Secondary Education in Uganda

New Survey Evidence

Sebastian Krantz,
BRAC Uganda Research Division

July 16, 2018

List of Acronyms and Abbreviations

A-level	Upper Secondary School (classes S5-S6)
BoG	Board of Governors
BTVET	Business, Technical, Vocational Education and Training
CV	(10-Fold) Cross-Validation (a technical term from machine-learning)
DOS	Director of Studies
EPDC	Education Policy and Data Center
ESSP	Education Sector Strategic Plan
GDI	Gender Development Index
GER	Gross Enrolment Rate
GII	Gender Inequality Index
GIR	Gross Intake Rate
GNI	Gross National Income
GPE	Global Partnership for Education
GPI	Gender Parity Index
HDI	Human Development Index
HD	Human Development
IHDI	Inequality-Adjusted Human Development Index
LASSO	Least Absolute Shrinkage and Selection Operator (regression shrinkage method)
MII	Multidimensional Institutions Index
MoES	Uganda Minsitry of Education and Sports
MPI	Multidimensional Poverty Index
O-level	Lower Secondary School (classes S1-S4)
OOB	Out-of Bootstrap-Aggregation (sample or variance)
PCA	Principal Components Analysis
PCR	Pupil-Classroom Ratio
PIRLS	Progress in International Reading Literacy Study
PLE	Primary Leaving Examination Scores: 4 (best) - 36 (worst)
PPP	Purchasing Power Parity Adjusted
PTA	Parent-Teacher Association
PVE	Percent Variance Explained
RF	Random Forest (predictive machine-learning algorithm)
SACMEQ	The Southern and Eastern Africa Consortium for Monitoring Educational Quality
SDI	World Bank Service Delivery Indicators
SSA	Sub-Saharan Africa
TIMSS	Trends in International Mathematics and Science Study
UACE	Uganda Advanced Certificate of Education (National A-level leaving exam) Scores: 20 (best) - 0 (worst)
UCE	Uganda Certificate of Education (National O-level leaving exam) Scores: 8 (best) - 72 (worst)
UPE	Universal Primary Education
USE	Universal Secondary Education
UNEB	Uganda National Examinations Board
WDI	World Bank World Development Indicators
WGI	World Bank Worldwide Governance Indicators

List of Tables

List of Figures

Table of Contents

Executive Summary

This study reviews the state of secondary education and government policy in secondary education in Uganda and presents findings from statistical analysis of recent a survey of 450 secondary schools, executed by BRAC Uganda in January and February 2018, which it contextualizes against recent policy trajectories.

The first part of the study is a broad policy review. After providing a statistical overview of the secondary subsector, important challenges and the government policy approaches taken towards tackling these challenges are discussed. Key challenges highlighted are (I.) Increasing the primary to secondary transition rate, which at 63.2% is still far below the governments declared goal of achieving USE; (II.) Increasing the lower-secondary completion rate, which remained very low over the last couple of years, and is at currently 36.2%; (III.) Ensuring the quality and relevance of secondary education to adequately prepare students for the job-market and higher-education. Critical aspects are enhanced teacher training and monitoring, adjustment of the curriculum and provision of adequate scholastic materials; (IV.) Bridging the rural-urban divide, especially in the Northern and North-Western Regions, and managing the chasms in terms of access, performance, curriculum, and cost, between government secondary schools and the large and still growing private sector. The policy priorities of the government, as recalibrated in the Education and Sports Sector Strategic Plan (ESSP) 2017-2020, are overall well positioned for tackling these issues. The focuses of the ESSP is on quality, which is necessary, but for achieving USE the government should not neglect the lingering access issues in terms of primary to secondary transition and especially the low lower-secondary completion rates, both of which are slightly underemphasized in the new plan. The plan also allocates larger budget shares to BTVET and tertiary education, while slightly reducing the secondary budget. Larger investments in tertiary education might not be optimal when secondary completion is low, especially if, as statistics show, the job market still cannot accommodate many university graduates.

The empirical analysis of a detailed survey of 450 secondary schools presented in the second part aims to aid the setting of policy priorities and their implementation by determining the secondary school characteristics most closely associated with educational success in various dimensions. The first part of this analysis examines determinants of the aggregate academic performance and excellence of secondary students. Very robust empirical results indicate that at a minimum, teacher education, measures to aid poor students, good boarding schools, and a better technological infrastructure are highly conducive to increased student performance on the national UCE exam. The findings also highlight that above all other predictors, high performing incoming students and the amount of financial resources the school receives are dominant in explaining the aggregate performance of students on the UCE, hence these variables need to, at a minimum, be controlled for when comparing schools or educational policies. The second part of the analysis investigates determinants of dropout, repetition and completion in secondary schools. Despite being less robust, the findings indicate that obstacles hindering instruction and a high teacher attrition rate, both likely reflecting larger structural issues and low-self esteem in some schools, are associated with higher dropout rates. It was also found that dropout rates are lower if parents are frequently notified about the performance of their child. Likewise, boarding schools and secondary schools with a cantine were found to have significantly lower dropout rates. The latter is indicative that students lacking basic needs when in school are more likely to drop out. For repetition rates, the analysis results suggest that in addition to some of the just mentioned factors, lessons of appropriate (i.e. moderate) length are conducive towards reducing repetition rates. Apart from reduced rates of absenteeism, more science teachers and more teachers with a pedagogic qualification relate positively to students passing the UCE and completing S4. The third part of the analysis focuses on gender equality in performance. The findings here were not very robust, the only possibly robust effects being that increased teacher gender parity increases student-gender parity, that sporting facilities increase gender parity, and that both student poverty and elitism are negatively correlated with gender parity. A final part of the analysis examines predictors of students career path. The key findings is that teacher education is a crucial input variable in determining the share of students making it to university. In addition, pedagogically skilled teachers, more internal monitoring and evaluation and a good condition of the school are associated with more graduates enrolling in university, while bad teacher behavior, student absenteeism, a low technological infrastructure and

a low minimum teacher salary are characteristic of schools sending a large share of their students directly into the labor market.

1 Introduction

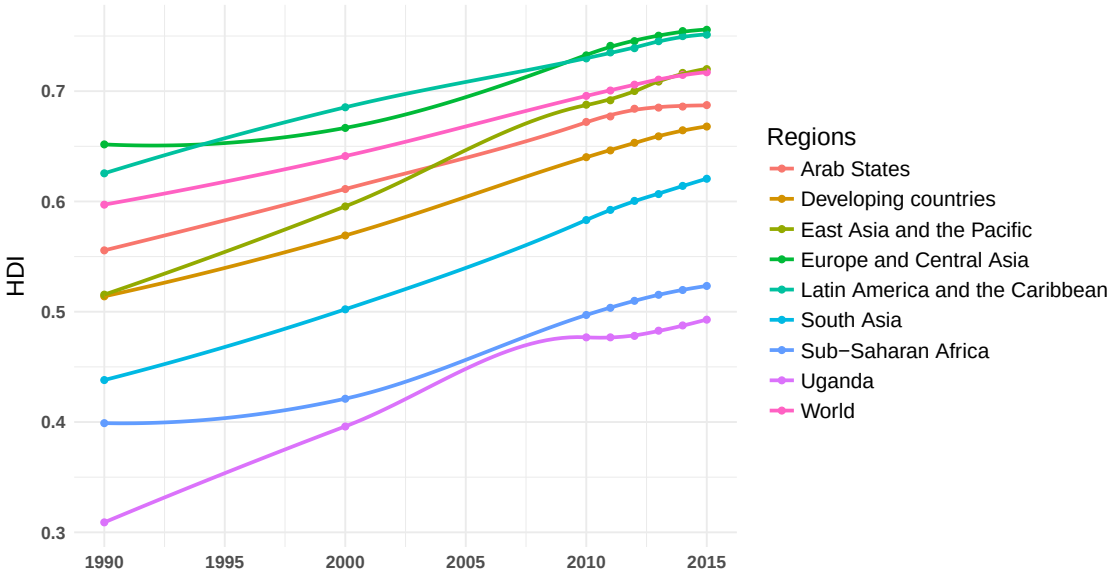
This study presents the findings of the statistical analysis of a survey of 450 secondary schools in Uganda collected by BRAC Uganda in January and February 2018 and contextualizes the findings within the broader Ugandan educational landscape and recent policy trajectories. The survey features detailed information (captured in about 200 questions to the Headmaster or Director of Studies (DOS)) on general school characteristics, school inventory, organizational effectiveness, the teaching approach, and the student as well as the teacher body. In addition, a differentiated set of educational outcome measures along the lines of access, quality and equality of outcomes across gender and income status are obtained from survey questions and official UCE test-score data for the years 2015 and 2016, published by the Uganda National Examinations Board (UNEB). The two main aims of this project are (1) to review the state of secondary education in Uganda, highlight current issues, and delineate the governments policy agendas put forth in the past and in the present towards solving these issues (2) to empirically distill the secondary school characteristics that are likely most conducive to secondary educational success along heterogenous dimensions, and in need of further policy emphasis in the coming years.

The remainder of this document is structured as follows: Section 2 provides a statistical overview of the country and the state of secondary education in Uganda. Section 3 reviews the current government policy agenda and key issues. Section 4 gives provides a brief excerpt on education finance and government budget allocations in Uganda. Section 5 introduces the survey data and describes the construction of predictors and outcome measures from the collected data. Section 6 presents the empirical analysis, and section 7 concludes by contextualizing the findings within local policy debates.

2 Statistical Overview

With an HDI of 0.493, UNDP ranks Uganda in 2016 as 163rd in terms of human development out of 185 participating countries. Like many developing countries, its HDI increased faster than the world average over the 1990 to 2015 period, with an HDI gain of approximately 0.2 compared to the World average gain of 0.12.

Figure 1: HDI Trends, 1990-2015

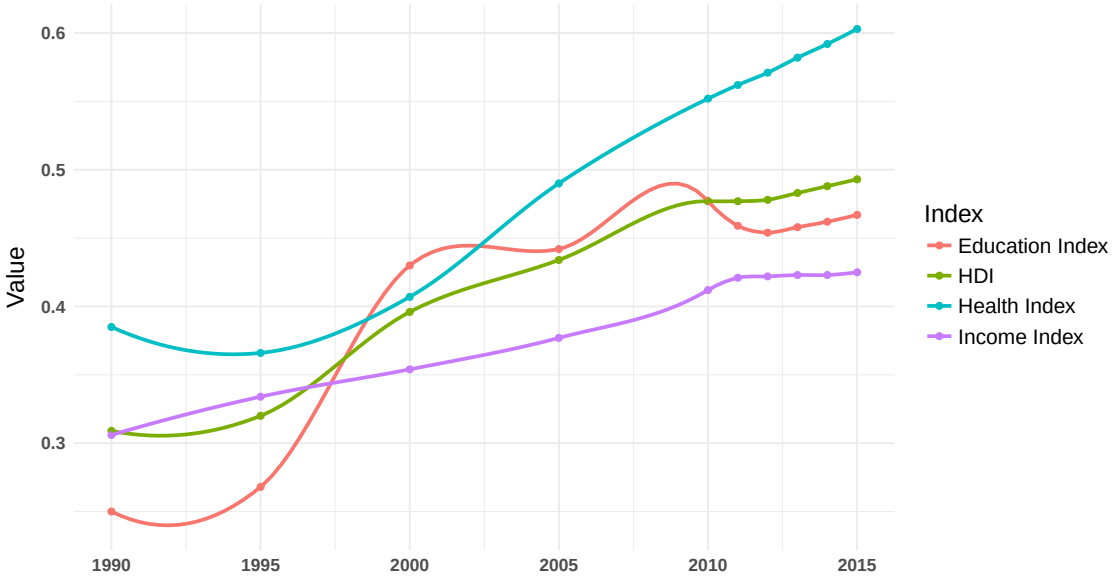


Source: UNDP Human Development Data (1990-2015)

Figure (2) decomposes this HDI trend into the dimension indexes of which it is the geometric mean. The health index is a linear transformation of life-expectancy at birth, the education index is a geometric average of mean years of schooling and expected years of schooling indexes computed in a similar way, and the income index is a linear transformation of GNI per capita (PPP $). Figure

(2) shows that most of Uganda's positive HDI trend is accounted for by improvements in health and education, both indexes gaining more than 0.2 over the sample period. It is also worthwhile to note that the trends in health and income have been very steady, while educational performance has been subject to much greater fluctuations, and even a brief decline between 2010 and 2012.

Figure 2: HDI Component Trends, Uganda 1990-2015

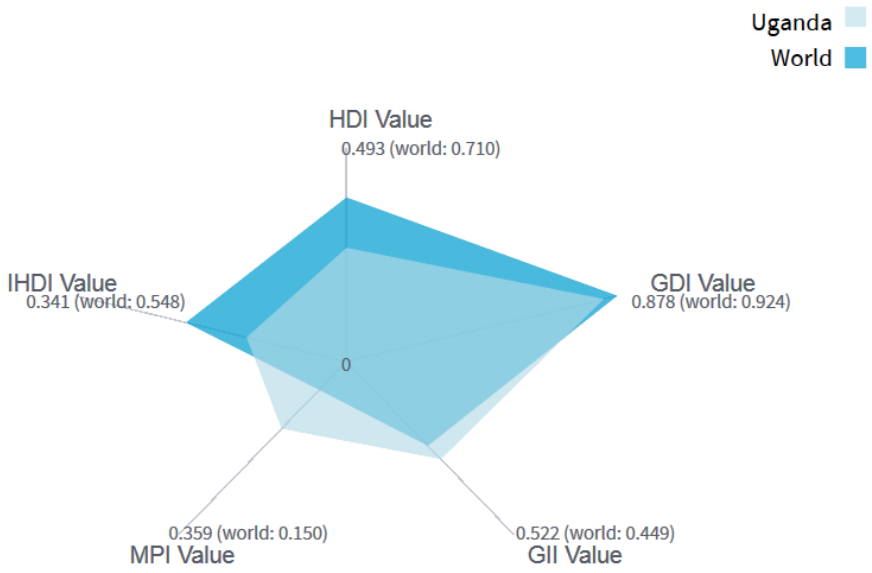


Source: UNDP Human Development Data (1990-2015)

2.1 Key Development Indicators

Figure (3) shows Uganda's performance on key development indices. It is outperformed by the World on all dimensions (Where it must be noted that the MPI and the GII are inverted progress indicators) but does surprisingly well on matters of gender inequality (GII) and gender development (GDI).

Figure 3: Uganda's Performance on Key Development Indicators (2016)



Source: UNDP Human Development Data (1990-2015)

Table (1) shows an abbreviated form of the UNDP country profile for Uganda (UNDP, 2016).

Most statistics are from 2016. Uganda has experienced formidable progress on all human development dimensions in recent decades. Life expectancy has risen, poverty and malnutrition have declined, income per capita has risen steadily, and the mobile internet revolution and gradual opening of the economy have provided Uganda with new forms of businesses and economic opportunities[1].

Table 1: Exerpt from the UNDP Uganda Country Profile

Indicator	*Value*	*Indicator*	*Value*
Health		Mean years of schooling, female (years)	4.5
Life expectancy at birth (years)	59.2	Mean years of schooling, male (years)	6.8
Deaths due to malaria (per 100,000 people)	57.9	Population with at least some secondary education, female (% ages 25 and older)	25.9
Deaths due to tuberculosis (per 100,000 people)	12	Population with at least some secondary education, male (% ages 25 and older)	32.1
HIV prevalence, adult (% ages 15-49), total	7.1	Share of seats in parliament (% held by women)	35
Infant mortality rate (per 1,000 live births)	37.7	Unemployment rate (total), female to male ratio	1.3
Public health expenditure (% of GDP)	1.8		
Stunting (moderate or severe) (% under age 5)	34.2	**Poverty**	
		Multidimensional Poverty Index (MPI), HDRO specifications	0.359
Education		Population in multidimensional poverty, headcount (%)	70.3
Expected years of schooling (years)	10	Population in multidimensional poverty, intensity of deprivation (%)	51.1
Adult literacy rate (% ages 15 and older)	73.9	Population in severe multidimensional poverty (%)	33.3
Government expenditure on education (% of GDP)	2.2	Population living below income poverty line, PPP $1.90 a day (%)	34.6
Gross enrolment ratio: pre-primary (% of preschool-age children)	11	Population near multidimensional poverty (%)	20.6
Gross enrolment ratio, primary (% of primary school-age population)	110	Working poor at PPP$3.10 a day (% of total employment)	60.6
Gross enrolment ratio, secondary (% of secondary school-age population)	28		
Gross enrolment ratio, tertiary (% of tertiary school-age population)	4	**Work, employment and Vulnerability**	
Mean years of schooling (years)	5.7	Child labour (% ages 5-14)	16
Population with at least some secondary education (% aged 25 and older)	30.8	Employment in agriculture (% of total employment)	71.9
Primary school dropout rate (% of primary school cohort)	75.2	Employment in services (% of total employment)	20.2
Primary school teachers trained to teach (%)	95	Total unemployment rate (% of labour force)	3.6
Pupil-teacher ratio, primary school (number of pupils per teacher)	46	Vulnerable employment (% of total employment)	78.9
		Youth not in school or employment (% ages 15-24)	5.9
Income/Composition of Resources		Youth unemployment rate (% ages 15-24)	6
Gross national income (GNI) per capita (2011 PPP$)	1,670		
Domestic credit provided by financial sector (% of GDP)	17.9	**Trade and Financial Flows**	
Gross domestic product (GDP), total (2011 PPP $ billions)	67.1	Exports and imports (% of GDP)	46.9
		External debt stock (% of GNI)	19.8
Inequality		Foreign direct investment, net inflows (% of GDP)	4
Inequality-adjusted HDI (IHDI)	0.341	Net official development assistance received (% of GNI)	6.2
Income inequality, Gini coefficient	41	Private capital flows (% of GDP)	-3.1
Inequality in education (%)	29.4	Remittances, inflows (% of GDP)	3.98
Inequality in income (%)	27.3		
Inequality in life expectancy (%)	35.7	**Mobility and Communication**	
Overall loss in HDI due to inequality (%)	30.9	Internet users (% of population)	19.2
		International student mobility (% of total tertiary enrolment)	7.2
Gender			
Gender Development Index (GDI)	0.878	**Demography**	
Adolescent birth rate (births per 1,000 women ages 15-19)	111.9	Population, total (millions)	39
Expected years of schooling, female (years)	9.9	Dependency ratio, young age (0-14) (per 100 people ages 15-64)	97.3
Expected years of schooling, male (years)	10.1	Median age (years)	15.9
Gender Inequality Index (GII)	0.522	Population, ages 15–64 (millions)	19.3
Human Development Index (HDI), female	0.459	Population Growth (%)	3.3
Human Development Index (HDI), male	0.523	Population, under age 5 (millions)	7.3
Labour force participation rate, female (% ages 15 and older)	82.3	Population, urban (%)	16.1
Labour force participation rate, male (% ages 15 and older)	87.7	Sex ratio at birth (male to female births)	1.03

Source: UNDP Uganda Country Profile (2016)

Table (1) shows amongst other things the present (2016) values of the statistics underlying the HDI. Life expectancy at birth is 59.2 years, expected years of schooling is 10 years and mean years of schooling is 5.7 years. The PPP adjusted GDP per capita is 1670$, at a moderate level of inequality (Gini 0.41). Uganda achieves, on a comparison of low-income countries, a quite high level of gender development (GDI 0.88). While the present mean years of schooling in the population still shows a noticeable disparity (male 6.8, female 4.5), the numbers are almost equalized for current expected years of schooling (male 10.1, female 9.9). Poverty is however still prevalent: UNDP estimates that 70.3% of the population live in multidimensional poverty, and 20.6% are near multidimensional poverty. 60.6% of the population earn less than 3.1$ a day. Despite a rapid growth of business in the metropolitan area around Kampala, the structure of the economy is still largely agricultural, with 71.9% of the population working in agriculture and 84% of the population living in rural areas. Uganda has a remarkably low youth unemployment rate of just 6% among the 15-24 year olds. The trading sector has also increased in recent years. Uganda's largest export commodity is coffee with share of 19% in exports. Its larges import commodity is petroleum (16% of imports). The top export destinations of Uganda are Kenya ($412M), South Sudan ($234M), Rwanda ($230M), the Democratic Republic of the Congo ($152M) and Italy ($119M) while it imports mostly from India ($1.09B), China ($863M), Kenya ($534M), the United Arab Emirates ($383M), and Japan ($322M) (**?**). Uganda currently runs a large trade deficit of 11.1% of its GDP and a current account imbalance of 8.6% of GDP (World-Bank, 2018c). With around 3.3% per

[1]For a more detailed overview of human development trends in Uganda, I refer the reader to the human development indicators accessible in Google's public data explorer.

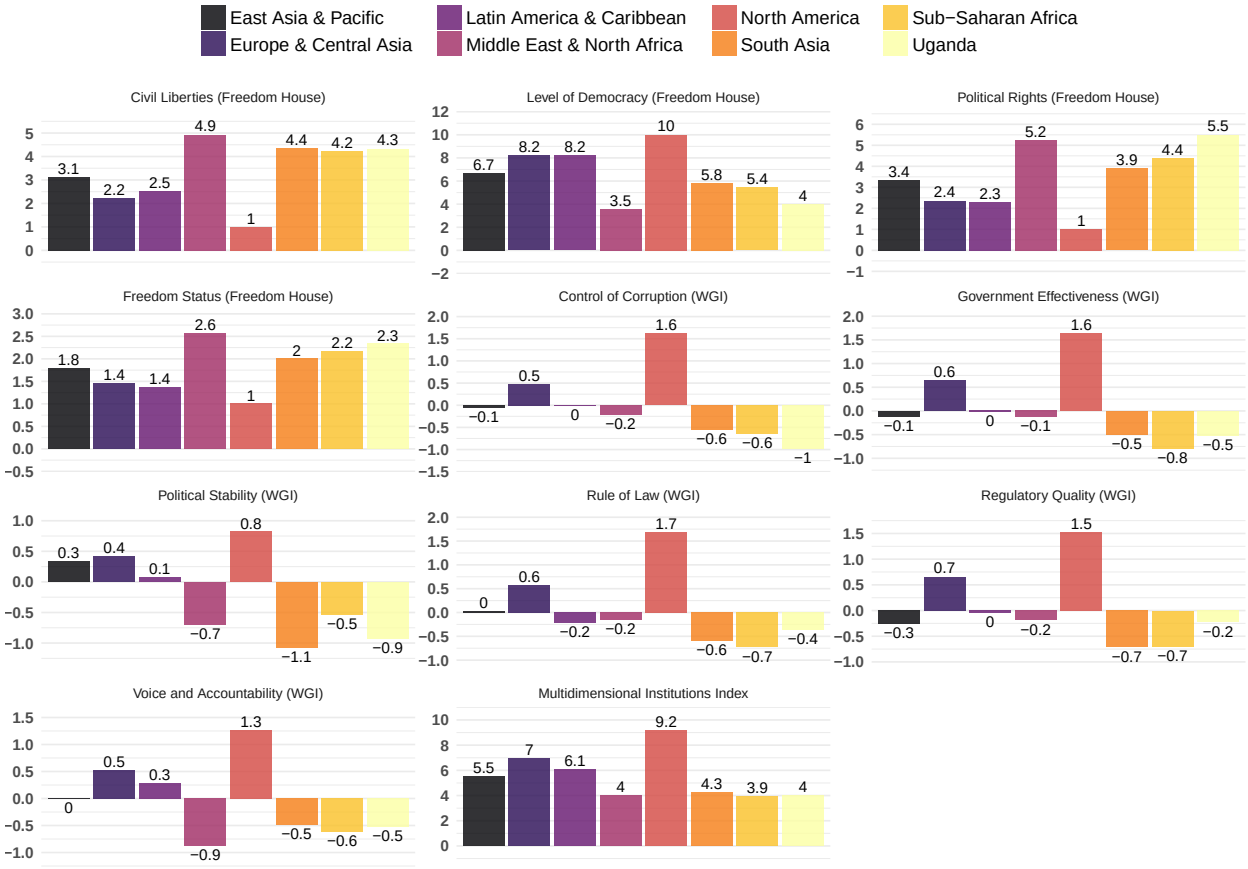
annum, Uganda experiences moderate population growth. Nearly half of the population (48% in 2014) is under 15 years of age (World-Bank, 2018c).

2.2 Quality of Governance

Figure (4) shows governance statistics from the World Bank's Worldwide Governance Indicators (WGI) and Freedom House (World-Bank, 2018d; Freedom-House, 2018). The WGI indexes all run from -2.5 to 2.5 and measure the quality of governance in six key dimensions. The four Freedom House indices included capture the level of democracy and political freedoms, but are measured on different scales (Civil Liberties, Political Rights, and Freedom Status are measured on an inverted scale)[2]. Figure (4) reports averages of these indices over the 2010-2016 period[3]. In addition, a multidimensional institutions index (MII) is computed as the factor score of the 10 indexes and mapped to a 0-10 scale. This factor score index explains 78% of the variance in the underlying 10 indices. The lowest factor loading in absolute value is on Political Stability, with $\rho = 0.77$, which is still quite high and suggests that the 10 indexes are highly collinear.

According to the MII, the overally quality of governance in Uganda is 4, slightly higher than the Sub-Saharan Africa (SSA) average of 3.9, and equal to the Middle-East & North Africa average. On the indices itself, Uganda performs comparatively best on Rule of Law and Regulatory Quality, and comparatively worst on Political Rights and Control of Corruption.

Figure 4: Governance Statistics (2010-2016 Averages)



Source: World Bank Worldwide Governance Indicators, Freedom House, Authors Computations

[2]More on the indexes can be read from the codebook of the QGS standard dataset from which all of these indicators are taken https://qog.pol.gu.se/data/datadownloads/qogstandarddata. For interpretation, the comparative aspect is most important.

[3]First, the regional aggregates are obtained for every year by taking the (unweighted) mean across all countries of a given region (World Bank classification). Then, for each region, an average over the 2010-2016 period is taken.

2.3 Secondary Education in Uganda

The educational system in Uganda is divided into Nursery/Kindergarten (3 years), primary school (7 years), secondary school (6 years), which divides itself into lower secondary school (4 years) and upper secondary school (another 2 years). From lower secondary school graduates can proceed to technical training centers, primary school teacher training or government department training, whereas upper secondary school enables students to attend university (UIA, 2010). At the end of primary school, students sit the nationally held primary leaving examination (PLE). The examination subjects are English, math, science and social studies. Primary school is free of cost in Uganda since 1997, but still pupils from rural areas in particular face major obstacles (like lacking materials (books, pens), lack of school meals and poor teaching quality). Next to government schools, there are many private schools where wealthier parents send their children, including low-cost private schools like Bridge Academies (URDT, 2018).

Pupils who pass their PLE proceed to secondary school. Secondary school is divided into O(Ordinary)-level and A(Advanced)-level. The first 4 years (S1-S4) constitute the O-level period (lower secondary school), at the end of which students sit the national Ugandan O-level exam. Students are examined in 8-10 subjects, and upon successful completion receive the Uganda Certificate of Education (UCE). Students passing this exam can proceed to 2 more years of higher secondary school (A-level, S5 and S6) or proceed to technical training. At the end of S6, students sit the A-level exam in at least 3 subjects. Upon successful completion, students receive the national Uganda Advanced Certificate of Education (UACE), enabling them to attend University (Nuffic, 2010; Kavuma, 2018). Since 2007, secondary school has also be freed from tuition, but only students with a 3rd division grade may attend free public secondary school. Primary school graduates may also choose to attend a three-year technical school instead of lower secondary school. Graduates from lower secondary school may choose to go on to 2-3 year technical institutes, 2-year primary teacher colleges (PTC's) or government department training colleges (DTC's) instead of attending upper secondary school (UIA, 2010). Figure (2) shows a mapping of the Ugandan education system produced by the UNESCO in 2011.

Although the education sector in Uganda has seen considerable improvements in recent decades, there remain substantial challenges. Some of the largest remaining challenges to quality secondary education in public schools are a high level of teacher and student absenteeism, weak school-level management structures, inadequate availability of learning materials, and large class sizes. A major issue is also the availability of teachers in disadvantaged areas and a lack of accommodation for teachers in rural, hard to reach areas (GPE, n.d.).

The following figures are taken from 2 documents published by the Education Policy and Data Center entitled "EPDC Education Trends and Projections 2000-2025" (EPDC, 2013), and "Uganda National Education Profile 2014" (EPDC, 2014). These publications are both a bit old but nicely show some general information about the state of education in Uganda. I will complement these figures with up-to-date trends and predictions specific to secondary education, taken from the World Bank EdStats database and the Uganda Ministry of Education Education Sector Strategic Plan (ESSP) 2017-2020 (World-Bank, 2018a; MoES, 2017).

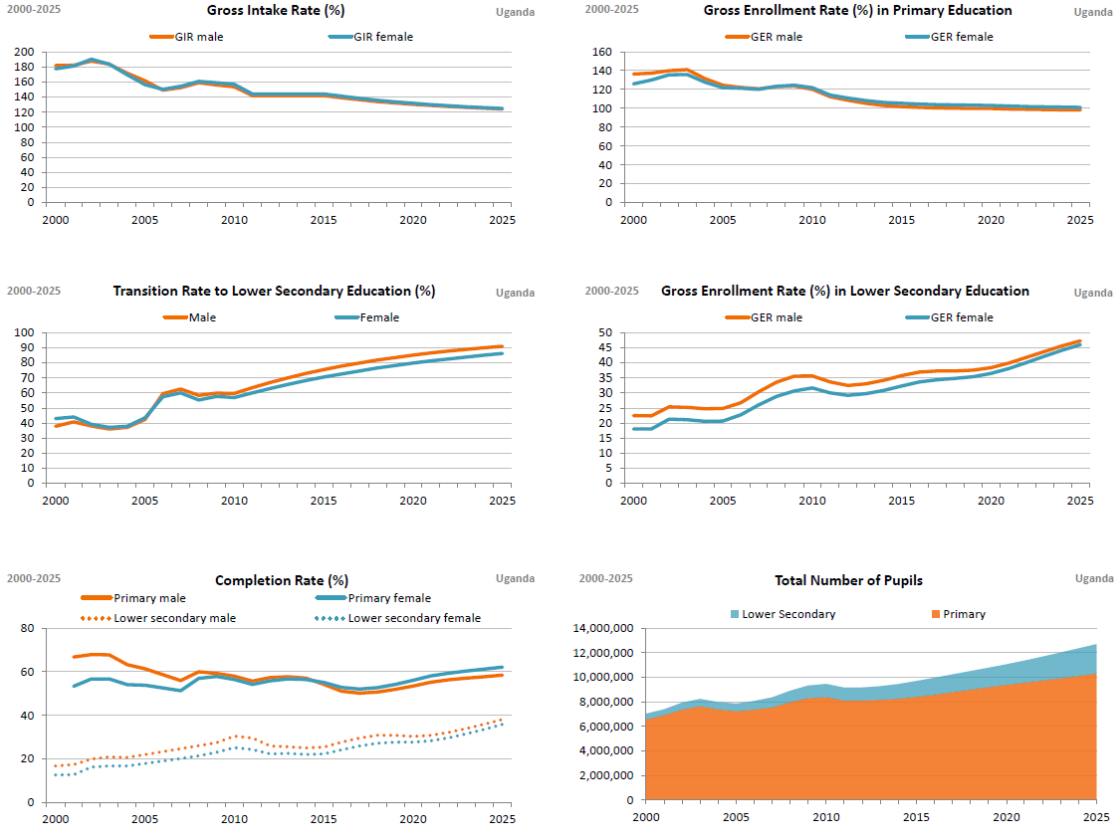
In 2011 Uganda has a total of 9,428,000 pupils enrolled in primary and secondary education. Of these pupils, about 8,098,000 (86%) were enrolled in primary education (EPDC, 2014). Figure (5) shows the EPDC general trends and prediction in Uganda, 2000-2025 (EPDC, 2013). The second row of the graph shows that access to lower secondary education, as captured by the transition of pupils from primary school and the rise in enrolment, has improved in recent years. Due to a population growth of around 3.3% per year, the total number of pupils has also increased, as shown in the bottom row of Figure (5). The completion rate was around 60% for primary and around 25% for secondary school. In 2014, 100% of students enrolled in primary school, while only around 25-30% of eligible students enrolled in lower secondary school.

Table 2: UNESCO ISCED Mapping of the Ugandan Education System (2011)

Name of the education programme	*Minimum entrance requirements*	*Main diplomas, qualifications or certificates awarded at end of programme*	*Theoretical entrance age*	*Theoretical duration (years)*	*ISCED 2011 level*		*Notes*
Early childhood education (pre-primary)	na	na	3	3	Early childhood education	0	All private schools
Primary	na	Primary leaving examination (PLE) certificate	6	7	Primary education	1	
Lower secondary (O' level)	Primary leaving examination (PLE) certificate	Uganda certificate of education (UCE)	13	4	Lower secondary education	2	
Post-primary, (vocational)	Primary leaving examination (PLE) certificate	Certificates 1	13	3	Lower secondary education	2	This includes technical schools, farm schools and vocational training. The programme can last two to three years.
Upper secondary (A' level)	Uganda certificate of education (UCE), Certificates 1	Uganda advanced certificate of education (UACE)	17	2	Upper secondary education	3	
Upper secondary (other, e.g. Business, Technical Vocational Education and Training (BTVET) and Primary Teachers Colleges programmes)	Uganda certificate of education (UCE), Certificates 1	Certificates 2	17	2	Upper secondary education	3	This includes technical institutes, community polytechnics and primary teachers colleges (PTC).
Diploma programmes (After UACE)	Uganda advanced certificate of education (UACE)	Diploma	19	2	Short-cycle tertiary education	5	This is a general bridging programme and is delivered by colleges, e.g. teachers colleges (NTC), technical colleges, business colleges, Health colleges and Instructor Training Colleges.
Diploma programmes (After certificate 2)	Certificates 2	Diploma	19	3	Short-cycle tertiary education	5	
Bachelor's degree programme	Uganda advanced certificate of education (UACE), Diploma	Bachelor's degree	19	4	Bachelor's or equivalent level	6	The programme can last three to four years.
long Bachelor's degree programme	Uganda advanced certificate of education (UACE), Diploma	Long Bachelor's degree (Medicine, Pharmacy, Dental Surgery, Veterinary, Architecture, etc.)	19	5	Bachelor's or equivalent level	6	
Postgraduate studies	Bachelor's degree	Postgraduate diploma (PGD)	22	1	Bachelor's or equivalent level	6	
Master's degree	Bachelor's degree, Postgraduate diploma (PGD)	Master's degree	22	2	Master's or equivalent level	7	
Doctoral studies	Master's degree	Doctoral degree	24	3	Doctoral or equivalent level	8	Theoretical duration is variable, depending on subject.

Source: UNESCO ISCED

Figure 5: EPDC Education Trends and Projections 2000-2025

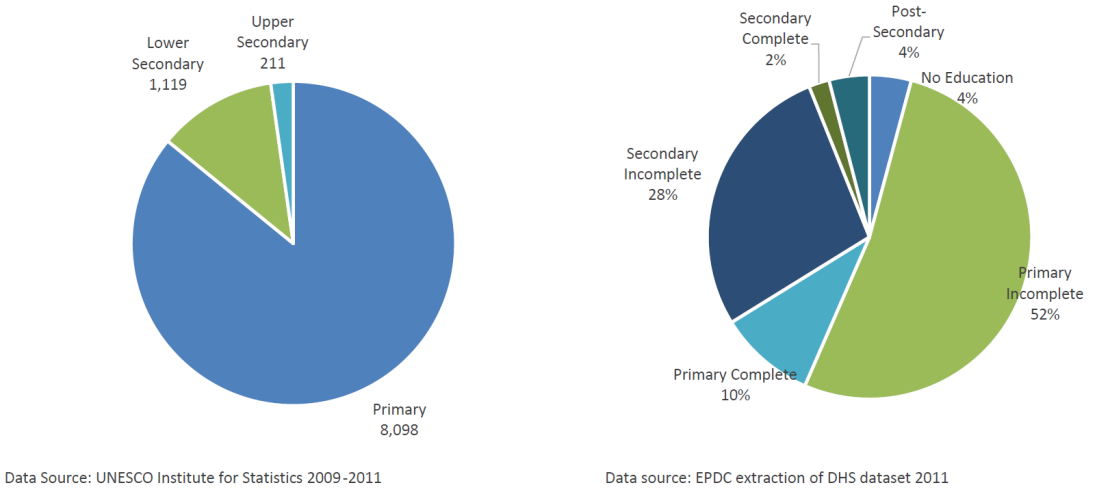


Source: Education Policy and Data Center (2013)

Figure (6) shows that in 2011, the number of primary school pupils outweighed the number of secondary school pupils by about 5:1. 50% of youth had incomplete primary education and 20% an incomplete secondary education, with only 2% having completed secondary and only 4% in post-secondary education.

