

FEASIBILITY ANALYSIS FOR THE CONSTRUCTION OF NEW INDICATORS USING HOUSEHOLD SURVEYS

Prepared by Miguel Castillo, consultant, under the supervision of Xavier Mancero, Chief of the Social Statistics Unit, Statistics Division, ECLAC.

UNITED NATIONS ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN (ECLAC)
STATISTICS DIVISION

November 22, 2022

1 Introduction

2 Conclusions

3 Appendix

- ECLAC´s Statistical Division regularly estimates regionally comparable education indicators (completion rates, net attendance rates, out-of-school rates among others) from BADEHOG.
 - Repository of household surveys for Latin America and the Caribbean.
 - Harmonized variables for comparison between countries.
 - Use of income variables for disaggregation.
- The main objective of this report was the feasibility analysis for the construction of new indicators using household surveys
 - Percentage of children over-age for grade (primary education, lower secondary education) (4.1.5)
 - Educational attainment rates by age group and level of education (4.4.3)
 - Youth/adult literacy rate (4.6.2)
 - Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months(4.3.1)

The process employed is the following:

- ① Review of the indicator metadata and required inputs.
- ② Review of the variables necessary to estimate the indicator.
 - ① Assess whether the harmonized variables already available in BADEHOG adjust to what is required.
 - ② If not, detailed review of survey questionnaires to find new variables to harmonize.
- ③ For those indicators that are possible to estimate, the corresponding processing codes were developed and the results obtained are presented:
 - ① Point estimation by country, year and different disaggregations.
 - ② Measures of precision of the point estimate (Standard error, confidence interval and coefficient of variation), taking into consideration the survey sampling design (when available).

What is BADEHOG?

BADEHOG is the ECLAC's Household Survey Data Bank. It has more than 250 household survey databases from 18 Latin American and Caribbean countries, for which we harmonize income, demographic, and social variables, among others. These databases are used to generate indicators of different dimensions (i.e poverty, inequality, labour, education and others) that are uploaded to the CEPALSTAT platform.

Country/Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total	
Argentina	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Bolivia (Est. Pluri. De	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	19
Brazil		✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	19
Chile	✓			✓			✓			✓			✓			✓		✓				✓		9
Colombia			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	17
Costa Rica	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	21
Ecuador		✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
El Salvador	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	19
Guatemala	✓						✓								✓									3
Honduras		✓			✓	✓				✓	✓	✓	✓	✓	✓	✓	✓		✓	✓				13
Mexico	✓		✓		✓		✓		✓				✓		✓					✓		✓		11
Nicaragua		✓				✓				✓					✓									4
Panama	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	18
Paraguay		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Peru	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	21
Dominican Republic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	21
Uruguay	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Venezuela (Rep. Bol. de)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓									15

Percentage of children over-age for grade (primary education, lower secondary education) (4.1.5)

Definition

UNESCO DEFINITION: Percentage of pupils in each level of education (primary and lower secondary general education) who are at least 2 years above the intended age for their grade.

Feasibility analysis

This indicator requires three variables, age, years of study and entrance age. The first two are available in the surveys and the third is based on ISCED 2011 for each country (see appendix)

Calculation method

$$OA_i = \begin{cases} 1 & \text{if } age_i - e_c - yearsstudy_i \geq 2 \\ 0 & \text{si } age_i - e_c - yearsstudy_i < 2 \end{cases} \rightarrow POAG_n = \frac{\sum_{g=1}^{d_n} OA_i}{E_n}$$

Where:

- age_i : age of the pupil i .
- e_c : entrance age at primary education.
- $yearsstudy_i$: approved years of study.
- $POAG_n$: Percentage of children over-age for grade in level n of education.
- E_n : Total enrollment in level n of education (all grades combined)

Percentage of children over-age for grade (primary education, lower secondary education) (4.1.5)

Results

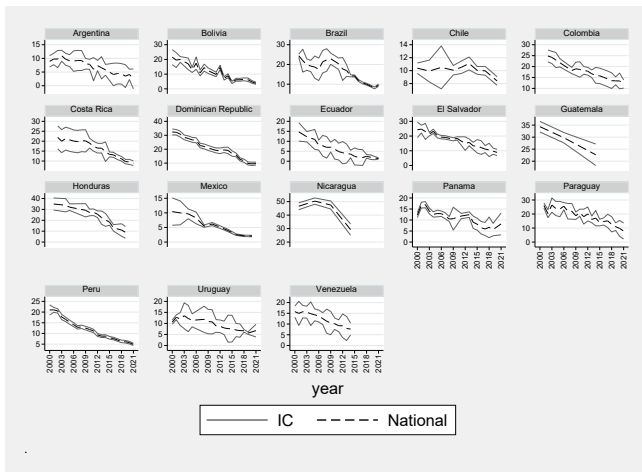


Figure: Primary

Percentage of children over-age for grade (primary education, lower secondary education) (4.1.5)

