United Nations
Educational, Scientific and
Cultural Organization

EDUCATION FOR ALL GLOBAL MONITORING REPORT

2012/ED/EFA/MRT/PI/51

Background paper prepared for the
Education for All Global Monitoring Report 2012

Youth and skills: Putting education to work

Schooling profiles; between and within country differences

Marcos Delprato

2012

This paper was commissioned by the Education for All Global Monitoring Report as background information to assist in drafting the 2012 report. It has not been edited by the team. The views and opinions expressed in this paper are those of the author(s) and should not be attributed to the EFA Global Monitoring Report or to UNESCO. The papers can be cited with the following reference: "Paper commissioned for the EFA Global Monitoring Report 2012, Youth and skills: Putting education to work" For further information, please contact efareport@unesco.org

Education for All Global Monitoring Report 2012

Title of the Work: Schooling profiles; between and within country differences

Author: Marcos Delprato

Abstract

This report consists of a large data exercise based upon a household survey, yielding education indicators for an array of countries, at both national and sub-national levels. Specifically, the expected total intake into primary school at different ages, as well as repetition and dropout rates and survival rates for each grade are calculated, allowing the reconstruction of the expected path of children entering the education system. Because schooling profiles may vary significantly by children's background, disaggregated calculations are also obtained for dimensions within a country (e.g., gender, wealth, regions, ethnicity), as well as overlapping dimensions (e.g., wealth and gender). The reconstructed cohort model is adopted for this exercise. Furthermore, a preliminary analysis of the main drivers of expected cohort completion rates is included.

1. Introduction

Most basic skills are acquired during primary school. Failure to complete a basic cycle of primary school, not only limits future opportunities for children, but also represents a significant drain on the limited resources of countries for the provision of primary education (Sabates, Akyeampong, Westbrook, & Hunt, 2011). Policies tackling the reasons driving the lack of completion of primary school and high drop out rates require detailed data, so that they can focus on more vulnerable and affected groups (e.g., UNESCO, 2007, 2010; Wils, Sylla, & Oliver, 2009). Data from household surveys (such as the Demographic and Health Survey) generally provides repetition and drop out rates at an individual level, as well as by sub-national and regional levels. Logically, the higher the repetition and drop out rates for the poorest population or certain ethnic or regional group, the less likely those people may complete primary school (e.g., Hunt, 2008; Lewin & Little, 2011). For instance, in a study for a large group of countries, Ingram, Wils, Carrol, &Townsend (2007) find that the correlation of school attendance with income is one of the strongest, with a child of the wealthiest quintile being between 30% to 40% more likely to be in school than a child from the poorest quintile.

Completion rates of primary education, however, are also critically influenced by primary school intake at different ages. Indeed, it is argued that the completion rates decrease when the proportions of children entering late are high (see, for instance, Cameron, 2005; Wils, 2004). Therefore, the current dataset construction exercise accounts for this, altering the reconstructed cohort model by the chance of children entering the education system. In doing so, more accurate estimations and aimed policies to induce school engagement can be devised. The rest of the report is organised as follows. Section 2 contains definitions for the indicators calculated. Section 3 describes the data sources and variables used. The calculation steps are explained in Section 4. A brief analysis is included in Section 5. Finally, conclusions are offered in Section 6.

2. Indicators

The large data exercise is based on the pseudo-longitudinal cohort reconstruction model. A school cohort is defined as a group of pupils who enter the first grade of a given cycle or level of education in the same school year, and who subsequently experience the events of promotion, repetition, dropout, or successful completion of the final grade. The reconstructed cohort method is a common way to analyze the internal efficiency of an education system as it is less demanding on the availability of detailed data over time. The model uses data on enrolments by grade for two consecutive years and on repeaters by grade from the first to second year are sufficient to enable the estimation of three main flow rates: promotion, repetition, and dropout. These estimated rates can be analyzed to study the patterns of repetition and dropout by grade as well as to derive other indicators of internal efficiency (UNESCO, 2006).

The reliability of the estimated cohort's progression would depend on the validity of the following assumptions: i) that there will be no additional new entrants during the lifetime of the cohort; ii) the hypothesis of homogenous behaviour by which the same rates of repetition, promotion, and dropout apply, regardless of whether a pupil has reached that grade directly or after one or more repetitions; iii) the number of times any given pupil will be allowed to repeat is well defined; and iv) that flow rates for all grades remain unchanged as long as members of the cohort are still moving through the cycle (UNESCO, 2006).

As elements of the cohort reconstruction model, the following indicators (UIS, 2009) are incorporated in the constructed dataset.

a) Repetition rate by grade (RR).

- Definition: proportion of pupils from a cohort enrolled in a given grade at a given school year who studies in the same grade in the following school year.
- Formula:

$$RR_i^t = \frac{RR_i^{t+1}}{E_i^t} \qquad (1)$$

where RR_i^t is the repetition rate at grade *i* in school year *t*, RR_i^{t+1} is the number of pupils repeating grade *i* in school year *t+1*, and E_i^t is the number of pupils enrolled in grade *i* in school year *t*.

b) Dropout rate by grade (DR).

- Definition: proportion of pupils from a cohort enrolled in a given grade at a given school year who are no longer enrolled in the following school year.
- Formula:

$$DR_i^t = \frac{D_i^{t+1}}{E_i^t} \qquad (2)$$

where DR_i^t is the dropout rate at grade *i* in school year *t*, D_i^{t+1} is the number of pupils dropping out from grade *i* in school year *t+1*.

c) Promotion rate by grade (PR).

- Definition: proportion of pupils from a cohort enrolled in a given grade at a given school year who studies in the next grade in the following school year. It is calculated by the identity below.
- Formula:

$$PR_i^t = 100 - (RR_i^t + DR_i^t) \qquad (3)$$

where PR_i^t is the promotion rate at grade *i* in school year *t*.

d) Expected intake at or before the appropriate age (or Adjusted net intake rate: ANIR).

- Definition: is the total intake of the population of entry age *a* or before in year *t* as a proportion of the population of entry age in the reference year.
- Formula:

$$ANIR_t = \frac{TI^{EAt}}{PEA_t} \qquad (4)$$

where TI^{EAt} is the total intake of the population of entry age *a* in year *t*, and PEA_t is the population (fixed) of entry age *a* in the reference year.

- Also, $PEA_t = TI^{EAt} + NE^{EAt} = \sum_{i=-n}^{0} PI_{EAt}^{t+n} + NEA^{EAt} + NE^{EAt}$. Thus, ANIR can be decomposed into the total intake on time (or earlier) and the aggregate number of non-entered at the appropriate age (NEA^{EAt}) and the number who hadn't entered primary school at all (NE^{EAt}).

e) Expected Intake at official age+*i*.

- Apply the most recent intake structure to the current ANIR. Need to estimate how much of the age specific cohort not entered on time will enter later.
- Definition: looking at the past cohort, it is the share of the remaining cohort that entered in year t+*i*.
- Formula:

$$IR_{t+i}^{EAt} = \frac{PI_{t+i}^{EAt}}{NEA^{EAt} + NE^{EAt}} \qquad (5)$$

f) Expected to never enter.

- Definition: is the share of the remaining cohort that never entered.
- Formula:

$$PNER^{EAt} = \frac{NE^{EAt}}{NEA^{EAt} + NE^{EAt}} \qquad (6)$$

g) Expected cohort c completion rate.

- Definition: proportion of children of school starting age who are expected to complete primary school (including those who start late and repeat primary school grades) given the current state of the education system (it is rarely an observed outcome).
- Formula: the IR in year t+*i* is estimated by the most recent observed value of the ANIR. That is,

$$\widetilde{TIR}^{EAt} = ANIR_t + \sum_{i=1}^{n} \widetilde{IR}_{t+i}^{EAt} + \widetilde{PNER}^{EAt} \qquad (7)$$

3. Data sources

Demographic and Health Surveys (DHS) for 24 countries are employed to construct the dataset for this report.[1] Data derives from the most recent DHS surveys from a selection of countries, though for most countries more than one round of data is used. Table 1 contains information on the list of countries, survey years and duration of the primary cycle.

Table 1. List of countries included in the report, dataset's years and primary cycle duration

Country	DHS years	Primary entry age	Primary duration
Azerbaijan	2006	6	4
Benin	2006, 2001	6	6
Bolivia	2008, 2003	6	6
Colombia	2010, 2005, 2000	6	5
Congo	2005	6	6
D. R. Congo	2007	6	6
Ghana	2008, 2003	6	6
Guinea	2005	7	6
India	2005	6	5
Kenya	2008	6	6
Liberia	2007	6	6
Madagascar	2008, 2003-04	6	5
Mali	2006, 2001	7	6
Moldova	2005	7	4
Namibia	2006	7	7
Nepal	2006, 2001	5	5
Niger	2006	7	6
Nigeria	2008, 2003	6	6
Rwanda	2005, 2000	7	6
Senegal	2005	7	6
Sierra Leone	2008	6	6
Uganda	2006	6	7
Zambia	2007, 2001-02	7	7
Zimbabwe	2005, 1999	6	7

It is standard practice to reduce the age of the children by one year if the survey took place six months after the beginning of the school year. If age is not adjusted, some children will be unfairly considered as entering late in school because the decision whether to enrol or not was taken when the child may have been below the appropriate age. For example, assume the school year coincides with the calendar year. If an enumerator visits a household with a 6-year old child in September and that child is not in Grade 1, this is not necessarily a case of delayed enrolment: when the decision was taken in January whether to enrol the child or not, it is more likely that the child was still 5 years old and therefore too young to go to school. Hence, this one-year adjustment to the age of each child should be

[1] Further details on the datasets and related documents can be found at: http://www.measuredhs.com/

made. For the adjustment one needs to know when the DHS fieldwork was undertaken and when the school year starts. Table 2 shows this information as well as whether or not the age adjustment was actually made in a given survey.

