0276289
private	.0498439	.0703152	.0541381
m4b_material	.0420311	.0480734	.0406758
average_student_breakfast	.0396402	.0411095	.0382467
m2_n_teachers	.0318875	.0412359	.0335512
m1_school_year	-.0271451	-.054491	-.0389113
average_langteacher_pedagogy	.0251018	.0311546	.0115724
teacher_student_ratio	-.020768	-.015701	-.0185276
m4b_work_displayed	.0162532	.0176643	.0180339
m4b_electricity	.0158926	.0172076	.016481
student_absence	-.0113024	-.0090868	-.0139505
m4c_teach_indch	.0112265	.0192542	.0140513
urban	.0111218	.0123633	.0141569
av_teacher_edu_bachelors_lang	-.0108794	-.0206959	-.0137988
m4b_chalk	.0100605	.0198401	.0154582
m1_school_type	.008717	.0052314	.0131907
average_male_lang_teacher	-.0080082	-.0145708	-.0101536
m4c_teach_creativity	.0070311	.0120802	.0087719
m2_n_cl_wchild	.0058725		.0056146
m1_streams_g4_pps	-.005741		-.0242859
m4c_teach_hw_corr	.0048261	.0174127	.0070756
m4c_teach_hit	-.0046529	-.010787	-.0074817
m1_sc	-.0030935	-.0016094	-.0069437
m1_days_in_session	-.0027849	-.0085204	-.0113041
av_teacher_edu_diploma_lang	.0024305	.0064564	.0111182
average_LT3_teacher	.0019114		.0104482
m4c_teach_apply	.0017326		.003301
m1_girls_n_toilets	.0012355		.0073422
average_LT4_teacher	.000605		.01083
m4c_teach_hw	.0003511		.006981
m4c_txtbook_teacher	.000034		.0052126
m1_streams_g3_pps			-.0201131
m1_n_girls_g4			-.0090086
perc_students_usedpaperpen			-.0087331
av_langteacheredu_gtss			-.0067551
m1_sc_minutes			.0062021
m1_n_total_g4			-.0058202
m1_gender_toilets			-.0047991
m4_lux_measure			-.0043421
m1_pta_meet_lastyr			.0037049
m4c_teach_intro			-.0026982
m4c_bookpen_n_pup			-.0024324
m4b_txtbook_class			.0017633
share_teaching			.0015946
m4d_lessonplan			.0014955
m4_light_front			.0013876
m1_toilets_accessible			-.0009793
m4c_teach_summary			-.0006965
average_lang_teacher_age			-.0004268
m4c_teach_sitting			-.0001095

Table 22: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Uganda sample when predicting language learning outcomes (108 potential covariates). Rounded to 7 decimal places.