Results

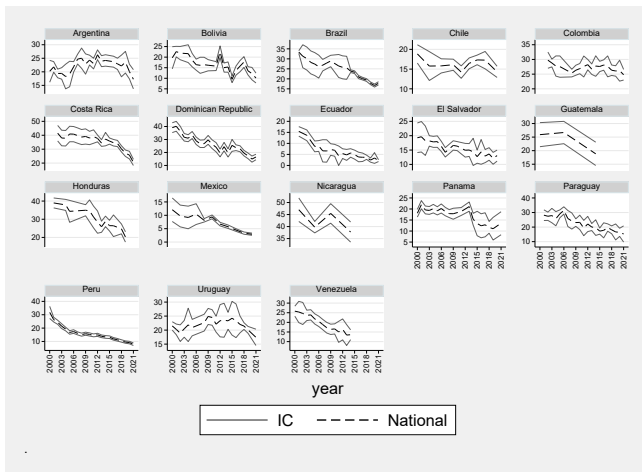


Figure: Lower Secondary

Definition

UNESCO DEFINITION: Distribution of the population aged 25 years and above according to the highest level of education attained or completed. This indicator is usually presented for age groups of at least 25 years and older. The indicator can be calculated for youth (15-24 years) if desired. The indicator measures the percentage of the population who completed at least each ISCED level of education.

Feasibility analysis

This indicators required three variables, years of study, age and entrance age (with years of study and entrance age we generate the variable ISCED (see appendix)

Calculation method

$$EA_{AG_t,b}^t = \frac{EAP_{AG_t,n}^t}{P_{AG_t}^t}$$

Where:

- $EA_{AG_t,b}^t$: Percentage of population in age group i that attained educational level n , in year t .
- $EAP_{AG_t,n}^t$: Population in age group i that attained educational level n , in year t
- $P_{AG_t}^t$: Population in age group i , in year t .

Educational attainment rates by age group and level of education (4.4.3)

Results 15 - 24 years old

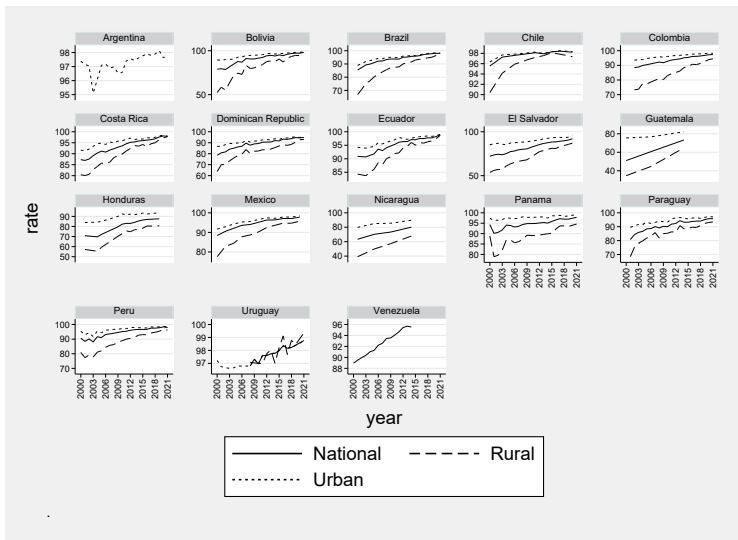


Figure: ISCED1

Educational attainment rates by age group and level of education (4.4.3)