Figure 6: Educational Attainment in 2011

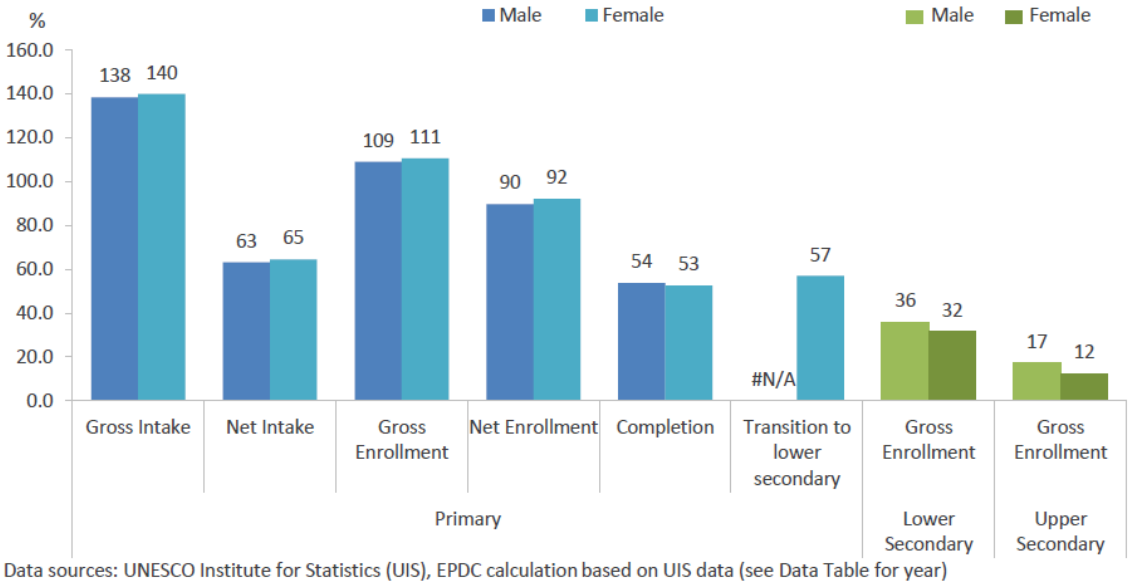
Left: Number of pupils by school level (in 1000's) | *Right*: Educational attainment, youth ages 15-24



Source: Education Policy and Data Center (2014)

Figure (7) shows enrolment and completion rates in primary schools and enrolment in secondary schools. About 35% of male and female students enroled in lower secondary school following primary school, and only about 15% of eligible students enroled in upper secondary school.

Figure 7: Primary to Secondary Transition in 2011



Source: Education Policy and Data Center (2014)

Figure (8) shows government expenditure and student-teacher ratio's by education sub-sector. Uganda's spending on secondary education is at 21% of GDP per capita per pupil twice as large as its spending on primary education, and above the low-income country average. With 19 pupils per teacher, Uganda is also far below the low-income country average. It is pleasing to find that the student-teacher ratio computed from the survey data presented below is 19.5, suggesting that

this figure is not that outdated, although this is noted with caution since according to the World Bank it has been increasing to around 24 between 2010 and 2015 (see Figure (11).

Figure 8: Expenditure and Student-Teacher Ratio in 2014

Left: Per pupil expenditure (% of per capita GDP) | *Right*: Pupil-teacher ratio

Source: Education Policy and Data Center (2014)

2.3.1 Access and Quality

Tables (3) shows the total enrolment figures in Ugandan education by subsector and gender, from 2007 to 2016, taken from the Ministry of Educations ESSP 2017-2020 (MoES, 2017). Enrolment has increased considerably in all subsectors over the time-period, with an average yearly growth rate of 4.7%, or 1.4% in excess of population growth (3.3%). The largest gain was realized in pre-primary education where enrolment increased 10-fold. In secondary education, total enrolment has increased 1.5-fold, from about 1 million to 1.5 million pupils, which makes for a yearly secondary enrolment growth of 8.8% or 5.5% yearly in excess of population growth. The increase in female enrolment has been slightly stronger than the increase in male enrolment, which is also graphically shown in Figure (10).

Table 3: Enrolment by Level of Education 2007-2016

Level	Sex	2007	2009	2011	2013	2015	2016
Pre Primary	Male	37,689	114,473	105,428	210,966	236,284	279,089
	Female	38,849	119,955	109,369	219,459	240,839	284,824
	Total	76,538	234,428	214,797	430,425	477,123	563,913
Primary	Male	3,779,338	4,150,037	4,039,734	4,219,523	4,122,663	4,294,473
	Female	3,758,633	4,147,743	4,058,443	4,240,197	4,141,654	4,361,451
	Total	7,537,971	8,297,780	8,098,177	8,459,720	8,264,317	8,655,924
Secondary	Male	517,254	648,014	662,003	727,212	657,163	765,406
	Female	437,074	546,440	596,081	635,527	608,845	691,871
	Total	954,328	1,194,454	1,258,084	1,362,739	1,284,008	1,457,277
Post-Primary (BTVET & PTCs)	Male	23,102	27,300	28,601	35,415	33,212	37,107
	Female	6,339	11,628	14,577	23,383	24,020	26,178
	Total	29,441	38,928	43,178	58,798	57,232	63,285
Tertiary (Diploma & Degree awarding)	Male	88,228	95,441	100,831	113688	143,212	144,314
	Female	66,854	74,035	78,738	87,572	114,643	114,552
	Total	155,082	169,476	179,569	201,260	257,855	258,866
Total		**8,753,360**	**9,935,066**	**9,793,805**	**10,512,942**	**10,340,535**	**10,999,265**

Source: Education and Sports Sector Strategic Plan 2017-2020 (MoES, 2017)

In Table (4), the government projects its forecasts of enrolment rates through 2020. The secondary education figures let one infer that the government expects secondary enrolment to continue increasing at an annual growth rate of 8.8%, while for total enrolment across sectors it projects an annual growth of 3.6%, which is roughly equal to the population growth rate.

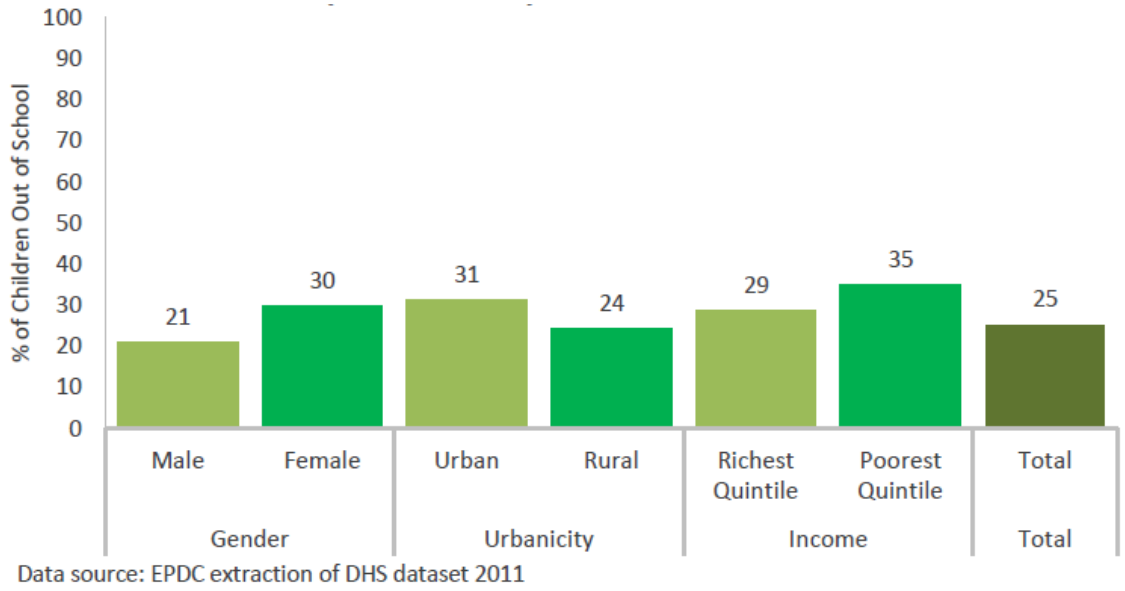
Table 4: Projected Enrolment by Subsector 2016-2020

Sub-sector	2015/16 (Actual)	2016/17	2017/18	2018/19	2019/20
Primary	**8,264,217**	**8,689,723**	**8,893,447**	**9,055,260**	**9,236,324**
Secondary	1,284,008	1,396,055	1,516,423	1,656,527	1,798,774
"O" Level	1,147,952	1,263,752	1,368,677	1,484,833	1,608,473
"A" Level	136,056	132,302	147,746	171,695	190,301
BTVET:	67,236	38,386	38,824	39,339	40,590
Post-P7	16,922	17,880	17,908	18,004	18,829
Post-S4	20,104	20,506	20,916	21,334	21,761
Post-S6/Tertiary	30,211	34,178	34,392	34,607	34,825
TIET:	59,133	59,711	60,469	61,237	62,017
Post-S4	16,495	16,825	17,161	17,505	17,855
Post-S6/Tertiary	42,638	42,887	43,308	43,733	44,162
University	198,338	204,289	210,417	216,730	223,232
Total	**9,872,932**	**10,388,164**	**10,719,580**	**11,029,093**	**11,360,937**

Source: Education and Sports Sector Strategic Plan 2017-2020 (MoES, 2017)

Yet, still around 70% if secondary school-aged children in Uganda are not enroled in secondary school. Figure (9) disaggregates the percentage of secondary school-aged children (13-18-year-olds) that are not in school (neither still in primary nor in secondary school) by gender, urbanicity and income. The statistics show that female, rural and poor students are at a clear disadvantage, although these effects are small to what one could expect in the East-African context (and progress has been made since 2011, more below).

Figure 9: Children of Secondary School Age (13-18) out of School 2011



Source: Education Policy and Data Center (2014)

Figure (10) shows further education trends since 1990, with World Bank income aggregates and the World average shown as comparison groups (World-Bank, 2018a). The series are truncated below 1990, and no-series had data for 2016 or 2017 available yet. The First row of Figure (10) shows that Uganda has not performed above the low-income country average in terms of increasing overall access to secondary education, but has made massive improvements in gender equality in access. Whereas in 1990 enrolment of 10 boys had been met by less than 6 girls, in 2015 this number has risen to 9 girls. A similar pattern is visible for lower secondary completion rates in the

second row, were a gradual increase in the lower-secondary completion rate from around 10% in 1995 to close to 30% in 2015 is visible in the left panel, whereas a massive jump in the GPI from 0.6 to 0.9 over the same period is evident in the right.

Rows 3 and 4 of Figure (10) present some incongruities. The third row shows that Uganda has a secondary school grade repeater rate of only 2% per year, below the high-income country average. This is quite unintuitive since the low-income country average is around 10%. Although these figures from the World Bank are surprisingly low, they appear to be right: The equivalent survey estimates yield a repeater rate of 2.5% for lower-secondary students and 1.8% for upper-secondary students, cf. Table (10). Gender equality in repetition is roughly balanced[4], despite showing a trend towards increased male repetition in 2005 and then again increased female repetition in 2013. The final two plots in row 4 of Figure (10) show that the expected years of schooling in Uganda has increased to 11 years through 2005, but then seen a decline again down to 9 years in 2013, while the GPI on this indicator has increased steadily and is now close to 1. The reason for this decline might be rooted in measures by the government towards reaching its goal of Universal Secondary Education (USE) by 2015 (EPDC, 2011). These strategies focused on reducing the unit costs of secondary education and increasing access and efficiency through:

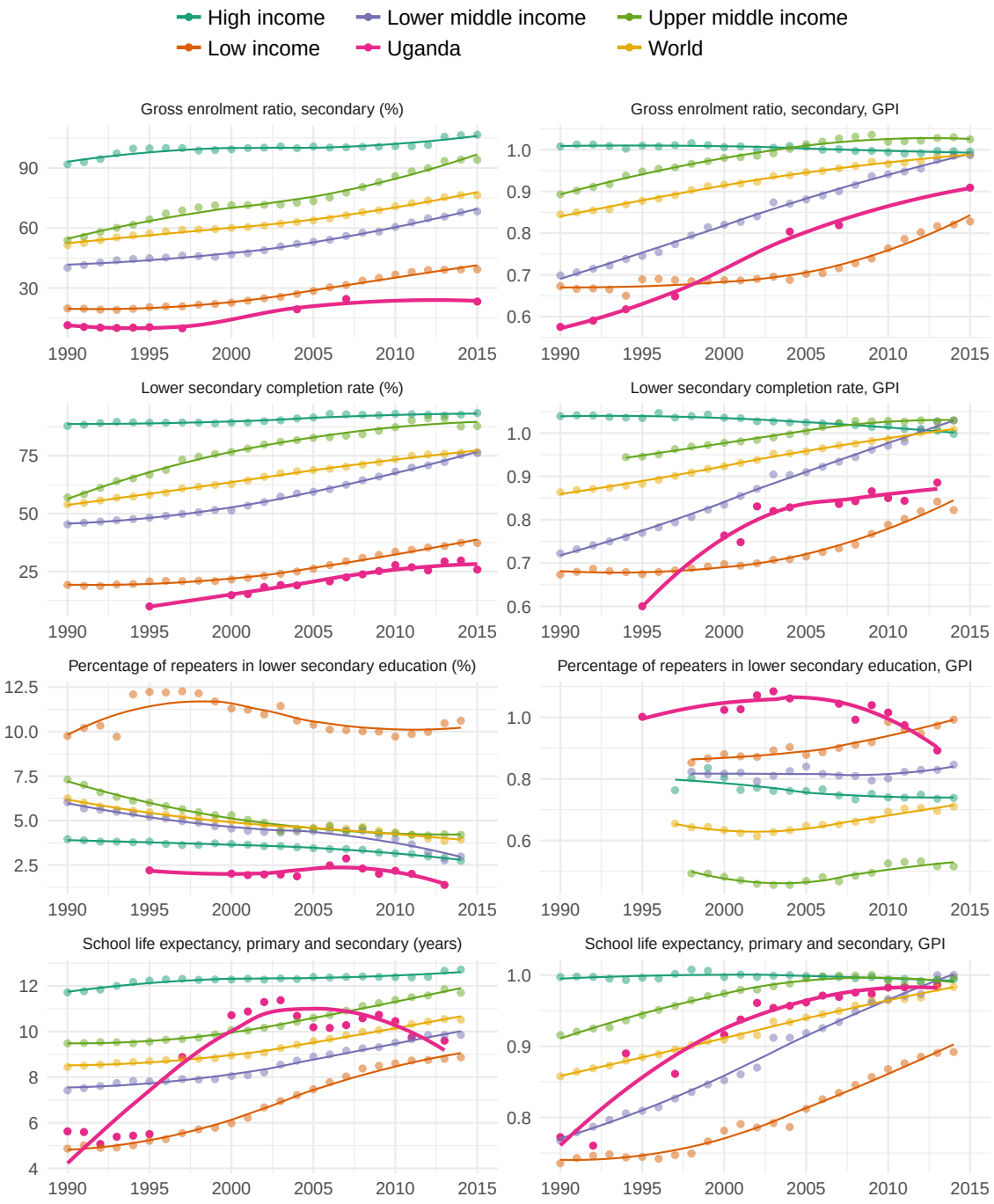
- Curriculum reduction and consolidation of subjects
- Increased teacher-pupil ratios and minimum class sizes
- Rationalization of teacher workloads and the number of subjects each teacher teaches
- Introduction of double shifts and multi-grade classrooms
- Decentralization of school management and curriculum planning to schools
- Redeployment of teachers to better meet demands

More information on past and current policy priorities of the government will be provided in section 3.

[4] Although it is probably not easy to get reliable estimates of the GPI when overall repetition rates are that low.

Figure 10: Secondary Education Trends in the International Context (1990-2015)

Fit: Lowess smoother with span $\lambda = 1$

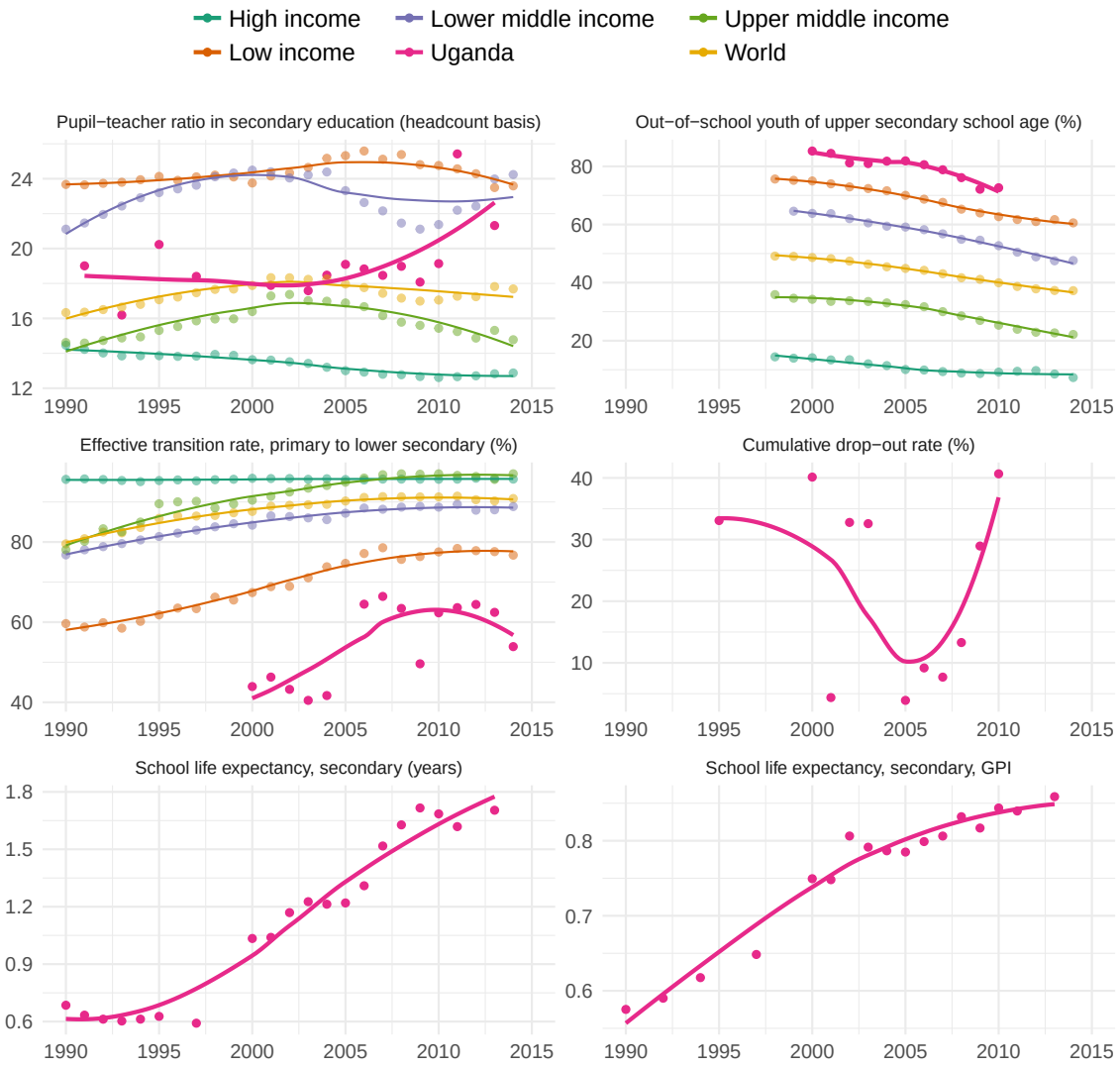


Source: World Bank Education Statistics (EdStats) (Accessed 11.04.2018)

Figure (11) shows some further secondary education trends over the 1990-2015 period. The first plot shows the pupil-teacher ratio in secondary schools, and affirms the hypothesis advanced regarding the impact of government USE reforms pursued from 2007 onwards: Whereas in 2005 the pupil-teacher ratio in Ugandan secondary schools matched the World average of 18, it rose again through 2013 to reach close to the low-income country average of 24 pupils per teacher. The reforms aimed at achieving USE by 2015 (which has very clearly not been achieved) have been criticized amongst other aspects on the grounds that dramatically increased student numbers coinciding with lack of professional development for teachers, head teachers, and local level administrators have led to a deterioration of educational quality (EPDC, 2011).

Figure 11: More Education Trends (1990-2015)

Fit: Lowess smoother, $\lambda = 1$



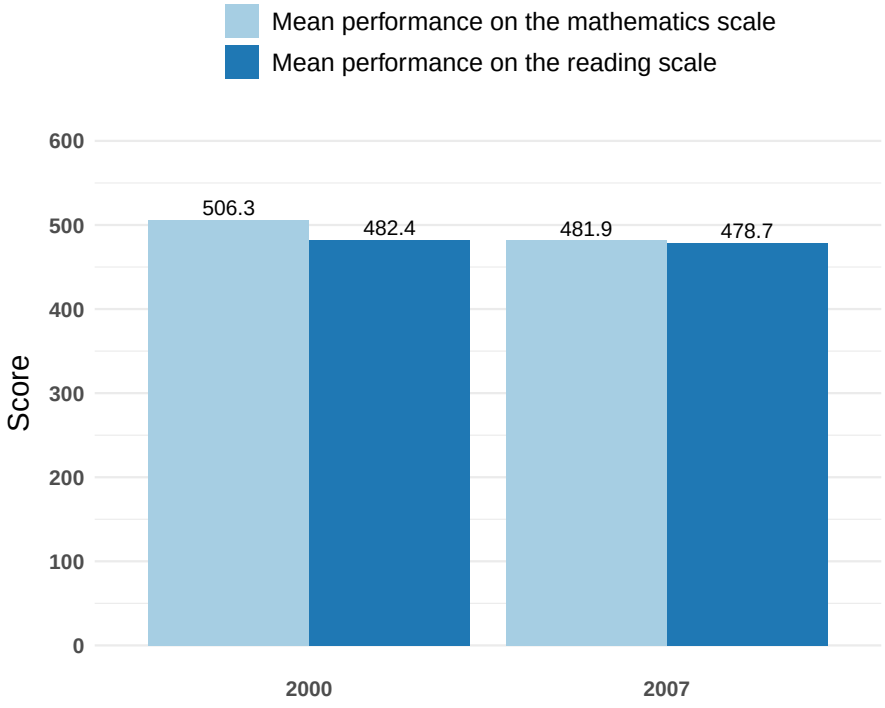
Source: World Bank Education Statistics (EdStats) (Accessed 11.04.2018)

The second plot of Figure (11) shows that the proportion of out-of-school youth of upper secondary age (17-19) has declined, suggesting that access to upper secondary education has also increased in recent years. The effective transition from primary to lower secondary education visualized in the third plot has increased but also shows a Kuznets-shape, with a peak in 2010 and decline thereafter. The decline is at odds with the early projections shown in Figure (5) and does not mingle well with a government policy aimed at increasing access since 2007. The remaining graphs of Figure (11) display Uganda specific trends. The 4th plot shows that the dropout rate shows a u-shaped trend, which is in line with the trends in school life expectancy in Figure (10) and effective transition rate. The 5th plot shows the trend in school life expectancy, but only for secondary schools. It is a steady upward trend, with a slight convexity in the 90's and slight concavity after 2000. This seems to shed some light on the U-shape puzzle: Since secondary life expectancy has been steadily increasing, the decline in Figure (10) has to be accounted for by a decrease in primary life expectancy, which is also what the effective transition rate from primary to lower secondary in Figure (11) suggests. The increase in the cumulative drop-out rate after 2005 could also well be accounted for by more pupils dropping out of primary school, lending to the conclusion that except for increasing the pupil-teacher ratio in secondary education, which is attributable to access-increasing government policies following 2007, the reforms did not have that large an impact on the quality of secondary education than originally inferred.

With further regards to secondary life expectancy, it is conspicuous how it increased from 0.6 years in 1990 to 1.7 years by 2013. The latter, however, is still very low when compared to the theoretical duration of 6 years for a completed secondary education, lending to the conclusion that USE still has a long way to go in Uganda.

A final set of statistics covers standardized tests that purely focus on educational quality. Uganda has not yet been part of any standardized international TIMSS or PIRLS assessment but has partaken in the assessments of The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) in 2000 (SACMEQ II) and 2007 (SACMEQ III). Figure (12) shows the mean performance in both years[5], a detailed breakdown of the scores by region, gender and income can be consulted here. Since the World Bank notes that scores might not be comparable across years, and I am unable to find more information about how the scores are obtained within a reasonable amount of effort, I do not interpret these test results any further.

Figure 12: SACMEQ Reading and Writing Scores, 2000 and 2007

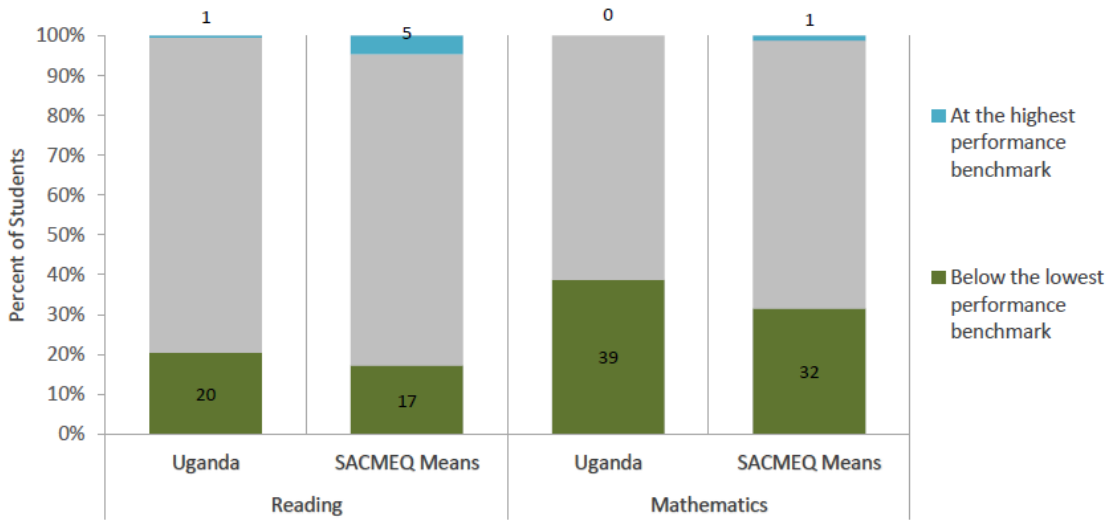


Source: World Bank Education Statistics (EdStats) (April 2018)

Figure (13) is more informative as it shows Uganda's performance on the SACMEQIII test in math and reading vis a vis other southern and eastern African countries. Uganda performs worse than average in both disciplines, with a larger tail at the low end of the performance distribution and a smaller tail in the high end of the performance distribution, in both reading and math.

[5]Mean performances are the scores for 6th-grade students. Mean scores are on SACMEQ scales for mathematics and reading, which have averages of 500 and standard deviations of 100. Data reflects country performance in the stated year according to SACMEQ, but may not be comparable across years or countries. Consult the SACMEQ website for more detailed information: http://www.sacmeq.org/. Taken from (World-Bank, 2018a).

Figure 13: Student Performance on SACMEQIII Learning Assessment (2007)



Data source: EPDC extraction of SACMEQ dataset 2007

Source: Education Policy and Data Center (2014)

3 Policy Trends and Priorities

The Ugandan Ministry of Education and Sports (MoES)recently recalibrated its policy priorities for the education sector in 2017 with the new ESSP 2017-2020 (MoES, 2017). The plan's three main strategic objectives are to:

(i) Achieve equitable access to relevant and quality education and training;

(ii) Ensure delivery of relevant and quality education and training; and

(iii) Enhance efficiency and effectiveness of education and sports service delivery at all levels.

These objectives differ markedly from those of the ESSP 2007-2015 which focussed predominantly on advancing primary education. The policies of the new ESSP continue to address remaining shortcomings in UPE, but focus more heavily on achieving USE, and promoting the BTVET and tertiary subsectors.

To achieve these strategic objectives, the Ministry plans to implement a number of priority interventions, the ones relevant to the secondary subsector are listed below:

(i) Achieve equitable access to relevant and quality education and training;

- A government secondary school per sub-county
- More classrooms in existing schools, target of 50:1 pupil classroom ratio
- Construction of teachers houses, to improve teacher welfare and motivation
- Programs enhancing participation of disadvantaged persons in primary, secondary and BTVET
- Lower costs to families: Improve the implementation of UPE, USE
- Formulate and implement a policy to rationalize levying of fees in public and private schools

(ii) Ensure delivery of relevant and quality education and training

- Strengthening the current inspection system, increase the frequency of inspection of schools
- Learning Assessment System that integrates development of skills and competences

- Develop a basic Competence and Skills Profile for each education level, in line with national aspirations for socio-economic transformation and labour market needs
- Improve teacher and instructor competencies (content, knowledge, skills and pedagogy)
- Recruit primary and secondary school teachers to meet standards of pupil-to-teacher ratios
- Recruit more Mathematics and Science teachers for secondary schools
- A National Teacher Policy: Establishment of a Teachers Regulatory Council (akin to other professional councils such as the Uganda Dental and Medical Practitioners Council), a National Institute of Teacher Education, continuous professional, and mandatory teaching accreditation in the professional development
- Reduce pupil-textbook ratios from 6:1 to 3:1 for primary and 3:1 to 1:1 for secondary schools
- Provide free scholastic materials such as mathematical geometry sets, exercise books, pens and pencils to pupils and students at primary and secondary level
- Revise Capitation Grants: Adopt a differentiated formula for allocation of Capitation Grants with the goal of increasing them so as to improve the quality of education

(iii) Enhance efficiency and effectiveness of education and sports service delivery at all levels.

- Establish a semi-autonomous body for inspection of education and training institutions
- Promote e-learning and computer literacy in secondary and tertiary education
- Policy for regulation of private providers: Development and implementation of a policy for regulation of private provision of pre-primary, primary and post primary education by non-state actors for efficient and effective delivery of education and sports services
- Rationalize and improve teacher recruitment, deployment and payroll management at primary and post-primary education levels.

To support the above listed strategic objectives and priority interventions and, the government set itself a set of quantitative targets. The ones pertaining to secondary education are listed below, where the baseline (from) refers to the 2015 indicator value and the endline (to) is the 2020 target.

Main Development Targets, ESSP 2017-2020

- Net Enrolment Ratio – Secondary: 21.8% in 2015 to 30.3% in 2020
- Secondary GPI from 0.9 to 1
- Transition rate to S1 from 63.2% to 70.5%
- Transition rate to S5 from 25% to 41.8%
- Student proficiency rate: Biology: 20.4% to 27.4%, Math: 41.8% to 43.5%, English: 50.3% to 55.3%
- Student-textbook ratio: 3:1 to 1:1
- S4 completion rate: 36.2% to 40.6%
- Perentage of secondary schools inspected (2 visits per term): 80% to 100%

3.1 Key Issues

Against the backdrop if this government agenda through 2020, a couple of key issues warrant further contextualization, which is briefly done in this section.

(1) Primary to Secondary Transition

Since its inception in 1997, the effects of UPE were dramatic. The primary student population rose from 3.1 million in 1996 to 5.2 million in 1997, an increase of 68% (UIA, 2010). At present, the primary net enrolment rate is near 100%. Although USE was announced in 2007, progress to absorb the large number of primary leavers has so far been limited. Between 2005 and 2010, secondary schools witnessed an unprecedented yearly enrolment growth of over 19% in some years, and total enrolment in secondary schools increased from 728,393 in 2005 to 1,194,454 at end of 2009 and 1,284,008 in 2015, and during the academic year 2010, 519,246 pupils who sat PLE examinations were competing for about 300,000 places available at S1 (UIA, 2010; MoES, 2017). A similar bottleneck applies to the advanced level where the number that graduates at senior four (264,000 sitting the UCE in 2010) is more than what the S5 intakes can handle (only around 100,000 sat the UACE in 2010). There has been considerable action on behalf of the government and private schools to cope with increasing enrolment: In 2006, there were 42,673 teachers with various qualifications but by end of 2009, the number had shot up to 65,045 teachers, an increase of almost 35% in a spate of 3 years (UIA, 2010). In 2010, the number of secondary schools had reached 3,149, an increase by 38% in three years from around 2250 in 2007, which reflects government engagement in USE, but predominantly the heightened level of private-sector expansion at this level (UIA, 2010). The ESSP 2007-2015 remarks that only about 14% of the increase in secondary school enrollment is traceable to policies aimed at expanding participation in secondary schools (MoES, 2008). During the ESSP 2007-2015, considerable efforts have been made to increase rural access and decrease gender disparities (the latter being ostensibly successful as Figure (11) shows), and a declared though unmet goal of the plan was to assure that all pupils successfully completing Primary 7 would have access to either academic secondary education or BTVET (MoES, 2008). The plan has also aimed at reducing the secondary curriculum (of typically around 18 subjects), and improving managerial efficiency through decentralization measures amongst other things, to better deal with the growing number of students. The ESSP 2017-20 continues the efforts of the ESSP 2007-15 in expanding existing schools and grant-aiding community schools, with the target of at least one government secondary school per sub-county (MoES, 2017). The largest part of education service delivery will continue to be undertaken by local governments in line with the decentralization policy.

Overall, the transition rate from P7 to S1 increased by 12.3% from 50.9% to 63.2% in the 2007 - 2015 period. The Ministry aims to raise it to 70.5% by 2020. Yet, in spite of the strides of the sector in increasing access by recruitment of required teachers, provision of materials and establishment of new schools, the quality of education remains an issue at both primary and secondary levels (MoES, 2017). The Present ESSP hence mostly focuses on increasing the quality of both primary and secondary education through deployment of significant additional human and material resources.

(2) Completion of Secondary School

Next to access which is a key issue, completion of (or survival rate in) secondary school is another towering obstacle towards achieving USE. One of the reasons for the low completion rate of 36.2% in 2015, reflecting a secondary school life-expectancy of around 2 years, is poverty (MoES, 2017). Parents still pay the bulk of secondary expenses, which are much higher than any expenses that typically incur for primary education. In addition to paying school fees, which under the current USE scheme are only completely alleviated for high performing students in government schools, many parents struggle to afford pens, exercise books, school uniforms and even school lunches for their children (RTF, 2015).

The ESSP 2007-2015 tried to raise the secondary completion rate to senior 4, which was initially at 35% in 2007, through different measures such as among other things:

- provision of Capitation Grants
- Increased access which reduced travel distances
- Increased monitoring and supervision
- Increased community participation through Boards of Governors (BoG's) & Parent-Teacher Associations (PTA's)

- Gender responsive policies improving the survival of girls, and
- Reduced Pupil-Classroom Ratio (PCR) to create a more favorable study environment.

Policy under the ESSP 2017-2020 continues with the implementation of these measures and aims to raise the S4 completion rate to 40.6% by the end of 2020. This is not a very ambitious target, and in spite of pledging in its policy priorities to lower the cost of secondary education to families through more efficient resource utilization, provision of scholastic materials and lower at public and private schools, the plan provides very few details on these policies and the overwhelming presence of quality measures in the above list of policy priorities suggests that the government has not yet devoted due attention to this large and lingering hidden access issue.