Tanzania Mathematics Post-Shrinkage Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
m2_n_teachers	.0184831	.0276956	.0085514
m4c_teach_indch_n	-.0164045	-.0233503	-.0016238
m1_toilets_clean	.0136482	.017764	.0073641
m4_lux_measure	-.0133737	-.0197216	-.008478
urban	.0107312	.0115089	.0064782
teacher_student_ratio	-.0093165	-.0108087	-.00893
m4c_teach_hw	.008607	.0135372	.0058298
m1_school_type	.0073571	.0126614	.0057721
exbook80	.0067755	.0144379	.0049565
m4b_txtbook_class	.0061328	.010853	.0047674
m4d_pup_grade_rec	-.0058872	-.008239	-.0047705
m4c_teach_intro	.0050138	.0062668	.0050516
m4b_board_contrast	.0047588	.0095732	.0053374
m4b_electricity	.004444	.0094758	.0051619
m4c_teach_summary	.0036784	.005743	.0044106
av_mathteacher_absenteeism	-.0035787	-.0070609	-.0044651
m4d_n_reg_pup	-.0033206	-.0150004	-.0025016
m4c_teach_task	.0027216	.00874	.0024382
student_absence	-.0019996	-.0006932	-.0030446
m2_n_classrooms	.0012309		.0044249
share_teaching	.0012111	.0012825	.0023357
m4c_teach_smile	.0009791		.002387
average_MT5_teacher	.0009557		.002228
m4c_teach_local_lang	.0006402		.002715
m4c_teach_feed_scold	-.0004271		-.0034939
m4b_work_displayed	.0000865		.003157
m4c_teach_indch_n			-.0100695
m4b_cornerlibrary			.0033191
m1_toilets_accessible			.0030146
m4d_workscheme_mt			-.0029096
average_math_teacher_score			.0028656
m1_toilets_private			.0028362
urb_rur_semi			-.0024472
m4b_material			-.0024439
m4c_teach_creativity			-.0023207
teacher_behaviour			.0022665
m4_tid			.0022631
m4b_chalk			-.0022185
m2_n_cl_wchild			.0021741
m4c_teach_hw_corr			.0021367
m4d_n_abs_pup			-.0020885
m4_light_back			.0020108
pencilpen80			.0020077
m2_n_cl_wchild_noteach			.0017927
m4c_board_pup			-.0016284
average_student_breakfast			-.0016185
m4b_classroom_hyg			-.0015962
m1_streams_g3			.0015801
perc_students_withtextbook			.0015109
m4_light_front			.0013827
m1_streams_g3_pps			-.0013113
average_mathteacher_pedagogy			.0012472
perc_students_usedpaperpen			.0009088
m4c_teach_feedback			.0009028
av_teacher_edu_bachelors_math			.000735
m1_school_year			.0006932
m4c_txtbook_pup			.0006765
perc_students_usedtextbook			.0006363
m1_sc_meet_lastyr			.0006041
av_mathteacheredu_gtss			.0005497
av_teacher_edu_secondary_math			-.0004395
perc_students_withexbook			.0003632
m4c_teach_feed_cor			.0003088
perc_students_withpen			.000246
private			.0002308

Table 23: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Tanzania sample when predicting mathematics learning outcomes (106 potential covariates). Rounded to 7 decimal places.

Tanzania Language Post-Selection Coefficients			
	CV Lasso	Adaptive Lasso	Elastic Net
m2_n_teachers	.0500568	.0755422	.0202803
m1_toilets_clean	.0283484	.0334873	.0149575
urban	.0150448	.0175273	.0135767
exbook80	.0137155	.018068	.0099585
m4b_electricity	.0137084	.0188512	.011509
private	.0135954	.0156727	.010295
average_male_lang_teacher	-.0125294	-.0155824	-.0116544
m4b_board_contrast	.0121461	.016731	.0099128
m1_school_type	.0111771	.0153564	.0087896
m1_n_boys_g4	-.0101446	-.030817	-.0027937
m4d_n_abs_pup	-.0099769	-.0086844	-.0059143
m1_streams_g4	-.0096482	-.0152636	-.0068289
m4_tid	.0090312	.0059462	.0120798
m4c_teach_summary	.0081787	.0093264	.0073789
m4b_txtbook_class	.0076297	.0134647	.0064151
m4b_chalk	-.0075376	-.01286	-.0064964
m4d_workscheme_mt	-.0068358	-.0104884	-.0059035
m4b_classroom_hyg	-.0063148	-.0059133	-.007418
perc_students_withtextbook	.006223	.0076393	.004257
m4d_n_reg_pup	-.0059139	-.0116197	-.0031574
m4c_teach_feedback	.0056061	.0088883	.0053445
teacher_student_ratio	-.0055536		-.0133656
m4c_teach_task	.0054291	.0081686	.0050102
m1_days_in_session	.005162	.0042072	.0067538
m4c_teach_indch_n	-.0047406	-.0085147	-.005626
m2_n_cl_wchild	.0045996	.0057192	.0064711
m4d_pup_grade_rec	-.003809	-.002272	-.0058644
average_lang_teacher_score	.002412		.0072719
m4c_teach_hit	.0021639		.003312
m4c_teach_smile	.001769		.0023356
m4d_offic_records	.0013371		.0030872
m4_light_front	.0010547		.0023242
m1_sc_minutes	.0006706		.0018233
m2_n_classrooms	.0004167		.0076005
m1_toilets_accessible			.0050027
student_absence			-.0046161
m1_toilets_private			.0039054
oprhaned_classrooms			.0022843
perc_students_usedpaperpen			.0018333
urb_rur_semi			-.0016805
average_LT4_teacher			-.0014759
teacher_behaviour			.0014534
m1_boys_toilets			.0013948
m4c_teach_hw			-.0013402
av_teacher_edu_bachelors_lang			-.0013185
av_teacher_edu_diploma_lang			-.0011936
av_langteacher_absenteeism			-.0010682
m1_n_total_g4			-.0010173
m4c_teach_feed_cor			-.0006763
m4c_teach_apply			.0004626
perc_students_withpen			.0004102
perc_students_withexbook			.0003679
average_LT3_teacher			-.0001594
m4d_lessonplan			.0000869
m2_n_cl_wchild_noteach			.0000583
m4_light_back			.0000068