Table 2. Fieldwork and school periods and survey age adjustment

Country	DHS years	Fiedwork period	School period	Age adjustment
Azerbaijan	2006	Late July to early November 2006	September to May	No
Benin	2006	03 August to 18 November 2006	October to July	No
	2001			
Bolivia	2008	13 February to 15 June 2008	February to November	No
	2003			
Colombia	2010	14 November to 18 December 2009 and 21 February to 14 November 2010	February to November	No
	2005			
	2000			
Congo	2005	8 July to 23 November 2005	October to June	No
D. R. Congo	2007	January to March 2007 and May to August 2007	September to July	Yes
Ghana	2008	Early September to late November 2008	September to August	No
	2003			
Guinea	2005	February to June 2005	October to June	Yes
India	2005	November 2005 to August 2006	April to March	No
Kenya	2008	13 November 2008 to late February 2009	January to December	No
Liberia	2007	25 December 2006 to April 2007	September to June	n/a
Madagascar	2008	23 November 2008 to 17 August 2009	September to June	No
	2003-04			
Mali	2006	May to December 2006	October to June	Yes
	2001			
Moldova	2005	13 June to 18 August 2005	September to May	No
Namibia	2006	10 December 2006 to 21 May 2007	January to December	n/a
Nepal	2006	5 February to 18 August 2006	May to April	n/a
	2001			
Niger	2006	02 January to 15 May 2006	October to June	Yes
Nigeria	2008	June to October 2008	September to July	Yes
	2003			
Rwanda	2005	28 February to 13 July 2005	January to December	No
	2000			
Senegal	2005	February to June 2005	October to July	Yes
Sierra Leone	2008	End of April to the end of June 2008		
Uganda	2006	5 May to 07 October 2006		
Zambia	2007	April to October 2007		
	2001-02			
Zimbabwe	2005	August 2005 to February 2006		
	1999			

Countries have been selected because they have the retrospective questions which allow the calculation of crucial education. The schooling variables' questions of the "household" roster dataset are used to construct the expected cohort completion rate.[2] The education part of the questionnaire is only asked to individuals aged from 5 to 24 years old. Moreover, calculation of promotion, repetition and drop out rates requires data on educational attainment of household members in two consecutive years. If any of the key variables to define these indicators are missing in either year, these observations are dropped. Final total samples at the national level include all grades of primary school level and the two first grades of secondary education. Key educational variables and dimensions are described in Table 3.

Table 3. Schooling variables and dimensions

Educational variables	Description
hv106	highest educational level
hv121	member attended school during current school year
hv122	educational level during current school year
hv123	grade of education during current school year
hv125	member attended school during previous school year
hv126	educational level during previous school year
hv127	grade of education during previous school year
hv129	school attendance status
Dimensions	**Description**
hv104	gender
hv207	wealth quintile
hv025	urban or rural residence
hv024	region
shlang	language
sv131	ethnicity (individual dataset)

4. Calculation steps

Promotion, repetition and dropout rates are calculated using the variables shown in Table 3. An individual is classified as a repeater if his/her grade in school year *t* is the same as in school year *t*+1. A dropout is constructed by comparing a person's attendance during the current and previous school year. If a person is not attending school in year *t*+1 but did attend school in year *t*, he/she is assigned to the dropout group.[3] Promotion is composed of pupils who had neither repeated nor dropped out.

Expected intakes at the official entry age (and later) are obtained by exporting tables by age from Stata into Excel and performing the calculations there. These tabulates are "first time in school'" (in grade 1), "total population'", "population already in school'", "has ever attended school but not in school anymore" and the "population of starting primary school age". The methodology consists of using the first entrants as a share of the age specific population. Specifically, by using these tables one is able to calculate the "remaining population to

[2] Only the individual datasets are used when ethnicity is not available in the household raw dataset. In these cases, the head of household's ethnicity is attached to each family member.

[3] In Stata code: `gen repetition = 1 if hv127~=. & hv123==hv127`, and `gen dropout=1 if hv121==0 & hv125==1`.

enter", the "first time intake as % of the remaining population who has never been to school", "population expected to enter", "% pupils entering grade 1 for the first time" and "remaining pop (est.)". The proportion by age of pupils entering grade 1 for the first time are equal to the expected intake at or before the appropriate age, and expected intake official age+*t*.

Calculations for the expected survival to grade *i* follow the cohort reconstruction model. Promotion, repetition and dropout rates by grade are copied from Stata into Excel and feed into the model. The starting number of pupils from the cohort reconstruction model is adjusted by the expected never to enter rate (say, 1000 students - NE^{EAt} is the initial number of students of the cohort model). Some of the assumptions behind the model are: i) there will be no additional entrants during the life-time of the cohort other than the original 1000 pupils, ii) the same repetition, dropout and promotion rates apply regardless of whether a pupil has reached that grade previously repeating or directly, iii) pupils are allowed to repeat at most three times, otherwise they form part of the dropout group.[4] In addition, the expected cohort completion rate (primary) is equivalent to the survival of the last grade of primary. A snapshot of the built dataset is displayed in Figure 1.

[4] For details, see: www.uis.unesco.org/i_pages/indspec/cohorte.htm.

Figure 1. Snapshot of the constructed database.

Country	Year	Sex	On time	Wealth quintile	Residence	Region	Ethnicity	Language	Religion	grade	Non-weighted sample size	Weighted sample size	Promotion rate	Repetition rate	Dropout rate	Expected Intake at or before the appropriate age	Expected Intake official age +1	Expected intake official age +2	Expected intake official age +3	Expected intake official age +4 or more	Expected never to enter	Expected Survival to grade
Senegal	2005	0	0	0	0	0	0	0	0	0	10011	10011									0.3666	
Senegal	2005	0	0	0	0	0	0	0	0	1	2065	2065	88.58	10.67	0.75	0.3687	0.1390	0.0596	0.0266	0.0395		0.6334
Senegal	2005	0	0	0	0	0	0	0	0	2	1763	1737	91.09	7.57	1.34							0.6273
Senegal	2005	0	0	0	0	0	0	0	0	3	1562	1608	92.47	6.31	1.23							0.6180
Senegal	2005	0	0	0	0	0	0	0	0	4	1322	1330	90.91	7.82	1.26							0.6098
Senegal	2005	0	0	0	0	0	0	0	0	5	1446	1457	88.97	8.46	2.58							0.6011
Senegal	2005	0	0	0	0	0	0	0	0	6	792	778	71.05	23.93	5.03							0.5839
Senegal	2005	0	0	0	0	0	0	0	0	7	582	564	93.59	4.5	1.91							0.5379
Senegal	2005	0	0	0	0	0	0	0	0	8	479	472	92.98	5.87	1.15							0.5270
Senegal	2005	0	0	1	0	0	0	0	0	0	1493	1493									0.4673	
Senegal	2005	0	0	1	0	0	0	0	0	1	448	468	84.08	14.36	1.56	0.2683	0.1051	0.0677	0.0359	0.0558		0.5327
Senegal	2005	0	0	1	0	0	0	0	0	2	311	309	88.86	9.16	1.99							0.5215
Senegal	2005	0	0	1	0	0	0	0	0	3	250	243	91.96	5.16	2.89							0.5097
Senegal	2005	0	0	1	0	0	0	0	0	4	195	201	92.77	6.92	0.31							0.4942
Senegal	2005	0	0	1	0	0	0	0	0	5	189	179	87.91	8.9	3.19							0.4924
Senegal	2005	0	0	1	0	0	0	0	0	6	68	65	58.89	36.4	4.71							0.4748
Senegal	2005	0	0	1	0	0	0	0	0	7	22	19	98	2	0							0.4184
Senegal	2005	0	0	1	0	0	0	0	0	8	10	9	81.16	18.84	0							0.4184
Senegal	2005	0	0	2	0	0	0	0	0	0	2177	2177									0.3905	
Senegal	2005	0	0	2	0	0	0	0	0	1	470	491	85.44	13.54	1.02	0.3345	0.1512	0.0572	0.0209	0.0458		0.6095
Senegal	2005	0	0	2	0	0	0	0	0	2	437	451	89.47	9.57	0.96							0.6008
Senegal	2005	0	0	2	0	0	0	0	0	3	363	371	92.41	5.99	1.61							0.5939
Senegal	2005	0	0	2	0	0	0	0	0	4	314	311	92.24	6.73	1.04							0.5837
Senegal	2005	0	0	2	0	0	0	0	0	5	335	328	90.87	6.35	2.78							0.5770
Senegal	2005	0	0	2	0	0	0	0	0	6	135	121	64.48	26.86	8.66							0.5598
Senegal	2005	0	0	2	0	0	0	0	0	7	65	55	92.68	5.27	2.06							0.4839
Senegal	2005	0	0	2	0	0	0	0	0	8	58	47	92.17	6.81	1.02							0.4734
Senegal	2005	0	0	3	0	0	0	0	0	0	2659	2659									0.3202	
Senegal	2005	0	0	3	0	0	0	0	0	1	533	546	90.15	9.26	0.59	0.3672	0.1793	0.0652	0.0366	0.0314		0.6798
Senegal	2005	0	0	3	0	0	0	0	0	2	471	478	91.38	6.86	1.76							0.6748
Senegal	2005	0	0	3	0	0	0	0	0	3	419	423	92.88	6.47	0.65							0.6619
Senegal	2005	0	0	3	0	0	0	0	0	4	362	360	90.9	6.48	2.62							0.6571
Senegal	2005	0	0	3	0	0	0	0	0	5	368	353	84.43	12.31	3.26							0.6385
Senegal	2005	0	0	3	0	0	0	0	0	6	241	253	70.24	23.53	6.23							0.6136
Senegal	2005	0	0	3	0	0	0	0	0	7	146	129	92.51	4.63	2.87							0.5563
Senegal	2005	0	0	3	0	0	0	0	0	8	119	116	99.07	0	0.93							0.5396
	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....							