(3) Quality and Relevance of Secondary Education

The Education for All Global Monitoring Report (2013-14) states that Uganda needs to speed up the expansion of its teacher force. In order to meet current demand, Uganda would need to expand recruitment by 6%, compared with a current average increase of 3% per year (UNESCO, 2014). Such teacher shortages imply larger class sizes and adversely impact the quality of education delivered. Teacher salaries are low: Teachers are the lowest paid civil servants in the country, yielding a lack of incentive for teachers to attend school and to commit to lessons. Absenteeism is therefore a major issue and exacerbates the problem of teacher shortages: The World Bank Service Delivery Indicators (SDI) (which are based on random visits to primary schools) showed that roughly 1 in 4 primary school teachers were absent from school, and of those present in school, 1 in 3 were not teaching. As a result, 40% of public primary school classrooms did not have a teacher teaching in them when the data were collected. In the survey we estimate the equivalent figure to lie between 5-10% for secondary schools, based on the Director of Studies, the true figure might be higher. Availability of school textbooks is also poor, particularly in rural areas (RTF, 2015).

Enhancing the quality and relevance of secondary education, together with increased resource utilization and managerial efficiency, is the heart of the new ESSP. This is of due course since in the 2007-2015 period the pass Rate at O-level dropped from 95.3% to 91%, and performance indexes at O and A-level indicate a decline in performance of 14.2% and 11.7% percentage points respectively (MoES, 2017). The ESSP 2017-20 asserts that the drop in performance is a consequence, predominantly, of reduced levels of proficiency in English and Mathematics. These, in turn, it attributes to inadequately trained teachers, especially in Science, Mathematics, and English, especially in rural schools, and ineffective utilization of instructional materials. Key targets of the new ESSP are to increase quality by lowering the student-teacher ratio, and improving the performance of secondary pupils in key subjects (as shown in the development targets above) through enhanced teacher training and recruitment of more teachers in Science and English, and by lowering the student-textbook ratio from 3:1 to 1:1. Some measures like the provision of instructional materials, more in-service support to teachers, and supply of housing for teachers and head-teachers in rural areas are continued from the MoES (2008). The previous ESSP also identified teacher training colleges lack of adequate instructional materials, infrastructure and facilities in-terms of laboratories, libraries and ICT facilities, combined with a shortfall in college graduates and low esteem of the teaching profession as a whole as causes of low performance in the sector. Together with improvements in teacher training facilities, the new ESSP adds increased national assessment and welfare gains for teachers to the pool of measures aimed at improving the quality of teaching and learning and attracting bright students to the teaching profession.

Aside from general quality concerns, the previous ESSP also noted that students are not acquiring the skills and knowledge they need for either the world of work or further education, and that the bifurcated system between academic preparation for higher education and vocational training for technician jobs is not appropriate for Uganda's national development needs (MoES, 2008). Even during the 2007-2015 period, only a minority of students achieved what is expected, and many left leaving school without the knowledge and skills they need to participate as citizens and productive workers (MoES, 2017). The previous ESSP already introduced a set of quality and efficiency increasing measures which are unaltered in the current policy statement and include amongst other things:

- Introduction of the eighteen subject curriculum from the former 42 subject curriculum with a school menu of ten subjects
- Reform of teacher deployment and utilization practices
- Introduction of double-shift teaching in over enrolled schools
- Provision of an initial stock of textbooks and instructional materials
- Provision of support for USE students in non-government schools
- Construction of new facilities in National Teachers Colleges
- Reducing students' course load to five or six per term
- Consolidating subject matter into fewer courses
- Introducing a nationally approved limited list of core textbooks as opposed to school-based lists
- Redeploying teachers from schools where there is a surplus to those where there are deficiencies
- Requiring all teachers to have the skills to teach at least two subjects so as to enhance efficiency
- Building small "seed" schools in each sub-county where there is none or where the need is great
- Rehabilitating existing facilities and expanding them to accommodate growing numbers
- Establish "centre of excellence" in each district - a school with full libraries and facilities for teaching Science and ICT

The new ESSP adds to this list of measures improved teacher training, teacher quality standards through a National Teacher Policy, increased monitoring, free scholastic materials, and enhanced science, ICT and e-learning facilities. The government furthermore pledges to improve the management capacity at all levels, promote community engagement, increase and improve infrastructure and ensure better resource utilization (MoES, 2017). It also mentions a number of lessons learned during the ESSP 2007-2015 period regarding educational quality that will be given more attention under the new plan:

- The presence of strong PTAs is very important in improving school performance
- Establishment of educational institutions without carrying out proper needs assessment may lead to creation of redundant or underutilized facilities
- The shortage of female teachers in the school system presents serious challenges to retention of girls due to the absence of role models

In short, the new ESSP presents a very ambitious and comprehensive programme to combat quality issues through 2020. It remains to be seen how many of these measures can eventually be implemented at sufficient scale in the coming years.

(4) The Rural-Urban, Private-Public Divide

Regional imbalances, especially between the North Eastern region and the rest of the country have persisted for years, with the highest inaccessibility to USE being concentrated in The North Eastern region with a proportion of less than 2% of the whole USE population (MoES, 2015). This shortfall is attributable largely to its terrain, prevailing negative attitudes towards education, and the nomadic nature of the people living there coupled with the cattle rustling problem that has traditionally plagued the Karamoja and surrounding sub-regions (MoES, 2015). With the cessation of hostilities in the region and heightened government and humanitarian intervention, enrolment

numbers in the 19 USE schools in the area are increasing though. Next to negative attitudes, domestic obligations and lack of funds, early pregnancy is another reason for the high drop out of the female students in rural areas, especially in the Northern and North Eastern regions (MoES, 2015).

The ESSP 2007-2015 pledged to improve equity in the participation of girls and needy students in rural areas by equipping and improving post-primary school facilities, and improving equity in participation of girls and needy students by targeting grants to schools in needy areas and bursaries to individual students, especially in the civil war-ravaged area of Northern Uganda (MoES, 2008). Under the 2017 ESSP the government also pledged to tailor schools to the communities they serve, whence rural schools may alter the calendar to meet farming seasons, and they should be able to accommodate students who drop out and return. This is facilitated through community "seed" schools, which are small and may offer multi-grade courses, while urban schools are larger and may use double-shifts to make efficient use of facilities (MoES, 2008). As mentioned before, the government also aimed at establishing centers of excellence in all districts to attract diligent students.

Commenting on rural access, the Rafiki-Thabo Foundation notes that the availability of school textbooks particularly poor, and retention rates are also lower in rural regions of Uganda, like Kabale (RTF, 2015). The quality of education is on average tangibly higher in Kampala.

The large role of the private sector in providing secondary education is also a particularity of the Ugandan system that warrants special policy consideration. As shown below in Figure (15), 62% of secondary schools in Uganda are currently privately run. Generally, the secondary sub-sector has three types of schools (Government-owned, private sector-owned and community-owned), next to a very small number of international schools that deliver foreign curricula. However, although the majority of existing schools (over 80%) are either privately or community-owned, they cater for a smaller proportion of less than 50% of school enrollments (MoES, 2008). The participation of the private sector is also more limited in rural areas, and government schools are yet scarce in these areas.

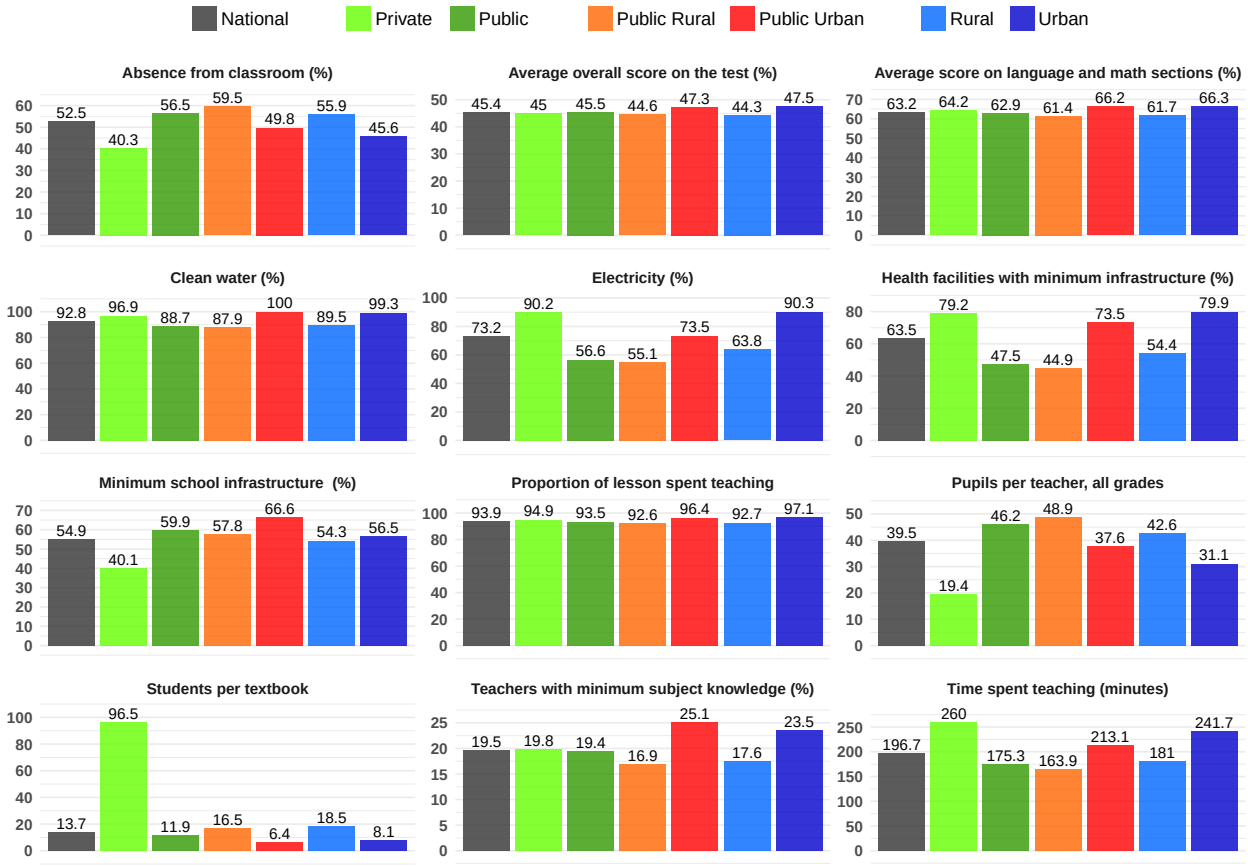
The private sector expansion in Ugandan secondary education over the recent years can be explained by a number of factors (MoES, 2017; UIA, 2010):

- Insufficient government secondary schools to accomodate cohorts of UPE graduates in the early 2000's
- Relatively expensive governmet secondary shools, even under USE scheme which only alleviates school fees for bright students
- Under Value Added Tax Statute (1996), education services are treated as an exempt supply. This means that education services are not subject to VAT
- Education materials such as textbooks and laboratory equipment are zero-rated, which means that investors in the education sector can claim for a refund from Government of any VAT that they pay on inputs (items purchased as education materials)
- To encourage interest in ICT and computer literacy, the government has removed all forms of taxes on computers to make them affordable to users in the country

In 2010, the Ministry of Education and Sports indicated that private investment was highest in secondary education. While currently (2017) 62% of secondary schools are privately run, already in 2009 the private sector owned about 58% of the secondary schools, and 47% in 2006. The UIA estimates that entrepreneurs will continue to dominate this sector for the foreseeable future as long as the policies continue to encourage private sector investment and USE remains far from realized (UIA, 2010). The high fraction of private secondary schools necessary limits the governments possibilities in refashioning the secondary education sector according to its desires (exempting curriculum guidelines where the government still seems to enjoy uncontested authority as most private schools simply follow the national curriculum. In the survey data only 4% of schools added own elements to the curriculum and 52% of schools in the survey are purely private).

To compare the delivery of educational services by private and public schools, in rural and urban areas, Figure (14) shows an excerpt from the World Bank Service Delivery Indicators (SDI) (World-Bank, 2018b). The indicators were last collected in 2013, and are unfortunately only based on random visits to primary schools and healthcare center. The relative performances in secondary schools must thus be cautiously interpolated from these indicators, and from survey estimates in Figure (36)

Figure 14: Service Delivery 2013



Source: World Bank Service Delivery Indicators, 2013

Figure (14) shows that private schools tend to outperform government schools in terms of facilities such as clean water, electricity and health facilities, but hardly provide any textbooks to students and lack overall infrastructure (such as libraries etc.). They however spend the most hours teaching, have the lowest pupil-teacher ratio, and the lowest ratio of student absenteeism (first plot). Rural schools perform worse on all counts, especially on student absenteeism, electricity and health infrastructure, pupils per teacher, hours taught, and teachers with minimum subject knowledge. Surprisingly, despite all these differences, the average test performances of private, public, rural and urban primary schools don't differ markedly. Private and public schools have quite equal test performances on average, and rural schools perform slightly worse, especially in Math and English.

4 Education Finance

Table (5) shows the governments proposed GDP allocation ratio to education (across sectors) (MoES, 2017). The figure for 2016/17 should read 2.46% and corresponds to the realized GDP share allocated to education in 2016/17. The government plans to increase the share of GDP in education by 0.5% through 2020, which amounts to an extra amount of around 350 million in 2011 PPP $.

Table 5: Proposed GDP Allocation Ratio 2016-2020

FY	FY2016/17	FY2017/18	FY2018/19	FY2019/20	FY2020/21
% GDP (excluding donor)	246%	2.6%	2.74%	2.88%	3.02%

Source: Education and Sports Sector Strategic Plan 2017-2020 (MoES, 2017)

Table (6) shows the share of the governments budget in education, 2012 through 2017. The share has declined from 14.6% to 12% during this period, although absolute expenditure has increased[6]. Table (7) displays the sub-sectoral budget shares in the governments expenditure on

Table 6: Education Expenditure as Share of National Budget 2012-2016

FY	2012/13	2013/14	2014/15	2015/16	2016/17
% Share	14.6	13.5	13.4	11.8	11.08
Amount allocated (Bn. Shs.)	1,687.63	1,868.06	2,275.45	2,009.61	2,634.31
Difference in amount allocated (Bn) in comparison with previous FY	210.68	180.43	407.39	-265.84	624.70

Source: Education and Sports Sector Strategic Plan 2017-2020 (MoES, 2017)

education, 2015-2018. The largest share remains in primary education, steadily eating up around 50% of the annual budget. The share of secondary education has been at 18% in 2015/16, but dropped to 16% in 2017/18. Primary spending also dropped by 4.5% during this timeframe. Expansions are noticeable especially in tertiary education, whose budget share has increased by 4.4% from 14.5% to 18.9%, making it the largest subsector after primary education in spending terms, and in Business, Technical, Vocational Education and Training (BTVET), whose budget share increased by around 2% from 9.4% to 11.3%.

Table 7: Sub-Sectoral Shares of Government Expenditures on Education 2015-2018

Subsector	FY 2015-16(Ushs.bn)	% share	FY 2016-17 (Ushs.bn)	% share	FY 2017-18 (Ushs.bn)	% share
Primary	1,094.63	53.95%	1,226.66	49.97%	1,224.22	49.47%
Secondary	363.63	17.92%	373.41	15.21%	396.06	16.01%
BTVET	190.16	9.37%	258.09	10.51%	279.98	11.31%
Tertiary	293.29	14.45%	479.96	19.55%	466.72	18.86%
Sports	11.10	0.59%	12.213	0.49%	12.213	0.49%
Others	75.28	3.7%	97.34	3.9%	107.51	4.34%
Total	**2,029.08**	**100.00%**	**2,454.61**	**100.00%**	**2,474.49**	**100.00%**

Source: Education and Sports Sector Strategic Plan 2017-2020 (MoES, 2017)

[6]When considering absolute amounts in Uganda shilling (UGX), one has to consider that Uganda had an inflation rate of 25% in 2012, which dropped to 5% in 2013 and oscillates around 5% since then.

Table (8) projects the ministry of education's recommended sub-sectoral budget shared through 2020[7]. It provides for a simple continuation of the status quo in 2017/18.

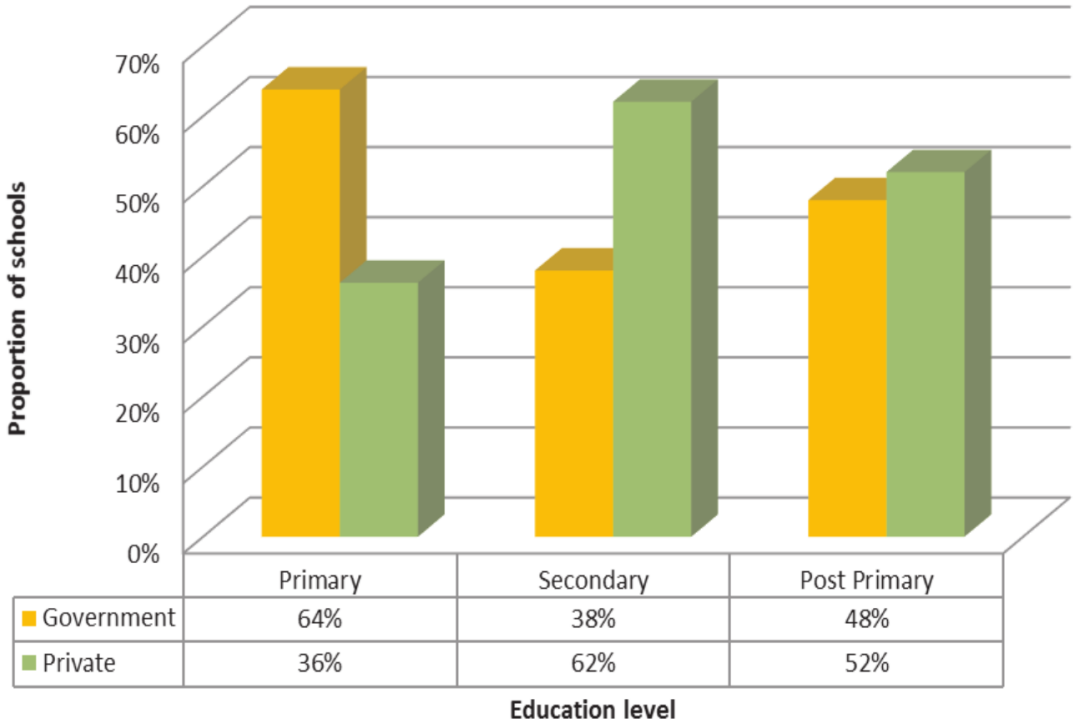
Table 8: Recommended Sub-Sectoral Budget Shares 2018-2020

Sub-sector	FY 2015-16	FY 2016-2017	FY 2017-2018	FY 2018-2019	FY 2019-2020
Primary	53.95%	49.97%	49.47%	49.47%	49.47%
Secondary	17.92%	15.21%	16.01%	16.01%	16.01%
BTVET	9.37%	10.51%	11.31%	11.31%	11.31%
Tertiary	14.45%	19.55%	18.86%	18.86%	18.86%

Source: Education and Sports Sector Strategic Plan 2017-2020 (MoES, 2017)

Figure (15) finally shows the share of schools by education and funding source. At primary level 2/3 of schools are government schools, but at the secondary level, this figure is reversed, with 62% of secondary schools being privately run. Due to bad learning conditions in many government schools, the sparse geographic coverage of public schools, and because secondary education has traditionally catered more to the wealthier sectors of societies, the educational landscape in Uganda has provided, and continues to provide, lucrative opportunities for private educational entrepreneurs (EPDC, 2011).

Figure 15: Schools By Funding Source in 2016



Source: Education and Sports Sector Strategic Plan 2017-2020 (MoES, 2017)

[7]This is the ministry of educations recommendation to the ministry of finance, thus it is not set in stone.

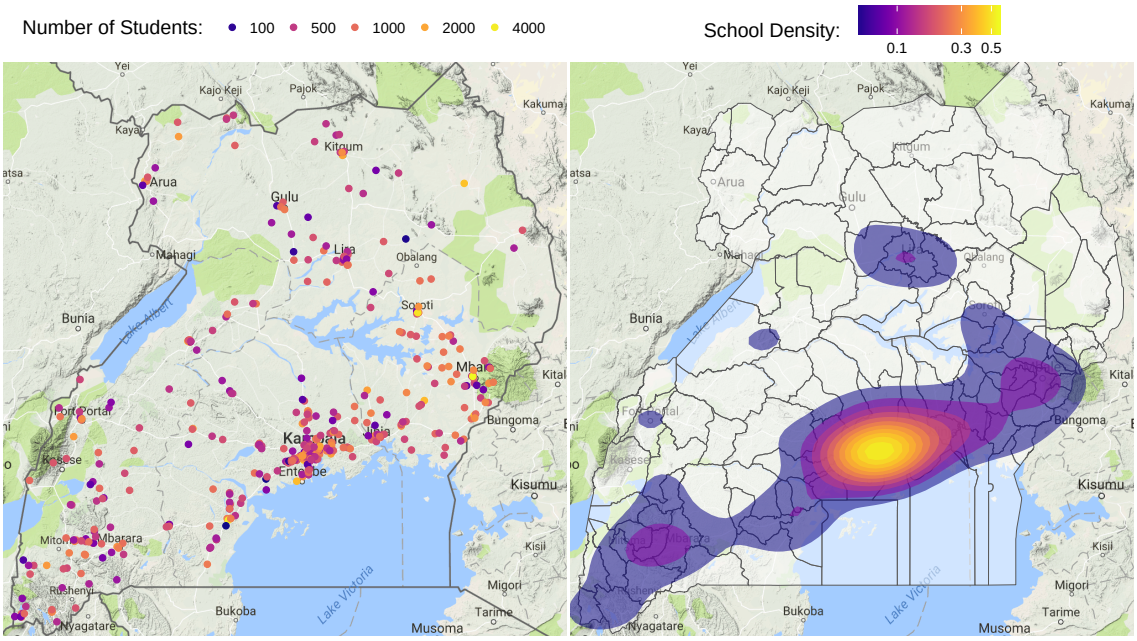
5 Data

This section presents the data collected and used in this study. All figures, tables, and pieces of statistical information that follow are based on this data (BRAC (2018) and authors computations) unless indicated otherwise, and are therefore not referenced.

5.1 Data Collection

Between 2nd of February and the 13th of March 2018, around 20 BRAC enumerators were sent to 450 secondary schools across the country to collect general data on school quality. The schools were randomly selected from a list of schools of applicants to the BRAC & Mastercard Foundation secondary scholarship, which is given out annually since 2014 and has yearly application rates of more than 10.000 students. The geographical distribution and approximate size of the schools in the sample are shown in Figure (16). With 92 schools, the Kampala area has the largest concentration of schools in the sample.

Figure 16: Geography and Size of Schools in the Sample



Notes: Green areas represent national parks and are very scarcely populated. The plot on the right shows the 112 districts that Uganda had in 2014 (meanwhile it has 121 districts). The left plot also shows the countries 4 administrative regions (light grey demarcation line, "Western", "Eastern", "Northern" and "Central" region).

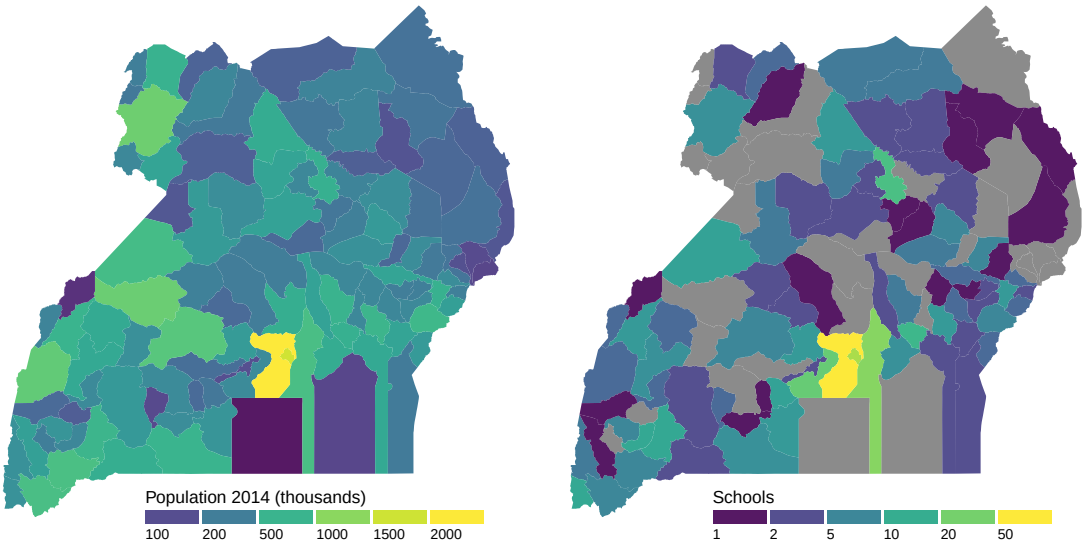
5.2 Representativeness of the Sample

Figure (17) shows that the sample of schools approximately reflects the underlying population density in Uganda, which is scarce in the north and denser around the lake and at the border to Rwanda. The correlation coefficient for the population in 2014 and the number of schools sampled per district is 0.8[8]. Figure (18) compares the sampled schools on key UCE test statistics with a pool of 3000 schools whose test statistics are published up to 2016 by the Uganda National Examinations Board. 394 of the sampled schools could be matched in this dataset, and are represented by the blue density in Figure (18). The UNEB does not publish aggregate UCE scores, but the fraction of students scoring in each of 10 divisions. The metrics employed for sample comparison in Figure (18) are the % of students scoring in the first (best) division of the test (a UCE score below 12), the mean division scored[9], the % of candidates scoring in the last (worst) division and the total number of students examined. The UNEB dataset also provides

[8]For districts with no schools in the sample (grey areas in Figure (17)), missing values were replaced by '0' before computing the correlation coefficient, otherwise, the correlation coefficient would be 0.82.

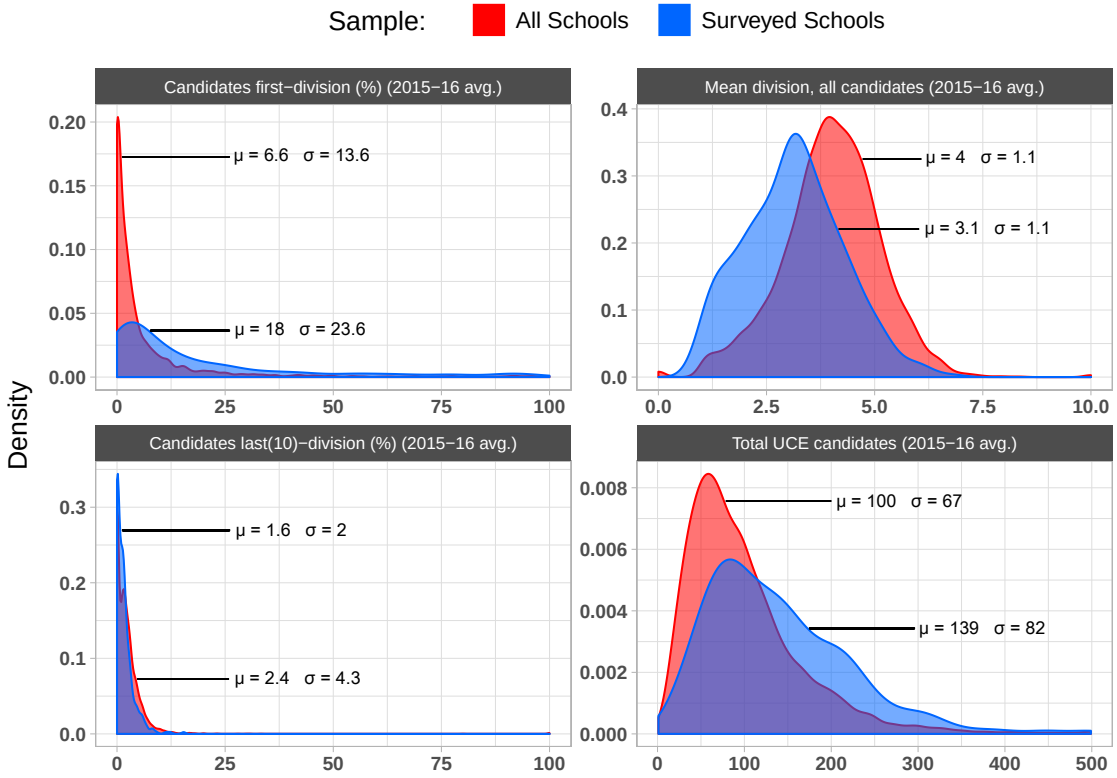
[9]Which is simple to compute as the sum of the divisions weighted by the proportion of students scoring in them.

Figure 17: School Concentration and Population (2014) by District, $r = 0.80$



gender-disaggregated statistics, but it turns out that the sampled schools do not differ much from the aggregate sample in terms of gender equality, and the average GPI is in any case close to 1 in both sets. Figure (18) shows that the sampled schools differ slightly from the aggregate: Scholarship applicants typically come from schools with more excellent candidates, as shown in the top left plot. The top right plot further shows that pupils from scholarship applicant schools generally score about one division better than the average Ugandan secondary school.

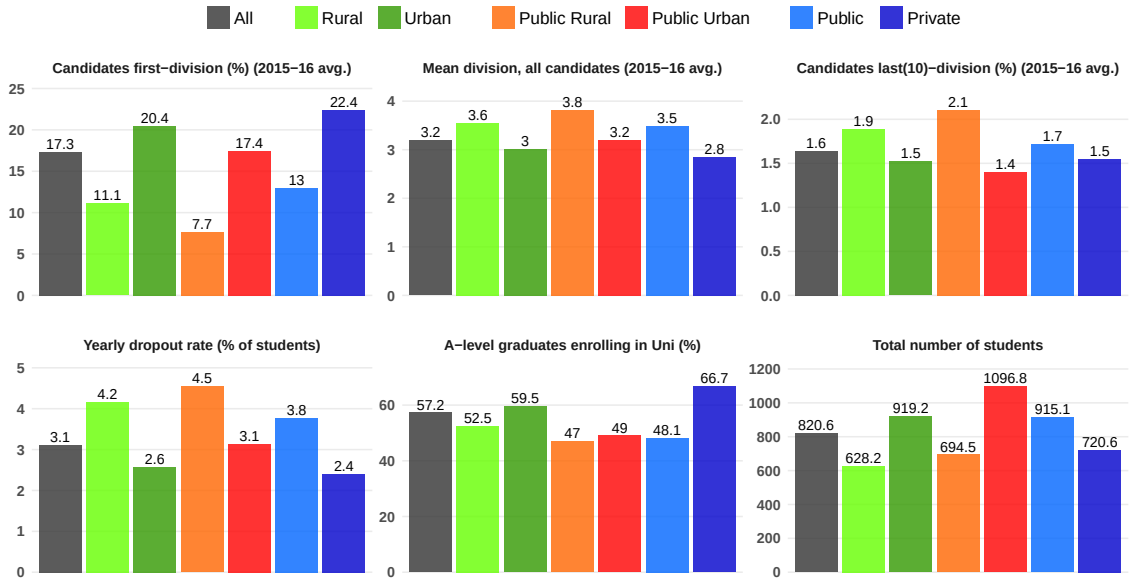
Figure 18: Representativeness of Sample



The bottom left plot, however, reveals that at the lower end of the grade-distribution this difference is much less pronounced. The bottom-right plot shows that scholarship-applicant schools are larger than the average school. The latter suggests that, presumably due to the limited advertising capabilities of BRAC and Mastercard Foundation[10], the sample is biased towards some larger and higher-performing urban and suburban schools.

The survey confirms this hypothesis: of the 450 surveyed schools, 148 respondents described the school's location as rural, 203 described it as suburban, and 99 as urban. It is also evident in the survey data that more urban and suburban schools are private: 57% of urban and suburban schools in the sample are fully private compared to only 31% of rural schools in the sample. Figure (19) provides a disaggregated view of the sampled schools. The top 3 plots are based on UNEB data[11], and the bottom 3 are based on the survey.

Figure 19: A Disaggregated View of Sampled Schools



The sample is thus, at the district level, roughly reflective of the population density, but includes a disproportional share of urban and suburban schools, which on average are larger and perform better than their rural counterparts. This bias could certainly be corrected with probability weights, but due to the small sample size, the rather complex nature of the statistical analysis that follows, and because such weights would need to be computed from the UNEB data which only features UCE performance statistics and also constitutes an imperfect representation of the population, probability weights are not implemented. It remains therefore merely to emphasize that the sample and the results produced are slightly urban-biased, and it is possible that at the rural level the ordering of the most important factors for improving secondary education is slightly shifted.

5.3 The Survey

The survey used to collect the data is provided in Table (26) in the Appendix. Due to space constraints, the choices for the multiple choice questions (the ones where the type starts with 'select_one' or 'select_multiple') were omitted. The survey has around 200 questions used to broadly capture features of the school and educational outcomes. Questions are grouped into 6 broad categories:

[10]The scholarship was mainly advertised in the national newspapers and some larger schools

[11]Where 394 of the 450 sampled schools could be matched, the top plots in Figure (19) thus represent disaggregations of the information already presented in the blue densities in Figure (18).

Table 9: Genealogy of Questions

Prefix	*Label*
Gen	General Questions about the School
Eq	Equipment / Inventory Questions
Eff	School Effectiveness
Ta	Teaching Approach
St	Student Body
Teach	Body of Teachers
other prefixes	Survey Administration

The preferred survey respondent was either the Head Teacher or the Director of Studies (DOS), and, for the most part, the enumerators were successful in locating and questioning these people (there are 68 Head Teachers, 86 Deputy Head Teachers, and 240 Directors of Studies in the cleaned sample of 422 schools). Following data collection, the survey was cleaned. Around 10 respondents refused to answer the survey in spite of a data confidentiality note at the beginning. Other frequent problems with the survey were the misunderstanding of some questions or entry of information in a unit other than the unit asked (e.g. some enumerators entered "1" for the lunch break meaning hours, whereas "60" minutes was asked). The design-aims were to make the survey as accessible as possible, and towards that end as many integrity constraints on questions as considered suitable were put in place. These measures could nevertheless not prevent minor data inconsistencies (e.g. between the shares of funding sources which did not add up to 100% in some cases). Some respondents (usually teachers answering when the Head Teacher or DOS was not available) could not answer a large number of questions. These observations, therefore, had to be dropped together with the refusals, so that of around 450 incoming surveys a final sample size of 422 is retained. A visual summary of the cleaned survey[12] is shown in Figure (36) in the Appendix.

5.4 Data Preparation

With brevity and parsimony guiding the design, most survey questions are only partially suitable or interesting for quantitative analysis. Therefore, a large part of the responses is recomputed into economically interesting variables for the analysis to follow (e.g. the survey asks for the number of students and the number of teachers, and the student-teacher-ratio is computed in post). Most survey questions ask about features of the school and its educational approach that will serve as predictor variables in the analysis. Some questions, however, address the schools performance in the last year and are recomputed to yield outcome measures. The latter are complemented by official UCE test-score data available online.

5.4.1 Predictor Space

Recomputing the questions into relevant measures yields 121 predictor variables. Two further problems surface at this point: (1) 1.3% of the data in the 422×121 predictor matrix, denoted $\mathbf{X}$, is missing, amounting to only 144 complete cases; (2) The dimensionality of $\mathbf{X}$ is too high to conduct meaningful quantitative analysis, and some measures are highly correlated (32 pairwise correlations are > 0.7, some close to unity). To both problems, technical solutions are sought. For (1) the 1.3% missing data is imputed using a nonparametric missing-value imputation algorithm for mixed-type data based on Random Forests, developed by Stekhoven (2015). Problem (2) is solved by performing factor analysis on variables measuring similar constructs in the imputed dataset, and obtaining factor scores using the regression scoring method. The overall factorability of $\mathbf{X}$, as measured by the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, is 0.73. The factorability varies greatly for different groups of variables. For the equipment variables, the factorability is moderate, with a KMO of 0.7. The large number of equipment variables is thus partialled up into various categories (e.g. technology or library etc.) for which separate factor scores are obtained. For student-poverty questions, on the other hand, factorability is high, with a

[12] It's a vectorized figure, the idea is to zoom-in in order to read it and have it side-by-side with Table (26). The variable names correspond to column 2 in Table (26).

