

Education Attainment and Enrollment around the World: An International Database
<http://www.worldbank.org/en/research/brief/edattain>

Deon Filmer
DECRG, The World Bank
February 2023

This document describes edattainxtry_5.dta, the fifth release of an international database with summary information on education attainment and enrollment derived from household surveys from around the developing world. The household survey data sets used to generate the summary statistics are:

- Demographic and Health Surveys (DHS) (<http://www.measuredhs.com>);
- the Multiple Indicator Cluster Surveys (MICS2) (<http://www.childinfo.org>);
- selected Living Standards Measurement Study Surveys (LSMS) (<http://www.worldbank.org/lsmis>);
- other selected country-specific Integrated Household Surveys (IHS) such as Socio-Economic Surveys.

The database, in Stata dta format, format contains the following indicators:

- Percentage of the population ages 15 to 19 that has completed each grade (grades 1 to 9);
- Percentage of the population ages 20 to 29 that has completed each grade (grades 1 to 9);
- Percentage of the population ages 30 to 39 that has completed each grade (grades 1 to 9);
- Percentage of the population ages 40 to 49 that has completed each grade (grades 1 to 9);
- Median grade attained by the population ages 15 to 19;
- Survival function estimates of the proportion of the population ages 6 to 19 that has completed each grade;
- Percentage of the population ages 6 to 17, by age in single years, that has ever attended school;
- Percentage of the population ages 6 to 17, by age in single years, that is currently in school;
- Percentage of the population ages 6 to 11 (as a group) that is currently in school;
- Percentage of the population ages 12 to 14 (as a group) that is currently in school;
- Percentage of the population ages 6 to 14 (as a group) that is currently in school.
- Concentration index (and its standard error) of the percentage of the population ages 15 to 19 that has completed grade 6.
- Concentration index (and its standard error) of the percentage of the population ages 6 to 14 that is currently in school.
- Percentage of the population ages 6 to 24, by age in single years, that is currently in school, by level of school being attended (primary, secondary, tertiary).

Each indicator (except the concentration indexes) is calculated for a variety of subgroups within the population. These are:

- All;
- Rural and Urban;
- Male and Female;
- Rural/Urban and Male/Female interacted;
- Poverty Quintiles (poorest quintile, quintiles 2, 3, 4 and richest quintile) with welfare rankings derived from an index of consumer durables and housing characteristics;
- Poverty Groups (Poorest 40 percent; Middle 40 percent; Richest 20 percent) with welfare rankings derived from an index of consumer durables and housing characteristics;
- The interaction of Poverty Groups and Male/Female

For surveys with a consumption-expenditures aggregate variable that can be used to rank households (i.e. the Integrated Household Surveys), the following indicators are also calculated:

- Expenditure Poverty Quintiles (poorest quintile, quintiles 2, 3, 4 and richest quintile) with welfare rankings derived from per capita household consumption-expenditures;
- Expenditure Poverty Groups (Poorest 40 percent; Middle 40 percent; Richest 20 percent) with welfare rankings derived from per capita household consumption-expenditures;
- The interaction of Expenditure Poverty Groups and Male/Female

The naming conventions for the indicators are:

Indicator	Naming convention
Percentage of the population ages 15 to 19 that has completed each grade (grades 1 to 9)	Variables starting with “a” and end with “_number” where <i>number</i> is the grade under consideration
Percentage of the population ages 20 to 29 that has completed each grade (grades 1 to 9)	Variables starting with “a2” and end with “_number” where <i>number</i> is the grade under consideration
Percentage of the population ages 30 to 39 that has completed each grade (grades 1 to 9)	Variables starting with “a3” and end with “_number” where <i>number</i> is the grade under consideration
Percentage of the population ages 40 to 49 that has completed each grade (grades 1 to 9)	Variables starting with “a4” and end with “_number” where <i>number</i> is the grade under consideration
Median grade attained by the population ages 15 to 19	Variables starting with “m1519”
Survival function estimates of the proportion of the population ages 6 to 19 that has completed each grade	Variables starting with “s” and end with “_number” where <i>number</i> is the grade under consideration
Percentage of the population ages 6 to 17, by age in single years, that has ever attended school	Variables starting with “e” and end with “_number” where <i>number</i> is the age under consideration
Percentage of the population ages 6 to 17, by age in single years, that is currently in school	Variables starting with “c” and end with “_number” where <i>number</i> is the age under consideration
Percentage of the population ages 6 to 11 (as a group) that is currently in school	Variables starting with “c611”
Percentage of the population ages 12 to 14 (as a group) that is currently in school	Variables starting with “c1214”
Percentage of the population ages 6 to 14 (as a group) that is currently in school	Variables starting with “c614”
Estimate of the concentration index (and its standard error*) of the percentage of the population ages 15 to 19 that has completed grade 6 – where the welfare ranking used to calculate the concentration index is based on an index of consumer durables and housing characteristics	mci_a6 (sci_m6)
Estimate of the concentration index (and its standard error*) of the percentage of the population ages 6 to 14 that has is currently in school – where the welfare ranking used to calculate the concentration index is based on an index of consumer durables and housing characteristics	mci_c614 (sci_c614)
Estimate of the concentration index (and its standard error*) of the percentage of the population ages 15 to 19	mcix_a6 (scix_m6)

that has completed grade 6 – where the welfare ranking used to calculate the concentration index is based on per capita household consumption-expenditures	
Estimate of the concentration index (and its standard error*) of the percentage of the population ages 6 to 14 that has is currently in school – where the welfare ranking used to calculate the concentration index is based on per capita household consumption-expenditures	mcix_c614 (scix_c614)
Percentage of the population ages 6 to 24, by age in single years, that is currently in school, by level of school being attended (primary, secondary, tertiary).	Variables starting with “ cl_p ” (primary) “ cl_s ” (secondary) and “ cl_t ” (tertiary) and end with “_number” where <i>number</i> is the age under consideration
* Standard errors are calculated from between 200 and 1000 bootstrap replications of the concentration index calculation.	