5. Analysis

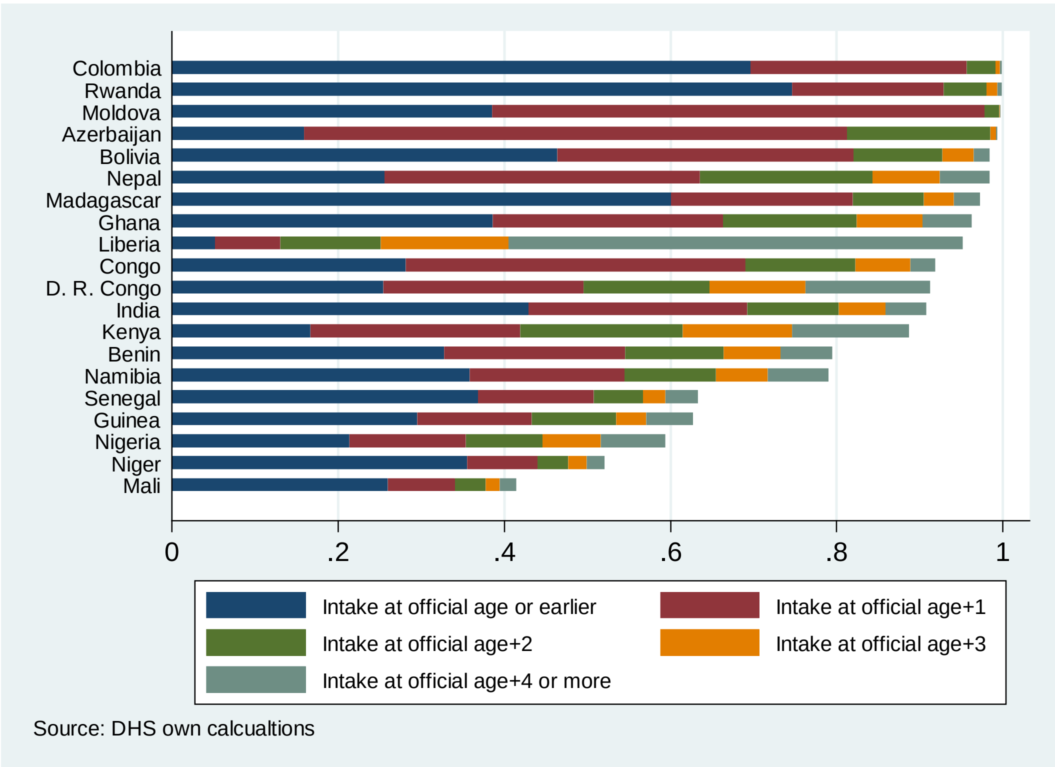
This section offers a preliminary analysis of some of the determinants of expected completion and exit points within the primary education system.

5.1. Intake into the first grade of primary

Figure 2 shows the expected intake into primary school at different ages. Although the sample average of ANIR at the official entry age or earlier is rather low across the sample ($\overline{anir}_0$= 0.35), most countries have a total intake above 75\%. The expected intake is below 65\% for only five out of the total 20 countries incorporated in the dataset. In particular, within the group of countries composed by Mali, Niger, Nigeria, Guinea and Senegal, at least 1 in 3 individuals will never attend school. Perhaps surprisingly, these countries have expected intakes (at the official entry age or earlier) which are higher than countries which perform better in terms of proportions of students out of school. For instance, Kenya, Congo, D. R. Congo, Liberia have lower ANIR than Niger and Senegal. What contributes to these five countries performing poorly is the proportionally less number of individuals who enter late (particularly one year later after the official entry age).

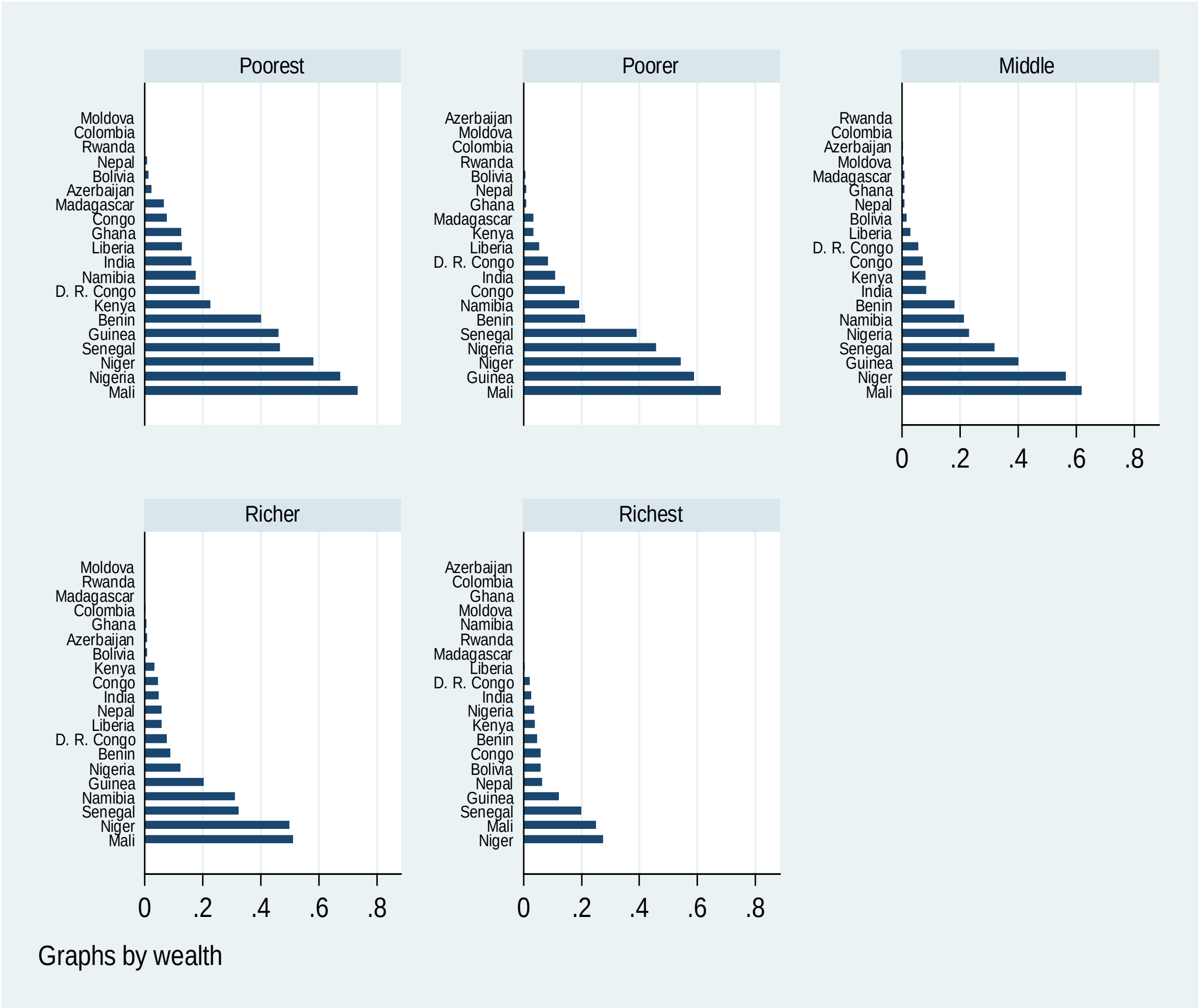
Furthermore, there is a decreasing tendency for older students to be expected to enter the education system; these can be seen in the Figure by the shorter bars' length across ages. Specifically, the total expected intake mean for the different entry ages are: $\overline{anir}_0$= 0.35 > $\overline{anir}_1$=0.26>$\overline{anir}_2$=0.11>$\overline{anir}_3$=0.05<$\overline{anir}_4$=0.07.

Figure 2. Expected intakes at different ages.



With regards to the dimensions explaining the distribution of the expected intake rate, level of wealth plays a key role in the likelihood of "never attend school". As can be seen in Figure 3, the richer people are, the more likely they would be to attend school. Consider, for example, two cases at the bottom of the distribution: Mali and Nigeria. Intuitively, one should expect that in countries where most of the population is out of school, the level of income is a significant determinant of school attendance. Indeed, this is true for these two countries. Whereas for the poorest quintile of Mali, the expected proportion of `never enter primary school' is approximately 75%, this figure goes down to 25% for the richest quintile. Similarly, for Nigeria, the likelihood to never being in school is reduced from 65 percent to less than 10 percent by comparing the poorest and richest groups. Equivalent results hold for countries such as Niger, Senegal, Guinea, Benin and Kenya. Even for some countries in the middle of the rank in Figure 3, if a pupil belongs to the top quintile of wealth, the probability of them never attending school is nearly zero (see, e.g., Namibia, Madagascar and Ghana). Note, on the contrary, there is a lack of explanatory power from income in countries with very high total ANIR (e.g., Azerbaijan, Colombia, and Moldova).

Figure 3. Expected proportion to never enter primary school by wealth quintile.



Figures 4 and 5 clearly demonstrate that the likelihood of "never enter school" also varies by gender and type of residence.[5] Females are a disadvantaged group as far as the expected proportion of "never enter primary school" is concerned. Figure 4 shows that they have a higher percentage of around 30% of never enter school than males, especially in D. R. Congo, Congo, and Niger. Additionally, living in rural areas leads to lower levels of "ever attending school", being at least twice as likely to never enter school as someone who comes from a urban location (see Figure 5). This could be due to the lack of school facilities and infrastructure in rural regions of the chosen countries. Females living in rural sites, therefore, are a particularly disadvantaged and high risk group.

[5] Only countries with a total ANIR less than 95\% are shown.

Figure 4. Expected proportion to never enter primary school by gender (for countries with ≥ 5%).