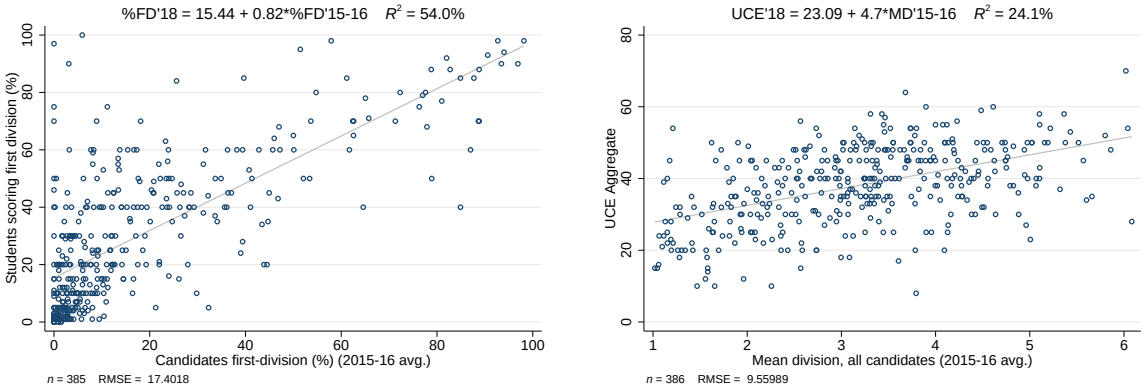
KMO of 0.86. Proceeding in this manner (i.e. factoring similar variables with the aim of obtaining a single representative factor score) for all the different groups of questions, the dimensionality of $\mathbf{X}$ is reduced from 422×121 to 422×55. All indexes obtained by factor scoring are designated as such and rescaled so that the 5'th percentile takes the value 1 and the 95'th percentile takes the value 9[13]. Dimensionality reduction also helped curtail the multicollinearity problem. In the reduced data matrix, of the $55^2 - 55 = 2970$ pairwise correlations, only 10 have a Pearsons r above 0.5, the maximum being 0.66. The Appendix provides summary statistics (Table 24), a correlation matrix (Figure 22), a cluster dendrogram (Figure 23) and Principal Components Analysis (PCA) (Table 25) for the dimension reduced predictor matrix. Histograms of the entire analysis dataset are shown in Figure (35).

5.4.2 Outcome Measures

The survey also asks several questions related to the performance of the school. The respondent was asked the average UCE and UACE test scores and the fraction of students scoring first-division in the school. He was also asked about the number of dropouts and grade-repeaters, and about differences in performance and dropout rates between male and female, rich and poor students. For A-level students, additional questions asked what most graduates are pursuing upon graduation. Since these answers might be biased in favor of the school, especially if the school is low performing, publicly available UCE performance data administered by the Uganda National Examinations Board (UNEB), recording the performance of more than 3,000 schools over the years 2011-2016 is also consulted. It was possible to match 394 of the 422 schools in the cleaned sample in this dataset. The data records performance as the percentage of candidates scoring in each UCE grading level. The gender breakdown is also indicated. The proximity of the survey outcome measures to this official data is assessed by computing 2015-2016 averages of the UNEB data for all schools, and then regressing the % students scoring first division against the equivalent survey measure. In addition, the mean-division scored by students from each school is computed from the UNEB data and regressed against the average UCE score from the questionnaire.

Figure (20) shows the results of this exercise. The left plot indicates that for a few schools at the low end of the spectrum there probably is some reporting bias, but this does not compromise the overall relationship. In general, there are few schools with a high proportion of students scoring first-division. The second plot shows a relationship that is even noisier but does not show any signs of reporting bias. It is helpful to note that these relationships are noisy not only because the respondent provided inexact answers, or because (as in the second plot) the metrics differ slightly, but also because two years separate the most recent UCE examinations on which the survey is based, and those of 2015-16.

Figure 20: Comparison: Survey vs. UNEB UCE Performance Data



[13]This is done so that a normally distributed indexes take on values between 0 and 10. This is almost never the case in practice, quite some factor scores have either positive or negative outliers. The distributions can be checked in Figure (35).

Table (10) shows summary statistics for the main outcome measures. Measures obtained from the UNEB data are indicated by '(2015-16 avg.)'. The 'UCE performance Index' is the factor score of 'UCE Aggregate' and 'Students scoring first division (%)'. Similarly the 'Performance equality Index' is the factor score of the two performance equality indexes below. The dropout gender equality index is obtained in a similar way (Its constituent variables are not of direct interest). Noteworthy is also that the GPI is computed slightly different than the classical GPI (which is simply the female figure divided by the male figure). Since quite some schools in the sample have girls performing significantly better than boys, the GPI computed here measures the deviation from parity (e.g. it equals female/male if female/male < 1 and male/female if male/female < 1).

Table 10: Summary of Main Outcome Measures (Survey and UNEB)

	N	Mean	Median	SD	Min	Max
UCE Aggregate	420	38.01	39.5	10.99	8	70
Students scoring first division (%)	419	29.59	23	25.62	0	100
UCE performance Index (factor score)	418	4.23	3.86	2.35	-1.21	10.62
UCE pass Index (factor score)	420	7.06	8.1	2.92	-9.47	9
Performance equality Index (factor score)	422	6.38	7.14	2.75	-4.97	9
Performamce equality, income (GPI)	422	0.88	1	0.19	0	1
Performamce equality, gender (GPI)	422	0.79	0.9	0.24	0	1
Yearly dropout rate (% of students)	418	3.11	1.52	4.58	0	45.98
Yearly repeater rate, O-level (%)	348	2.47	1.4	3.42	0	26.88
UACE Aggregate	369	11.2	11	4.1	0	24
A-level graduates enrolling in Uni (%)	354	57.25	60	26.68	0	100
A-level graduates beginning to work (%)	337	21.42	15	22.57	0	95
Yearly repeater rate, A-level (%)	345	1.77	0	3.33	0	24.53
Candidates first-division (%) (2015-16 avg.)	388	17.34	7.75	22.8	0	98.15
Mean division, all candidates (2015-16 avg.)	388	3.19	3.24	1.15	1.02	6.08
Candidates last(10)-division (%) (2015-16 avg.)	388	1.64	1.1	2.65	0	41.25
Candidates first-division (GPI) (2015-16 avg.)	230	0.53	0.53	0.24	0.09	1
Mean division (GPI) (2015-16 avg.)	341	0.87	0.88	0.07	0.55	1

Figure (21) shows a correlation matrix of the educational outcomes in Table (10). Correlated variables are grouped together in the matrix using hierarchical clustering. The matrix shows that there are several clusters of highly correlated outcomes and a lot of insignificant correlations. The 4 large blocks in the matrix show the positively coded (e.g. more is better) performance metrics and the negatively coded (e.g. more is worse) performance metrics that load positively onto themselves (red-blocks) and negatively onto each-other (blue blocks). It is important to note that the UCE aggregate score is negatively coded and takes values from 8 (best) to 72 (worst), while the UACE score is positively coded from 0 (worst) to 20 (best). Students in the first division have a UCE score below 12. Aside from the large blocks, the middle of Figure (21) shows 3 smaller blocks on the diagonal corresponding to grade repetition and gender equality. The fractions of grade-repeaters at O- and A-level are correlated with each other and weakly correlated with the dropout rate, but uncorrelated with everything else. Of the remaining two blocks, the larger one in the center of the matrix shows the loadings of income[14] and gender-parity indexes computed from the survey onto each other, and the smaller block below represents the correlations of GPI measures obtained from UNEB data. These measures obtained from different sources are surprisingly completely uncorrelated with each other, casting doubt on the headmaster's ability to estimate gender-equality.

A PCA on the outcome space reported in Table (11) yields very similar results. The first 4 components roughly correspond to i. performance metrics, ii. performance equality computed from the survey, iii. dropout and repetition rates and iv. the gender-parity measures computed from the UNEB data. The first 4 PC's explain 30%, 12%, 9% and 8%, respectively, of the outcome space, together amounting to a cumulative variance share of 59%.

[14]For the income performance equality index GPI is added in brackets because it is coded in the same way (1= equality, 0=Total inequality), the index has, of course, nothing to do with gender.

Figure 21: Correlation Matrix of Main Outcome Measures

Note: Correlations insignificant at the 5% level are crossed out, variables are grouped by hierarchical clustering with complete linkage.

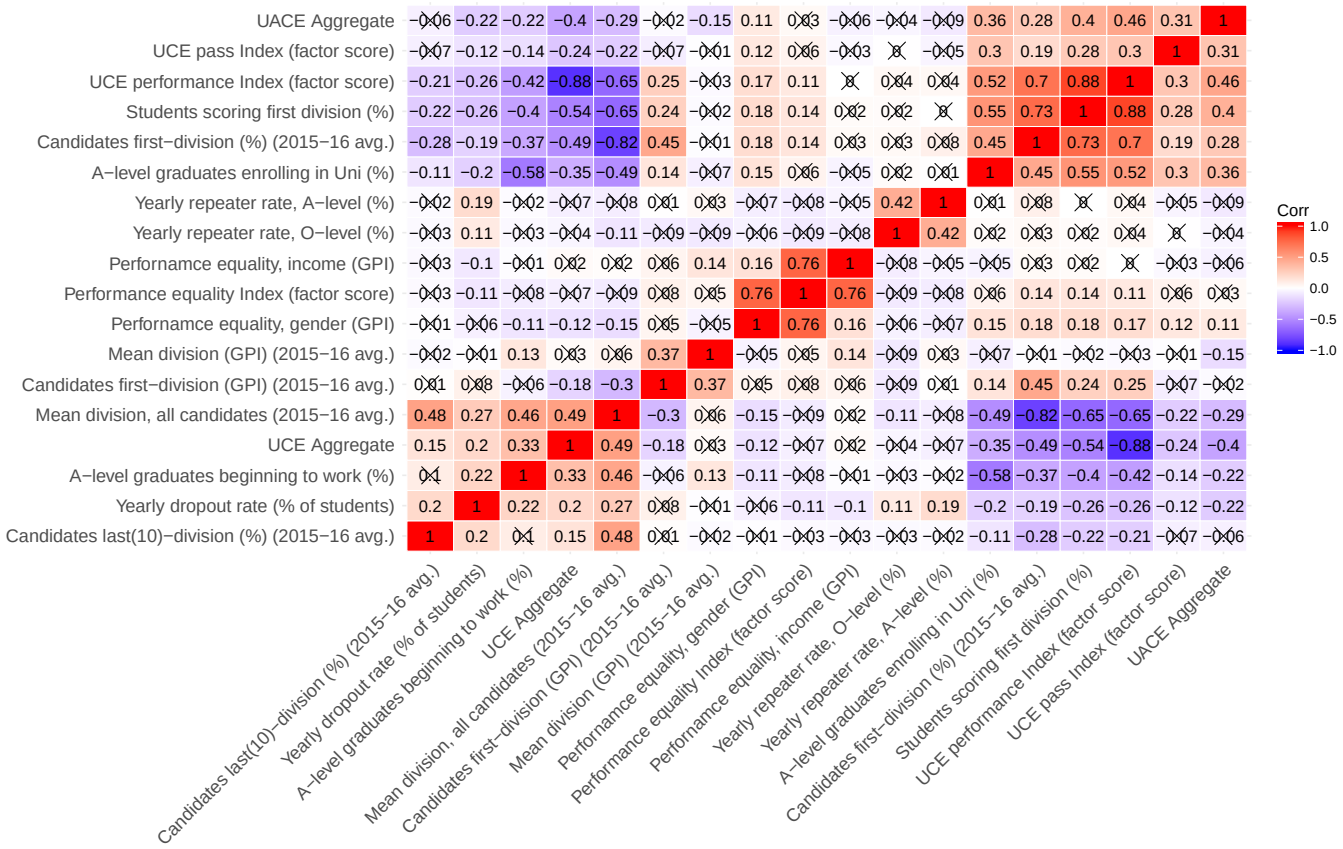


Table 11: PCA on Main Outcome Measures

Varimax rotated, loadings < 0.4 omitted, KMO = 0.69

Variables	PC1	PC2	PC3	PC4
PVE	30%	12%	9%	8%
UCE Aggregate	0.74			
Students scoring first division (%)	-0.86			
UCE performance Index (factor score)	-0.91			
UACE Aggregate	-0.58			
A-level graduates enrolling in Uni (%)	-0.74			
A-level graduates beginning to work (%)	0.63			
Candidates first-division (%) (2015-16 avg.)	-0.80			
Mean division, all candidates (2015-16 avg.)	0.83			
Performance equality Index (factor score)		0.99		
Performamce equality, income (GPI)		0.75		
Performamce equality, gender (GPI)		0.76		
Yearly dropout rate (% of students)			-0.50	
Yearly repeater rate, O-level (%)			-0.77	
Yearly repeater rate, A-level (%)			-0.81	
Candidates first-division (GPI) (2015-16 avg.)				-0.79
Mean division (GPI) (2015-16 avg.)				-0.73
UCE pass Index (factor score)				
Candidates last(10)-division (%) (2015-16 avg.)				

5.4.3 Data Integrity Checks

It remains to conduct a few additional integrity checks on the data before beginning with the main analysis. The first check is whether the characteristics of the respondent significantly influence any of the answers. In its preambular part, the survey records the respondent's relationship to the school, highest level of education, years of employment at the school, years of employment in other schools, years of teaching experience, years occupying the current position, gender, age and marital status. Regressing all predictors and outcomes in the not-dimension-reduced dataset on the respondents characteristics and collecting the R^2 of each regression yields 4 variables for which the R^2 is greater than 0.1: How private the school is ($R^2 = 0.15$), the age of the school ($R^2 = 0.13$), the median teacher salary ($R^2 = 0.11$) and the number of O-level students ($R^2 = 0.1$). In the dimension-reduced data set, the teacher experience factor score index is added to this group ($R^2 = 0.11$). These are all formal characteristics of the school, thus the respondent's characteristics can be assumed unrelated to performance and other sensitive questions, and therefore not a source of bias.

This exercise is repeated for geographic characteristics of the school by regressing all variables in turn on the latitude and longitude of the school and dummies for each of the 86 districts in the sample. Here the R^2's are expectedly much higher. All R^2 are greater than 0.1, but in the reduced dataset there are only 3 variables for which the R^2 exceeds 0.5: A dummy indicating whether the teacher has the only textbook or not ($R^2 = 0.55$), a dummy indicating a pedagogic qualification of the teacher ($R^2 = 0.55$), and a dummy indicating whether the school has a cantine ($R^2 = 0.51$). Other variables that are well predicted by these geographic characteristics are the number of teacher vacancies, the fraction of students scoring first-division and the corresponding GPI, the teacher education index and the fraction of A-level graduates continuing on to university. These results are not surprising given what is known about regional inequality in education in Uganda. Under the assumption that geographic characteristics do not cause educational outcomes except through their effects on the included predictors (e.g. there is no direct relationship between geography and educational outcomes), omitting district dummies will not bias the coefficients. In any case, the greater worry with this data is that including such fixed-effects would bias the size of the coefficients downwards, because the 422 schools in the sample spread across 86 districts, some districts only represented by a single school, so putting district fixed-effects almost amounts to putting school fixed-effects. Hence, geographic controls are not included in the analysis.

The last integrity check investigates whether the predictors are well-separated from the outcomes. If this is the case then there should be no variable in $\mathbf{X}$ that proxies for or is collinear with any of the outcome measures. This is checked by regressing each outcome measure on $\mathbf{X}$ and capturing the R^2. The result is shown in Table (12).

Table 12: Predictable Variance in Main Outcome Measures (Survey and UNEB)

Name	R^2	N	*Name*	R^2	N
Mean division, all candidates (2015-16 avg.)	0.654	388	Yearly dropout rate (% of students)	0.330	418
Students scoring first division (%)	0.631	419	Candidates first-division (GPI) (2015-16 avg.)	0.313	230
Candidates first-division (%) (2015-16 avg.)	0.628	388	Yearly repeater rate, O-level (%)	0.278	348
UCE performance Index (factor score)	0.625	418	Perfornamce equality, gender (GPI)	0.255	422
A-level graduates enrolling in Uni (%)	0.463	354	Perfornamce equality, income (GPI)	0.247	422
UCE Aggregate	0.456	420	Performance equality Index (factor score)	0.228	422
UACE Aggregate	0.434	369	Mean division (GPI) (2015-16 avg.)	0.207	341
A-level graduates beginning to work (%)	0.430	337	Candidates last(10)-division (%) (2015-16 avg.)	0.192	388
UCE pass Index (factor score)	0.355	420	Yearly repeater rate, A-level (%)	0.181	345

As the table indicates, not more than 65% of the variance in any outcome measure can be predicted with the 55 predictors at hand, suggesting that no variable in $\mathbf{X}$ closely proxies for any of the outcome measures.

6 Analysis

The main aim of this research is to find out which school characteristics best predict educational success in its various dimensions. Integrity checks already revealed that geography and regional inequality is a major hurdle to equitable educational access that the Government is aware of. To gain an impression of what other key issues may be at stake, the empirical approach taken in the following is one of careful variable selection. As previous empirical work shows, just regressing each outcome on the 55 predictors and then examining the $t-$statistics is likely to be of little information, one reason being the problems of joint multicollinearity and overfitting (particularly in small datasets like this one), another that statistical significance is an imperfect proxy for effect-size.

An alternative is to use various procedures for selecting an appropriate model that includes less than the full set of predictors. This is the approach adopted. To guarantee empirical robustness, three such methods are executed in parallel. The first is the well-known Forward-Stepwise Selection algorithm. Forward Selection starts with a constant and adds variables one by one, each time adding the variable that gives the greatest increase in R^2. Of the three methods, it is the most interpretable but also the most deterministic method (each variable chosen at one stage influences which variable get's chosen afterwards).

The second (and better) method employed is performing regularized regression using the Least Absolute Shrinkage and Selection Operator (LASSO). The LASSO is a so-called shrinkage method which is performed on standardized data (so as to yield standardized coefficients) and minimizes

$$\min_{\beta} \quad \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{k} |\beta_j|. \tag{1}$$

The first term in Eq. (1) is just the residual sum of squares and the second term is a penalty on the sum of the absolute values of the standardized coefficients. The effect of increasing the penalty parameter λ is to shrink the coefficient estimates towards 0. Shrinkage methods are mainly used to improve the fit of a regression as shrinking the coefficient estimates can significantly reduce their variance (James et al., 2013). They have however also proven themselves a powerful variable selection tool since the coefficients of less important variables shrink to 0 earlier as λ increases. In practice, the LASSO regression will be performed multiple times starting at $\lambda = 0$ (which amounts to linear regression) and then increase λ in about 80 intervals until the last coefficient is shrunk to 0. Based on the value of λ at which a coefficient first hits 0, predictors can be ranked (e.g. coefficients surviving a larger penalty belong to stronger predictors). Since after each incremental increase in λ the optimization is repeated, the selection result is non-deterministic and arguably superior to Forward Selection.

The third method employed is the variable importance ranking provided by the popular non-linear Random Forest algorithm, a variant of which was already used to impute the predictors. This algorithm, first propagated by Breiman (2001), is a so-called ensemble machine-learning algorithm based on trees. It works by building a large number of random decision trees (the Random Forest, with 1000 trees in this case), where each tree is built on a random subset of variables (about 1/3 of the 54 variables) and on a bootstrap sample taken from the data (about 30% of the sample is duplicated). It then predicts the outcome using all of the 1000 trees, and averages those predictions, yielding very accurate aggregate predictions. Since each tree is built from a random subset of data and the available variables, the forest is constructed non-deterministically. Furthermore, since decision trees are highly non-linear structures, the algorithm is able to take account of complex interactions between various predictors in determining the outcome (e.g. the algorithm does not rely on the additive separability of predictors like linear regression). An unbiased measure of the predictive power of the forest is obtained by using each three to predict the data that was not in the bootstrap sample on which the tree was built (around 30% of the data), and then, for each observation in the dataset, averaging the predictions across all of the trees in which this observation was not in their bootstrap sample. The importance of a predictor can then be obtained by looking at how much the overall prediction error of the forest increases (% increase in MSE) when the data for that variable is permuted while all others are left unchanged Liaw et al. (2002). For more information about decision tree's and Random Forests including implementing

examples, the reader is advised to consult James et al. (2013) or Liaw et al. (2002).

In order to save space, for each of the three methods, only the top 10 predictors will be reported. Greatest credence in terms of the ranking is afforded to the LASSO results. The Random Forest serves mainly as a cross-check for whether there are significant non-linear relationships in the prediction leading to a different ranking. For each method, a linear model with the top 10 variables suggested by the method is estimated. For Forward Selection, these coefficients are reported, while for LASSO and Random Forest this regression only serves to determine the sign (indicated by a '+' or '−' prefix) and significance (indicated by a '∗' at the end indicating significance at the 5% level) of the predictors[15]. Additional statistics like the Random Forest and LASSO outputs, information criteria like the BIC and the adjusted R^2 for Forward Selection, and 10-fold cross-validation results for LASSO and Forward Selection are provided on a separate page for each outcome table in the Appendix. It should also be noted that if the overall predictive power, as indicated in Table (12) is low, variable importance rankings are also less robust[16]. The following four sections present the empirical results, in turn, for performance, dropout and repetition, gender equality, and after-school placement.

6.1 Predictors of Performance

Tables (13) and (14) show the variable selection results for the mean division scored and the UCE aggregate. The Forward R^2 in Tables (13) shows that an optimally selected model with 10 variables reaches an R^2 of 60.2%, which is almost the R^2 of 65.5% reached when predicting with all 54 variables (cf. Table 12).

Table 13: Variable Ranking: Mean division, all candidates (2015-16 avg.)

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	Admission PLE score requirement	0.354	0.023*	− Boarding students (%)*	0.682	+ Admission PLE score requirement*	36.870
2	Log average attendance cost	0.471	-0.146*	+ Admission PLE score requirement*	0.682	− Boarding students (%)*	27.040
3	Student poverty Index (factor score)	0.503	0.058*	− Log average attendance cost*	0.621	− Log average attendance cost*	20.750
4	Boarding students (%)	0.526	-0.006*	+ Student poverty Index (factor score)*	0.516	+ Student poverty Index (factor score)*	20.690
5	Total number of students	0.555	0*	− Teacher education level Index (factor score)*	0.428	− Teacher education level Index (factor score)*	20.420
6	Teacher education level Index (factor score)	0.574	-0.06*	− Total number of students*	0.269	− Total number of students*	19.210
7	Student-Teacher and Class-Size Index	0.583	0.015	+ Student absenteism (%)*	0.223	− Technological equipment Index (factor score)*	18.860
8	Fraction teachers leaving every year	0.589	1.52*	− Minimum teacher monthly salary (1000 UGX)	0.223	+ Student absenteism (%)*	16.610
9	Government funding share (%)	0.595	0.005*	− Technological equipment Index (factor score)	0.185	− Minimum teacher monthly salary (1000 UGX)	12.720
10	Sanitary facilities Index (factor score)	0.600	-0.044*	− Sanitary facilities Index (factor score)	0.185	+ Government funding share (%)*	11.650

The coefficients corresponding to this 10-variable model are shown in the third column of Table (13). Columns 4 and 5 of Table (13) show the LASSO selection result with the corresponding critical λ values[17] and sign and significance at the 5% level are indicated as described. Columns 6 and 7 of Table (13) show the Random Forest selection with the corresponding % increase in the MSE after permuting the predictor. Table (13) provides a very robust selection result for the top 10 variables impacting the mean UCE division scored by students in 2015-16. The union of the top 10 predictors from all 3 methods yields only 13 distinct predictors, while they all intersect on 6 variables being most important: The PLE score requirement of incoming students, the average attendance cost, student poverty, the faction of boarding students, the size of the school, and the education level of teachers. The signs are all in the anticipated direction where it must be noted that the best division is 1 and the worst 10, and the PLE score is also inverse-coded from 4 (best) - 36 (worst). In addition to these 7 intersection variables, LASSO and the Random Forest rank technological equipment and the minimum teacher salary among the top 10, and the LASSO also gauges good sanitary facilities as very important for high performance. While the PLE requirement and cost function as natural proxies for the selectivity and elitism of the school, and larger schools are more likely to be located in urban areas and offer a better curriculum, it is less obvious that

[15]It is clear that in a Random Forest predictors neither really have a sign (since their effect could be non-linear and the forest is non-parametric), nor is classical inference possible. Thus the sign and significance indications for the Random Forest variables based on a linear regression should be interpreted with caution and treated as what they are: The output of a non-linear model that is linearized and then classically assessed.

[16]This becomes particularly problematic for the Random Forest which is geared to predict well out-of-sample. More on this later.

[17]Again these are the values for the penalty parameter λ at which the coefficient is shrunk to 0. They are used to create the variable ranking.

boarding schools generally perform better. A plausible explanation could involve a more encompassing academic and social curriculum and fewer domestic problems hindering boarding-students learning progress. Another factor is that boarding schools are generally more expensive than regular schools and attract students from wealthier families. The correlation coefficient between the % boarding students and the student poverty index is -0.41. Besides these 'environmental' variables, the result is affirmative of the policy review in suggesting that teacher training and financial incentives, poverty of students and absenteeism, and technological and sanitary facilities in secondary schools are the main cornerstones influencing student performance. Performance metrics and additional statistics for Table (13) are provided in Figure (24) in the Appendix. 10-fold cross-validation (CV) suggest an optimal model size of 22, and with an R^2 of 53% the CV predicting performance is close to the in-sample fit.

Table (14) presents analogous results for the UCE aggregate score from the survey, which Figure (20) showed to be mildly correlated with the mean division scored. The in-sample fit is worse than for the mean division scored, with the 10-variable Forward selected model only reaching an R^2 of 37.1%. This also renders the selection result slightly less robust: The union of the top 10 predictors from all 3 methods contains 16 predictors, while they intersect on 5 predictors being most important: Admission PLE score requirement, cost, teacher bad behavior[18], the number of weekly homework hours and student absenteeism[19]. However, both LASSO and the Random Forest (RF) intersect on 9 predictors, adding the fraction of boarding students, student poverty, the education level of teacher and technological equipment to the former 5. These are the top 9 according to both LASSO and RF. The signs are all in the anticipated direction, thus at second glance Table (14) is very much in line with Table (13) in pointing out the most important issues. Performance metrics are again provided in Figure (25). The CV predictive performance of the linear model is 22% of the variance, while the RF explains 35.3% of the out-of-sample (out-of bootstrap-aggregation (OOB)) variance.

Table 14: UCE Aggregate

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	Admission PLE score requirement	0.191	0.33*	− Log average attendance cost	4.798	− Log average attendance cost	23.980
2	Log average attendance cost	0.256	-2.316*	− Boarding students (%)	4.798	+ Admission PLE score requirement*	22.750
3	Teacher bad behavior Index (factor score)	0.282	0.453*	+ Admission PLE score requirement*	4.798	− Boarding students (%)	19.690
4	Homework hours (per S4 student per week)	0.305	-0.281*	+ Student poverty Index (factor score)	3.013	+ Teacher bad behavior Index (factor score)*	18.510
5	Student absenteism (%)	0.319	0.179*	+ Number of issues hindering instruction	2.745	+ Student poverty Index (factor score)	14.760
6	Number of administrative student records kept	0.333	1.162*	+ Student absenteism (%)	2.745	− Homework hours (per S4 student per week)*	12.990
7	Students strikes (N/last 3 years)	0.343	3.432*	− Teacher education level Index (factor score)	2.502	− Teacher education level Index (factor score)	12.510
8	Teacher experience Index	0.355	-0.548*	+ Teacher bad behavior Index (factor score)*	2.279	− Technological equipment Index (factor score)	12.410
9	Number of assessments (per term per subject)	0.364	-1.204*	− Homework hours (per S4 student per week)*	1.892	+ Student absenteism (%)*	12.150
10	Fraction teachers leaving every year	0.371	14.702*	− Technological equipment Index (factor score)	1.724	− Student misconduct Index (factor score)	11.120

To continue the discussion along the lines of the comparison posed in Figure (20), Tables (15) and (16) show the best predictors for the % of candidates scoring first-division from survey and UNEB 2015-16 data. Tables (15) shows that with an R^2 of 58.4%, the predictive power of the 10-variable Forward model is high. The selection result is more precise again with 13 variables in the union of the three methods and 6 in the intersection. Next to the usual PLE requirement and cost variables, the intersection includes student poverty, teacher education, the government funding share and the total number of students. The intersection between LASSO and RF includes only one extra variable, the fraction of boarding students. The LASSO also ranks among the top 10 the number of issues hindering instruction[20], and indexes for the extent of internal monitoring and bad teacher behavior. The RF therewhile also considers the number of subjects taught, technological equipment and the minimum monthly teacher salary very important. While most of these are variables seen already in Tables (13) and (14), it is interesting to observe that the government funding share more strongly relates to the % of excellent students than to the UCE aggregate, affirming the heuristic that private schools are more likely to be centers of excellence. A variable that first appears in Table (15), in both the Forward and LASSO rankings, is the internal monitoring index,

[18]This index is obtained as a factor score from 3 variables: A variable recording the count of bad teacher behaviors observed in the school (e.g. arriving late in school, absenteeism, lack of preparation or harassing students), and two variables recording the fraction of classes that had to be cancelled or taught by someone else because of teacher absenteeism.

[19]Simply the proportion of students absent from class at any given moment.

[20]This variable counts the marks on a checklist where the respondent had to tick off problematic areas, i.e. lack of scholastic materials, lack of laboratories or library materials etc...

which measures how often student performance is communicated to the principal, and how often major teacher and staff meetings take place. These routines appear to be important for cherishing excellence. The CV predictive performance of the 10-variable model is about 51% and the RF explains 63% of the OOB variance.

Table 15: Students scoring first division (%)

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	Student poverty Index (factor score)	0.361	-2.562*	− Student poverty Index (factor score)*	15.385	− Government funding share (%)*	33.910
2	Log average attendance cost	0.450	3.145*	− Admission PLE score requirement*	14.019	− Student poverty Index (factor score)*	29.130
3	Admission PLE score requirement	0.498	-0.52*	+ Log average attendance cost*	11.638	+ Log average attendance cost	28.960
4	Teacher education level Index (factor score)	0.523	1.464*	+ Teacher education level Index (factor score)*	10.604	− Admission PLE score requirement*	28.610
5	Internal monitoring Index (factor score)	0.537	0.822*	+ Boarding students (%)	7.309	+ Teacher education level Index (factor score)*	25.760
6	Teacher bad behavior Index (factor score)	0.550	-0.914*	− Government funding share (%)*	6.068	+ Total number of students*	17.880
7	Government funding share (%)	0.560	-0.148*	− Number of issues hindering instruction*	6.068	+ Number of subjects taught*	15.930
8	Total number of students	0.570	0.005*	+ Internal monitoring Index (factor score)*	4.590	+ Boarding students (%)	15.130
9	Number of subjects taught	0.577	0.884*	+ Total number of students*	3.811	+ Technological equipment Index (factor score)*	14.750
10	Number of issues hindering instruction	0.584	-1.373*	− Teacher bad behavior Index (factor score)*	3.811	+ Minimum teacher monthly salary (1000 UGX)	11.010

The variables selected in Table (16) differ slightly from those of Table (15). In Table (16) there are 16 variables in the union and 5 in the intersection of the three methods. The intersecting variables are the PLE requirement, student poverty, the total number of students, the fraction of boarding students and the teacher education level. LASSO and RF also include the cost and the age of the school in their intersection, and LASSO ranks the number of assessments, teacher GPI, and sanitary facilities among the top 10 predictors. Overall the ranking provided in Table (16) emerges as less intuitive than the one presented in Table (15). Particularly remarkable is that the first two predictors in Table (16), PLE requirement and student poverty index, outperform the other predictors by quite an edge (e.g. λ values of 14 vs. 5 for the next best predictor, and %IncMSE's of 38 and 28 vs. 15 for the next best predictor). The performance ranking of Table (15) on the other hand is a lot more gradual and balanced than that of Table (16). The CV predictive performance of the model in Table (16) is at 48%, and the RF explains 58% of the OOB variance.

Table 16: Candidates first-division (%) (2015-16 avg.)

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	Admission PLE score requirement	0.400	-0.937*	− Admission PLE score requirement*	14.400	− Admission PLE score requirement*	37.820
2	Student poverty Index (factor score)	0.506	-2.37*	− Student poverty Index (factor score)*	14.400	− Student poverty Index (factor score)*	27.710
3	Number of assessments (per term per subject)	0.523	3.047*	+ Boarding students (%)	5.175	− Government funding share (%)	14.830
4	Total number of students	0.536	0.007*	+ Log average attendance cost	4.715	+ Total number of students*	14.580
5	Boarding students (%)	0.549	0.079*	+ Total number of students*	4.296	+ Log average attendance cost	12.310
6	Teacher gender parity (1=parity, 0=Only 1 gender)	0.559	9.397*	+ Teacher education level Index (factor score)	4.296	+ Boarding students (%)	11.820
7	Sporting facilities Index (factor score)	0.565	0.837*	+ Number of assessments (per term per subject)*	3.567	+ Teacher education level Index (factor score)	11.720
8	1[A-level offered]	0.571	-8.845*	+ Teacher gender parity (1=parity, 0=Only 1 gender)*	3.250	+ Age of school (years)	10.590
9	Teacher education level Index (factor score)	0.577	0.842*	+ Age of school (years)	2.961	+ Technological equipment Index (factor score)	7.880
10	External school inspections per year	0.582	-0.521*	+ Sanitary facilities Index (factor score)	2.961	− Student absenteism (%)	7.010

Both tables are nevertheless similar. To guard against the slight discrepancy introduced by the source of data and the time of collection, it is worthwhile to compute the intersection of both tables. The intersection of all predictors selected by either of the three methods in Tables (15) and (16) includes 8 variables that surface as most-important: Attendance cost and admission requirement, student poverty and teacher education, government funding share, total number of students, share of boarding students and technological equipment. Of these 8, 4 variables, the admission requirement, student poverty, teacher education and the total number of students, were in the top 10 of each method in both tables[21].

At last, after considering both measures separately, Table (17) reports the results for the UCE performance index, computed as the factor score of the aggregate UCE and the % of students scoring first division (both taken from the survey). The table holds no surprises, it's interpretation is therefore left to the reader. Additional statistics accompanying the results presented in Table (17), but also in Tables (15) and (16) are again provided in Figures (28), (26) and (27) respectively.

As a final step in analyzing predictors of performance, it is worthwhile to look at the predictors that figure prominently for all of the 5 performance measures considered so far. The intersection of all predictors suggested in Tables (13), (14), (15), (16) and (17) contains 6 predictors: The

[21] In the intersection of the intersections so to speak.

Table 17: UCE performance Index (factor score)

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	Admission PLE score requirement	0.333	-0.056*	− Admission PLE score requirement*	1.357	+ Log average attendance cost*	30.540
2	Log average attendance cost	0.431	0.421*	− Student poverty Index (factor score)*	1.357	− Admission PLE score requirement*	29.650
3	Student poverty Index (factor score)	0.482	-0.209*	+ Log average attendance cost*	1.236	− Student poverty Index (factor score)*	26.930
4	Teacher bad behavior Index (factor score)	0.505	-0.112*	+ Boarding students (%)	1.026	− Government funding share (%)	20.820
5	Teacher education level Index (factor score)	0.525	0.12*	+ Teacher education level Index (factor score)*	0.852	+ Teacher education level Index (factor score)*	20.800
6	Number of issues hindering instruction	0.541	-0.148*	− Number of issues hindering instruction*	0.644	+ Boarding students (%)	19.930
7	Students strikes (N/last 3 years)	0.549	-0.555*	− Teacher bad behavior Index (factor score)*	0.488	− Teacher bad behavior Index (factor score)*	17.210
8	Boarding students (%)	0.556	0.008*	− Student absenteism (%)	0.405	− Student absenteism (%)	14.860
9	Groupwork & leadership Index (factor score)	0.563	-0.055*	+ Technological equipment Index (factor score)	0.369	− Number of issues hindering instruction*	14.200
10	Fraction teachers leaving every year	0.568	-2.771*	− Government funding share (%)	0.336	+ Technological equipment Index (factor score)	13.830

PLE admission requirement and log attendance cost, the student poverty and teacher education indexes, the % of boarding students and the technological equipment index. In addition, and this is quite remarkable, the intersection across all tables, of the intersections of the LASSO and RF rankings within each of the 5 tables, contains all the just mentioned variables except for the technological equipment index. Thus, as a bare minimum, the very robust message of this analysis of determinants of performance is: (1) Inputs in terms of bright students and financial resources matter a lot; (2) Educated teachers are key; (3) Aggregate poverty negatively affects educational success in the Ugandan context; (4) Boarding schools perform better than non-boarding schools, and (5) technological infrastructure matters[22].