Table 24: Shows the selected and standardised post-shrinkage coefficients from CV Lasso, Adaptive Lasso and Elastic Net methods for the Tanzania sample when predicting language learning outcomes (106 potential covariates). Rounded to 7 decimal places.

A.5 Variable Selection Summaries

All Countries Variable Selection Methods Summary

Panel A: Mathematics	N	p	λ	*alpha*	Non-zero coefficients	Out of sample R^2	CV mean prediction error
CV Lasso	1,313	106	.0042345	N/A	46	0.4572	.0130678
Adaptive Lasso	1,313	106	.0020516	N/A	38	0.4879	.012328
Elastic Net	1,313	106	.2586326	0.006	81	0.4630	.0129273
Panel B: Language							
CV Lasso	1,208	106	.0040341	N/A	50	0.5883	.0320406
Adaptive Lasso	1,208	106	.0039107	N/A	42	0.6062	.0306479
Elastic Net	1,208	106	.0213314	0.186	71	0.5884	.0320284

Table 25: Shows the regularisation methods used to estimate the best predictors of learning outcomes. For full lists of the initial (p) and selected covariates p please refer to the appendix. Also $N = 2,069$ before the analysis, but missing observations for some covariates led to exclusion which reduced the samples (N). Note that for the Elastic Net models, the alpha selection was limited to 0.001 increments between 0 and 1.

Kenya Variable Selection Methods Summary

Panel A: Mathematics	N	p	λ	*alpha*	Non-zero coefficients	Out of sample R^2	CV mean prediction error
CV Lasso	189	108	.0038026	N/A	37	0.3476	.0058093
Adaptive Lasso	189	108	.0119422	N/A	30	0.4705	.004715
Elastic Net	189	108	.29442	0.01	89	0.3580	.0057168
Panel B: Language							
CV Lasso	194	108	.0083454	N/A	32	0.3282	.0173272
Adaptive Lasso	194	108	.0008605	N/A	31	0.4304	.0146913
Elastic Net	194	108	.0502605	0.158	33	0.3311	.0172534

Table 26: Shows the regularisation methods used to estimate the best predictors of learning outcomes in Kenya. For full lists of the initial (p) and selected covariates p please refer to the appendix. Also $N = 306$ before the analysis, but missing observations for some covariates led to exclusion which reduced the samples (N). Note that for the Elastic Net models, the alpha selection was limited to 0.001 increments between 0 and 1.

Mozambique Variable Selection Methods Summary

Panel A: Mathematics	N	p	λ	*alpha*	Non-zero coefficients	Out of sample R^2	CV mean prediction error
CV Lasso	103	106	.008222	N/A	11	0.1090	.0040209
Adaptive Lasso	103	106	.0006958	N/A	10	0.2753	.0032702
Elastic Net	103	106	.1086902	0.075	11	0.1142	.0039971
Panel B: Language							
CV Lasso	103	106	.0313867	N/A	8	0.2188	.0352774
Adaptive Lasso	103	106	.0457933	N/A	5	0.3237	.030539
Elastic Net	103	106	.0313867	0.972	8	0.2188	.0352739

Table 27: Shows the regularisation methods used to estimate the best predictors of learning outcomes in Mozambique. For full lists of the initial (p) and selected covariates p please refer to the appendix. Also $N = 203$ before the analysis, but missing observations for some covariates led to exclusion which reduced the samples (N). Note that for the Elastic Net models, the alpha selection was limited to 0.001 increments between 0 and 1.