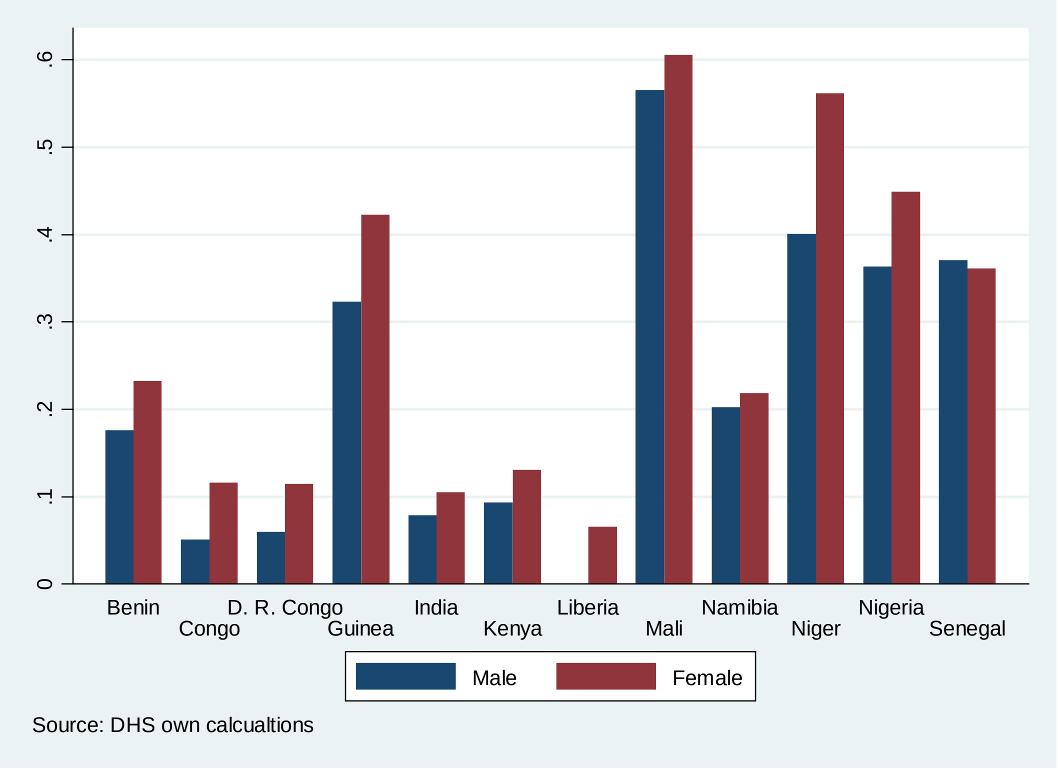
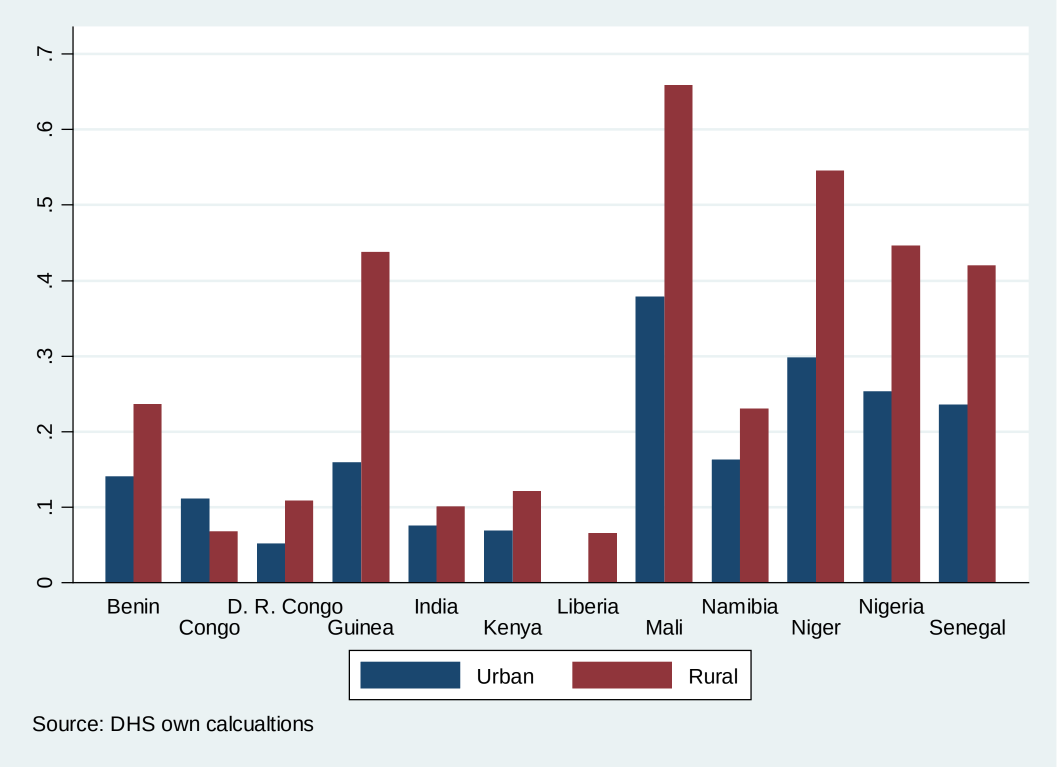


Figure 5. Expected proportion to never enter primary school by residence (for countries with ≥ 5%).



A close regional examination of the group of five countries attaining the lowest expected intake into primary school in Figure 6 reveals significant differences across regions. The odds for the total ANIR between the less favoured region to the most favoured region is nearly 1 to 4. For example, the North West region of Kenya has a mean total intake of just above 20% and, in southern regions, the total ANIR is nearly 100%. Likewise in Mali, the "expected to enter rate" for the Sikasso and Mopti regions are around 30% and, for the Bamako region twice as much. Equally, whereas in the Agadez region of Niger the mean of ANIR is over 80%, it is just 30% in the Zinder and Maradi regions. In other words, there are crucial differences between regions in expected total intakes.

Figure 6. Expected proportion to enter primary school by region (for countries with ≤70%).

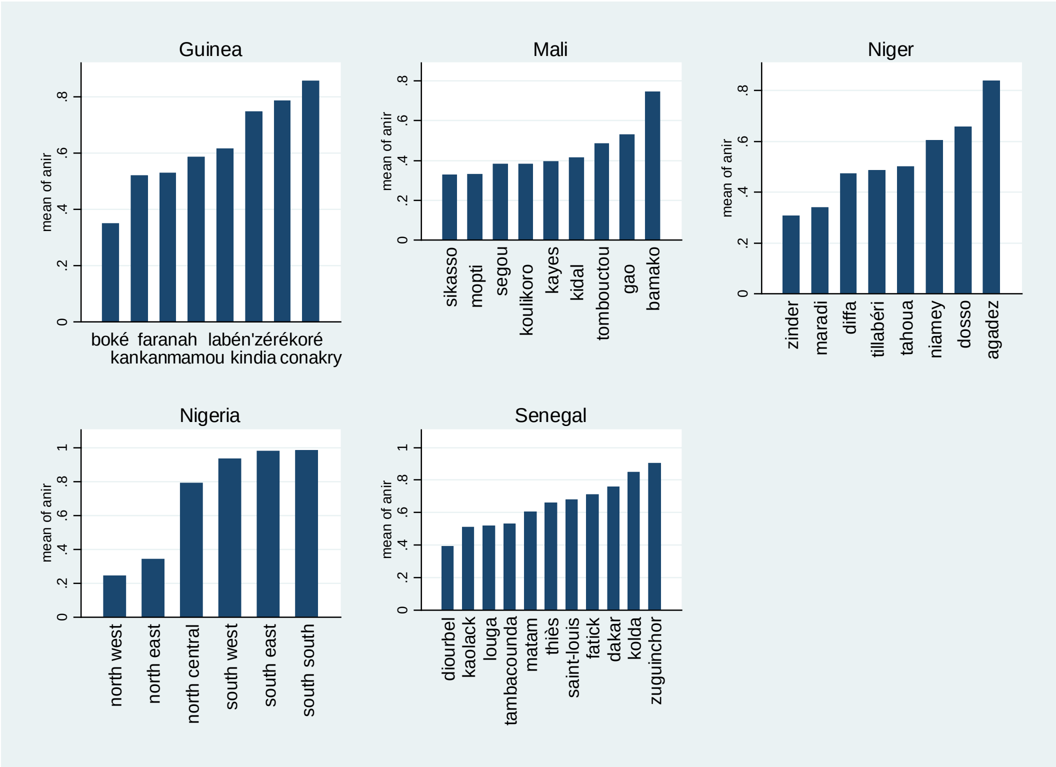


Figure 7 plots the same information by ethnicity. Although, estimations are not as dissimilar as the previous regional analysis, there is still an important variation on the total ANIR by ethnical groups. In Senegal, for instance, "wolof" people are half as probable as the "diola" ethnic group to ever enter school. In Mali, the groups of "tanacheck" and "sonrai" are three times more likely to attend primary school than the "bobo" ethnic. In short, differences in the expected level of ever being in school can be accounted for by the ethnical group in which each pupil belongs.

Table 4 to Table 9 shed further light on the determinants of the expected intake into primary school. To begin with, Tables 4 and 5 present correlations of expected intake with wealth, for both the total sample and for each country. For the whole sample, only the two extreme wealth quintiles are significantly correlated (p-values $\leq$ 0.05) with the mean value of "expected to never enter" indicator. There is a positive correlation between belonging to the poorest income quintile and "never to enter school" of 0.18 and a negative association for the richest quintile ($\hat{\rho}$ = -0.13). Yet, for the middle quintiles, associations are not statistically significant.

Figure 7. Expected proportion to enter primary school by ethnicity (for countries with ≤70%).