6.2 Predictors of Dropout, Repetition and Completion

Aside from enhancing aggregate performance, Uganda faces other great issues in its education system that relate to dropouts and over-age pupils. This section, therefore, seeks empirical answers to the question: What does a school need to satisfy for its students to stay in school, and complete it without having to repeat grades? Table (18) shows the top 10 determinants of the aggregate dropout rate. The 10-variable Forward selection model reaches an R^2 of 27.7%, about half the fit typically reached in the previous section, therefore these results need to be treated with more caution. In particular, the accompanying Figure (29) shows that the CV predictive performance is low and explains only around 5% of the out-of-sample variance and only when predicting with the top 3 variables. The RF also only explains 7% of the OOB variance in the dropout rate.

Table 18: Yearly dropout rate (% of students)

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	1[A-level offered]	0.100	-2.426*	− 1[A-level offered]*	1.449	− Teacher education level Index (factor score)	9.630
2	Number of issues hindering instruction	0.156	0.53*	+ Number of issues hindering instruction*	1.449	− Total number of students*	9.250
3	Fraction teachers leaving every year	0.193	11.983*	+ Fraction teachers leaving every year*	1.320	+ Fraction teachers leaving every year*	8.160
4	Parents informed on student performance (N/year)	0.210	-0.73*	− 1[Cantine available]*	0.999	+ Number of issues hindering instruction*	7.640
5	Technological equipment Index (factor score)	0.226	0.411*	− Parents informed on student performance (N/year)*	0.829	+ Government funding share (%)	5.360
6	Boarding students (%)	0.243	-0.019*	− Boarding students (%)	0.755	− Log average attendance cost	4.600
7	Students strikes (N/last 3 years)	0.255	1.442*	− Total number of students*	0.688	− Boarding students (%)	4.230
8	1[Cantine available]	0.264	-1.656*	+ Students strikes (N/last 3 years)*	0.627	+ Median teacher monthly salary (1000 UGX)	4
9	External school inspections per year	0.271	0.155*	+ Government funding share (%)	0.521	+ Students strikes (N/last 3 years)*	3.990
10	Number of technical aids in use	0.277	-0.308	+ 1[Externally-supported initiatives or projects]*	0.474	+ Minimum teacher monthly salary (1000 UGX)	3.790

The low fit is reflected in the selection result: The union of the three methods contains 18, and they only intersect on 4: The number of issues hindering instruction, yearly teacher attrition rate, the share of boarding students and the number of student strikes in the last 3 years. LASSO and the RF also consider the total number of students and the government funding share to be very important. LASSO and RF intersect on these 6 variables, but LASSO and Forward Selection intersect on 7 variables: The 4 aforementioned variables and a dummy indicating whether A-level classes are offered, how often parents are informed on student performance in a given year, and a dummy indicating whether the school has a cantine. Since the RF only explains 7% of the OOB variance and its ranking is very imprecise (as Figure (29) shows), greater attention is afforded to the two linear methods which at least reach an in-sample fit of around 30%. A couple of interesting findings surface in the variables suggested by the LASSO and Forward Selection. First, the best predictor is whether the school offers A-level classes, which alone explains 10% of the variance in dropout rates. The coefficient indicates that in schools taking students up to S6, the aggregate dropout rate is 2.4% lower. This is presumably the case because the dropout rate at A-level is

[22]This index combines computers, science laboratories, beamers and internet access, per student respectively.

much lower than in O-level (i.e. once students made it to upper secondary, they tend to finish their education). Variables ranked 2 and 3, the number of issues hindering instruction and the teacher attrition rate, seem to proxy for both the capacity and motivation of the school to provide an adequate and stimulating learning environment to its students. It seems intuitive that if looming issues hinder instruction on a daily basis, and teachers expend efforts to secure themselves a job at a different school, it does not motivate students to remain in school. The 4'th predictor is how often parents are informed about the performance of their child. If parents do not receive such information on a regular basis, it is easily imaginable that performance will tend to deteriorate leading eventually to the student dropping out. Next, having a cantine appears to reduce the dropout rate by 1.7%. This is a key finding already well known from the literature on primary education. If children cannot take meals in school, they tend to go hungry and henceforth are more likely to drop out. The findings confirm that a cantine is also an important factor preventing drop-out in secondary schools. Boarding schools also have significantly fewer dropouts, probably because everything the children need is provided at school and they do not need to travel long distances to school. Lastly, both methods suggest student strikes are an important determinant of dropout rates. Presumably, the direct effect of such strikes on dropout rates is small, but strikes proxy for both untenable conditions in the school itself, inviting student protest, and also possibly for a lack of strict control-over and discipline of the student body. The coefficient indicates that one such event in the last 3 years is associated with a 1.4% higher dropout rate in concerned schools.

Table (19) shows the results for the yearly repeater rate of O-level students. The in-sample fit is lower than for the dropout rate, with the 10-variable Forward model reaching an R^2 of 20%. This time, however, the out-of-sample predictive performance for both the linear model and the RF hit 0%, as shown in Figure (30). The results for the RF model were omitted since the algorithm is geared towards predicting well out of sample, and provides a spurious variable ranking when the predictive power reaches 0%. Forward Selection and LASSO still give similar rankings in this case, but it is important to be aware that these methods are only geared towards explaining the in-sample variance.

Table 19: Fraction O-level repreaters (%)

#	Forward	F-R^2	F-Coef	LASSO	λ
1	Student misconduct Index (factor score)	0.084	0.289*	+ Student misconduct Index (factor score)*	0.990
2	Government funding share (%)	0.112	-0.015*	− Government funding share (%)	0.516
3	Length of lessons (minutes)	0.130	0.022*	+ Length of lessons (minutes)*	0.516
4	Parents informed on student performance (N/year)	0.145	-0.535*	− Parents informed on student performance (N/year)*	0.470
5	Groupwork & leadership Index (factor score)	0.157	0.12*	+ Log average attendance cost	0.428
6	1[Externally-supported initiatives or projects]	0.168	-0.852*	− Age of school (years)*	0.428
7	Library facilities Index (factor score)	0.180	0.077*	+ Groupwork & leadership Index (factor score)*	0.428
8	Female students (%)	0.189	0.011	+ Library facilities Index (factor score)*	0.390
9	Age of school (years)	0.195	-0.018*	− Student-Teacher and Class-Size Index	0.356
10	Sanitary facilities Index (factor score)	0.204	0.167*	− 1[Externally-supported initiatives or projects]*	0.356

This being noted, LASSO and Forward Selection intersect on 8 variables ranked most important in explaining repetition rates: Student misconduct, the government funding share, the length of lessons, the frequency with which parents are informed on student performance, library facilities, a dummy indicating externally-supported initiatives or projects, an index measuring whether group work and/or student leadership is emphasized in the teaching approach and the age of the school. It is noteworthy that misconduct seems to be associated with more students having to repeat grades, and that public schools have higher repetition rates. Furthermore, longer classes appear to increase repetition (presumably because students retention is lower if classes are too long, classes in the surveyed schools range between 40 min and 120 min, the latter being clearly too long). Both methods also suggest that if parents are frequently notified about student performance, then this decreases repetition, and older schools have lower repetition rates. The coefficients suggest that more group work and better library facilities increase repetition rates, which clearly is a spurious result as both indexes relate significantly and in the intended direction to the outcome measures considered in the previous section and only very slightly to repetition rates. It remains to conclude that although some of the top variables in Table (19) appear plausibly to be important variables impacting repetition rates, the lower part of the table contains some spurious and unintuitive results, and the low in-sample fit and lacking out-of-sample predictive power should prompt the reader to treat the table with extreme caution.

Akin to Table (19), there is a set of results for A-level repetition. These results are omitted because as Table (10) and Figure (35) show, the average repeater rate at A-level is only 1.8%, and with an R^2 of 10% even the in-sample fit of the resulting Forward model is extremely low.

A final indicator to consider in this section is the UCE Pass Index, which is obtained as a factor score from two questions asking what % of students who registered for S4 actually sit the UCE at the end of the year, and what fraction of students sitting the exam pass it on the first attempt. The in-sample R^2 for the 10-variable Forward model is 25%, and the CV results in Figure (31)

Table 20: UCE Pass Index (factor score)

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	Teacher education level Index (factor score)	0.112	0.352*	+ Teacher education level Index (factor score)*	0.975	+ Teacher education level Index (factor score)*	20.350
2	Frequency of breaks	0.140	-3.705*	− Frequency of breaks*	0.558	+ Subject-teacher-composition Index (factor score)*	17.670
3	1[Cantine available]	0.160	1.306*	− Student absenteism (%)	0.558	+ Log average attendance cost	12.880
4	Homework hours (per S4 student per week)	0.179	-0.104*	+ 1[Cantine available]*	0.508	− Student absenteism (%)*	12.210
5	Student absenteism (%)	0.195	-0.025	+ Number of technical aids in use	0.508	− Median teacher monthly salary (1000 UGX)*	11.380
6	Minimum teacher monthly salary (1000 UGX)	0.211	-0.003*	+ Fraction teachers with pedagogic qualification*	0.463	− Homework hours (per S4 student per week)*	11.040
7	Fraction teachers with pedagogic qualification	0.222	0.967*	− Teacher bad behavior Index (factor score)	0.385	− Government funding share (%)	10.420
8	Teacher experience Index	0.232	0.141*	− Open teacher vacancies (% of teachers)	0.385	+ 1[Cantine available]*	9.290
9	Open teacher vacancies (% of teachers)	0.241	-0.023*	− Homework hours (per S4 student per week)*	0.350	+ Student-Teacher and Class-Size Index*	8.590
10	Teacher bad behavior Index (factor score)	0.249	-0.098*	+ Subject-teacher-composition Index (factor score)	0.350	− Frequency of breaks*	8.510

indicate a low out-of-sample predictive performance. The RF, on the other hand, explains 23% of the OOB variance in the index, signifying that the predictors do carry some predictive power. The 3 methods in Table (20) have a union of 16 predictors and intersect on 5: Teacher education and a cantine, which positively relate to passing, and the frequency of breaks, student absenteeism and the number of homework hours per week per S4 student which surprisingly all negatively relate to passing. The latter is especially curious since the mean number of homework hours per S4 student in the sample is only 6.3 h/week. In addition, LASSO and RF intersect on the Subject-teacher-composition Index, which proxies for the number of science teachers vis-a-vis arts and humanities or social science teachers, and LASSO and Forward selection also intersect on the fraction of teachers with a pedagogic qualification, the share of open teacher vacancies[23], and the teacher bad-behavior index. The signs of the latter are all the anticipated direction, although only the fraction of teachers with a pedagogic qualification is statistically significant at the 5% level.

6.3 Predictors of Gender Equality

This part of the analysis studies determinants of gender equality in performance. To begin with, the results presented in this section are a lot less robust than the results presented for performance. One reason why this might be the case is that Uganda has already more-or-less reached gender parity (GPI = 0.9 in official figures).

For the schools in the sample, the official UNEB statistics indicate that almost as many have girls performing better than boys than the other way around. For this reason, a GPI that measures the deviation from parity, 1 = parity, 0 = total non-parity was computed in place of the traditional GPI. Thus in terms of gender, the analysis essentially reduces to studying why some schools strike parity closer than others. This might be difficult to determine, it is conceivable that the mechanisms leading to parity or non-parity in educational outcomes are much more complex than the mechanisms leading to good aggregate performance.

In the survey, the respondent was also asked questions about equality between students from rich and poor family backgrounds. In particular, the respondent was requested to estimate gender and income-inequality in performance by assigning percentage scores to the relative performances of male and female, rich and poor students. As Figure (21) and Table (11) however demonstrated, the GPI indexes computed from the survey are completely unrelated to those computed from the official UNEB UCE test data. This finding strongly suggests that the survey measures based on the respondent's perception are too noisy to be considered useful at all for quantitative analysis. Consequently, the following analysis will restrict itself to two GPI's computed from the UNEB test

[23]Number of open vacancies divided by number of teachers.

data, hence forfeiting all research on the determinants of income parity.

An additional problem, visible in Table (12), is the very low fit observed for all GPI indexes. If predicting with 55 variables in a dataset with around 350 observations obtains an R^2 of 20% it suggests many spurious rather than substantive empirical relationships, which poses great problems for any model selection algorithm. This is visible in the results: Of the two GPI indexes computed from UNEB data, the mean division GPI yields an R^2 of 12.3% for the 10-variable Forward model. The CV predictive power of this model is 0% and the RF also predicts 0% of the OOB variance in the GPI. The three methods have a union of 20 variables and intersect on two variables whose coefficients are both statistically insignificant. In plain English, these results appear entirely spurious and are not worth reporting. The GPI calculated on the % of students scoring first division yields better results. Here the in-sample fit of the 10-variable model is 23%, and, as Figure (32) shows, CV yields a predictive power of 5% and the RF explains 13.4% of the OOB variance.

Table (21) reports these results.

The selection result is quite imprecise, the union of the three methods includes 17 predictors and 4 are in the intersection: Sporting facilities, student poverty, the admission PLE score requirement and teacher gender parity. It is interesting that the number of sporting facilities are positively correlated with gender parity in excellent students, and that more selective schools in terms of the PLE entry requirement score worse on gender parity. The negative correlation between poverty and gender-, on the other hand, poses no surprises, and the positive effect of teacher-gender parity on students gender parity is well studied in the literature (the most important mechanism being role-model effects, especially for girls). The government has also highlighted the need for more female teachers to serve as role-models for girls in its ESSP 2020 (MoES, 2017)).

Table 21: Candidates first division (GPI) (2015-16 avg.)

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	Sporting facilities Index (factor score)	0.084	0.027*	+ Sporting facilities Index (factor score)*	0.068	– Admission PLE score requirement	13.400
2	Student poverty Index (factor score)	0.134	-0.019*	– Student poverty Index (factor score)	0.068	– Student poverty Index (factor score)*	12.400
3	1[A-level offered]	0.150	-0.243*	– Admission PLE score requirement	0.052	+ Teacher education level Index (factor score)	10.520
4	Frequency of breaks	0.166	0.216	– 1[A-level offered]*	0.030	+ Sporting facilities Index (factor score)*	9.860
5	Admission PLE score requirement	0.181	-0.006*	+ Teacher education level Index (factor score)	0.030	+ Teacher gender parity (1=parity, 0=Only 1 gender)	7.500
6	Open teacher vacancies (% of teachers)	0.192	0.002	+ Teacher gender parity (1=parity, 0=Only 1 gender)	0.027	– Log average attendance cost	7.020
7	Teacher gender parity (1=parity, 0=Only 1 gender)	0.201	0.13	+ 1[Externally-supported initiatives or projects]	0.025	+ 1[School has own curriculum elements]	5.350
8	Number of technical aids in use	0.211	-0.02	+ Frequency of breaks	0.022	+ 1[Externally-supported initiatives or projects]	5.050
9	Teacher bad behavior Index (factor score)	0.219	0.009	+ Subject-teacher-composition Index (factor score)	0.022	– Number of main assessment modes	3.750
10	Fraction teachers leaving every year	0.226	-0.39	– Fraction teachers leaving every year	0.020	+ School location (1. Urban, 2. Sub-Urban, 3. Rural)	3.670

6.4 Predictors of Career Path

This final section reviews predictors of the career path of A-level students. The two variables considered, % of A-level graduates enrolling in university and % beginning to work directly are based on simple estimates of the respondent but are surprisingly well correlated with other performance measures, and with each-other ($r = -0.58$). The two measures capture in a sense both ends of the spectrum of possibilities a Ugandan A-level student has, or might not have, after graduation, but leave out a large array of possibilities in the middle. In particular, it is reasonable to expect that most students that do not enroll in university or start working will enroll in a TVET programme, teacher training college, apprenticeship or some other form of non-university further education.

Table (22) presents the results for % enrolling in university. The 10 variable model reaches an R^2 of 41.2%, which is again a reasonable in-sample fit, and the accompanying statistics in Figure (33) show that 30% of the variance can be predicted out-of-sample using 9 variables, and the RF explains 34% of the OOB variance. Consequentially the selection results are again reasonably precise, with 14 variables in the union and 6 in the intersection of the three methods. The 6 intersection variables are: The % of boarding students, teacher education, student poverty, the government funding share, the fraction of teachers with a pedagogic qualification and the admission PLE score requirement. The signs are all in the anticipated direction and all 6 variables are significant at the 5% level. Next to these 6 intersection variables, LASSO and the RF intersect on the attendance cost and the number of issues hindering instruction as important predictors of university enrollment, and the LASSO also considers internal monitoring and the school being in a good condition as important.

Table 22: A-level graduates enrolling in university (%)

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	Boarding students (%)	0.225	0.143*	+ Boarding students (%)*	12.639	+ Boarding students (%)*	22.500
2	Teacher education level Index (factor score)	0.306	2.016*	+ Teacher education level Index (factor score)*	11.516	+ Log average attendance cost	21
3	Student poverty Index (factor score)	0.332	-1.252*	− Student poverty Index (factor score)*	10.493	− Government funding share (%)*	18.210
4	Government funding share (%)	0.350	-0.215*	− Admission PLE score requirement*	8.712	− Admission PLE score requirement	15.500
5	1[School in good condition]	0.372	10.685*	− Government funding share (%)*	7.938	+ Technological equipment Index (factor score)	14.720
6	Student-Teacher and Class-Size Index	0.385	0.447*	+ Log average attendance cost	6.590	− Student poverty Index (factor score)	14.700
7	Fraction teachers with pedagogic qualification	0.395	9.048*	+ 1[School in good condition]*	4.542	+ Teacher education level Index (factor score)*	12.990
8	Internal monitoring Index (factor score)	0.402	0.861*	+ Internal monitoring Index (factor score)	4.542	+ Fraction teachers with pedagogic qualification*	12.760
9	Admission PLE score requirement	0.408	-0.363	− Number of issues hindering instruction	3.771	− Number of issues hindering instruction	7.740
10	1[Student appeal system]	0.412	4.903	+ Fraction teachers with pedagogic qualification*	3.771	+ Minimum teacher monthly salary (1000 UGX)	7.570

Table (23) shows the results for the % of A-level students beginning to work directly following graduation. The in-and out-of-sample fit is about 4-5% less than for Table (22), as Figure (34) confirms. The union of the three methods contains 15 variables, and 5 variables populate the intersection: The % of boarding students, teacher education, teacher bad behavior, teacher minimum salary and student absenteeism. In addition, LASSO and the RF intersect on student poverty, the government funding share and technological equipment. LASSO also selects the admission PLE requirement and the frequency of breaks.

Table 23: A-level graduates beginning to work (%)

#	Forward	F-R^2	F-Coef	LASSO	λ	Random Forest	%IncMSE
1	Boarding students (%)	0.195	-0.144*	− Boarding students (%)*	9.942	− Boarding students (%)*	24.660
2	Teacher education level Index (factor score)	0.270	-1.906*	− Teacher education level Index (factor score)*	9.059	− Log average attendance cost	17.090
3	Frequency of breaks	0.303	34.822*	+ Admission PLE score requirement	6.244	− Teacher education level Index (factor score)*	16.870
4	Teacher bad behavior Index (factor score)	0.332	1.073*	+ Student poverty Index (factor score)	6.244	+ Student poverty Index (factor score)	11.780
5	Student absenteism (%)	0.346	0.424*	+ Student absenteism (%)*	6.244	− Technological equipment Index (factor score)	10.740
6	Fraction teachers leaving every year	0.360	-48.005*	+ Frequency of breaks*	4.723	+ Teacher bad behavior Index (factor score)*	9.580
7	Admission PLE score requirement	0.370	0.313	+ Teacher bad behavior Index (factor score)*	4.304	+ Student absenteism (%)*	8.980
8	Minimum teacher monthly salary (1000 UGX)	0.376	-0.013	+ Government funding share (%)	3.921	− Extracurricular activities Index (factor score)	7.720
9	Number of main assessment modes	0.381	-1.934*	− Minimum teacher monthly salary (1000 UGX)	3.573	+ Government funding share (%)	7.530
10	Number of administrative student records kept	0.387	1.359	− Technological equipment Index (factor score)	3.256	− Minimum teacher monthly salary (1000 UGX)	7.310

Across the two tables, all methods single out the % of boarding students and teacher education as the key predictors of career path, with boarding schools and schools with highly educated teachers sending significantly more students to university and less directly into the labor market. In addition, the LASSO's of both tables intersect on student poverty, admission PLE requirement (i.e. internal ability or previous education), and the government funding share as key environmental variables shaping the career paths of students. It is interesting to notice that aside from these variables on which Table (22) and Table (23) intersect, Table (22) predominantly emphasizes variables indicating the quality of education, such as the fraction of teachers with a pedagogic qualification, education cost, the extent of internal monitoring end evaluation, the condition of the school, while Table (23) rather emphasizes structural issues in education like teacher bad behavior, student absenteeism, minimum teacher salary and technological infrastructure.

7 Conclusion

This study set out, in the first part, to review the state of secondary education in Uganda and expose key challenges and the government policy approaches taken towards tackling these challenges. The key challenges distilled where to (I.) Increase the primary to secondary transition rate, which at 63.2% is still far below the governments declared goal of achieving USE; (II.) Increase the lower-secondary completion rate, which remained very low at currently 36.2% over the last couple of years; (III.) Ensure the quality and relevance of secondary education to adequately prepare students for the job-market and higher-education, critical aspects here are enhanced teacher training and monitoring, adjustment of the curriculum and provision of adequate scholastic materials, and (IV.) To bridge the rural-urban divide, especially in the Northern and North-Western Regions, and to manage the chasms in terms of access, performance, curriculum, and cost, between government secondary schools and the large and still growing private sector. The policy priorities of the government, as set in the ESSP 2020, appear to be roughly in the right place in tackling these issues. The focus is on quality, which is necessary, but for achieving USE the government should not neglect the lingering access issues in terms of primary to secondary transition and especially the low lower-secondary completion rates, both of which are underemphasized in the new plan. The

plan also allocates larger budget shares to BTVET and tertiary education, while slightly reducing the secondary budget. Larger investments in tertiary education might not be highly efficient when secondary completion is still that low, especially if, as official statistics have indicated, the job market still cannot straight-away accommodate a condiserable fractin of the university graduates.

In the second part, an empirical analysis of a detailed survey of 450 secondary schools, collected by BRAC Uganda in January and February 2018, was presented to aid the setting of priorities and their implementation by determining the secondary school characteristics most closely associated with educational success in various guises. The first part of this analysis looked at determinants of the aggregate academic performance and excellence of secondary students. The very robust findings indicated that at a minimum teacher education, measures to aid poor students, good boarding schools, and a better technological infrastructure are highly conducive to increased student performance on the national UCE exam. The findings also indicated that above all bright incoming students and the amount of financial resources the school receives are dominant in explaining the aggregate performance of students in the UCE, thus these variables need to, at a minimum, be controlled for when comparing schools or educational approaches. The second part of the analysis researched determinants of dropout, repetition and passing the UCE in secondary schools. Although less robust, the findings here indicate that obstacles hindering instruction and a high teacher attrition, probably proxying for larger structural issues and low-self esteem in some schools, are associated with higher dropout rates. It was also found that dropout rates are lower if parents are frequently notified about the performance of their child. Likewise, secondary schools with a cantine and boarding schools were found to have significantly lower dropout rates, which may be rationalized as indicating that students lacking basic needs when in school are more likely to drop out. For repetition, the analysis results suggest that in addition to some of the just mentioned factors, lessons of appropriate length are conducive to reducing repetition rates (which are very low already at around 2% and thus not a big policy priority). The results also suggest that student absenteeism relates negatively, and more science teachers and more teachers with a pedagogic qualification relate positively to students passing the UCE. The third part of the analysis focused on gender equality in performance. The findings here were not very robust, the only possibly robust effects being that increased teacher gender parity increases student-gender parity, that sporting facilities increase gender parity, and that both student poverty and elitism seem to be negatively correlated with gender parity. Gender parity is however already quite high already in Uganda, with an aggregate GPI of 0.9 which is reflected in the sample. The final part of the analysis examined predictors of students career path. Here again, teacher education emerged as the crucial input variable in determining the share of students making it to university. In addition, pedagogically skilled teachers, more internal monitoring and evaluation and a good condition of the school are associated with more graduates enrolling in university, while bad teacher behavior, student absenteeism, low technological infrastructure and low minimum teacher salary are characteristics of schools sending a large share of their students directly into the labor market.

As the author of this study, it is my hope that these findings will receive recognition by BRAC, and possibly by the Ministry of Education and Sports, and be put to appropriate and careful use. Further research could focus more strongly on determinants of income-inequality in performance within schools and regional imbalances in Uganda. Both research agendas are likely to require substantially more and better data.

References

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

EPDC. (2011). *Secondary education: A guide to education project design based on a comprehensive literature and project review.* Education Policy Data Center. Retrieved from https://www.epdc.org/sites/default/files/documents/EQUIP2%20SOAK%20-%20Secondary%20Education.pdf

EPDC. (2013). *Education trends and projections, uganda, 2000-2025.* Education Policy and Data Center. Retrieved from https://www.oecd.org/dev/emea/40578334.pdf

EPDC. (2014). *National education profile, uganda.* Education Policy and Data Center. Retrieved from https://www.epdc.org/sites/default/files/documents/EPDCNEP_Uganda.pdf%0Ahttp://www.epdc.org/sites/default/files/documents/EPDCNEP_Nigeria.pdf%0Ahttp://www.epdc.org/sites/default/files/documents/EPDCNEP_Uganda.pdf

Freedom-House. (2018). *Freedom house political indicators.* Freedom House, data by Quality of Government Institute. Retrieved from https://qog.pol.gu.se/data/datadownloads/qogstandarddata

GPE, u. a. y., title = Education in Uganda. (n.d.).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Kavuma, R. (2018). *Explainer: The education system in uganda.* the Guardian. Retrieved from https://www.theguardian.com/katine/2010/feb/08/education-system-explainer

Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. *R news*, *2*(3), 18–22.

MoES. (2008). Revised education sector strategic plan 2007-2015. *Uganda Ministry of Education and Sports*. Retrieved from http://planipolis.iiep.unesco.org/sites/planipolis/files/ressources/uganda_revised_education_sector_strategic_plan_2007-2015.pdf

MoES. (2015). A comprehensive report on the universal post primary education & training (uppet /use) & universal post o' level education & training (upolet) national headcount exercise 2015. *Uganda Ministry of Education and Sports*. Retrieved from http://www.education.go.ug/files/downloads/UPPETandUPOLETNationalHeadcount2015WorkingReportMay2016(Fine).pdf

MoES. (2017). *Education and sports sector strategic plan 2017-2020.* Uganda Ministry of Education and Sports. Retrieved from http://www.education.go.ug/files/downloads/EDUCATION%20AND%20SPORTS%20SECTOR%20STRATEGIC%20PLAN.pdf

Nuffic. (2010). *Description of ugandas education system and comparison with the dutch system.* Nuffic. Retrieved from https://www.nuffic.nl/en/publications/find-a-publication/education-system-uganda.pdf

RTF. (2015). *Factsheet: Education in uganda.* Rafiki Thabo Foundation. Retrieved from http://www.rafiki-foundation.org.uk/wp-content/uploads/2011/08/Education-factsheet-Uganda.pdf

Stekhoven, D. J. (2015). missforest: Nonparametric missing value imputation using random forest. *Astrophysics Source Code Library*.

UIA. (2010). *Uganda education profile.* Uganda Investment Authority. Retrieved from http://www.ugandainvest.go.ug/uia/images/Download_Center/SECTOR_PROFILE/Education_Sector_profile.pdf

UNDP. (2016). *Uganda country profile.* United Nations Development Programme. Retrieved from http://hdr.undp.org/en/countries/profiles/UGA#

UNESCO, T. (2014). *Learning: Achieving quality for all, efa global monitoring report 2013/14.* Paris, UNESCO.

URDT. (2018). *Background to the education sector in uganda.* Uganda Rural Development & Training Programme. Retrieved from http://urdt.net/index.php/urdt-schools/184.html

World-Bank. (2018a). *Education statistics (edstats) database.* The World Bank. Retrieved from http://databank.worldbank.org/data/reports.aspx?source=education-statistics-~-all-indicators

World-Bank. (2018b). *World bank service delivery indicators (sdi).* The World Bank. Retrieved from http://datatopics.worldbank.org/sdi/

World-Bank. (2018c). *World development indicators.* The World Bank. Retrieved from https://data.worldbank.org/products/wdi

World-Bank. (2018d). *The worldwide governance indicators (wgi) project.* The World Bank. Retrieved from http://info.worldbank.org/governance/WGI/#home

Appendix

The Appendix contains first a summary and additional analysis of the predictor variables, then some performance metrics and additional statistics for each of the results tables in the analysis section, and finally histograms of both the analysis dataset and the cleaned survey, followed by the survey itself. Additional information may be requested from the author at sebastian.krantz@gmx.de.

Summary of the Predictors

Table 24: Summary of Predictors ($\mathbf{X}$), $N = 422$, $K = 55$

Predictor Name	N	Mean	Median	SD	Min	Max
Student-Teacher and Class-Size Index	422	20.5	19.97	6.57	4.68	45
Log average attendance cost	422	6.76	6.89	0.92	3	9.21
1[A-level offered]	422	0.89	1	0.31	0	1
Total number of students	422	820.6	700	574.83	80	4500
Boarding students (%)	422	59.48	70	36.72	0	100
Female students (%)	422	51.99	51.07	20.4	0	100
School location (1. Urban, 2. Sub-Urban, 3. Rural)	422	2.12	2	0.74	1	3
Age of school (years)	422	29.2	21.5	21.85	2	117
Teacher gender parity (1=parity, 0=Only 1 gender)	422	0.52	0.5	0.22	0	1
1[School in good condition]	422	0.9	1	0.3	0	1
Government funding share (%)	422	18.6	2	25.45	0	100
1[Externally-supported initiatives or projects]	422	0.35	0	0.48	0	1
Admission PLE score requirement	422	23.15	25	7.53	5	36
Number of admission criteria	422	2	2	0.99	1	8
Number of fee-reduction criteria	422	1.28	1	0.95	0	6
Technological equipment Index (factor score)	422	4.58	4.22	2.59	0.27	16.63
Sanitary facilities Index (factor score)	422	4.32	3.83	2.24	-1.53	14.17
Library facilities Index (factor score)	422	4.51	3.46	6.61	1	94.17
Recreational environment Index (factor score)	422	2.96	1	2.58	1	12.15
Sporting facilities Index (factor score)	422	4.46	4.13	2.59	1	23.73
1[Cantine available]	422	0.92	1	0.27	0	1
1[Teacher posesses the only textbook]	422	0.5	0.5	0.5	0	1
Number of issues hindering instruction	422	2.12	2	1.78	0	8
Professional development Index (factor score)	422	4.24	3.45	2.76	1	14.42
Internal monitoring Index (factor score)	422	3.44	2.26	2.73	-0.26	14.5
PTA meetings per school year	422	2.01	2	1.81	0	18
External school inspections per year	422	3.92	3	2.91	0	25
Parents informed on student performance (N/year)	422	1.96	2	0.7	1	6
1[School has own curriculum elements]	422	0.04	0	0.2	0	1
1[Student appeal system]	422	0.85	1	0.36	0	1
Number of administrative student records kept	422	3.51	3	1.36	1	7
Extracurricular activities Index (factor score)	422	4.26	3.67	2.8	-1.35	21.12
Daily time spent in school	422	7.36	7.21	0.87	3.57	10.64
Length of lessons (minutes)	422	52.01	40	18.27	30	120
Frequency of breaks	422	0.28	0.25	0.12	0.12	1
Groupwork & leadership Index (factor score)	422	5.17	6.33	3.54	-4.33	20.03
Number of main assessment modes	422	2.11	2	1.05	1	5
Number of assessments (per term per subject)	422	2.69	3	0.91	2	9
Homework hours (per S4 student per week)	422	6.27	5	4.76	0	25
Number of subjects taught	422	14.88	15	2.66	7	26
Students strikes (N/last 3 years)	422	0.1	0	0.35	0	2
Number of technical aids in use	422	2.75	3	1.49	0	5
Lowest grade to start specializing	422	2.92	3	0.28	2	4
Student poverty Index (factor score)	422	5.65	6.08	2.52	-2.73	9.65
Student misconduct Index (factor score)	422	4.55	4.04	3.46	1	47.09
Student absenteism (%)	422	6.24	5	7.22	0	50
Subject-teacher-composition Index (factor score)	422	4.22	3.6	3.02	-0.16	26.44
Teacher bad behavior Index (factor score)	422	3.83	3.04	2.72	1	19.6
Teacher experience Index	422	4.39	4.45	2.4	-0.05	11.9
Teacher education level Index (factor score)	422	5.64	6.08	2.58	-0.89	9.26
Open teacher vacancies (% of teachers)	422	6.01	0	13.01	0	80
Minimum teacher monthly salary (1000 UGX)	422	347.73	300	180.9	80	1950
Median teacher monthly salary (1000 UGX)	422	582.6	525	442.12	135	6125
Fraction teachers with pedagogic qualification	422	0.64	0.65	0.33	0.01	1
Fraction teachers leaving every year	422	0.06	0.04	0.07	0	0.48

Figure 22: Correlation Matrix of Predictors ($\mathbf{X}$), $N = 422$, $K = 55$

Correlations insignificant at the 5% level set to 0. Predictors Grouped using Hierarchical Clustering.