Results 15 - 24 years old

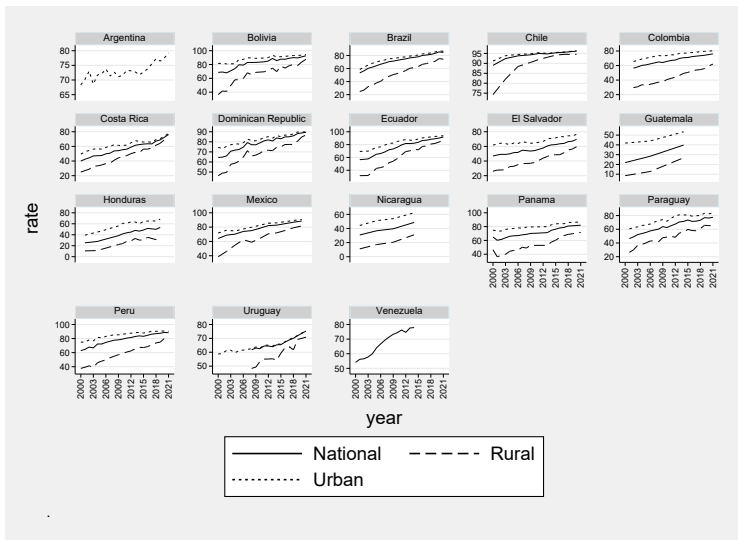


Figure: ISCED2

Definition

UNESCO DEFINITION: The youth literacy rate is defined by the percentage of the population aged 15 to 24 years that can read and write. It is typically measured according to the ability to comprehend a short simple statement on everyday life. Generally, literacy also encompasses numeracy, and measurement may incorporate a simple assessment of arithmetic ability. The literacy rate and number of literates should be distinguished from functional literacy, a more comprehensive measure of literacy assessed on a continuum in which multiple proficiency levels can be determined. The adult literacy rate is defined by the percentage of the population aged 15 years and over that can read and write.

Calculation method

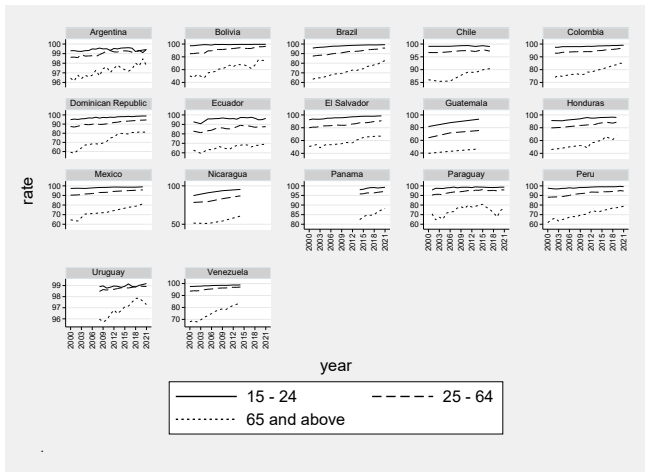
$$LR_{AG_i} = \frac{LP_{AG_t}}{P_{AG_t}}$$

Where:

- LR_{AG_i} : literacy rate of population in age group i .
- LP_{AG_t} : literate population in age group i .
- P_{AG_t} : population in age group i , excluding persons with unknown literacy status.
- i : 15 to 24 years old (youth), 15 years and older (adults).

Youth/adult literacy rate (4.6.2)

Results



Youth/adult literacy rate Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months(4.3.1)

Definition

UNESCO DEFINITION:Percentage of youth and adults in a given age range (e.g. 15-24 years, 25-64 years, etc.) participating in formal or non-formal education or training in a given time period (e.g. last 12 months)

Calculation method

$$PR_{AG_i} = \frac{E_{AG_i}}{P_{AG_i}}$$

Where:

- PR_{AG_i} : participation rate of the population in age group i in formal and non-formal education and training.
- E_{AG_i} : enrolment of the population in age group i in formal and non-formal education and training.
- P_{AG_i} : population in age group i .
- i : 15-24 years, 15 years and above, 25-64 years, etc.

Youth/adult literacy rate Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months(4.3.1)

Feasibility analysis

The harmonized variable on attendance available in BADEHOG refers only to the formal education of children and young people. Therefore, a detailed review of the survey questionnaires was carried out, to identify if additional information on attendance to non-formal education or training was available.

Limitations for the generation of this indicator

- Only attendance at formal education
- Only reported as a reason for absence from work:
 - Mexico 2016: What is the main reason you were absent from work last month? - *Attendance at training courses*
- Information only for employees.
 - Chile 2017: In the last 12 months, have you participated or are you currently participating in any **job training** of at least 8 hours duration?
- Does not include all types of specialization.

Youth/adult literacy rate Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months(4.3.1)