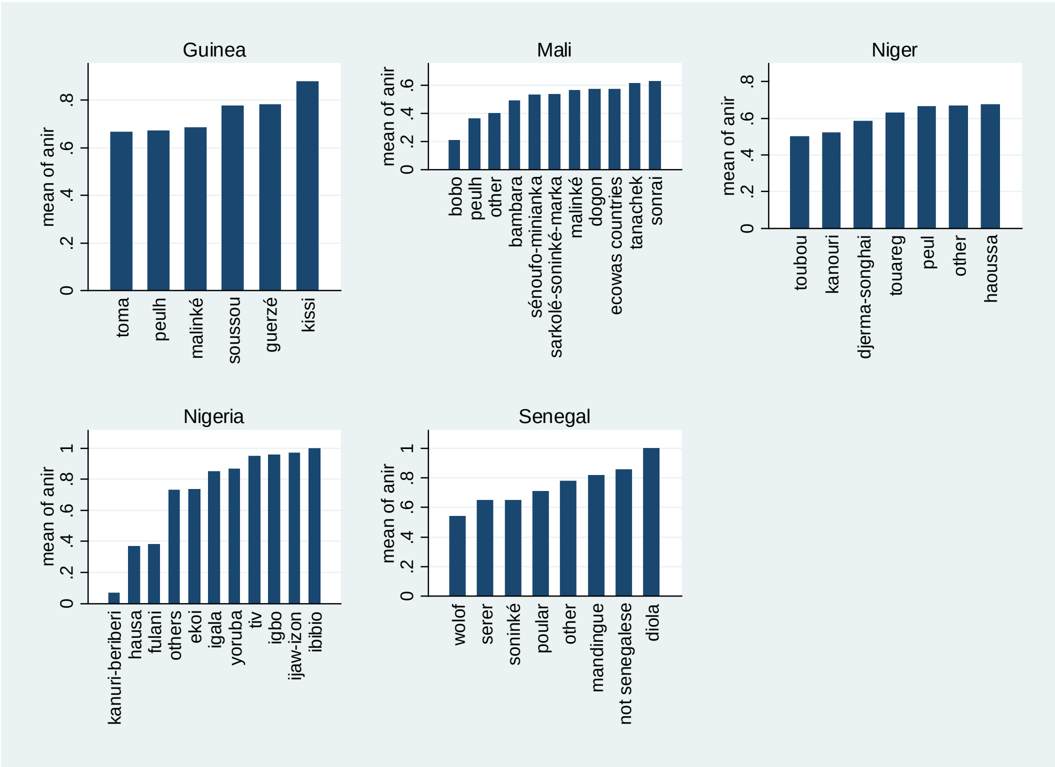


Table 4. Correlation of expected to never enter primary school by wealth quintile.

	Wealth quintile									
	poorest	p-value	poorer	p-value	middle	p-value	richer	p-value	richest	p-value
Never to enter	0.1847	0.0000	0.022	0.5206	-0.0242	0.4804	-0.0558	0.1025	-0.1252	0.0002

Table 5. Correlation of expected intake by age and wealth.

	Expected intake by age									
	anir0	p-value	anir1	p-value	anir2	p-value	anir3	p-value	anir4	p-value
Total sample	0.2987	0	0.0704	0.0396	-0.0814	0.0173	-0.128	0.0002	-0.1363	0.0001
Azerbaijan	0.6358	0	-0.0889	0.5755	-0.2311	0.1409	-0.2513	0.1084	-0.3263	0.035
Benin	0.9537	0	0.7417	0	-0.4816	0.0008	-0.6418	0	-0.8073	0
Bolivia	0.0122	0.9379	0.1426	0.3617	-0.5878	0	-0.3497	0.0215	-0.0469	0.7652
Colombia	-0.0173	0.9169	0.0124	0.9405	-0.2874	0.0761	0.0433	0.7937	0.2657	0.102
Congo	0.6624	0	0.5664	0.0001	-0.3699	0.0173	-0.4864	0.0013	-0.5657	0.0001
D. R. Congo	0.7844	0	0.7211	0	-0.4658	0.0019	-0.4679	0.0018	-0.8303	0
Ghana	0.8821	0	0.0804	0.6039	-0.4902	0.0007	-0.6368	0	-0.4997	0.0006
Guinea	0.8306	0	0.3864	0.0151	-0.1551	0.3459	-0.3494	0.0293	-0.6965	0
India	0.8784	0	0.6899	0	-0.2018	0.1838	0	0.0079	-0.805	0
Kenya	0.7334	0	0.3764	0.0128	0.3543	0.0197	0	0.0453	-0.4767	0.0012
Liberia	0.3549	0.0181	0.7555	0	0.4464	0.0024	0.3763	0.0118	-0.7336	0
Madagascar	0.8257	0	-0.0254	0.8682	-0.6551	0	-1	0	-0.4541	0.0017
Mali	0.8533	0	0.5732	0	0.0567	0.7115	0	0.2669	-0.0004	0.9977
Moldova	0.4808	0.0019	-0.1355	0.4107	-0.2398	0.1414	-1	0.0011	.	0
Namibia	-0.386	0.0097	-0.0282	0.8558	-0.5272	0.0002	0	0.8491	0.078	0.6147
Nepal	-0.6774	0	0.0233	0.879	0.5254	0.0002	0	0.1115	0.0804	0.5995
Niger	0.7573	0	-0.248	0.1448	-0.0298	0.8629	0	0.614	-0.6616	0
Nigeria	0.9353	0	0.9242	0	0.8342	0	0.0839	0.5836	-0.4867	0.0007
Rwanda	0.6098	0	-0.578	0	-0.555	0.0001	0	0.0017	-0.2711	0.0717
Senegal	0.7115	0	-0.1538	0.3188	-0.1355	0.3805	0	0.0135	-0.4178	0.0048

When inspecting the associations by country and by entry age in Table 5, an obvious relationship emerges among wealth and intake rate by entry age. On the one hand, wealth is statistically and positively related to the $\overline{anir}_0$ (expected intake at official age or earlier) and $\overline{anir}_1$ (expected intake at official entry age+1). On the other hand, there are negative associations with $\overline{anir}_2$, $\overline{anir}_3$ and $\overline{anir}_4$. As expected, pupils from richer backgrounds are probable to start on time or, at most, only one year later.

Further support to this hypothesis is contained in the regressions of expected intake entry age of Tables 7, 8, and 9, which yield a decreasing estimated wealth coefficient. That is, $\hat{\beta}^{anir0}_{wealth}$=0.0401>$\hat{\beta}^{anir4}_{wealth}$=-0.0107. Yet, the general relationship between wealth and the total ANIR is positive, as it is indicated by $\hat{\beta}^{ANIR}_{wealth}$=0.0236 in Table 6; and also when one controls for gender and residence.

Table 6. Total ANIR regressions

	model1	model2	model3	model4
gender	-0.0224			-0.0227
	(0.0226)			(0.0223)
wealth		0.0236**		0.0221**
		(0.0081)		(0.0082)
residence			-0.0441	-0.0370
			(0.0225)	(0.0224)
cons	0.8699***	0.7655***	0.9028***	0.8596***
	(0.0356)	(0.0269)	(0.0357)	(0.0560)
N	366	366	366	366
r2	0.0027	0.0226	0.0105	0.0327

Standard errors in parentheses. * p < 0.05, ** p<0.01, *** p<0.001

Results from this Table further suggest that gender and residence do not have any significant impact upon the total ANIR.[6] This is reflected in the overall fit for each specification. For instance, in the case of $\overline{anir}_0$ the $\hat{r}$ equals 0 or 0.01 for gender and residence respectively, whereas it increases to more than 6% in the specification which includes wealth as a covariate.

Table 7. ANIR regressions by age (official entry age or earlier and entry age+1).

	model10	model20	model30	model40	model11	model21	model31	model41
gender	0.006			0.006	-0.015			-0.015
	(0.022)			(0.022)	(0.018)			(0.018)
wealth		0.040***		0.039***		0.007		0.006
		(0.008)		(0.008)		(0.007)		(0.007)
residence			-0.048*	-0.036			-0.021	-0.019
			(0.022)	(0.022)			(0.018)	(0.019)
cons	0.354***	0.243***	0.437***	0.294***	0.276***	0.234***	0.285***	0.286***
	(0.035)	(0.026)	(0.035)	(0.054)	(0.029)	(0.022)	(0.029)	(0.046)
N	366	366	366	366	366	366	366	366
r2	0.000	0.067	0.013	0.074	0.002	0.003	0.003	0.007

[6] Even though the dummy residence is significant at 10%.

In summary, wealth is not only a key determinant of the probability to enter school, but also at the time in which pupils enter primary education.

Table 8. ANIR regressions by age (official entry age+2, official entry age+3)

	model12	model22	model32	model42	model13	model23	model33	model43
gender	-0.014			-0.014	0.005			0.006
	(0.011)			(0.011)	(0.008)			(0.008)
wealth		-0.005		-0.005		-0.008*		-0.008*
		(0.004)		(0.004)		(0.003)		(0.003)
residence			0.002	0.001			0.001	-0.001
			(0.011)	(0.011)			(0.008)	(0.008)
cons	0.128***	0.122***	0.104***	0.141***	0.039**	0.071***	0.046***	0.065**
	(0.018)	(0.013)	(0.018)	(0.028)	(0.013)	(0.010)	(0.013)	(0.021)
N	366	366	366	366	366	366	366	366
r2	0.004	0.004	0.000	0.008	0.001	0.017	0.000	0.019

Table 9. ANIR regressions by age (official entry+4 or more)

	model14	model24	model34	model44
gender	-0.006			-0.006
	(0.013)			(0.012)
wealth		-0.011*		-0.010*
		(0.005)		(0.005)
residence			0.021	0.018
			(0.013)	(0.013)
cons	0.072***	0.095***	0.031	0.074*
	(0.020)	(0.015)	(0.020)	(0.031)
N	366	366	366	366
r2	0.001	0.015	0.008	0.021

5.2. Repetition, dropout and expected schooling patterns

Figure 8 to Figure 11 display survival rates by grade, grouping countries by primary school length. Recall that the constructed dataset adds to extra grades after the official duration of primary school. For instance, for Azerbaijan and Moldova six survival grades are displayed in Figure 8 as their primary duration is 4 years. For the two countries, nearly the full cohort of students is retained within education as the total intake is very similar to the primary completion rate and beyond. Hence, there is stable behaviour between entry and exit points in systems with four years cycles, as well as very high values for total intake. On the contrary, Colombia, India, Madagascar and Nepal whose primary duration is 5 years, are more likely to lose more students during the whole cycle. Figure 9 certainly shows that for these countries, a steady decrease across grades' survival is apparent (between 10 to 20 percent), which becoming clearer after completion of primary (between grades 5 to 7), where dropping out is a more generalized phenomenon.