Nigeria Variable Selection Methods Summary

Panel A: Mathematics	N	p	λ	α	Non-zero coefficients	Out of sample R^2	CV mean prediction error
CV Lasso	513	108	.0042132	N/A	47	0.3795	.0157335
Adaptive Lasso	513	108	.0216007	N/A	31	0.4242	.0146
Elastic Net	447	108	.3401735	0.006	73	0.389	.015493
Panel B: Language							
CV Lasso	433	108	.0106069	N/A	32	0.5437	.0363554
Adaptive Lasso	433	108	.0055054	N/A	27	0.5747	.0338807
Elastic Net	433	108	.102592	0.087	45	0.5455	.0362108

Table 28: Shows the regularisation methods used to estimate the best predictors of learning outcomes in Nigeria. For full lists of the initial (p) and selected covariates p please refer to the appendix. Also $N = 760$ before the analysis, but missing observations for some covariates led to exclusion which reduced the samples (N). Note that for the Elastic Net models, the alpha selection was limited to 0.001 increments between 0 and 1.

Uganda Variable Selection Methods Summary

Panel A: Mathematics	N	p	λ	α	Non-zero coefficients	Out of sample R^2	CV mean prediction error
CV Lasso	258	108	.0063541	N/A	28	0.4543	.0064803
Adaptive Lasso	258	108	.00251	N/A	23	0.5265	.0056225
Elastic Net	258	108	.3721617	0.014	42	0.4602	.0064092
Panel B: Language							
CV Lasso	242	108	.0119606	N/A	31	0.6176	.0264819
Adaptive Lasso	242	108	.0285689	N/A	23	0.6746	.0225338
Elastic Net	242	108	.0549599	0.115	50	0.6183	.0264331

Table 29: Shows the regularisation methods used to estimate the best predictors of learning outcomes in Uganda. For full lists of the initial (p) and selected covariates p please refer to the appendix. Also $N = 400$ before the analysis, but missing observations for some covariates led to exclusion which reduced the samples (N). Note that for the Elastic Net models, the alpha selection was limited to 0.001 increments between 0 and 1.

Tanzania Variable Selection Methods Summary

Panel A: Mathematics	N	p	λ	α	Non-zero coefficients	Out of sample R^2	CV mean prediction error
CV Lasso	230	106	.0080407	N/A	26	0.1825	.0114339
Adaptive Lasso	230	106	.0079694	N/A	20	0.3114	.0096303
Elastic Net	230	106	1.129236	0.002	65	0.2066	.0110967
Panel B: Language							
CV Lasso	220	106	.0081024	N/A	34	0.3007	.021913
Adaptive Lasso	220	106	.006991	N/A	26	0.3965	.0189106
Elastic Net	220	106	.6273374	0.008	56	0.3166	.0214167

Table 30: Shows the regularisation methods used to estimate the best predictors of learning outcomes in Tanzania. For full lists of the initial (p) and selected covariates p please refer to the appendix. Also $N = 400$ before the analysis, but missing observations for some covariates led to exclusion which reduced the samples (N). Note that for the Elastic Net models, the alpha selection was limited to 0.001 increments between 0 and 1.

A.6 OLS Descriptive Statistics

All Countries Descriptive Statistics - Mathematics

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Mathematics Score	.4462	.1637	0	1	2068
Teacher Subject Knowledge	.5374	.2392	0	1	1994
Teacher Pedagogical Skills	.2233	.1565	0	.8376	1994
Teacher Education	.4954	.4551	0	1	1995
Teacher Absenteeism	.2611	.3436	0	1	1995
Teacher Behaviour	8.755	2.81	1	14	1831
Student-Teacher Ratio	154.1	164.5	0	1492	2065
Share of Time Spent Teaching	.8749	.2302	0	1	2067
Facilities Index	7.573	2.153	0	13	2048
School Committee	.8811	.3237	0	1	2069
Students Per Stream	49.98	39.71	0	331	2045
Private	.2187	.4134	0	1	2067
Urban	.2146	.4106	0	1	2069
Student Breakfast	.7055	.3154	0	1	2068
Student Absenteeism	.246	.2396	0	1	2005

Table 31: Shows the descriptive statistics for mathematics in 'All Countries'. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