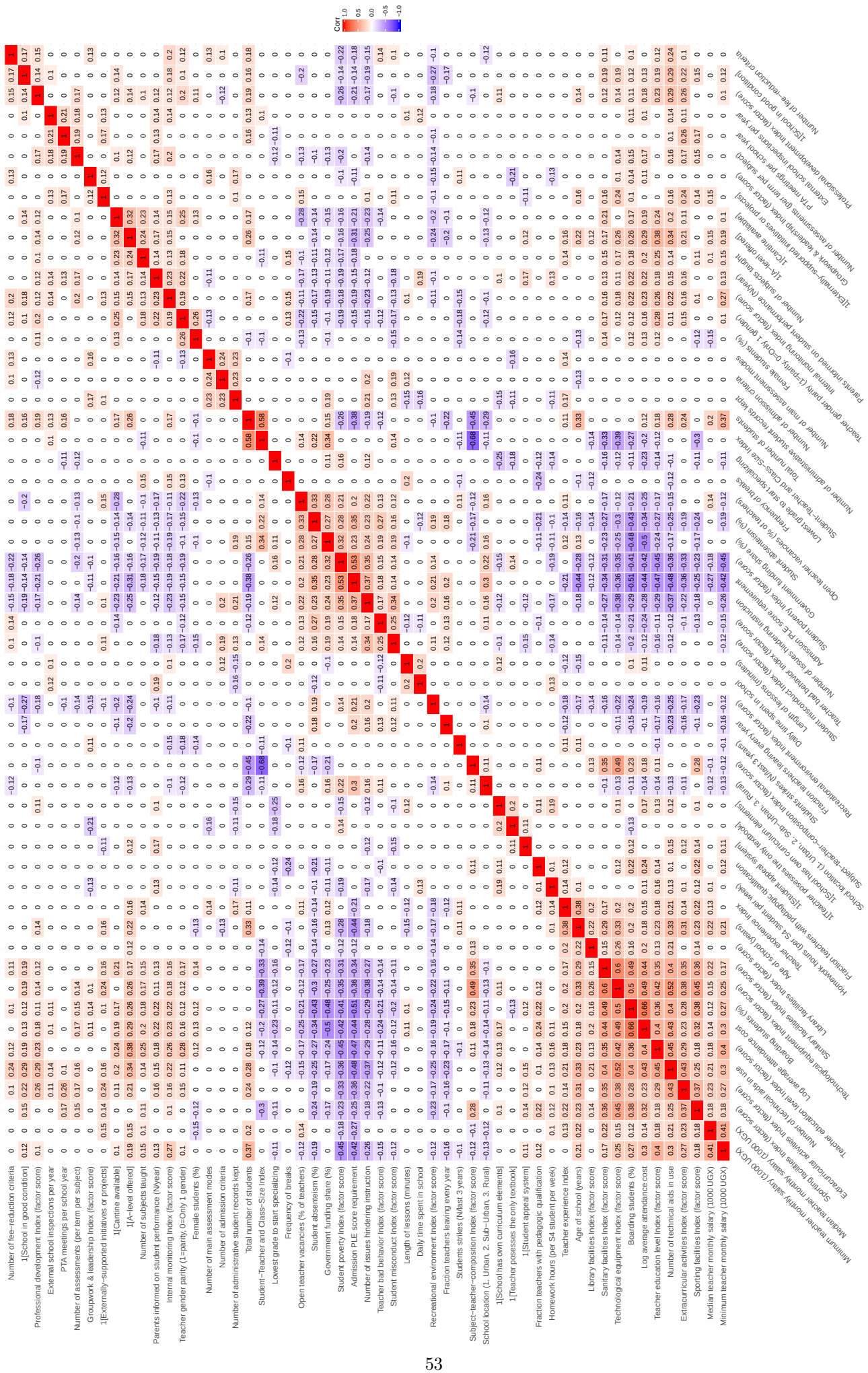


Figure 23: Hierarchical Clustering of Predictors ($\mathbf{X}$), $N = 422$, $K = 55$

Linkage = Complete | Distance Metric = 1 - Abs. Pearsons r

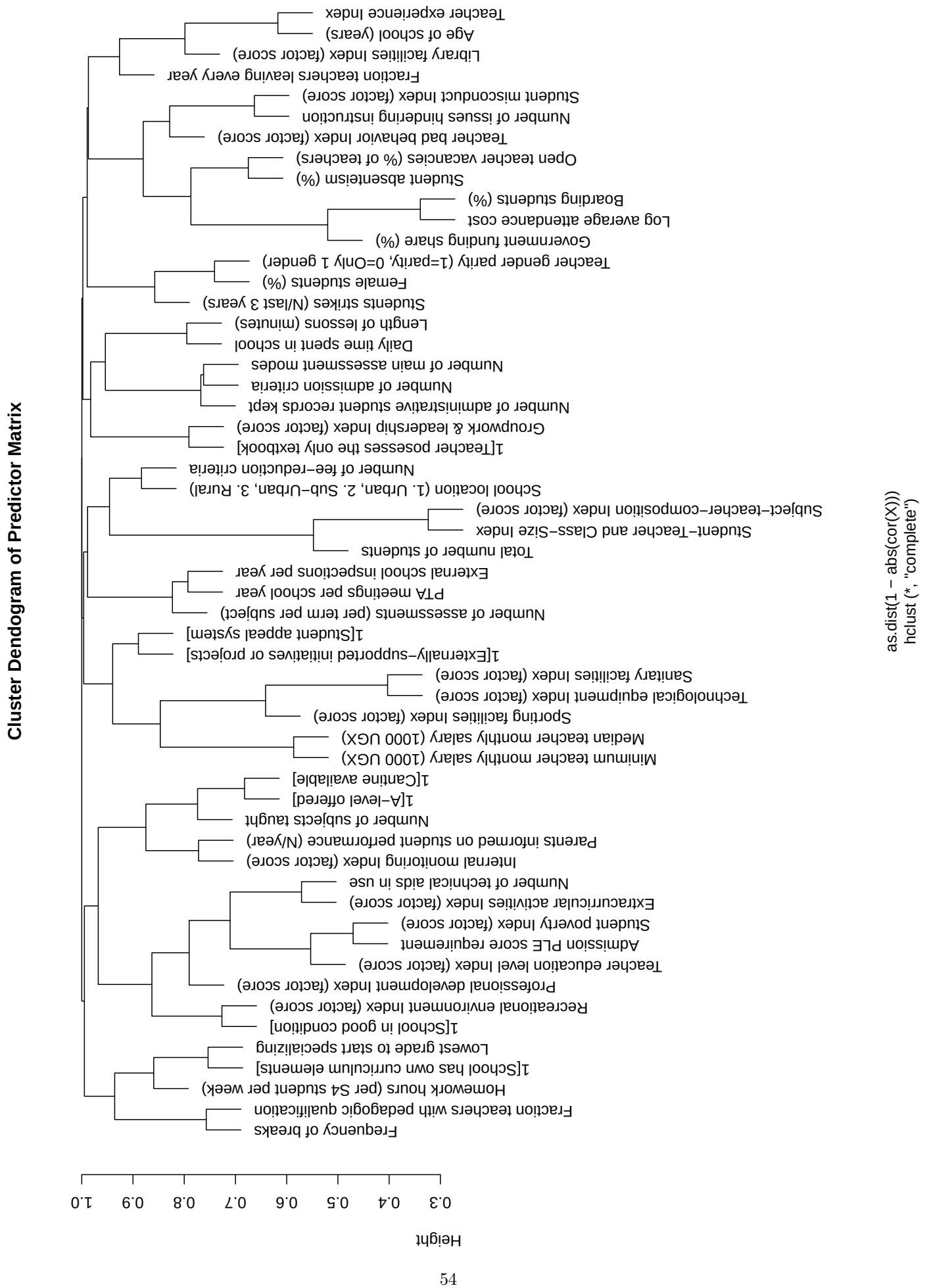


Table 25: PCA on Predictors ($\mathbf{X}$), $N = 422$, $K = 55$, KMO = 0.8

Varimax Rotated, Loadings < 0.4 Omitted, Predictors with no Loadings > 0.4 Omitted

Principal Components:	PC1	PC2	PC3	PC4	PC5
% Variance Explained:	14%	6%	5%	4%	4%
Admission PLE score requirement	0.730				
Number of technical aids in use	-0.660				
Student poverty Index (factor score)	0.630				
Teacher education level Index (factor score)	-0.620				
Total number of students	-0.590	-0.560			
Minimum teacher monthly salary (1000 UGX)	-0.560				
Extracurricular activities Index (factor score)	-0.530				
Number of issues hindering instruction	0.520				
1[A-level offered]	-0.500				
Age of school (years)	-0.500		0.500		
Log average attendance cost	-0.470	0.450			
Boarding students (%)	-0.470	0.540			
Technological equipment Index (factor score)	-0.470	0.640			
Sanitary facilities Index (factor score)	-0.430	0.550			
Professional development Index (factor score)	-0.430				
Student-Teacher and Class-Size Index		-0.830			
Subject-teacher-composition Index (factor score)		0.780			
Sporting facilities Index (factor score)		0.480			
Government funding share (%)		-0.470			
Teacher experience Index			0.480		
Open teacher vacancies (% of teachers)			0.470		
Teacher gender parity (1=parity, 0=Only 1 gender)			-0.430		
Female students (%)			-0.410		
Length of lessons (minutes)				0.480	
Lowest grade to start specializing				-0.400	
Daily time spent in school				0.430	
Number of admission criteria					-0.590
Number of administrative student records kept					-0.570
Groupwork & leadership Index (factor score)					-0.460
Number of main assessment modes					-0.530

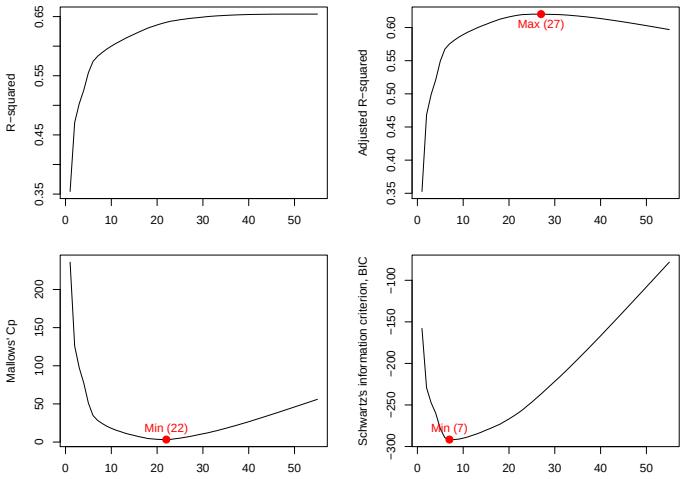
Performance Metrics and Additional Results

Figure 24: Performance Metrics and Additional Results for Table (13)

Mean division, all candidates (2015-16 avg.)

Forward Selection: R^2 and Information Criteria

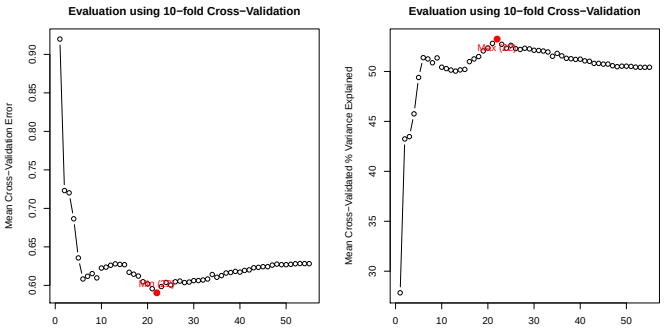
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

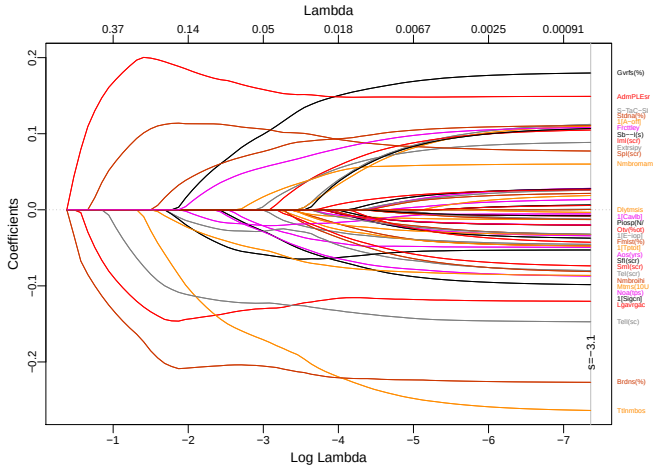
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a Forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 Forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



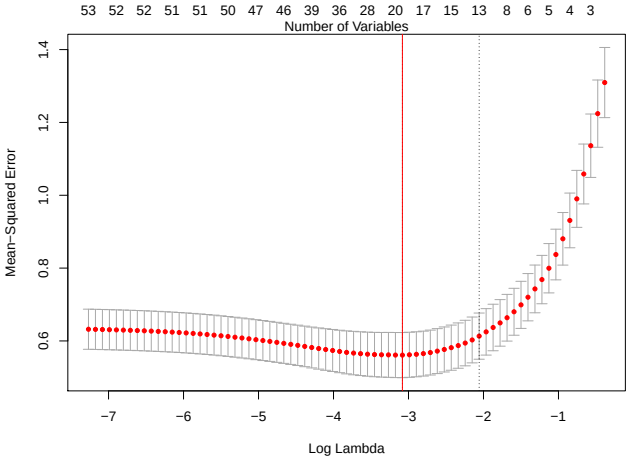
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

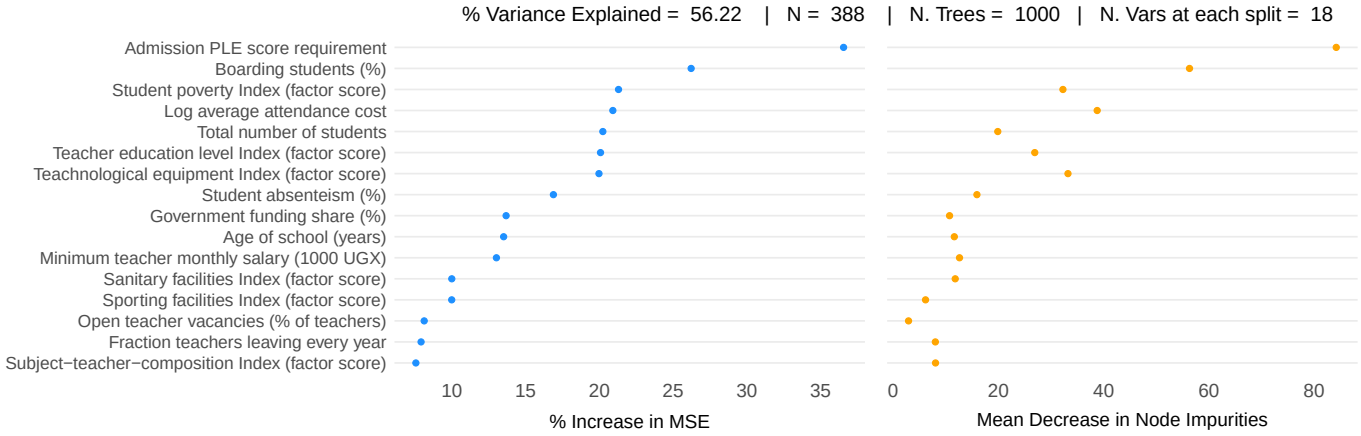
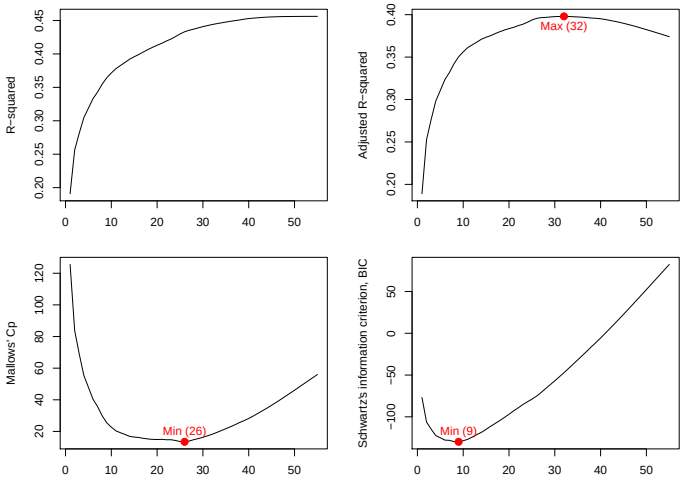


Figure 25: Performance Metrics and Additional Results for Table (14)

UCE Aggregate

Forward Selection: R^2 and Information Criteria

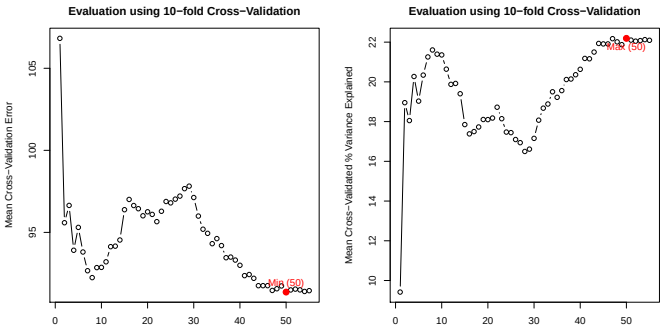
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

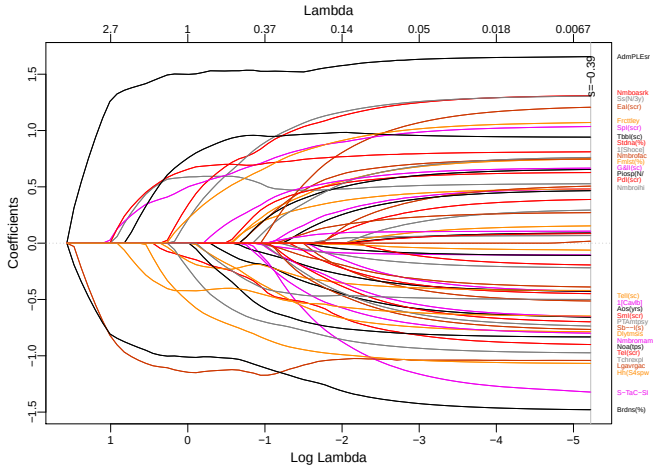
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a Forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 Forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



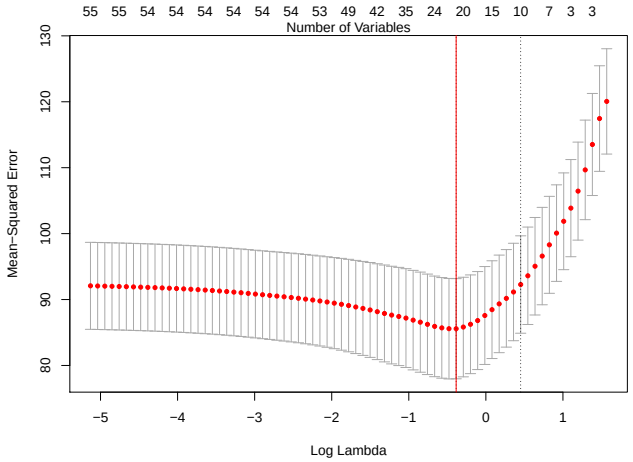
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

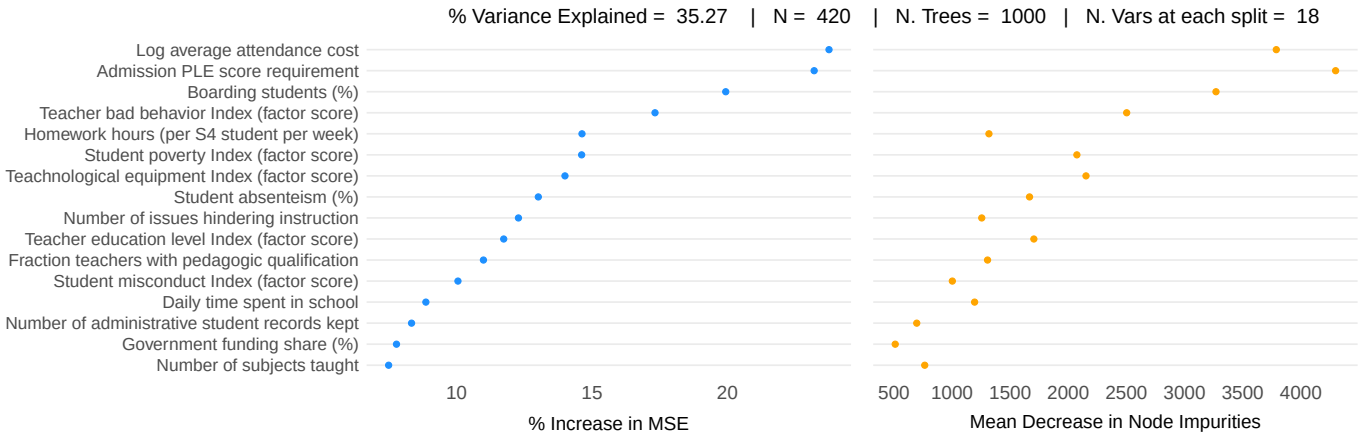
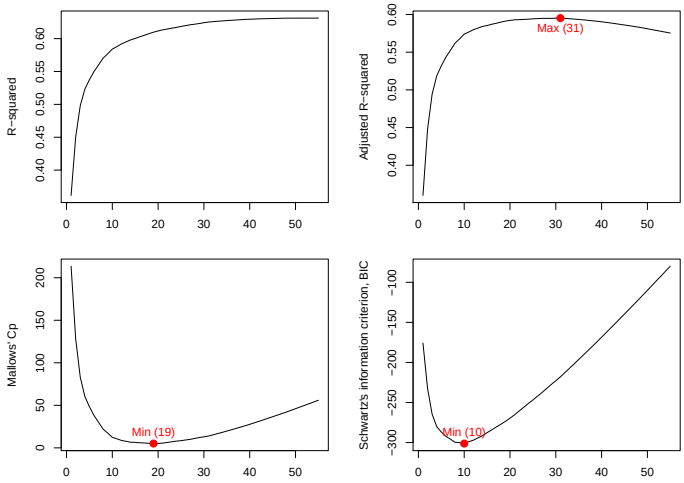


Figure 26: Performance Metrics and Additional Results for Table (15)

Students scoring first division (%)

Forward Selection: R^2 and Information Criteria

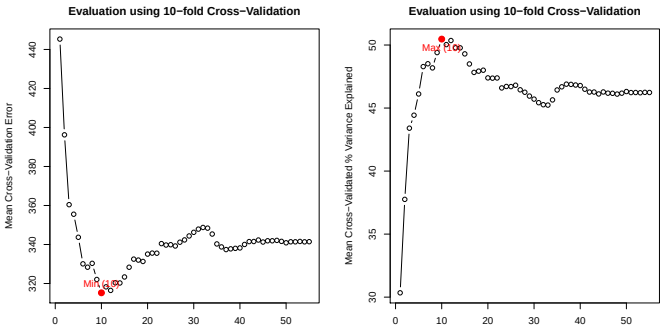
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

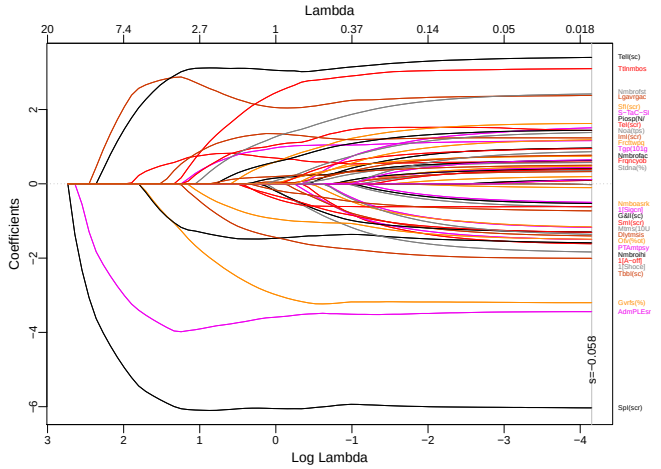
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a Forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 Forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



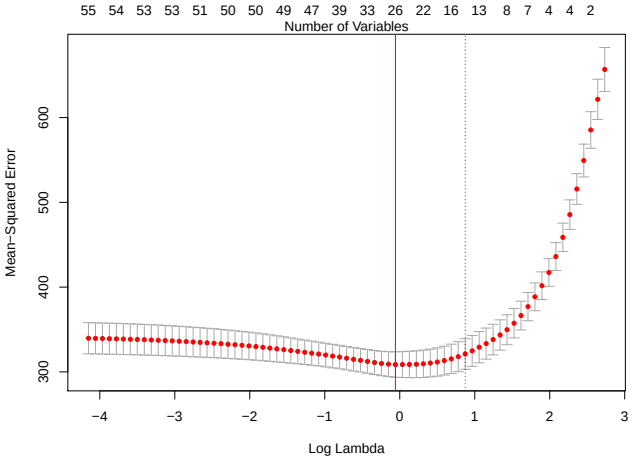
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

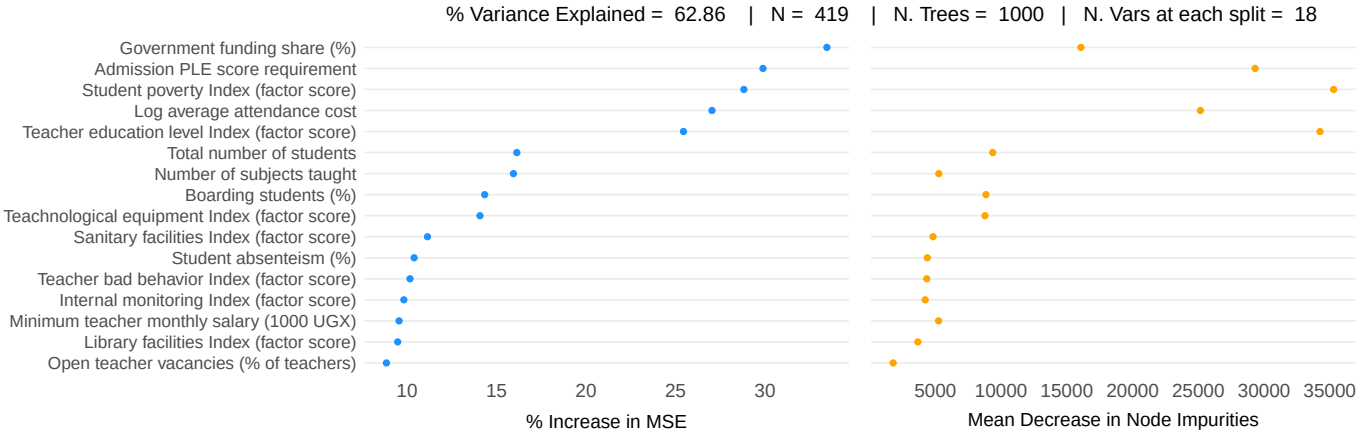
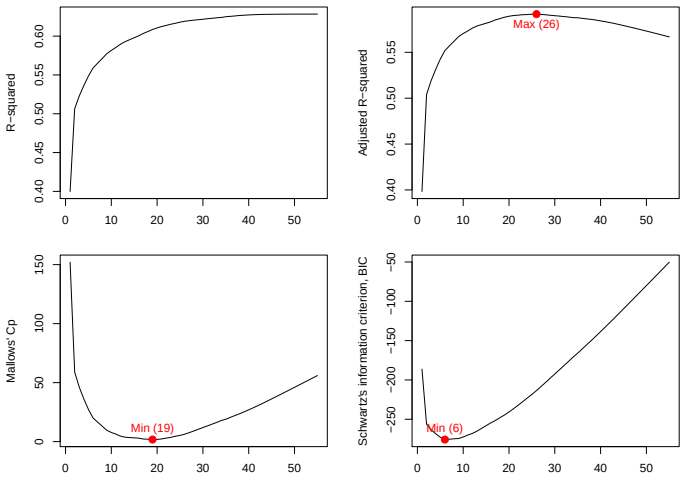


Figure 27: Performance Metrics and Additional Results for Table (16)

Candidates first-division (%) (2015-16 avg.)

Forward Selection: R^2 and Information Criteria

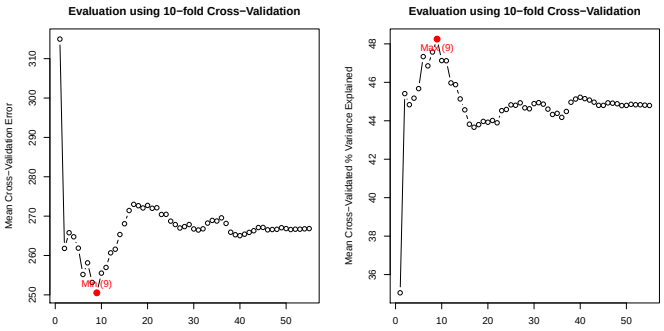
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

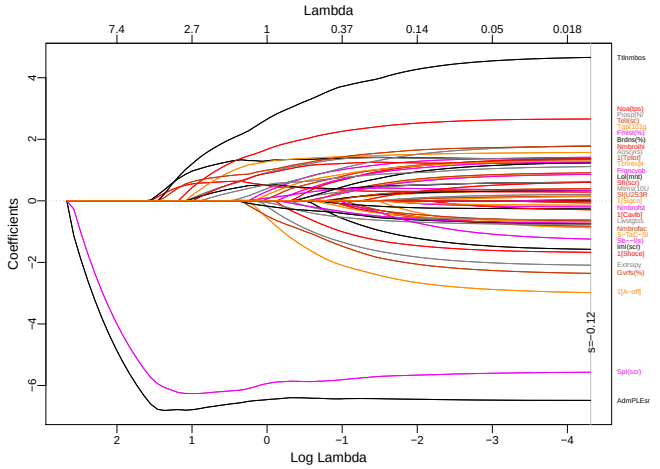
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a Forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 Forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



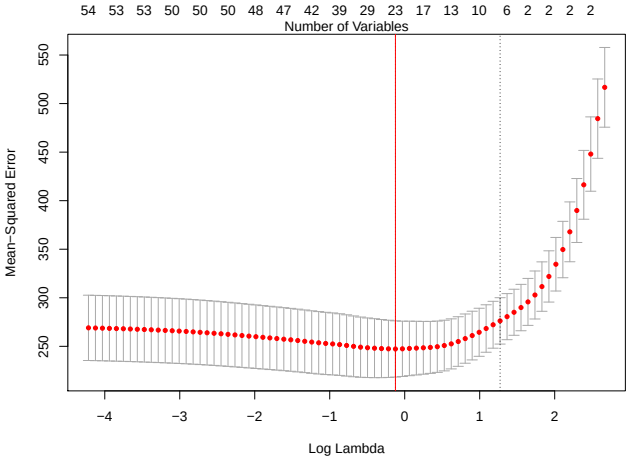
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

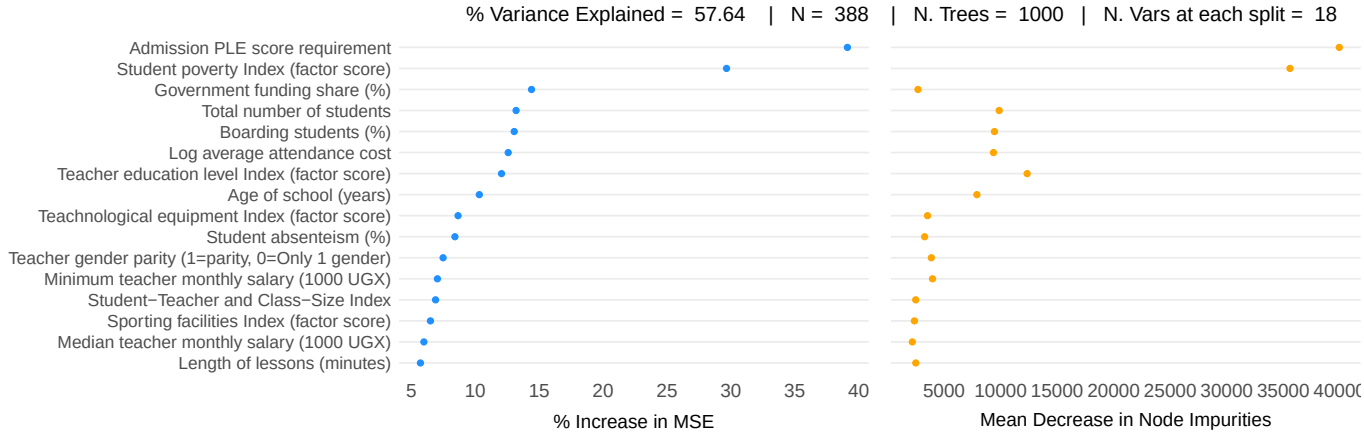
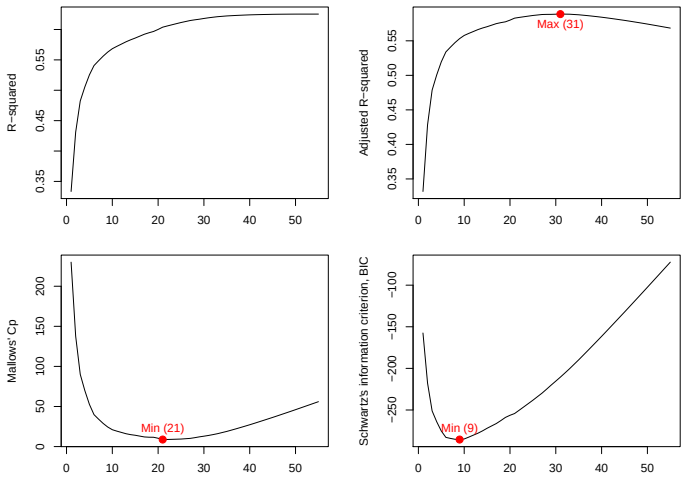


Figure 28: Performance Metrics and Additional Results for Table (17)

UCE performance Index (factor score)

Forward Selection: R^2 and Information Criteria

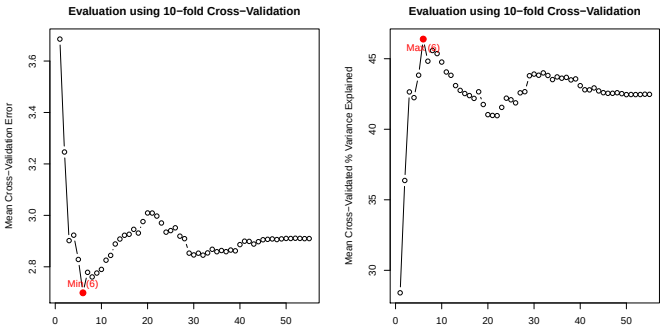
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

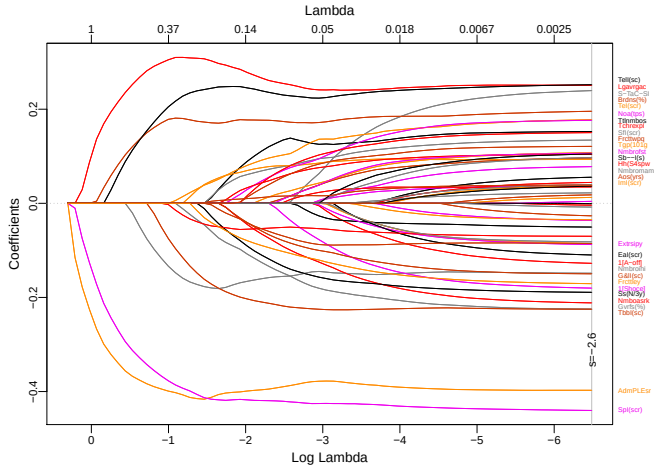
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a Forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 Forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



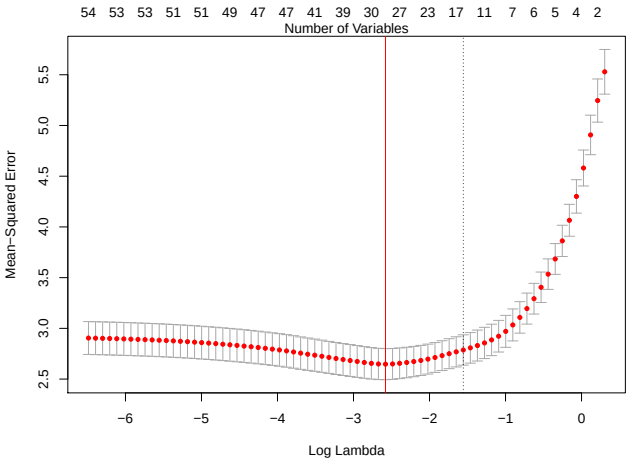
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