Figure 8. Expected cohort completion rate by grade (Primary cycle length = 4 years).

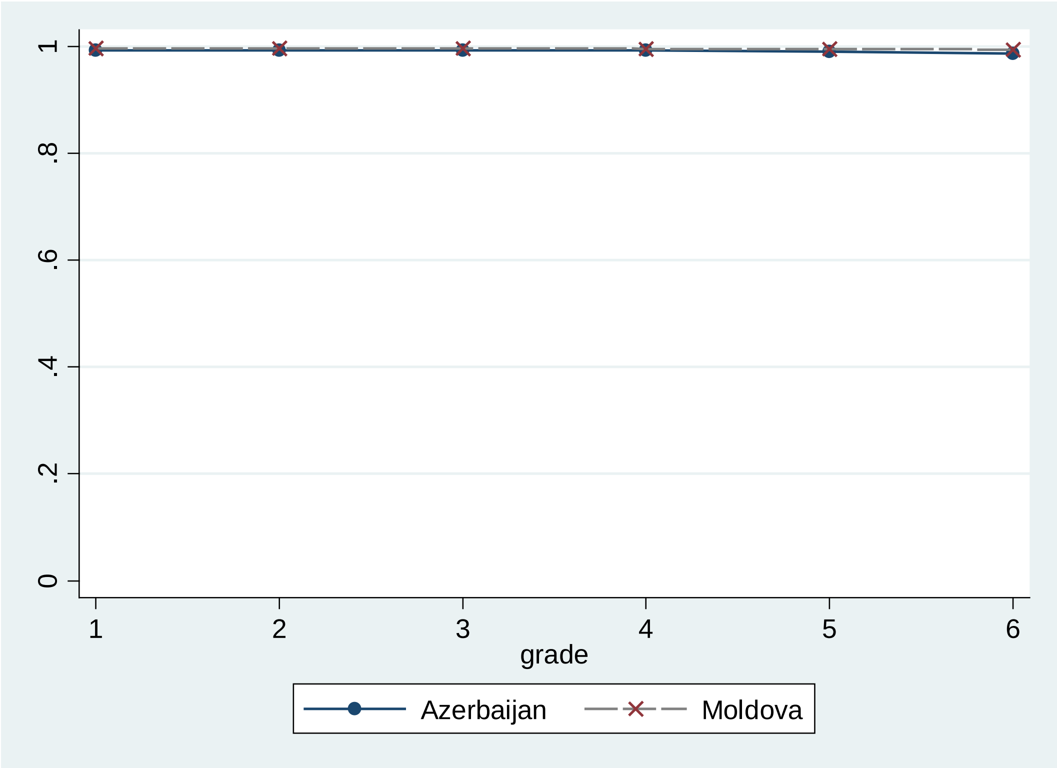


Figure 9. Expected cohort completion rate by grade (Primary cycle length = 5 years).

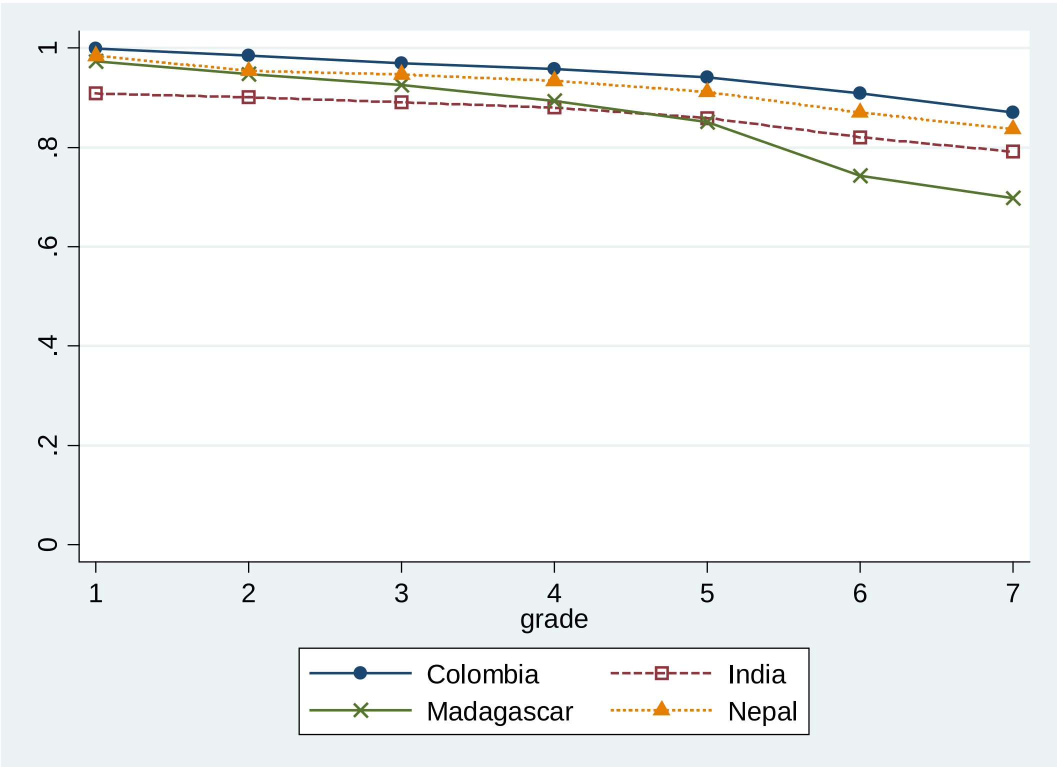


Figure 10. Expected cohort completion rate by grade (Primary cycle length = 6 years).

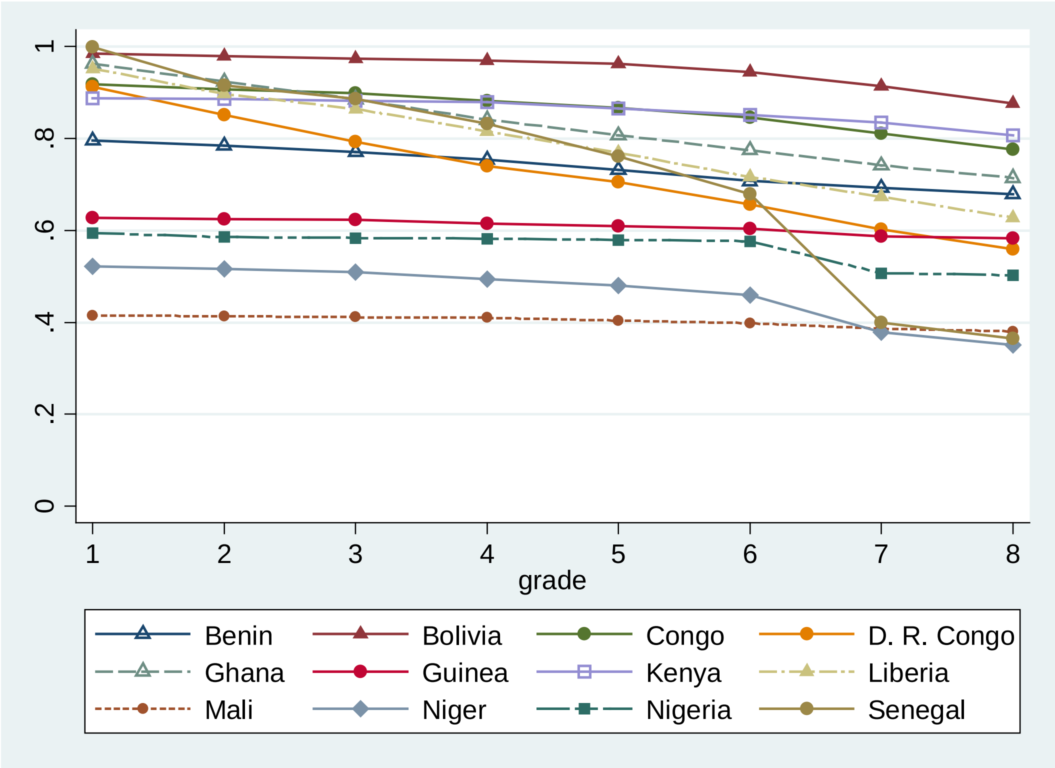
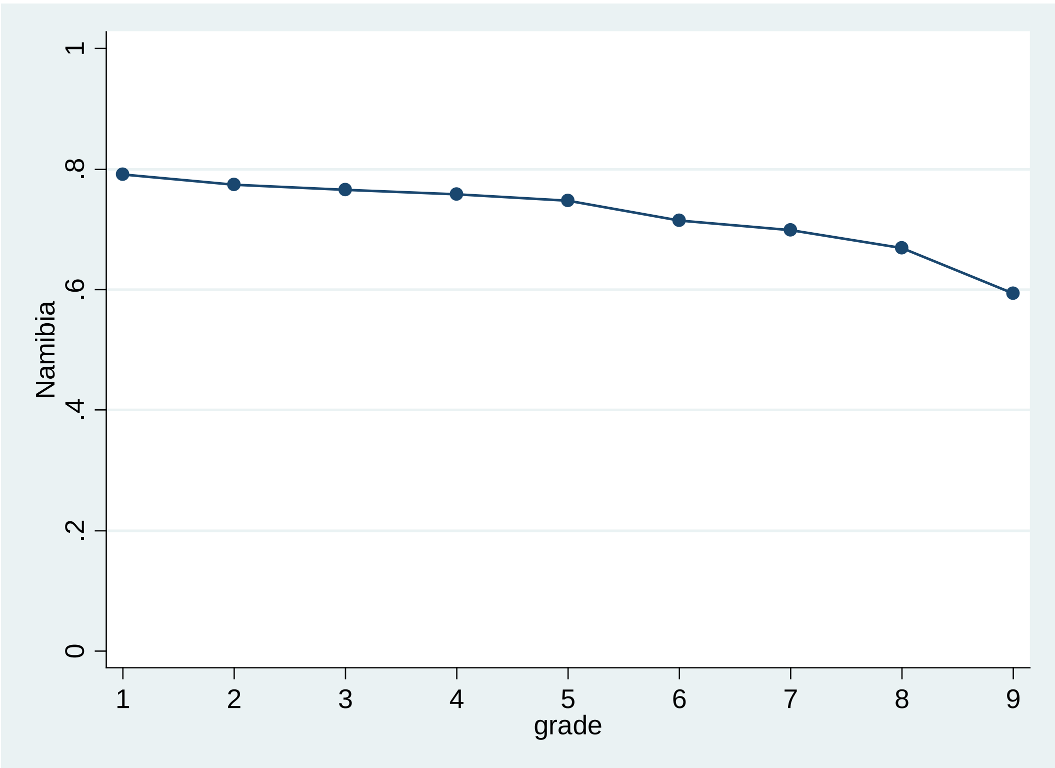


Figure 11. Expected cohort completion rate by grade (Primary cycle length = 7 years).