All Countries Descriptive Statistics - Language

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Language Score	.503	.2916	.0012	1	2068
Teacher Subject Knowledge	.4366	.1773	0	.8444	1879
Teacher Pedagogical Skills	.2314	.1607	0	.7051	1879
Teacher Education	.4623	.4615	0	1	1879
Teacher Absenteeism	.2457	.3655	0	1	1879
Teacher Behaviour	8.755	2.81	1	14	1831
Student-Teacher Ratio	154.1	164.5	0	1492	2065
Share of Time Spent Teaching	.8749	.2302	0	1	2067
Facilities Index	7.573	2.153	0	13	2048
School Committee	.8811	.3237	0	1	2069
Students Per Stream	49.98	39.71	0	331	2045
Private	.2187	.4134	0	1	2067
Urban	.2146	.4106	0	1	2069
Student Breakfast	.7055	.3154	0	1	2068
Student Absenteeism	.246	.2396	0	1	2005

Table 32: Shows the descriptive statistics for language in 'All Countries'. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Kenya Descriptive Statistics - Mathematics

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Math Score	.6057	.0971	.3471	.9294	306
Teacher Subject Knowledge	.7998	.1333	.2717	1	297
Teacher Pedagogical Skills	.3638	.1459	0	.7479	297
Teacher Education	.6782	.3489	0	1	297
Teacher Absenteeism	.3431	.3719	0	1	297
Teacher Behaviour	10.32	2.631	1	14	229
Student-Teacher Ratio	91.11	108.1	7.364	654	306
Share of Time Spent Teaching	.7959	.2959	0	1	306
Facilities Index	8.661	1.386	5	12	304
Parent-Teacher Association	.4020	.4911	0	1	306
School Committee	.9379	.2417	0	1	306
Students Per Stream	37.53	17.22	2	95	304
Private	.2190	.4142	0	1	306
Urban	.3235	.4686	0	1	306
Student Breakfast	.8717	.1698	.1	1	306
Student Absenteeism	.1203	.1488	0	1	278

Table 33: Shows the descriptive statistics for mathematics in Kenya. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Kenya Descriptive Statistics - Language

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Language Score	.8003	.1621	.0432	1	306
Teacher Subject Knowledge	.6495	.0952	.3333	.8444	296
Teacher Pedagogical Skills	.3602	.1461	0	.6906	296
Teacher Education	.7007	.3527	0	1	296
Teacher Absenteeism	.3275	.3698	0	1	296
Teacher Behaviour	10.32	2.631	1	14	229
Student-Teacher Ratio	91.11	108.1	7.364	654	306
Share of Time Spent Teaching	.7959	.2959	0	1	306
Facilities Index	8.661	1.386	5	12	304
Parent-Teacher Association	.4020	.4911	0	1	306
School Committee	.9379	.2417	0	1	306
Students Per Stream	37.53	17.22	2	95	304
Private	.2190	.4142	0	1	306
Urban	.3235	.4686	0	1	306
Student Breakfast	.8717	.1698	.1	1	306
Student Absenteeism	.1203	.1488	0	1	278

Table 34: Shows the descriptive statistics for language in Kenya. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Mozambique Descriptive Statistics - Mathematics

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Mathematics Score	.2542	.0745	.0588	.7059	203
Teacher Subject Knowledge	.2599	.1664	0	.8696	182
Teacher Pedagogical Skills	.1263	.1237	0	.6105	182
Teacher Education	.0663	.1942	0	1	182
Teacher Absenteeism	.3196	.4257	0	1	182
Teacher Behaviour	6.819	2.247	1	13	199
Student-Teacher Ratio	247	176.9	25	862	200
Share of Time Spent Teaching	.9617	.0779	.2683	1	203
Facilities Index	6.837	1.823	2	10	203
School Committee	.7734	.4197	0	1	203
Students Per Stream	46.18	17.77	6	98.5	203
Urban	.1527	.3606	0	1	203
Student Breakfast	.7308	.2684	0	1	203
Student Absenteeism	.51	.2718	0	.9815	202

Table 35: Shows the descriptive statistics for mathematics in Mozambique. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Mozambique Descriptive Statistics - Language

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Language Score	.2103	.1989	.013	.9351	203
Teacher Subject Knowledge	.2935	.1054	0	.6341	182
Teacher Pedagogical Skills	.1263	.1214	0	.6105	182
Teacher Education	.0681	.1956	0	1	182
Teacher Absenteeism	.3288	.4313	0	1	182
Teacher Behaviour	6.819	2.247	1	13	199
Student-Teacher Ratio	247	176.9	25	862	200
Share of Time Spent Teaching	.9617	.0779	.2683	1	203
Facilities Index	6.837	1.823	2	10	203
School Committee	.7734	.4197	0	1	203
Students Per Stream	46.18	17.77	6	98.5	203
Urban	.1527	.3606	0	1	203
Student Breakfast	.7308	.2684	0	1	203
Student Absenteeism	.51	.2718	0	.9815	202