The naming conventions for the disaggregations of the data are:

Subgroup	Naming convention (i.e. variable name contains the following expression)
All	All
Rural and Urban	Rural, Urban
Male and Female	Male, Female
Rural/Urban and Male/Female interacted	MalUrb, MalRur, FemUrb, FemRur
Poverty Quintiles (poorest quintile, quintiles 2, 3, 4 and richest quintile) with welfare rankings derived from an index of consumer durables and housing characteristics	Quint1, Quint2, Quint3, Quint4, Quint5
Poverty Groups (Poorest 40 percent, Middle 40 percent, Richest 20 percent) with welfare rankings derived from an index of consumer durables and housing characteristics	Poor40, Mid40, Rich20
The interaction of Poverty Groups and Male/Female	Poo40M, Poo40F, Mid40M, Mid40F, Ric20M, Ric20F
Expenditure Poverty Quintiles (poorest quintile, quintiles 2, 3, 4 and richest quintile) with welfare rankings derived from per capita household consumption-expenditures	Quint1x, Quint2x, Quint3x, Quint4x, Quint5x
Expenditure Poverty Groups (Poorest 40 percent, Middle 40 percent, Richest 20 percent) with welfare rankings derived from per capita household consumption-expenditures	Poor40x, Mid40x, Rich20x
The interaction of Expenditure Poverty Groups and Male/Female	Poo40Mx, Poo40Fx, Mid40Mx, Mid40Fx, Ric20Mx, Ric20Fx

In addition to the variables described above, an additional selected set of country/year variables have been merged in from the [World Development Indicators database](#) (accessed May 9, 2022). The variables included are:

countrycode	Country Code
countryname	Country Name
region	Region Code
regionname	Region Name
adminregion	Administrative Region Code
adminregionname	Administrative Region Name
incomelevel	Income Level Code
incomelevelname	Income Level Name
lendingtype	Lending Type Code
lendingtypename	Lending Type Name
exppsp	Expenditure per student, primary (% of GDP per capita)
exppss	Expenditure per student, secondary (% of GDP per capita)
exppst	Expenditure per student, tertiary (% of GDP per capita)
pcrf	Primary completion rate, female (% of relevant age group)

pcrm	Primary completion rate, male (% of relevant age group)
pert	Primary completion rate, total (% of relevant age group)
pyears	Primary education, duration (years)
studp	Primary education, pupils
gerp	School enrollment, primary (% gross)
nerp	School enrollment, primary (% net)
gers	School enrollment, secondary (% gross)
ners	School enrollment, secondary (% net)
syears	Secondary education, duration (years)
studs	Secondary education, pupils
gdppc2015	GDP per capita (constant 2015 US\$)
gdppcp2017	GDP per capita, PPP (constant 2017 international \$)
gnipca	GNI per capita, Atlas method (current US\$)
gnipcp	GNI per capita, PPP (current international \$)
gnipc2015	GNI per capita (constant 2015 US\$)
gnipcp2017	GNI per capita, PPP (constant 2017 international \$)
cpi2010	Consumer price index (2010 = 100)
pov1_190	Poverty headcount ratio at \$1.90 a day (PPP) (% of population)
pov2_320	Poverty headcount ratio at \$3.20 a day (PPP) (% of population)
povgap1_190	Poverty gap at \$1.90 a day (PPP) (%)
povgap2_320	Poverty gap at \$3.20 a day (PPP) (%)
gini	GINI index
pop014	Population ages 0-14, total
popt	Population, total

Note that these data are not available for all country/year combinations.

If using this database, a suggested reference is:

Filmer, Deon. 2023. "Education Attainment and Enrollment around the World: An International Database." <http://www.worldbank.org/en/research/brief/edattain>. Version 5plus. Accessed MM/DD/YYYY.

Further reading:

- Poverty:
 - “The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries”. *Population and Development Review*. 1999. 25(1). Deon Filmer and Lant Pritchett.
- Gender:
 - “Gender and Wealth Disparities in Schooling: Evidence from 44 Countries.” *International Journal of Education Research*. 2005. 43(6). Deon Filmer.
- Distance:
 - “If you build it, will they come? School availability and school enrolment in 21 poor countries” *Journal of Development Studies*. 2007. 43(5). Deon Filmer.
- Orphanhood:
 - “Inequalities in Children’s schooling: AIDS, Poverty, and Gender” *World Development*. 2006. 34(6). Martha Ainsworth and Deon Filmer.
- Disability:
 - “Disability, poverty and schooling in developing countries: Results from 14 household surveys” *World Bank Economic Review*. 2008. 22(1). Deon Filmer.
- Methods:
 - “Estimating Wealth Effects without Expenditure Data – or Tears: With an Application to Educational Enrollments in States of India”. *Demography*. 2001. 38(1). Deon Filmer and Lant Pritchett.
 - “Assessing Asset Indices.” World Bank Policy Research Working Paper No. 4605. April 2008. Deon Filmer and Kinnon Scott.
 - “Chapter 8: The Concentration Index” in *Analyzing Health Equity using Household Survey Data* by Owen O’Donnell, Eddy van Doorslaer, Adam Wagstaff, and Magnus Lindelow. <http://go.worldbank.org/LVSSZJX900>