It should be noted that the majority of countries from the built dataset have primary cycle length of 6 years. As can be seen in Figure 10, this is a heterogeneous group, but, nonetheless, with largely lower rates of both initial intake and completion of primary school than the two previous sets of countries. In fact, the mean value for the survival to grade 6 for these 12 countries is 0.67, and for the two earlier groups of countries of 0.98 and 0.86. Certain countries stand out by exhibiting a sharp decrease in the cohort completion rates and very low constant values of survival rates across the cycle. For the latter group, survival rates across grades between 0.4 and 0.6 are rather stable (for example, Mali and Liberia). However, for the former set of countries, they lose nearly half of their cohort before they reach the final grade of primary education (e.g., D. R. Congo, Liberia). Furthermore, Namibia tends to lose more students in the two years after primary school terminates, after grade 7 (Figure 11).

Wealth, a previously found key force driving the proportion of individuals entering school on time or slightly over-age (entry age+1), it also a major influence on the completion rates of primary school. Wealth has a significant and positive impact ($\hat{\beta}_{wealtj}$~0.04, but gender and residence are not, as is shown in Table 10. Progression of the cohort to each grade in Table 11 is directly affected by wealth too and it is increasing; that is, $\hat{\beta}^{grade1}_{wealth}$=0.045 <….< $\hat{\beta}^{grade8}_{wealth}$=0.068. When the cycle considered is 5 years, only the first two grades of this cycle are not related to income (Table 11). Therefore, in general, the wealthier a student is, the higher the probability that he/she will survive to each grade of the cycle and complete primary school and beyond.

Table 10. Expected completion of primary school regressions

	model1	model2	model3	model4
gender	-0.0279			-0.0287
	(0.0224)			(0.0216)
wealth		0.0424***		0.0416***
		(0.0079)		(0.0079)
residence			-0.0344	-0.0210
			(0.0224)	(0.0217)
cons	0.7798***	0.6108***	0.7899***	0.6877***
	(0.0353)	(0.0260)	(0.0355)	(0.0542)
N	366	366	366	366
r2	0.0042	0.0740	0.0064	0.0809

Table 11. Survival by grade regressions (primary cycle length = 6 years)

	grade1	grade2	grade3	grade4	grade5	grade6	grade7	grade8
gender	-0.032	-0.034	-0.028	-0.033	-0.038	-0.043	-0.066*	-0.072*
	(0.029)	(0.027)	(0.027)	(0.026)	(0.025)	(0.025)	(0.029)	(0.029)
residence	-0.027	-0.023	-0.020	-0.022	-0.021	-0.016	-0.030	-0.025
	(0.029)	(0.028)	(0.027)	(0.026)	(0.025)	(0.025)	(0.029)	(0.029)
wealth	0.045***	0.049***	0.052***	0.055***	0.057***	0.063***	0.066***	0.068***
	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)
cons	0.743***	0.702***	0.661***	0.641***	0.621***	0.572***	0.557***	0.527***
	(0.073)	(0.069)	(0.067)	(0.065)	(0.063)	(0.062)	(0.073)	(0.074)
N	237	237	237	237	237	237	235	233
r2	0.081	0.104	0.119	0.139	0.155	0.185	0.170	0.181

Table 12. Survival by grade regressions (primary cycle length = 5 years)

	grade1	grade2	grade3	grade4	grade5	grade6	grade7
gender	0.002	0.007	0.014	0.014	0.014	0.008	0.007
	(0.024)	(0.023)	(0.023)	(0.023)	(0.024)	(0.027)	(0.029)
residence	0.003	0.012	0.019	0.027	0.017	-0.015	0.006
	(0.024)	(0.023)	(0.023)	(0.023)	(0.024)	(0.027)	(0.029)
wealth	0.005	0.012	0.018*	0.027**	0.031***	0.038***	0.048***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.010)	(0.010)
cons	0.919***	0.854***	0.802***	0.745***	0.727***	0.714***	0.611***
	(0.058)	(0.055)	(0.055)	(0.056)	(0.058)	(0.065)	(0.071)
N	76	76	76	76	76	76	76
r2	0.005	0.035	0.078	0.141	0.161	0.184	0.228

The focus is now on dropout and repetition patterns. Firstly, at which point of the primary cycle are students more likely to drop out? Moreover, how do drop out patterns vary by country? Figures 12 and 13 attempt to provide an answer to these questions. For most countries, a share of 60% of the total cycle dropout occurs within the first two grades. Quite the opposite, in the middle part of the cycle the number of exits from the education system reduces considerably, to less than 15% in grades 3 and 4. The dropout share rises up again, just before where the primary school cycle concludes, with mean proportions of around 35% for grade 6. There are no specific dropout patterns by country since results appear to be linked to the length of each country's schooling cycle (large at the beginning and at the end of each cycle).

Figure 12. Dropout at each grade as share of total dropout cycle (grades 1 to 4).

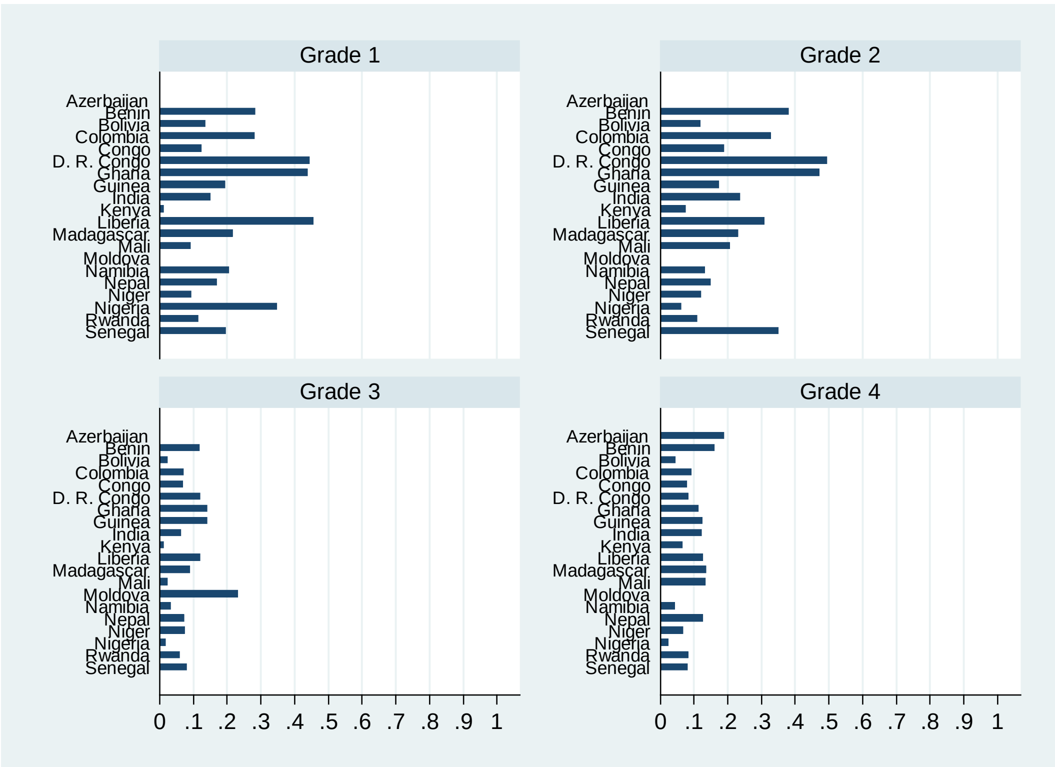


Figure 13. Dropout at each grade as share of total dropout cycle (grades 5 to 8).