Table 36: Shows the descriptive statistics for language in Mozambique. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Nigeria Descriptive Statistics - Mathematics

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Mathematics Score	.4046	.1727	0	1	760
Teacher Subject Knowledge	.4299	.2097	0	.9565	759
Teacher Pedagogical Skills	.1316	.1073	0	.4866	759
Teacher Education	.9338	.1449	0	1	760
Teacher Absenteeism	.1520	.2242	0	1	760
Teacher Behaviour	8.889	3.061	1	14	682
Student-Teacher Ratio	141.8	136.9	0	1492	760
Share of Time Spent Teaching	.8607	.2547	0	.9833	760
Facilities Index	6.814	2.69	0	13	760
Parent-Teacher Association	.9592	.1979	0	1	760
School Committee	.7776	.4161	0	1	760
Students Per Stream	26.67	17.63	0	157	738
Private	.3958	.4893	0	1	758
Urban	.2053	.4042	0	1	760
Student Breakfast	.8799	.1898	0	1	760
Student Absenteeism	.2054	.2433	0	1	743

Table 37: Shows the descriptive statistics for mathematics in Nigeria. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Nigeria Descriptive Statistics - Language

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Language Score	.4534	.306	.0012	1	760
Teacher Subject Knowledge	.3322	.14	0	.6444	662
Teacher Pedagogical Skills	.13	.1124	0	.6487	662
Teacher Education	.8446	.3325	0	1	662
Teacher Absenteeism	.1002	.2695	0	1	662
Teacher Behaviour	8.889	3.061	1	14	682
Student-Teacher Ratio	141.8	136.9	0	1492	760
Share of Time Spent Teaching	.8607	.2547	0	.9833	760
Facilities Index	6.814	2.69	0	13	760
Parent-Teacher Association	.9592	.1979	0	1	760
School Committee	.7776	.4161	0	1	760
Students Per Stream	26.67	17.63	0	157	738
Private	.3958	.4893	0	1	758
Urban	.2053	.4042	0	1	760
Student Breakfast	.8799	.1898	0	1	760
Student Absenteeism	.2054	.2433	0	1	743

Table 38: Shows the descriptive statistics for language in Nigeria. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Uganda Descriptive Statistics - Mathematics

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Mathematics Score	.4568	.1133	.2118	.7765	399
Teacher Subject Knowledge	.6014	.1729	0	.913	379
Teacher Pedagogical Skills	.2126	.1122	0	.5684	379
Teacher Education	.0648	.2035	0	1	379
Teacher Absenteeism	.3405	.3944	0	1	379
Teacher Behaviour	9.465	2.06	3	14	342
Student-Teacher Ratio	165.2	198	6.636	1154	399
Share of Time Spent Teaching	.8992	.1918	0	1	398
Facilities Index	8.49	1.428	5	13	390
Parent-Teacher Association	.89	.313	0	1	400
School Committee	.975	.1563	0	1	400
Students Per Stream	79.71	44.21	5	295	400
Private	.2025	.4024	0	1	400
Urban	.1875	.3908	0	1	400
Student Breakfast	.6221	.2596	0	1	399
Student Absenteeism	.2321	.1687	0	1	395

Table 39: Shows the descriptive statistics for mathematics in Uganda. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Uganda Descriptive Statistics - Language

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Language Score	.5255	.2697	.0432	.9988	399
Teacher Subject Knowledge	.5631	.1197	.1333	.8222	376
Teacher Pedagogical Skills	.2264	.1138	0	.6177	376
Teacher Education	.0815	.239	0	1	376
Teacher Absenteeism	.3568	.3908	0	1	376
Teacher Behaviour	9.465	2.06	3	14	342
Student-Teacher Ratio	165.2	198	6.636	1154	399
Share of Time Spent Teaching	.8992	.1918	0	1	398
Facilities Index	8.49	1.428	5	13	390
Parent-Teacher Association	.89	.313	0	1	400
School Committee	.975	.1563	0	1	400
Students Per Stream	79.71	44.21	5	295	400
Private	.2025	.4024	0	1	400
Urban	.1875	.3908	0	1	400
Student Breakfast	.6221	.2596	0	1	399
Student Absenteeism	.2321	.1687	0	1	395