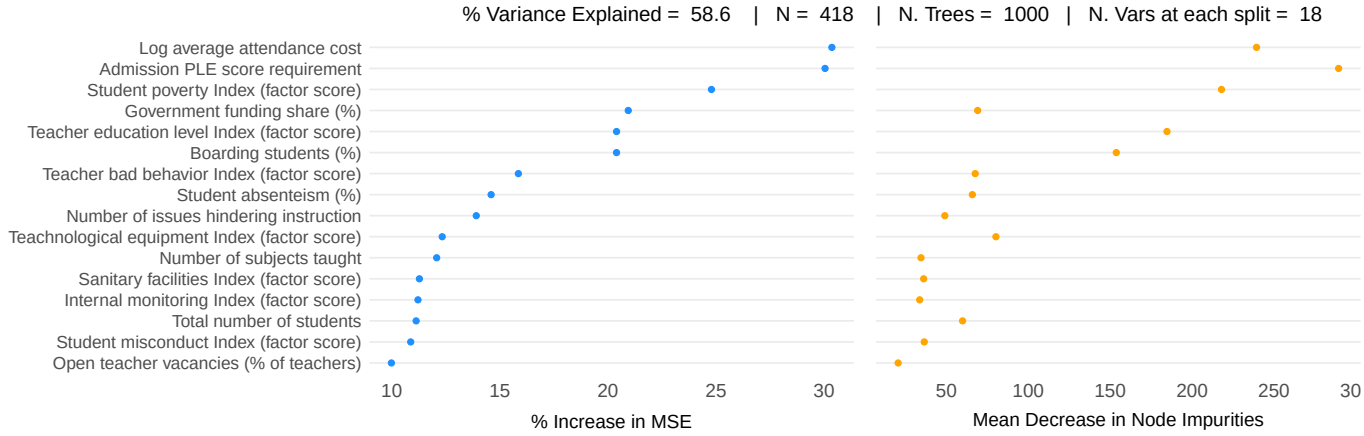
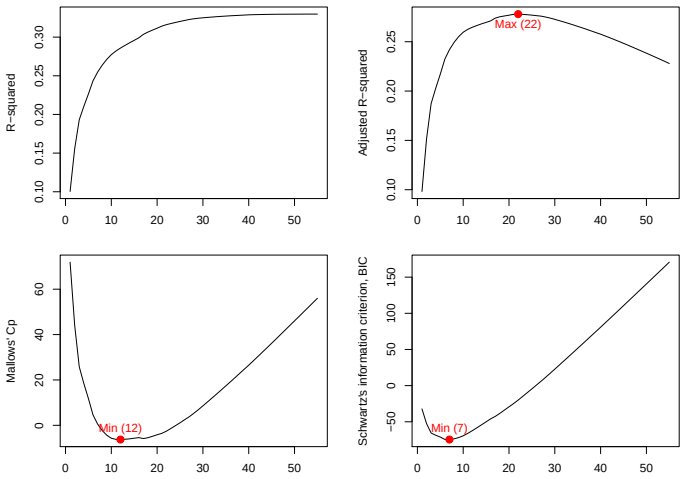


Figure 29: Performance Metrics and Additional Results for Table (18)

Yearly dropout rate (% of students)

Forward Selection: R^2 and Information Criteria

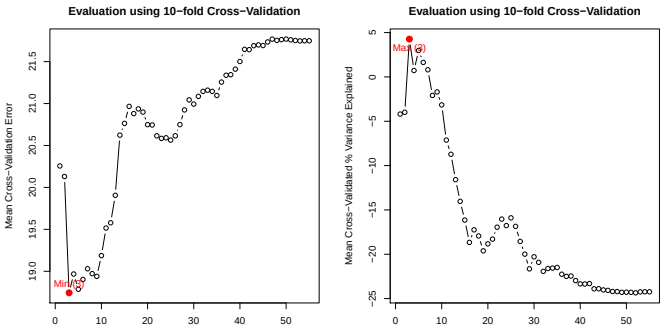
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

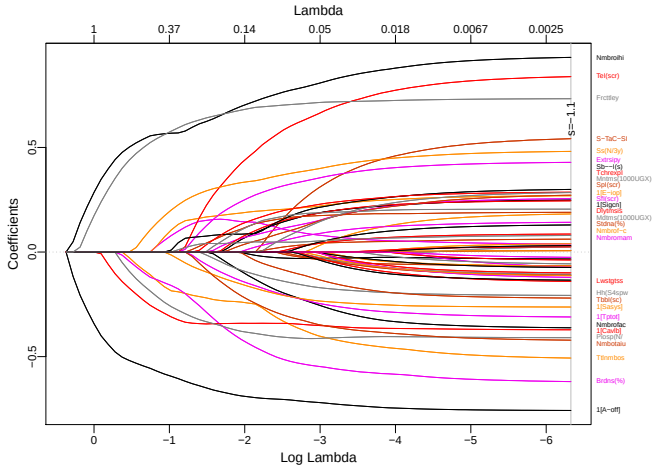
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a Forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 Forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



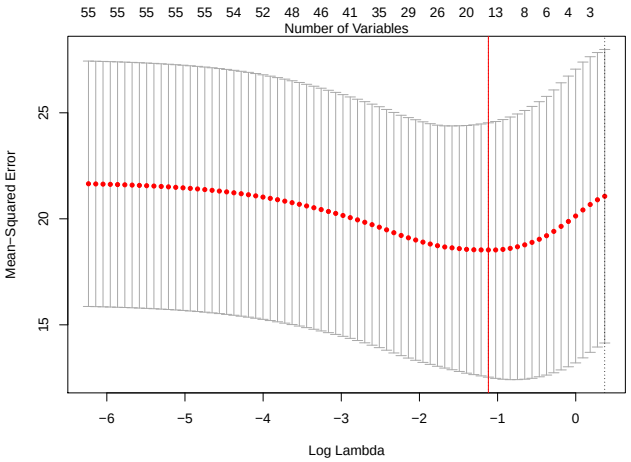
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

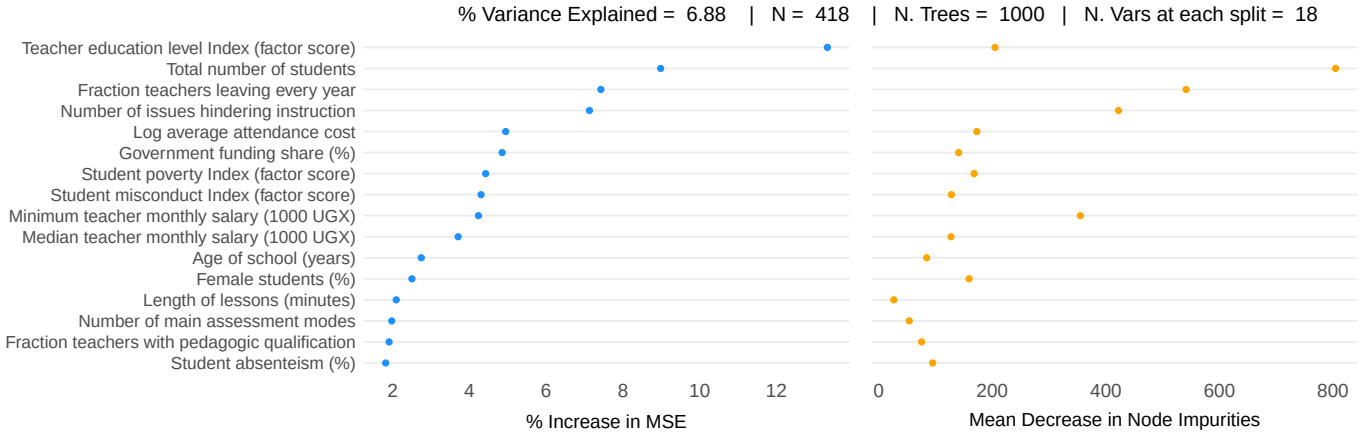
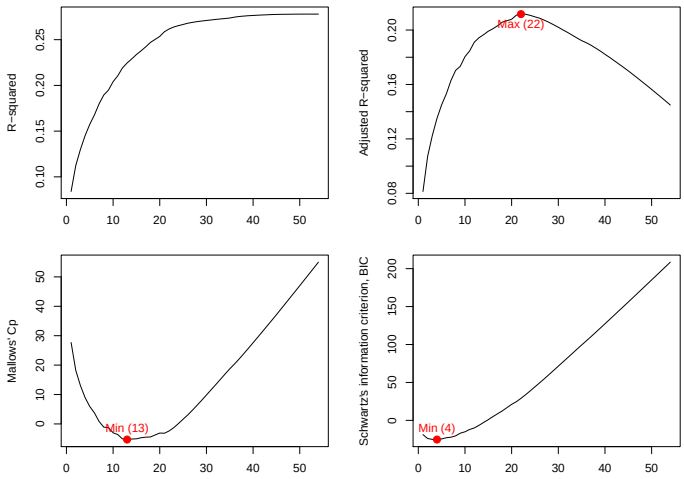


Figure 30: Performance Metrics and Additional Results for Table (19)

Yearly repeater rate, O-level (%)

Forward Selection: R^2 and Information Criteria

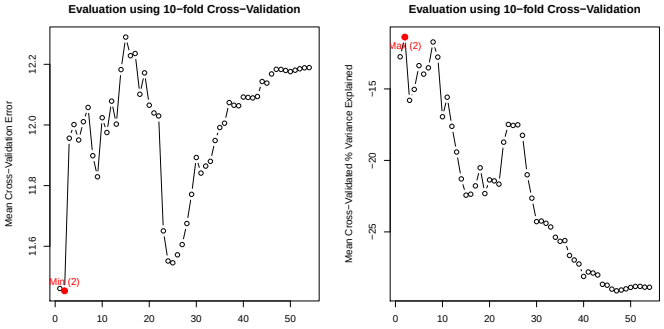
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

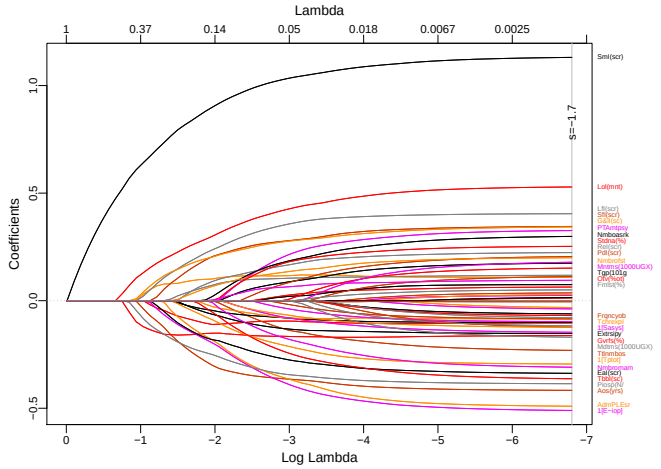
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a Forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 Forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



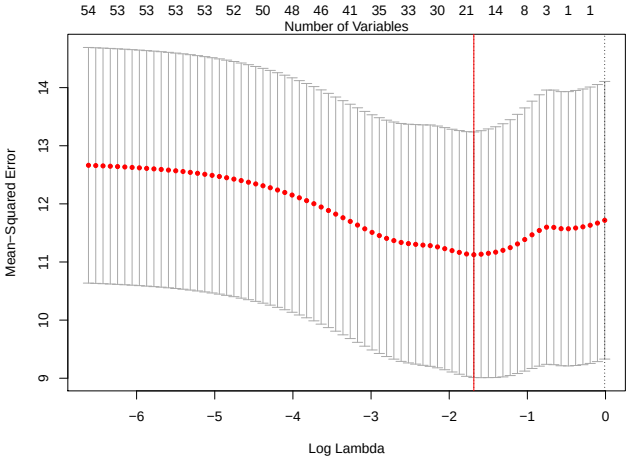
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

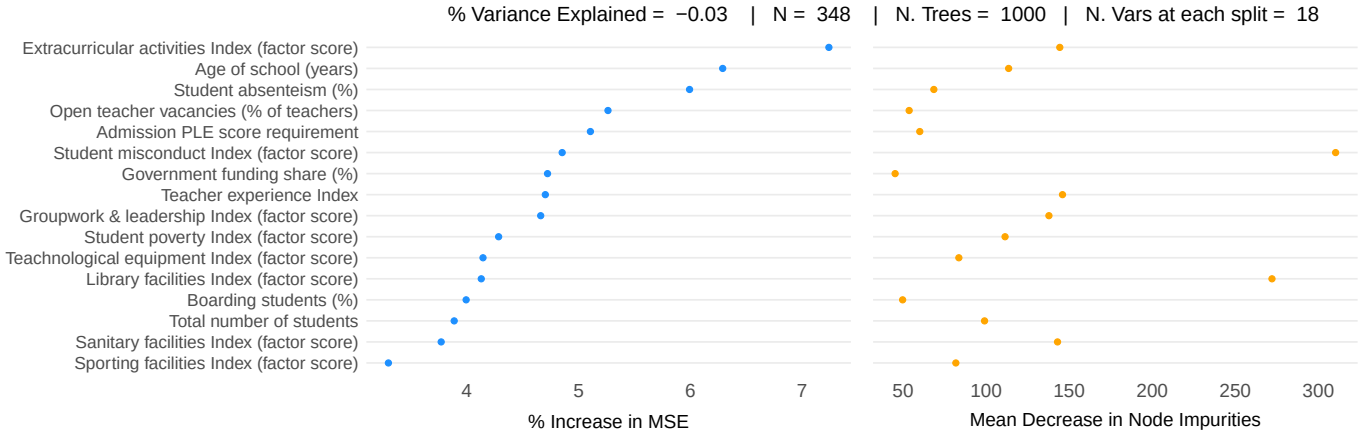
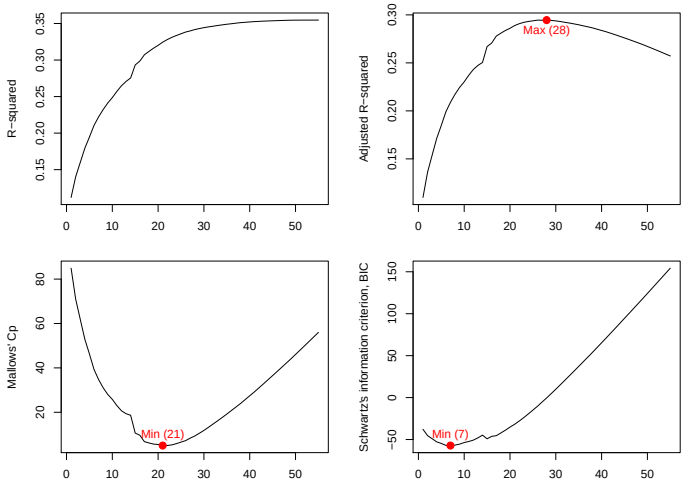


Figure 31: Performance Metrics and Additional Results for Table (20)

UCE Pass Index (factor score)

Forward Selection: R^2 and Information Criteria

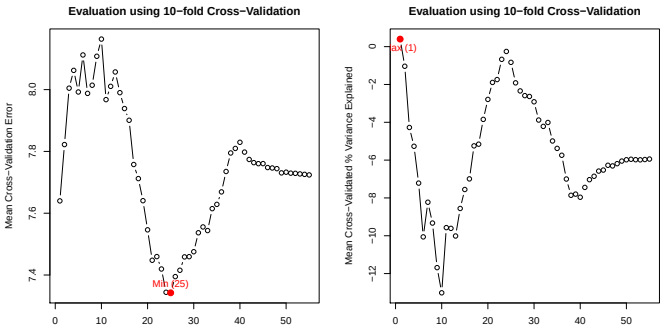
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

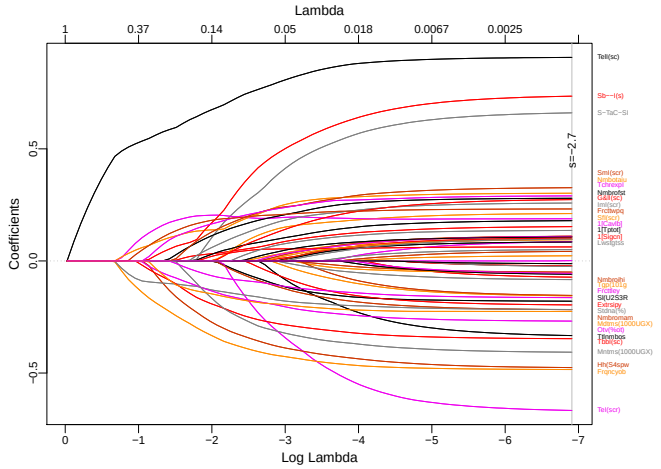
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a Forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 Forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



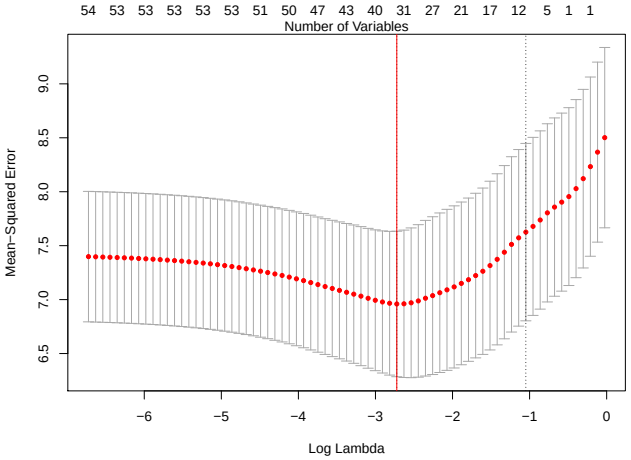
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

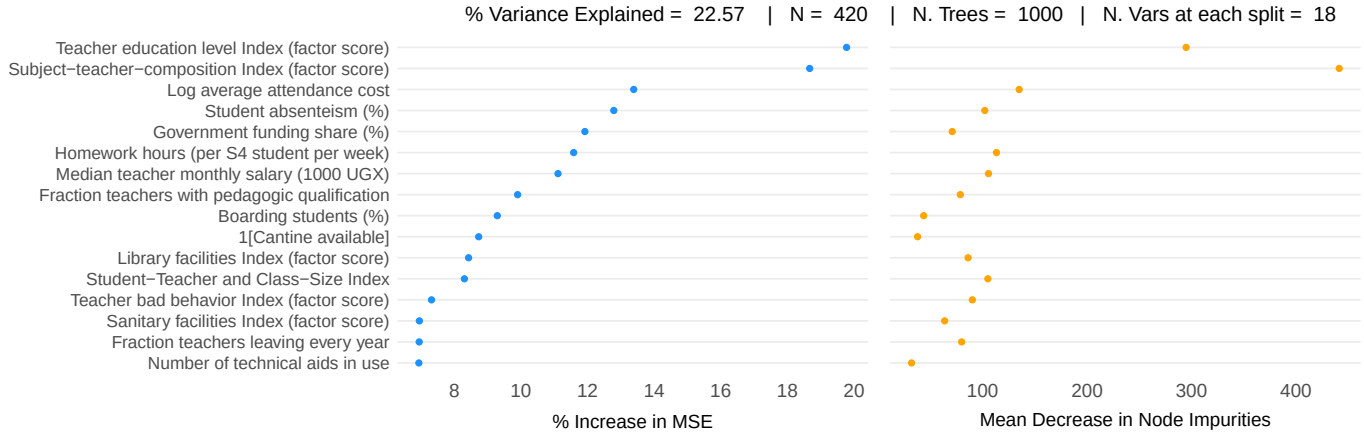
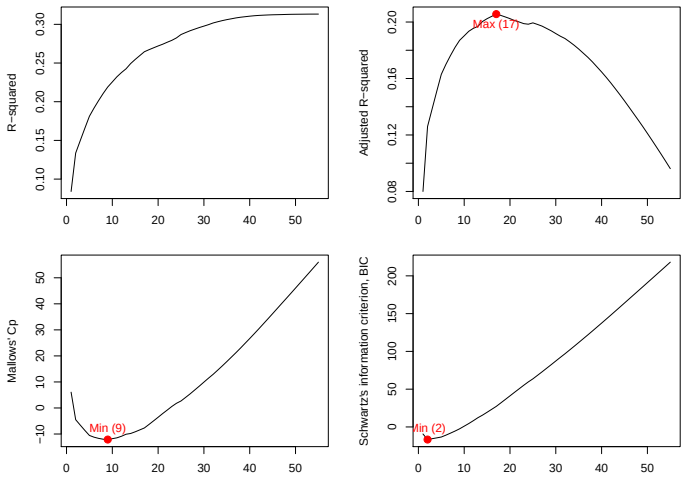


Figure 32: Performance Metrics and Additional Results for Table (21)

Candidates first-division, (GPI) (2015-16 avg.)

Forward Selection: R^2 and Information Criteria

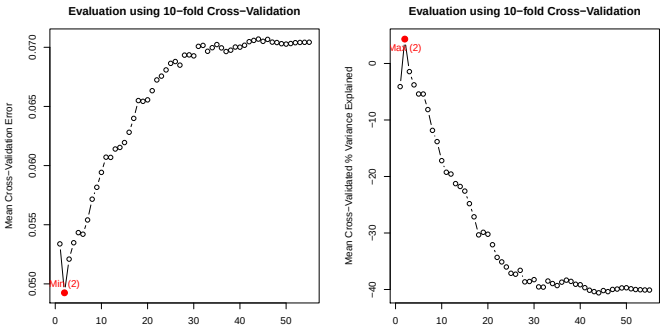
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

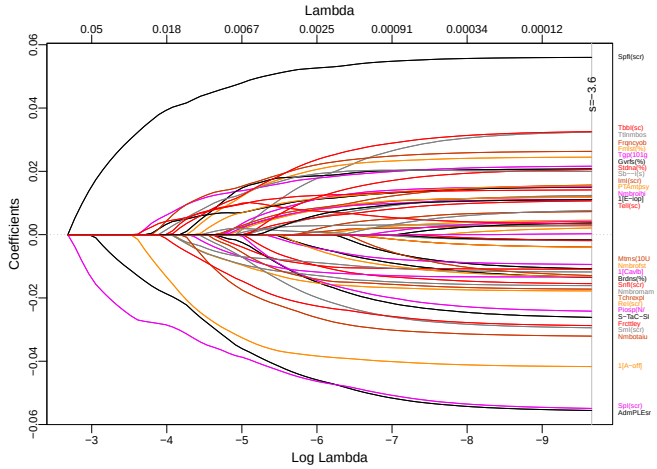
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



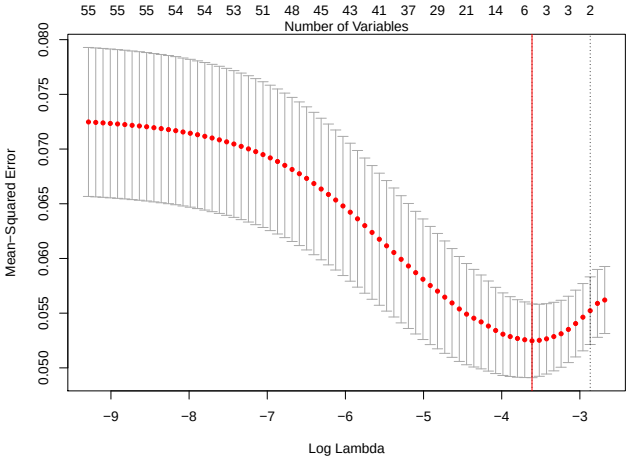
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

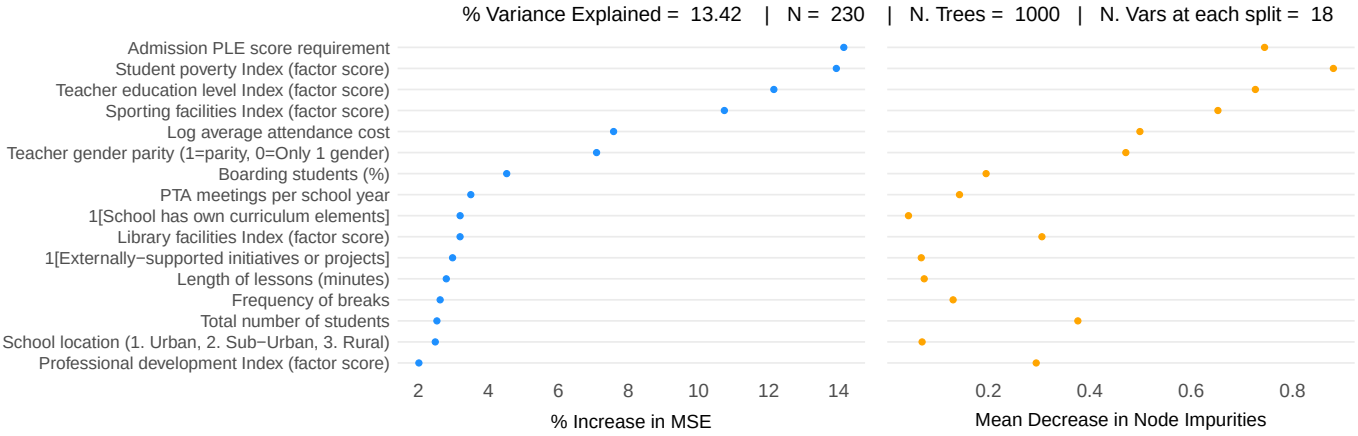
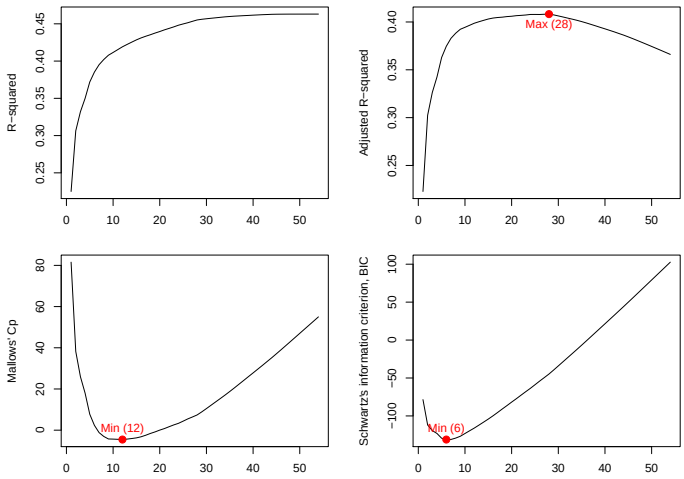


Figure 33: Performance Metrics and Additional Results for Table (22)

A-level graduates enrolling in university (%)

Forward Selection: R^2 and Information Criteria

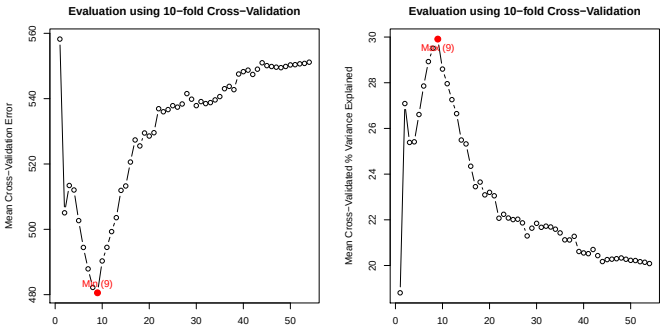
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

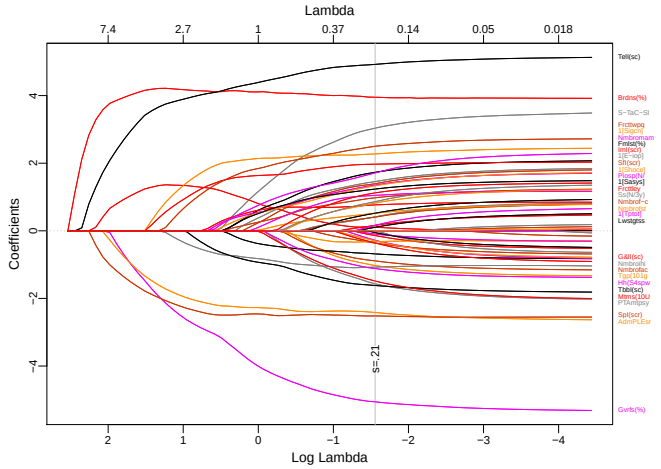
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



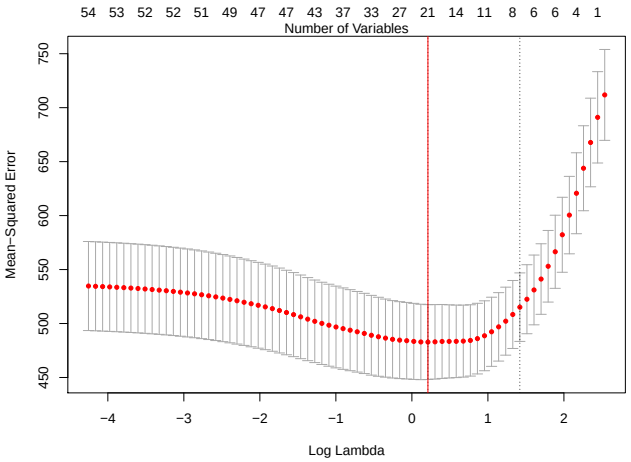
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.

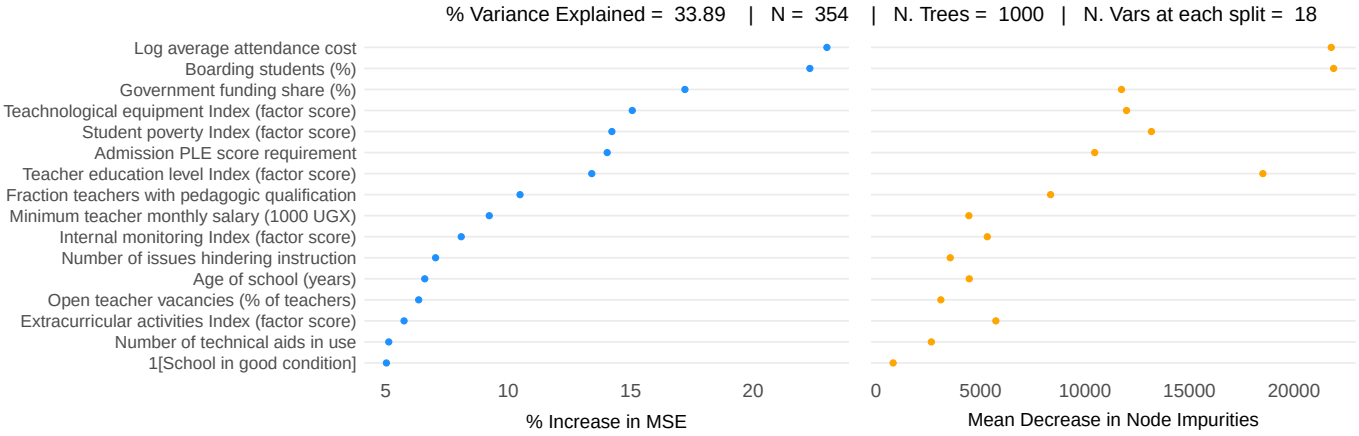
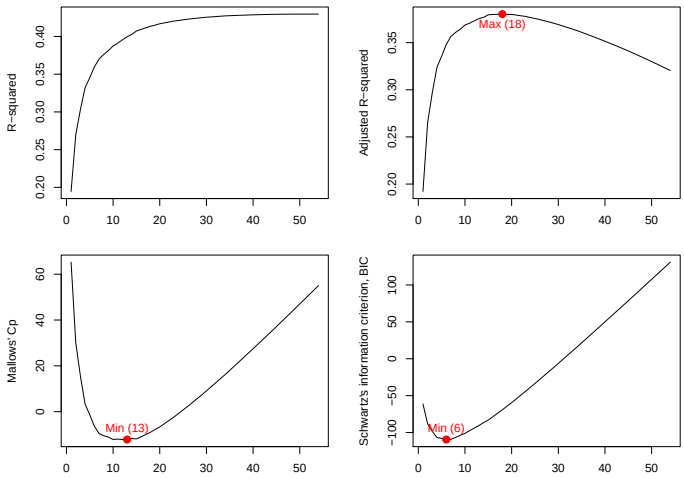


Figure 34: Performance Metrics and Additional Results for Table (23)

A-level graduates beginning to work (%)

Forward Selection: R^2 and Information Criteria

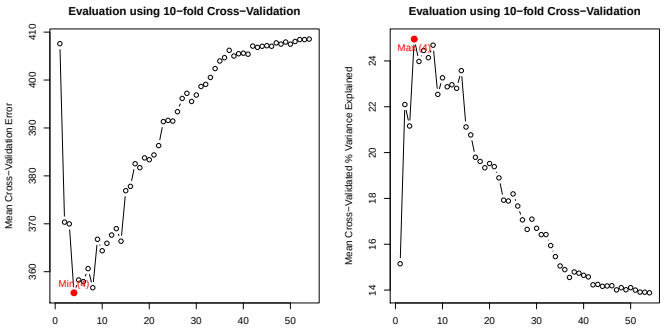
Number of variables (model size) on the x-axis



Forward Selection: 10-Fold Cross-Validation (CV)

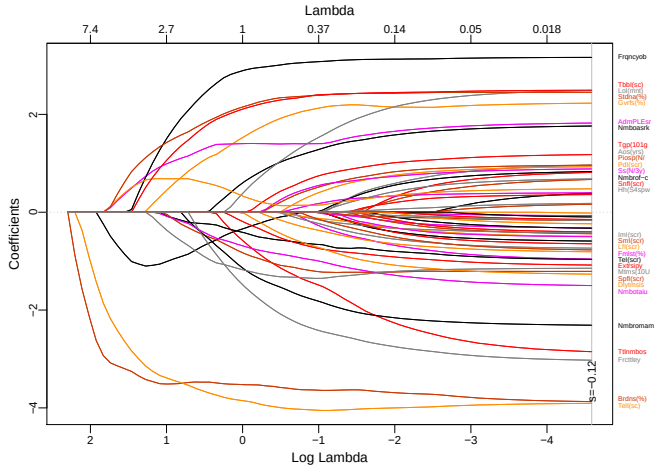
Number of variables (model size) on the x-axis

Note: Cross-validation (CV) and information criteria are used to select the optimal size (i.e. number of variables) of a predictive model so as to avoid the overfitting problem. For 10-fold CV, the observations are cut into 10 slices, a model is built on 9 slices and then used to predict the remaining slice. This is done in the 10 different ways one can do this, the errors are averaged. Thus for the plot below, a forward selected model of each size (1-54) is considered (e.g. size 2 means only the top 2 forward-selected predictors are used etc.) and its predictive power is evaluated using 10-fold CV (yielding one dot on the plot).