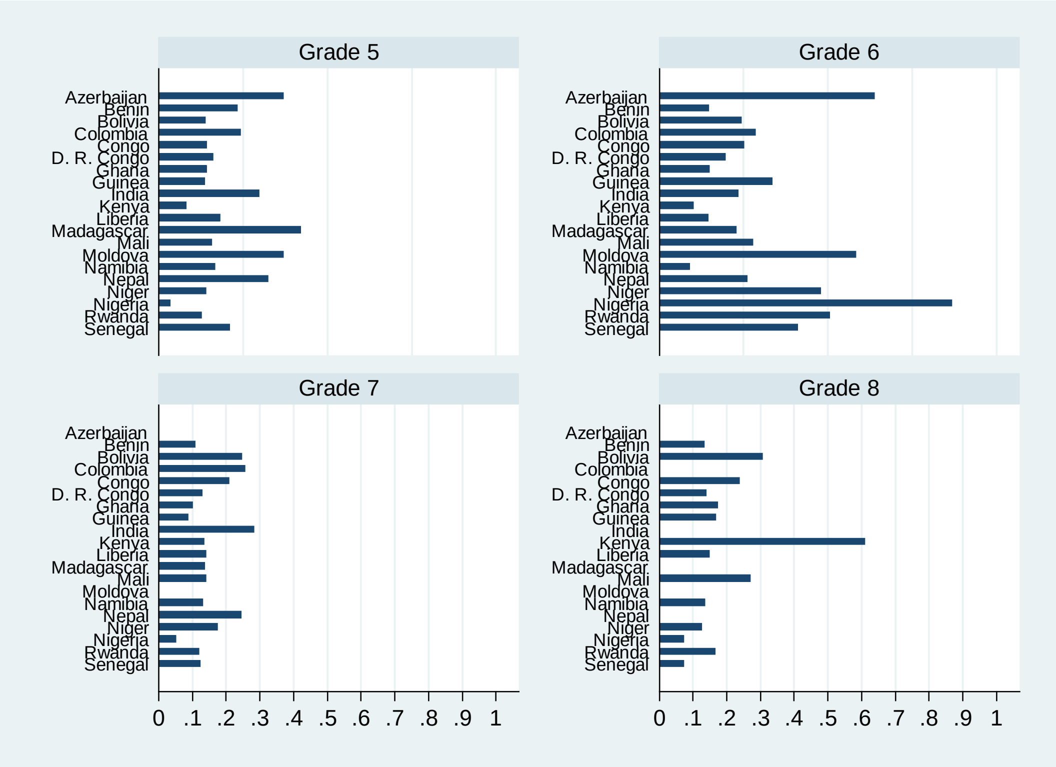
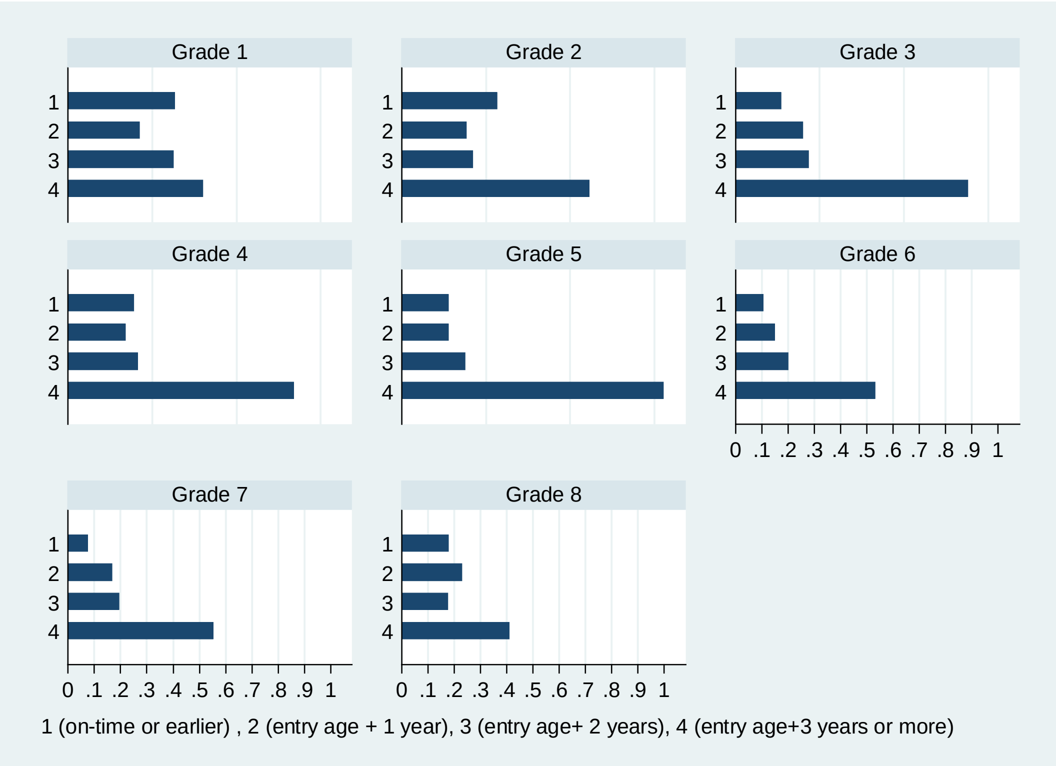


Figure 14. Proportion of total grade's dropout by over-age students categories.



The same analysis is carried out by entry age. Here, the total dropout per grade is divided by the dropout rates of each over-age student category. As shown in Figure 14, the older an individual is when entering primary school, the more likely he/she will exit the system at some point of the schooling cycle. Particularly, the group of students who enter at least 3 years later than the official entry age, is the one which performs very poorly. The other three groups are more alike in terms of dropout; though dropout rates for over-aged students by 2 years are mostly above dropout rates of 1 year over-age students.

Finally, regressions of dropout and repetition rates, controlling for country fixed effects, are displayed in Tables 13 and 14. The two dependent variables are measured in terms of deviation from their grades' averages. Likelihood ratio test on country fixed effects (not shown in the Tables) reveals that, for all regressions, specific impacts for each country are significant. The simpler models, which do not include the 19 countries dummies, are not nested within the full country effects models. This result holds across grades and for the two indicators. On the other hand, only wealth has a significant impact in dropout and repetition. Interestingly, the largest negative effect of income for dropout occurs in the last grade of the primary cycle and in the first grade of the next cycle for repetition rates. Consequently, country and wealth have a strong influence on how these indicators depart from their average values.

Table 13. Regression of dropout (deviations from its mean) by grade with country fixed effects.

	grade1	grade2	grade3	grade4	grade5	grade6	grade7	grade8
gender	-0.04	-0.48	0.31	-0.23	0.50	1.15	1.98*	2.71
	(-0.14)	(-1.25)	(1.10)	(-0.55)	(1.07)	(1.33)	(2.01)	(1.84)
residence	-0.36	-0.57	0.05	-0.26	0.42	-0.09	-0.58	-1.09
	(-1.40)	(-1.48)	(0.17)	(-0.63)	(0.89)	(-0.10)	(-0.59)	(-0.73)
wealth	-0.38***	-0.61***	-0.43***	-0.48**	-0.64***	-1.42***	-0.86*	-1.26*
	(-4.14)	(-4.44)	(-4.19)	(-3.17)	(-3.80)	(-4.58)	(-2.41)	(-2.37)
N	393	391	393	391	393	390	344	262

t statistics in parentheses. * p < 0.05, ** p<0.01, *** p<0.001

Table 14. Regression of repetition (deviations from its mean) by grade with country fixed effects.

	grade1	grade2	grade3	grade4	grade5	grade6	grade7	grade8
gender	-1.50	-1.21	-0.62	-0.05	0.20	1.71	0.75	-0.11
	(-1.78)	(-1.74)	(-1.14)	(-0.11)	(0.35)	(1.47)	(0.63)	(-0.11)
residence	0.82	-0.44	1.07*	0.21	0.50	2.54*	-1.37	-1.44
	(0.97)	(-0.63)	(1.97)	(0.47)	(0.89)	(2.17)	(-1.15)	(-1.39)
wealth	-0.92**	-0.70**	-0.61**	-0.51**	-0.53**	-0.79	-1.86***	-0.81*
	(-3.08)	(-2.86)	(-3.16)	(-3.26)	(-2.64)	(-1.89)	(-4.37)	(-2.18)
N	393	391	393	391	393	390	344	262

6. Conclusions

This report consisted of a methodological note and a preliminary analysis of a built dataset for an array of 24 countries. The data source employed was the Demographic and Health Surveys (DHS). The DHS contains enough schooling information to derive dropout, repetition and promotion rates per grade by comparing variables for two consecutive years. The main calculated components of the dataset are total intakes by entry age and survival rates by grade. These are used to directly obtain expected indicators of a cohort “to never to enter school” and “to complete primary school”. The main procedure consisted of exporting tabulates from Stata into Excel and relying on the methodology of the reconstructed cohort flow model. Inputs of this model are the matrix of promotion, repetition and dropout rates as well as the initial probability to “never enter school” by the cohort. Standard assumptions, such as the maximum number of times a student is allowed to repeat, are also incorporated into the model.

From a preliminary investigation of the dataset, numerous issues arise. Firstly, as far as the dimensions explaining the distribution of the different indicators is concerned, the level of wealth is not the only key determinant of the probability to enter school, and the survival to each grade and completion rates of primary school, but also the time at which pupils enter primary education. Secondly, the likelihood of “never enter school” also varies by gender and type of residence, but not the survival and completion rates. Furthermore, dropout patterns by country are high at the beginning and at the end of each primary school cycle, and over-age students being the highest dropout risk group.

References

Cameron, L., 2005. Primary completion rates. EPDC Technical Paper WP-09-01, Education Policy and Data Center, Washington, DC.

Hunt, F., 2008. Dropping out from school: A cross-country review of literature. Consortium for Research on Educational Access, Transitions and Equity CREATE Pathways to Access No 16, University of Sussex, Brighton.

Ingram, G., Wils, A., Carrol, B., Townsend, F., 2007. Educational inequality within countries: Who are the out of school children? Policy briefs, Education Policy and Data Center, Washington, DC.

Lewin, K., Little, A., 2011. Access to education revisited: Equity, drop out and transitions to secondary school in South Asia and Sub-Saharan Africa. *International Journal of Educational Development, CREATE Special Edition,* 31, 333-337.

Sabates, R., Akyeampong, K., Westbrook, J., Hunt, F., 2011. The hidden crisis: Armed conflict and education. Tech. rep., Paper commissioned for the EFA Global Monitoring Report.

UIS, 2009. Education indicators. Technical guidelines. Tech. rep., Montreal, UNESCO.

UIS, 2011. Questionnaire on statistics of education, pre-primary, primary, secondary and post-secondary non-tertiary education. Tech. rep., Montreal, UNESCO.

UNESCO, 2006. National education sector development plan: a result-based planning handbook. Paris:

UNESCO, 2007. Education for all by 2015: Will we make it? EFA global monitoring report 2008, Paris: UNESCO Publishing.

UNESCO, 2010. Reaching the marginalized. EFA global monitoring report 2010, Paris: UNESCO Publishing.

Wils, A., 2004. Late Entrants Leave School Earlier: Evidence from Mozambique. *International Review of Education*, 50, 17-37.

Wils, B., Sylla, B., Oliver, S., 2009. Pupil performance and age: A study of promotion, repetition, and dropout rates among pupils in four age groups in 35 developing countries. Tech. rep., Paper commissioned for the EFA Global Monitoring Report.
29