Table 40: Shows the descriptive statistics for language in Uganda. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Tanzania Descriptive Statistics - Mathematics

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Mathematics Score	.4902	.125	.1	.8294	400
Teacher Subject Knowledge	.6165	.1787	0	1	377
Teacher Pedagogical Skills	.3549	.123	0	.8376	377
Teacher Education	.1079	.2432	0	1	377
Teacher Absenteeism	.3085	.364	0	1	377
Teacher Behaviour	7.945	2.469	1	13	379
Student-Teacher Ratio	168.2	180.5	17.19	836	400
Share of Time Spent Teaching	.894	.1859	0	1	400
Facilities Index	7.67	1.352	4	11	391
School Committee	.995	.0706	0	1	400
Students Per Stream	74.65	50.39	12	331	400
Private	.01	.0996234	0	1	400
Urban	.2075	.406	0	1	400
Student Breakfast	.3113	.2842	0	1	382
Student Absenteeism	.2906	.2181	0	1	387

Table 41: Shows the descriptive statistics for mathematics in Tanzania. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

Tanzania Descriptive Statistics - Language

Variable	Mean	Standard Deviation	Minimum	Maximum	N
Average Student Language Score	.4959	.1814	.0432	.9309	400
Teacher Subject Knowledge	.394	.1177	.0667	.8222	363
Teacher Pedagogical Skills	.3692	.1253	0	7051	363
Teacher Education	.1629	.3018	0	1	363
Teacher Absenteeism	.2874	.3683	0	1	363
Teacher Behaviour	7.945	2.469	1	13	379
Student-Teacher Ratio	168.2	180.5	17.19	836	400
Share of Time Spent Teaching	.894	.1859	0	1	400
Facilities Index	7.67	1.352	4	11	391
School Committee	.995	.0706	0	1	400
Students Per Stream	74.65	50.39	12	331	400
Private	.01	.0996234	0	1	400
Urban	.2075	.406	0	1	400
Student Breakfast	.3113	.2842	0	1	382
Student Absenteeism	.2906	.2181	0	1	387

Table 42: Shows the descriptive statistics for language in Tanzania. Note that the numbers reported were rounded to 4 decimal points, or the first 4 digits where appropriate.

A.7 Highest and Lowest VIF Scores

The highest VIF for 'All Countries' Mathematics was the covariate for the percentage of teachers with an education greater than secondary school with 1.68 and the lowest was the mathematics teacher's absence rate covariate with 1.12.

The highest VIF for 'All Countries' Language was the covariate for student breakfast with 1.59 and the lowest was the language teacher's absence rate covariate with 1.09.

The highest VIF for Kenya Mathematics was the Private covariate with 1.88 and the lowest was the Student Absence covariate with 1.06.

The highest VIF for Kenya Language was the Private covariate with 1.76 and the lowest was the Student Absence covariate with 1.07.

The highest for Mozambique Mathematics was the Student Absence covariate with 1.48 and the lowest was the Share of Time Spent Teaching covariate with 1.04. The highest VIF for Mozambique Language was student absence with 1.50 and the lowest was the Share of Time Spent Teaching covariate with 1.4.

The highest VIF for Nigeria Mathematics was the Share of Time Spent Teaching covariate with 1.55 and the lowest was the Average Student Breakfast covariate with 1.08. The highest VIF for Nigeria Language was the Share of Time Spent Teaching covariate with 1.58 and the lowest was the PTA dummy covariate with 1.07.

The highest VIF for Uganda Mathematics was the Private dummy covariate with 1.58 and the lowest was the Average Student Breakfast covariate with 1.09. The highest VIF for Uganda Language was the Private dummy covariate with 1.55 and the lowest was the Teacher Behaviour covariate with 1.08.

The highest VIF for Tanzania Mathematics was the Urban dummy covariate with 1.34 and the lowest was the SC dummy covariate with 1.02. The highest VIF for Tanzania Language was urban dummy covariate with 1.42 and the lowest was the SC dummy covariate with 1.02.