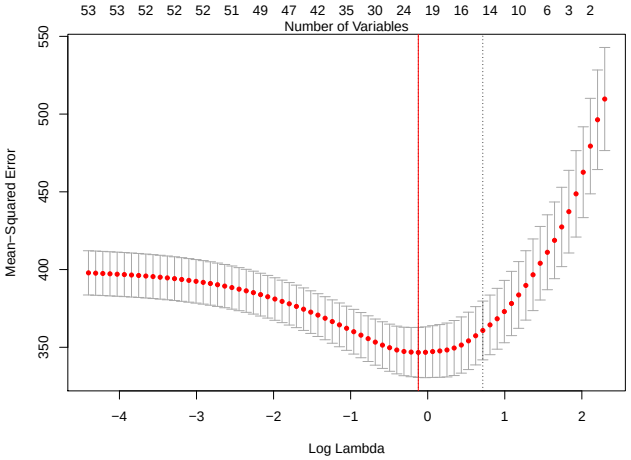
LASSO: Standardized Coefficeints vs. λ

Variable names written highly compressed on the right



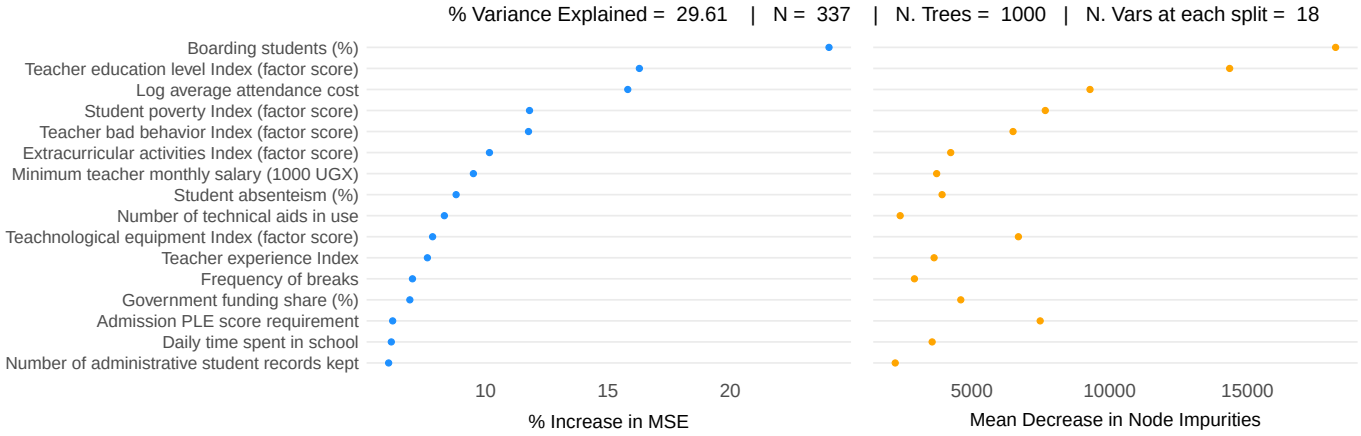
LASSO Optimal Model Size (λ) using 10-fold CV

Cross-validation to select value of λ the minimizes the MSE



Random Forest Result: Top 15 Predictors and % Variance Explained (PVE)

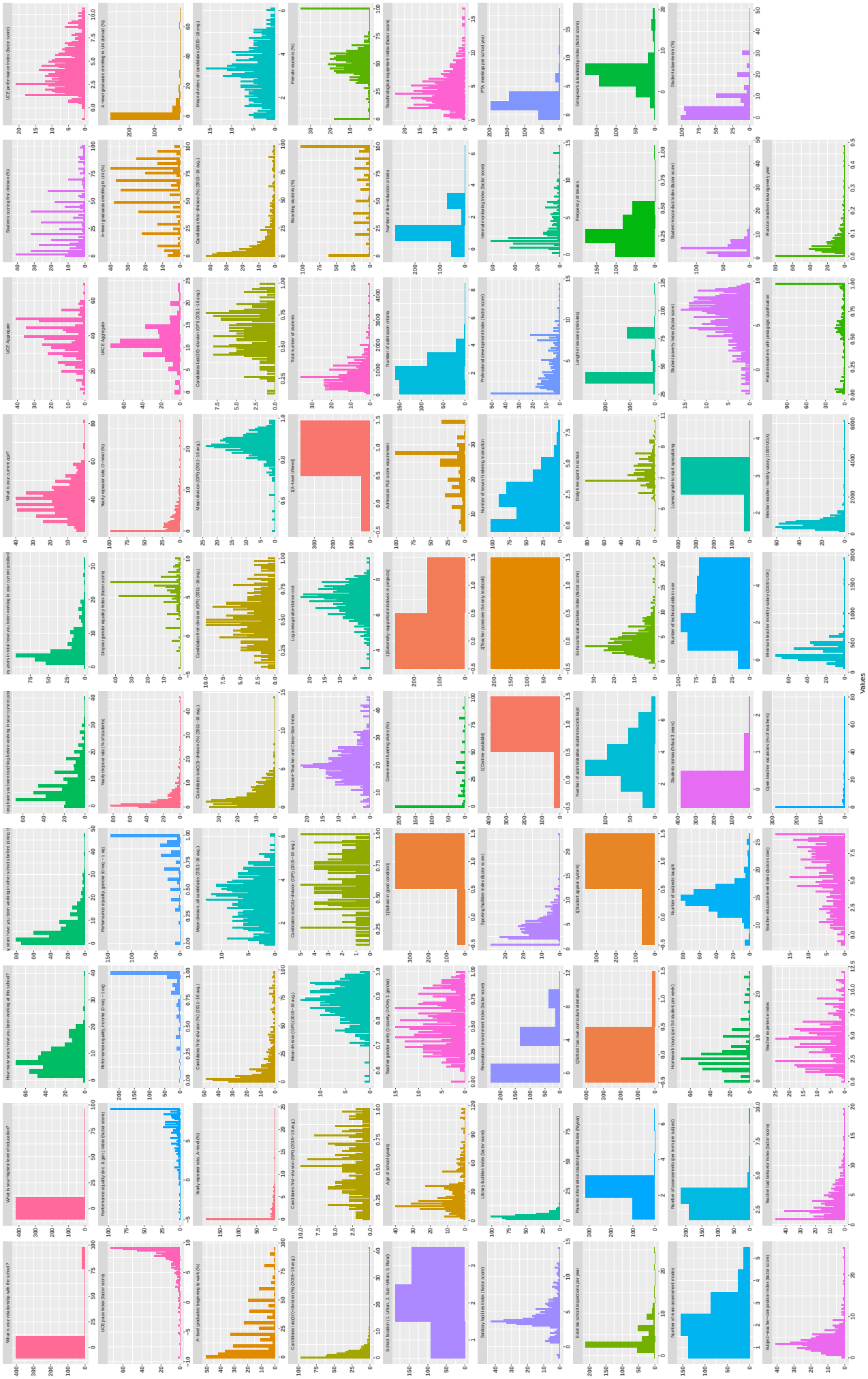
'Mean Decrease in Node Impurities' is another metric to rank predictors. A similar ranking on both metrics indicates a robust result. Naturally, if the overall predictive performance of the algorithm as measured by the PVE is very low, the ranking should be treated with extreme caution.



Histograms of the Analysis Dataset

Figure 35: Histograms of the Analysis Dataset. $N = 422$

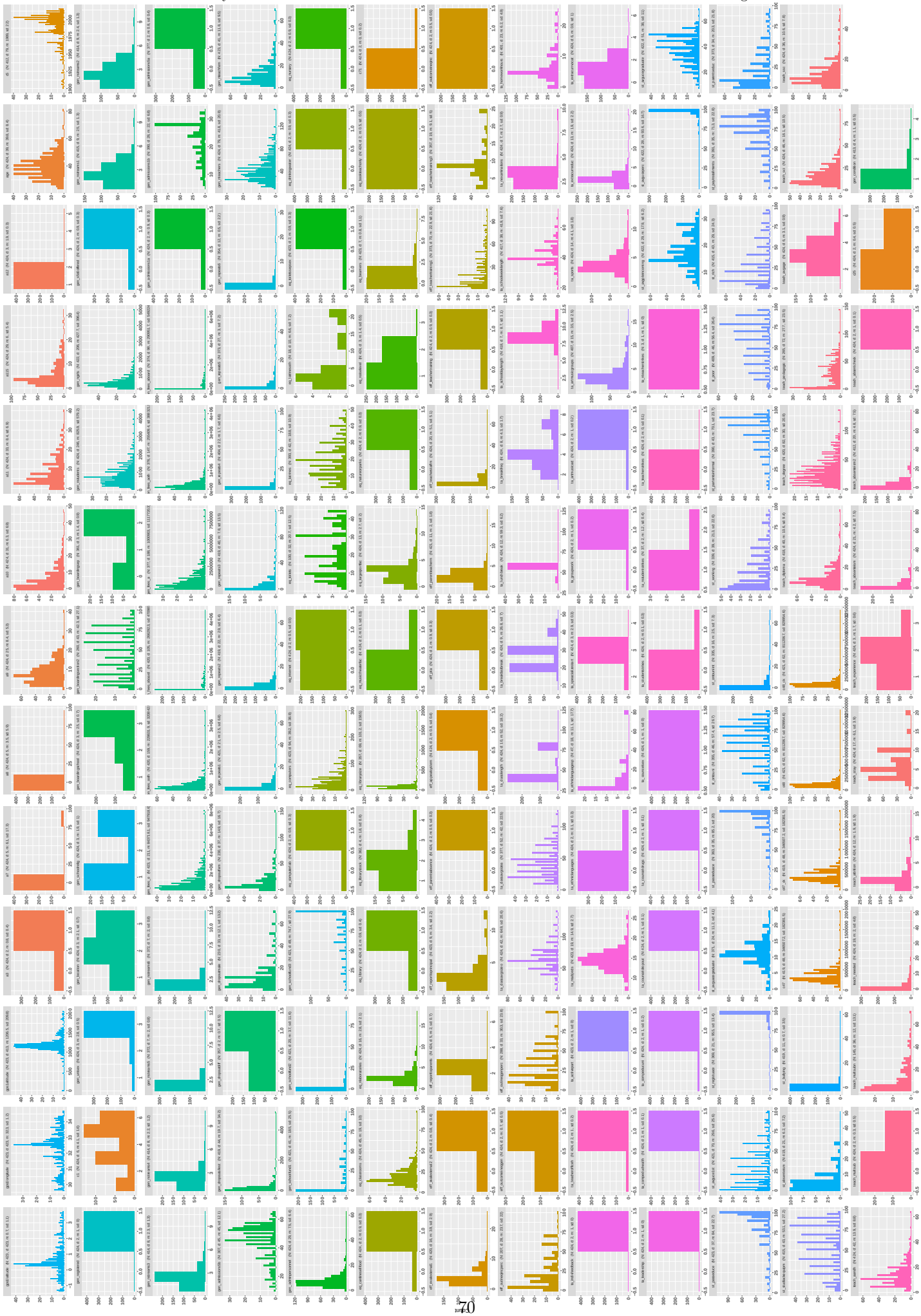
For Summary Statistics see Table (24)



Graphical Summary of The Cleaned Survey Dataset

Figure 36: Graphical Summary of the Cleaned Survey Dataset. $N = 424$

Statistics: N → Number of observations | d → Number of distinct values | m → Mean | sd → Standard deviation
The survey itself is attached in Table 26 below. Non-numerical variables are not shown in this figure.



The Survey

Table 26: **The Survey**

Type	*Variable Name*	*Question*	*Hint*
phonenumber	data_phone		
start	starttime	INTRO	
end	endtime		
today	today		
deviceid	devideid		
text	aa1	PLEASE WRITE YOUR NAME	
integer	aa2	PLEASE ENTER YOUR ID NUMBER	
date	aa3	INTERVIEW DATE	
select_one district	district	DISTRICT	
text	district_oth	SPECIFY OTHER	DISTRICT
select_one school_name	school	WHICH SCHOOL ARE YOU AT?	
text	sch_oth	SPECIFY OTHER	SCHOOL
geopoint	gps1	PLEASE CAPTURE THE GPS	
select_one yes_no	a04	Do you understand these terms and agree to participate in this survey?	
text	a1	Confirm Surname of Respondent	ASK FOR THE SURNAME WITH CORRECT SPELLING
text	a2	Confirm Other Names of Respondent	ASK FOR THE OTHER NAME WITH CORRECT SPELLING
select_one sex	a3	GENDER	OBSERVE THE SEX OF THE RESPONDENT
select_one relation	a7	What is your relationship with the school?	
text	a7_oth	OTHER SPECIFY	
select_one edu	a8	What is your highest level of education?	COMPLETED
text	a8_oth	OTHER SPECIFY	
integer	a9	How many years have you been working at this school?	YEARS
integer	a10	How many years have you been working in other schools before joining this school?	YEARS
integer	a11	How long have you been teaching before working in your current position?	YEARS
integer	a115	How many years in total have you been working in your current position? (at this and at other schools)	YEARS
select_one married	a12	What is your marital status?	
integer	age	What is your current age?	COMPLETE YEARS
text	village	What is the school's Village/LC1?	
integer	c5	What year did the school open?	
select_one opt_out	c5b	...	
select_one yes_no	Gen_registered	Is the school registered with the Ugandan government?	
select_one type1	c1	What kind of school is...	
text	c1_oth	SPECIFY OTHER	
select_one unisex	Gen_Unisex	Is this school...?	
select_one location	Gen_Location	Is the school...?	
select_one religion	Gen_SchoolRelig	Is this school...?	
text	Gen_SchoolReligOth	SPECIFY OTHER	
select_one xboarding	Gen_BoardingSchool	Is this school...?	
select_one yes_no	Gen_RegularSchool	So all scholars are day scholars?	
integer	Gen_BoardingSection2	What percentage of students are day scholars?	

Type	*Variable Name*	*Question*	*Hint*
select_one boarding_sep	Gen_BoardingSep	How are students separated in the boarding section?...	
select_one disttoschool	Gen_AvgHomeDist	Can you estimate the average distance of attending student from home in km?	KM
integer	Gen_Nstudents	What is the total number of students at this school?	TOTAL NUMBER OF STUDENTS
integer	Gen_Ngirls	How many girls attend this school?	TOTAL NUMBER OF GIRLS. ENTER 9999 IF DON'T KNOW.
integer	gen_rationgirls	Number of girls	ENTER THE NUMBER OF GIRLS TO BOYS. IF THERE ARE 2 GIRL TO EVERY 3 BOYS - YOU WRITE 2 FOR THE NUMBER OF GIRLS ON THIS LINE.
integer	gen_ratioboys	Number of boys	ENTER THE TO NUMBER OF BOYS. IF THERE ARE 2 GIRLS TO EVERY 3 BOYS - YOU WRITE 3 FOR THE NUMBER OF BOYS ON THIS LINE.
decimal	Gen_GenderRatio	ENTER IN THE RATIO	AS A NUMBER IN THIS FIELD. E.G. IF GIRLS/BOYS=2/1 ENTER 2, IF GIRLS/BOYS=1/2, ENTER 0.5.
select_one alevel	Gen_S5S6offered	Does this school teach classes at A-level (S5 and S6)?	
integer	Gen_NstreamS1	Number of streams in S1	E.g. the number of S1 classes typically taught parallel every year
integer	Gen_NstreamS2	Number of streams in S2	E.g. the number of S2 classes typically taught parallel every year
integer	Gen_NstreamS3	Number of streams in S3	E.g. the number of S3 classes typically taught parallel every year
integer	Gen_NstreamS4	Number of streams in S4	E.g. the number of S4 classes typically taught parallel every year
integer	Gen_NstreamS5	Number of streams in S5	E.g. the number of S5 classes typically taught parallel every year
integer	Gen_NstreamS6	Number of streams in S6	E.g. the number of S6 classes typically taught parallel every year
integer	Gen_Fees_O	School Fees	AVERAGE PER YEAR - NOT PER TERM.
integer	Gen_Fees_Ooth	Other fees (e.g. for school uniform, lunches, PTA, registration, examination, administration fees etc.)?	AVERAGE PER YEAR - NOT PER TERM.
integer	Gen_Fees_Oboard	Boarding fee (IF NOT INCLUDED IN SCHOOL FEE)	AVERAGE PER YEAR - NOT PER TERM.
integer	Gen_Fees_A	School Fees	AVERAGE PER YEAR - NOT PER TERM.
integer	Gen_Fees_Aoth	Other fees (e.g. for school uniform, lunches, PTA, registration, examination, administration fees etc.)?	AVERAGE PER YEAR - NOT PER TERM.
integer	Gen_Fees_Aboard	Boarding fee (IF NOT INCLUDED IN SCHOOL FEE)	AVERAGE PER YEAR - NOT PER TERM.
select_multiple typefin	Gen_FeeReduction	Does the school provide any scholarships, busaries, or financial assistance?	

Type	*Variable Name*	*Question*	*Hint*
select_multiple feereductions	Gen_FeeReductions	On what basis are scholarships or busaries granted to pupils?	SELECT ALL THAT APPLY
text	Gen_FeeReductionsOth	SPECIFY OTHER	
select_one yes_no	Gen_AdmissionS1a	Is there a cutoff (e.g. requirement) on the PLE for incoming students to be admittted into S1?	
integer	Gen_AdmissionS1b	What is the PLE aggregate cutoff?	
select_multiple admit	Gen_AdmissCriteria	On what criteria are pupils admitted to your school?	SELECT ALL THAT APPLY
text	Gen_AdmissCriteriaOth	SPECIFY OTHER	
select_one yes_no	Gen_AdmissionS5a	Is there a cutoff (e.g. requirement) on the UCE for incoming students to be admittted into S5?	
integer	Gen_AdmissionS5b	What is the UCE aggregate cut-off?	
integer	Gen_LEAVINGAllEst	What is the average number of students leaving the school without completing it successfully after either S4 or S6 every year?	NUMBER OF DROP OUT STUDENTS LEAVING EVERY YEAR (Dropouts + school switchers)
integer	Gen_DropoutAllEst	What is the average number of students dropping out every year? (OVERALL ESTIMATE)	NUMBER OF DROP OUT STUDENTS EVERY YEAR (ACROSS ALL GRADES)
select_one yes_no_dk	Gen_DropoutDiff	Was the dropout rate different for boys and girls in this school in 2017?	
integer	Gen_DropoutMale	If the dropout rate differs by gender, can you estimate the number of male students having dropped out?	Enter the NUMBER of boys dropping out of the school in 2017. Enter 99 if Don't Know. Enter 98 if Refuse to Answer.
integer	Gen_DropoutFem	And the number of female dropouts?	Enter the NUMBER of girls dropping out of the school in 2017. Enter 99 if Don't know. Enter 98 if Refuse to Answer.
integer	Gen_RepeatS1	Number of S1 grade repeaters	ENTER 99 IF DON'T KNOW. ENTER 98 IF REFUSED TO ANSWER.
integer	Gen_RepeatS2	Number of S2 grade repeaters	ENTER 99 IF DON'T KNOW. ENTER 98 IF REFUSED TO ANSWER.
integer	Gen_RepeatS3	Number of S3 grade repeaters	ENTER 99 IF DON'T KNOW. ENTER 98 IF REFUSED TO ANSWER.
integer	Gen_RepeatS4	Number of S4 grade repeaters	ENTER 99 IF DON'T KNOW. ENTER 98 IF REFUSED TO ANSWER.
integer	Gen_RepeatS5	Number of S5 grade repeaters	ENTER 99 IF DON'T KNOW. ENTER 98 IF REFUSED TO ANSWER.
integer	Gen_RepeatS6	Number of S6 grade repeaters	ENTER 99 IF DON'T KNOW. ENTER 98 IF REFUSED TO ANSWER.
integer	Gen_Nteachers	What is the total number of teachers teaching here?	Number of teachers
integer	Gen_NteachFem	How many teachers are female?	Number of female teachers
integer	Gen_AdminPersonnel	What is the number of school administrative or management personnel?	Number of Administrative Personnel (Including principals, assistant principals, other management staff, receptionists,secretaries, administration assistants whose main activity is administration ormanagement)
integer	Gen_SchoolFund1	What percentage of funding comes from Government	%
integer	Gen_SchoolFund2	What percentage of funding comes from Donations?	%

Type	*Variable Name*	*Question*	*Hint*
integer	Gen_SchoolFund3	What percentage of funding comes from Parents payments (of fees etc.)	%
select_one yes_no	Eq_ComputerLab	Is there a computer lab (for students/computer science course)?	
integer	Eq_Computers	How many computers does the school own? (Total, in labs, and in offices)	NUMBER OF COMPUTERS
select_one yes_no	Eq_Internet	Does the school have internet access?	
select_multiple toilets	Eq_Toilets_type	What kind of sanitary facilities are there for students?	SELECT ALL THAT APPLY
text	Eq_Toilets_typeoth	SPECIFY OTHER	
integer	Eq_Toilets	How many toilets for students are there?	NUMBER OF TOILETS
integer	Eq_Latrines	How many latrines for students are there?	NUMBER OF LATRINES
integer	Eq_Latrinesoth	How many other sanitary facilities for students are there?	NUMBER OF OTHER SANITARY FACILITIES
select_one yes_no	Eq_ToiletsSepSex	Are the toilets/latrines gender-separated?	SEPARATE FACILITIES FOR BOYS AND GIRLS
select_one yes_no	Eq_DrinkingWater	Is drinking water freely provided to students?	
select_one yes_no	Eq_Nursery	Are there medical facilities or a nursery with first aid equipment?	
select_one yes_no	Eq_CantineSellFood	Is there a cantine for students to purchase food outside of general meals?	
integer	Eq_Classrooms	What is the total number of classrooms (including special purpose classrooms like computer classrooms or laboratories)?	
integer	Eq_Nlaboratories	What is the total number of equipped laboratories (for physics, chemistry, electronics etc.)?	TOTAL NUMBER OF DISTINCT LABS (EQUIPED MEANS THAT THEY ARE USEABLE FOR THEIR INTENDED PURPOSE, E.G. FOR A CHEMISTRY LAB THERE MUST BE CHEMICALS)
select_one yes_no	Eq_Library	Does the school have a library?	
select_one lib_stock	Eq_LibraryStock	How well is the library stocked?	
integer	Eq_LibrarySize	What is the approximate size of the school library in m^2? (LET'S ASSUME A TYPICAL CLASSROOM HAS 40 M^2, IF YOU DON'T KNOW THE EXACT SIZE, ESTIMATE THE SIZE IN M^2 USING THIS BENCHMARK)	SIZE IN m^2
select_one yes_no	Eq_MusicTheaterFac	Are there separate musical or theatrical rooms or facilities?	
integer	Eq_LargeSportFac	How many large sporting facilities such as sporting halls, volleyball courts, gym's, soccer fields etc. does the school have?	NUMBER OF DISTINCT AREAS/FACILITIES
select_one yes_no	Eq_NatureOrParks	Does the school have any green-spaces, parks, or large-outdoor areas within it's compound?	OBSERVE IF POSSIBLE
select_one outside_noise	Eq_NoiseLevel	What was the level of outside noise on a typical workday in 2017?	
integer	Eq_Beamers	Does the school own any projectors?	IF YES, ENTER THE NUMBER OF SUCH DEVICES, IF NO, ENTER 0
select_one yes_no	Eq_BookTeachOnly	Did the teacher have the only textbook for classes?	
select_one yes_no	Eq_BookShare	For classes using textbooks, do students need to share?	
integer	Eq_BooksMath	What, approximately, is the average number of students sharing a Mathematics Textbook	Number of students sharing a math book (= 1 if every student has his own)
select_one books_optmath	Eq_BooksMathNO	...	

Type	*Variable Name*	*Question*	*Hint*
integer	Eq_BooksSCI	What, approximately, is the average number of students sharing a Science Textbook (across all sciences e.g. biology, physics, chemistry)	Number of students sharing a science book (= 1 if every student has his own)
select_one books_optSCI	Eq_BooksSCINO	...	
integer	Eq_BooksOth	What, approximately, is the average number of students sharing a Textbook in other (non-Math, non-Science) subjects	Number of students sharing a book (= 1 if every student has his own)
select_one books_optOth	Eq_BooksOthNO	...	
select_multiple hinderfact	Eq_HinderFactors	Is the school's capacity to provide instruction hindered by anything?	SELECT ALL THAT APPLY. PROBE BUT DO NOT READ ANSWERS.
text	Eq_HinderFactorsOth	SPECIFY OTHER	
select_one yes_no	c71	Does the school have its own written statement of the curriculum content to be taught (i.e. other than national curriculum guides)?	
text	c71_a	What curriculum is used?	
integer	Eff_EvalExternal1	How many external inspections are there per year?	Number of inspections per year
select_one yes_no	Eff_EvalExternal2	Is student performance information communicated to an educational authority?	Aside from the official UCE or UACE test scores
select_multiple teacheval	Eff_EvalInternal	How does the school evaluate teachers?	SELECT ALL THAT APPLY. PROBE BUT DO NOT READ ANSWERS.
integer	Eff_ReportToParents	How often student performance reported to the parents? PER TERM	MUST BE DIRECTLY COMMUNICATED TO PARENTS. ENTER 0 IF NO SUCH REPORTS TO PARENTS
integer	Eff_ReportToPrincipal	How often is student performance reported to the principal or head teacher? PER TERM	
select_one yes_no	Eff_ParentsAbsence	When students are often absent without an excuse, are parents notified?	
select_multiple bullysacntion	Eff_BullySanction	In the event of bullying among students, in what ways are bullies sanctioned?	SELECT ALL THAT APPLY
text	Eff_BullySanctionOth	SPECIFY OTHER	
select_multiple teachaward	Eff_TeachAward	Is any recognition given to good teachers in your school? If yes, what kind of recognition?	SELECT ALL THAT APPLY
text	Eff_TeachAwardOth	SPECIFY OTHER	
select_multiple keeprecords	Eff_KeepRecords	What student information/records are kept by the school?	SELECT ALL THAT APPLY
text	Eff_KeepRecordsOth	SPECIFY OTHER	
select_one yes_no	Eff_AppealSystem	Is there an internal system for students to appeal against educational decisions against them?	EXAMPLE - A GRADE THEY FEEL IS UNFAIR
select_one yes_no	Eff_PTA	Does the school have a parent-teacher association (PTA)?	
integer	Eff_ParentTeacherM	How often does the PTA meet in a school year?	NUMBER OF MEETINGS PER SCHOOL YEAR
integer	Eff_teacherStaffM	How many teacher/staff meetings to discuss the state of the school, general issues, and improvements?	NUMBER OF MEETINGS PER TERM
select_one yes_no	Eff_TeacherTraining	Do teachers receive training or professional development?	
integer	Eff_TeacherTraining1	How many teachers every year receive training or professional development?	
select_multiple workshop	Eff_TeacherTraining2	What type of professional development?	
text	Eff_TeacherTraining2_oth	SPECIFY OTHER	
integer	Eff_TeacherTraining3	On average, how many hours in total is a workshop?	

Type	*Variable Name*	*Question*	*Hint*
select_one yes_no	Eff_OutcomeIneqInc	Is there a noticeable difference in performance between rich and poor students?	
integer	Eff_OutIneqIncPerc	What is the difference in %? For example, a rich students score 10% better than poor students.	PERCENT IMPROVEMENT OF RICH STUDENTS OVER POOR STUDENTS.
select_one yes_no	Eff_OutcomeIneqGen	Is there a noticeable difference in performance between male and female students in this school?	
integer	Eff_OutIneqGenPerc	What is the difference in %? For example, a female students will score 10% better than male students.	PERCENT IMPROVEMENTS OF FEMALE STUDENTS OVER MALE STUDENTS. (e.g. if female students are twice as good as male students, enter 100 (%))
integer	Ta_ClassAvgSizeO	What is the average class size in O-level lessons (across subjects)?	Number of Students per lesson/class
integer	Ta_ClassAvgSizeA	What is the average class size in A-level lessons (across subjects)?	Number of Students per lesson/class
select_multiple groupclass	Ta_StudentSepClass	How are students grouped into streams?	CHECK ALL THAT APPLY
text	Ta_StudentSepClassOth	SPECIFY OTHER	
integer	Ta_ClassLength	On average, how long is a class lesson?	IN MINUTES
integer	Ta_BreakfastBreak	How long is the breakfast break?	IN MINUTES
integer	Ta_LunchBreak	How long is the lunch break?	IN MINUTES
integer	Ta_BreakFreq	How many consecutive lessons do students have before there is a longer break?	Number of classes
integer	Ta_SchooldayLength	How long is a typical school day?	IN HOURS
integer	Ta_Schoolweeklength	How many hours of lessons are there in a week?	IN HOURS
select_multiple mainassess	Ta_MainAssessModes	How do you measure student performance?	SELECT ALL THAT APPLY
text	Ta_AssessOTHER	PLEASE SPECIFY OTHER ASSESSMENT MODE	text
integer	Ta_Nexaminations	What is the average number of examinations or assessments per term per subject?	Number of exams/assessments per subject per term
integer	Ta_HomeworkHours	How many hours of homework do S4 students typically complete per week?	in hours
select_one yes_no	Ta_IndivFeedback	Do teachers give students feedback on individual work?	
select_one yes_no	Ta_TeacherOfficeH	Do teachers have consultation times for students?	OPEN OFFICE HOURS FOR STUDENTS TO DROP-IN TO ASK QUESTIONS
select_one yes_no	Ta_ActivePart	Is active participation in class encouraged?	(In the average class with the average teacher)
integer	Ta_Nsubjects	How many distinct subjects are taught at this school?	NUMBER OF SUBJECTS
select_one yes_no	Ta_EthnicLanguages	Are any classes (other than language classes) taught in local languages?	
integer	Ta_EthnicLanguagesP	What percentage of classes (other than language classes) is taught in local languages?	%
select_multiple techaids	Ta_TechAids	Are any of the following technical aids used?	SELECT ALL THAT APPLY
select_one grade	Ta_SpeacializeStart	What is the lowest class for pupils to start specializing and choosing their courses?	Enter the class (S1,S2,S3,S4,S5 or S6)
select_one yes_no	Ta_Groupwork	Do students do groupwork or complete groupwork assignments frequently?	(In the average class with the average teacher)
select_one yes_no	Ta_ArtMusicAct	Are activities offered in performing arts, theatre, arts, music, debating or poetry? (may be extracurricular or student run groups or societies)	

Type	*Variable Name*	*Question*	*Hint*
integer	Ta_ArtMusicGroups	How many activities in the arts and humanities are offered to students? Including extracurricular.	Enter the total number groups, activities or classes that fall in these categories (e.g. theatre groups, bands, orchestras, painting, debate clubs, political clubs etc.) (NOT PER WEEK, BUT IN GENERAL)
integer	Ta_Sports	How many sports activites are offered to students? Including extracurricular.	ENTER THE TOTAL NUMBER OF SPORTS/TEAMS
integer	Ta_ExtracurrEduc	How many extracurricular academic education activities are offered by the school (such as science workshops)?	ENTER THE TOTAL NUMBER OF ACTIVITIES
integer	Ta_ExtracurrVocat	How many extracurricular vocational training (e.g. gardening, woodwork) activities offered?	ENTER THE TOTAL NUMBER OF ACTIVITIES
select_one yes_no	Ta_Leadership	Are there opportunities for students to fill leadership roles?	
select_multiple leadership	Ta_Leadershiproles	What type of leadership roles?	
text	Ta_Leadershiprolesoth	SPECIFY OTHER	
select_one yes_no	Ta_CompEducHealth	Is (complementary) education in health or sexual health provided to students?	This might be embedded into the curriculum or provided in a workshop etc.
select_one yes_no	Ta_CompEducUni	Is (complementary) education about higher education/University and other options upon graduation provided to students?	This might be embedded into the curriculum or provided in a workshop etc.
select_one yes_no	Ta_Counceldropout	Are students counselled on matters like dropout prevention?	
select_one yes_no	Ta_Counceldrugs	Are students councelled on drug abuse/prevention?	
select_one yes_no	Ta_Councelsex	Are students conunselled on abstinence?	
select_one yes_no	Ta_OwnFoodAllow	Are students allowed to bring their own food to school?	
select_one yes_no	Ta_StudentStrikes	In the past three years were there any student strikes or riots?	
integer	Ta_NStudentStrikes	How many student strikes or riots?	Number of distinct causes for upheaval in the last three years
select_one yes_no	Ta_TeacherStrikes	In the past three years were there any teacher strikes or riots?	
integer	Ta_NTeacherStrikes	How many teacher strikes or riots?	Number of distinct causes for upheaval in the last three years
integer	St_AvgPLEincoming	What is the average aggregate PLE of incoming students?	PLE
integer	St_AvgPLEperc	What % of students who registered for S4 sit the UCE at the end of S4?	%
integer	St_AvgUCEgraduate	What is the aggregate UCE of students sitting it from this school?	Average UCE
integer	St_PassedUCE	What % of students attempting the UCE from this school pass it on the first attempt?	%
integer	St_AvgPLE1st	On average, what % of students score first division?	%, ON UCE
integer	St_registUACE	What % of students who registered for S6 sit for the UACE at the end of the year?	%
integer	St_AvgUACEgraduate	What is the aggregate UACE of students sitting it from this school?	Average UACE
integer	St_PassedUACE	What % of students attempting the UACE from this school pass it on the first attempt?	%
integer	St_OntoUNI	What % of A-level graduates go on to enrol in University?	%
integer	St_OntoUNIInt	What % of A-level graduates go on to enrol in University abroad?	%
integer	St_Working	What % of A-level graduates go directly on to work (no further education)?	%

Type	*Variable Name*	*Question*	*Hint*
integer	St_ParentsInformal	What % of students have parents working in the informal sector (e.g. as farmers or boda drivers)?	%
integer	St_Poor	What % of Students are Poor: Non-scholarship students where Parents have GREAT problems Paying school fees, books, equipment and food, and providing an adequate home-learning environment for their child?	%
integer	St_Rich	What % of Students are Rich: Non-scholarship students where Parents have NO problems at all Paying school fees, books, equipment and food, and providing an adequate home-learning environment for their child?	%
integer	St_PayInstallments	What % of parents pay their school feel in installments? (as opposed to paying for the whole year or term directly)	%
integer	St_ParentsEduc	What % of students parents have higher education (e.g. hold Bachelor, Master or PhD Degrees)?	%
integer	St_StudLackEquipm	What % of students have problems getting required school equipment (e.g. notebooks)?	%
integer	St_Absenteism	On a typical school day, what percentage of students are absent from school for any reason?	%
integer	St_Bullying	In a typical school week, how many incidences of bullying or other abuse are there?	NUMBER OF INCIDENCES
select_multiple badbehavior	St_BadBehavior	In the last year, has the learning progress of (some) students been substantially hindered by one of the following behaviours?	CHECK ALL THAT APPLY
text	St_BadBehaviorOth	SPECIFY OTHER	
integer	c47	What is the minimum monthly teacher salary at this school?	MONTHLY SALARY. DOUBLE CHECK NUMBER OF '0"S ENTERED
integer	c47_ck	ENUMERATOR PLEASE RE-ENTER THE AMOUNT	
integer	c48	What is the maximum monthly teacher salary at this school?	MONTHLY SALARY. DOUBLE CHECK NUMBER OF '0"S ENTERED
integer	c48_ck	ENUMERATOR PLEASE RE-ENTER THE AMOUNT	
integer	Teach_Diploma	Number of teachers with a diploma (Secondary school + non-university education)	NMBER OF TEACHERS
integer	Teach_Degree	Number of teachers with a university degree	NUMBER OF TEACHERS
integer	Teach_Pedagogic	Number of teachers with an education qualification (This refers to the number of teachers that have gone through pedagogic training and obtained a teaching diploma. For example: An engineer who later found employment as a math teacher at a shool does not have an education qualification.)	NUMBER OF TEACHERS WITH AN EDUCATION QUALIFICATION
select_one teacher-age	Teach_AvgAge	What is the average age of teachers?	
integer	Teach_SCI	Number of teachers in Math or Science subject area (Math, Physics, Chemistry, Biology, Computer-science)	NUMBER OF TEACHERS
integer	Teach_SSC	Number of teachers in Social Science subject area (politics, economics/business/accounting/entrepreneurship, geography, law/civil education)	NUMBER OF TEACHERS
integer	Teach_AandH	Number of teachers in Arts & Humanities subject area (languages, History, Art, Music, crafts, religion)	NUMBER OF TEACHERS
select_one yes_no	Teach_multsub	Are there teachers who teach across fields?	E.G Teaches Math and English
integer	Teach_multsubN	How many teachers teach across fields?	E.G Teaches Math and English
integer	Teach_Needed	How many unfilled teaching positions?	NUMBER OF VACANCIES
integer	Teach_Attrition	How many teachers are typically leaving the school every year (retirement or other reasons)?	NUMBER OF TEACHERS PER YEAR
integer	Teach_Stay	How long does a teacher, once employed, typically stay at this school?	NUMBER OF YEARS

Type	*Variable Name*	*Question*	*Hint*
select_one teachexperience	Teach_Experience	How many years of teaching experience (at this or at other schools) does the average teacher have?	NUMBER OF YEARS
select_multiple badbehaviorteach	Teach_BadBehavior	In the last year, has the learning progress of (some) students been substantially hindered by one of the following teacher behaviours?	
text	Teach_BadBehavOth	SPECIFY OTHER	
integer	Teach_Absenteism	In a typical term, what percentage of the class periods had to be cancelled because of absence of the assigned teacher?	% of classes (I't doesn't matter why the teachers are absent)
integer	Teach_Absenteism2	In a typical term, what percentage of the class periods had to be coverey by somebody else because of absence of the assigned teacher?	% of classes (I't doesn't matter why the teachers are absent)
select_one yes_no	Teach_AbsentCheck	Is teacher absenteism being monitored by someone?	
select_multiple teachabspunish	Teach_AbsentPenalty	How is teacher absenteism being punished?	SELECT ALL THAT APPLY
text	Teach_AbsentPenalOth	Please specify other punishment for teacher absenteism	
select_one yes_no	c25	Are there any externally-supported initiatives, programs, and projects are implemented at this school?	
begin repeat	g11	Externally-supported Projects	
text	c25_a	What is the name of the project/initative/program?	
select_multiple project	c25_b	What type of project/initative/program?	
text	c25_b_oth	SPECIFY OTHER	
select_multiple people1	c25_c	Who are the main key stakeholders involved in implementing the project/initative/program?	
text	c25_c_oth	SPECIFY OTHER	
end_repeat			
select_one condition	Gen_Condition	OBSERVE THE CONDITION OF THE SCHOOL	
text	e1	RECORD GENERAL NOTES ABOUT THE INTERVIEW AND ANY SPECIAL INFORMATION THAT WILL BE HELPFUL FOR THE DATA ANALYSIS.	REMEMBER EVERY SURVEY IS DIFFERENT AND YOUR NOTES SHOULD CAPTURE WHAT HAPPENED DURING THE INTERVIEW.
select_one language	e2	LANGUAGE OF INTERVIEW	
text	e2_oth	SPECIFY OTHER	