anio	ARG	BOL	BRA	CHL	COL	CRI	DOM	ECU	GTM	HND	MEX	NIC	PAN	PER	PRY	SLV	URY	VEN
2000	X	X	...	~	...	✓	~	...	~	...	X	...	X	✓	...	X	X	~
2001	X	X	X	✓	~	~	...	~	...	X	X	✓	~	X	X	~
2002	X	X	X	...	X	✓	~	X	...	X	✓	✓	X	X	~
2003	X	...	X	~	X	✓	~	~	X	✓	✓	X	X	~
2004	X	X	X	...	X	✓	~	X	X	...	X	✓	✓	X	X	~
2005	X	X	X	...	X	✓	~	~	...	X	...	X	X	✓	✓	X	X	~
2006	X	X	X	~	...	✓	~	~	~	...	X	...	X	✓	~	X	...	~
2007	X	X	X	✓	~	~	X	✓	~	X	X	~
2008	X	X	X	✓	~	~	X	✓	~	...	X	~
2009	X	X	X	~	~	✓	~	~	...	X	...	X	X	✓	~	X	X	~
2010	X	~	~	✓	~	~	...	X	~	✓	~	X	X	~
2011	X	X	X	X	~	✓	~	~	...	X	X	✓	~	...	X	~
2012	X	X	X	...	~	✓	~	~	...	X	~	✓	~	X	X	~
2013	X	X	X	~	~	✓	~	~	...	X	X	✓	~	X	X	~
2014	X	X	X	...	~	✓	~	~	~	X	~	X	X	✓	~	X	X	~
2015	...	X	X	~	~	✓	~	~	...	X	X	✓	~	X	X	...
2016	X	X	~	...	~	✓	~	~	...	X	~	...	X	✓	~	X	X	...
2017	X	X	~	...	~	✓	~	~	X	✓	~	X	X	...
2018	X	X	~	...	~	✓	~	~	...	X	~	...	X	✓	~	X	X	...
2019	X	X	~	...	~	✓	~	~	...	X	X	✓	~	X	X	...
2020	X	X	~	X	~	✓	~	~	X	✓	~	X	X	...
2021	X	X	~	...	~	✓	~	~	X	✓	~	...	X	...

X: not feasible

~ : feasible with limited information

... : no data

- Of the four indicators proposed, it is possible to calculate three of them (4.1.5, 4.4.3, 4.6.2) with their respective breakdowns and precision measures, for a long series.
- The fourth indicator (4.3.1) cannot be calculated, since the vast majority of the surveys do not include the necessary information.
- Regarding the indicators calculated, these are internally consistent and comparable between countries, within the limitations of comparing surveys of great diversity both between countries and in different waves within the same country.
- As a way of evaluating the quality of these indicators, it would be interesting to proceed in the future with an external validation exercise based on registry statistics in those countries where they are available.

Appendix: Information about educational cycles

Country	Official entrance age for primary school (e)	Primary cycle duration (db)	duration of the lower secondary cycle (dbs)	High school cycle duration (das)	Official entry age for lower secondary school (ebs)	Official entrance age to high school (eas)
Argentina	6	6	3	3	12	15
Bolivia	6	6	2	4	12	14
Brazil	6	5	4	3	11	15
Chile	6	6	2	4	12	14
Colombia	6	5	4	2	11	15
Costa Rica	6	6	3	2	12	15
Ecuador	6	6	3	3	12	15
El Salvador	7	6	3	2	13	16
Guatemala	7	6	3	2	13	16
Honduras	6	6	3	2	12	15
Mexico	6	6	3	3	12	15
Nicaragua	6	6	3	2	12	15
Panama	6	6	3	3	12	15
Paraguay	6	6	3	3	12	15
Peru	6	6	3	2	12	15
Dominican Republic	6	6	2	4	12	14
Uruguay	6	6	3	3	12	15
Venezuela	6	6	3	2	12	15

Information about education cycles by country based on ISCED 2011.

Appendix: Information about educational cycles

Education level	Description	Criteria	ISCED1	ISCED2	ISCED3
1	Incomplete primary	$ae < db$			
2	Complete primary	$ae = db$	0		
3	Incomplete low secondary	$ae > db \& ae < (db + dbs)$	0		
4	Incomplete upper secondary	$ae \geq (db + dbs) \& (ae < (db + dbs + das))$	0	0	
5	Complete secondary	$ae == (db + dbs + das)$	0	0	0
6	Incomplete tertiary	$ae > (db + dbs + das) \& ae < (db + dbs + das + 5)$	0	0	0
7	Complete tertiary	$ae \geq (db + dbs + das + 5)$	0	0	0

Information about the construction of educational levels and ISCED criteria

Appendix: Availability of information on sampling design

anio	ARG	BOL	BRA	CHL	COL	CRI	DOM	ECU	GTM	HND	MEX	NIC	PAN	PER	PRY	SLV	URY	VEN
2000	X	X	...	✓	...	X	✓	...	X	...	X	...	✓	✓	...	X	✓	X
2001	X	X	✓	X	✓	X	...	X	...	X	✓	✓	X	X	✓	X
2002	X	✓	X	...	X	X	✓	X	...	✓	✓	X	X	X	X
2003	X	...	X	X	X	X	✓	X	✓	✓	X	✓	X	X
2004	X	X	X	...	X	X	✓	X	X	...	✓	✓	X	✓	X	X
2005	X	✓	X	...	X	X	✓	X	...	X	...	X	✓	✓	X	✓	X	X
2006	X	X	X	X	...	X	✓	X	X	...	X	...	✓	✓	X	✓	...	X
2007	X	✓	X	X	✓	X	✓	✓	X	✓	X	X
2008	X	✓	X	...	✓	X	✓	X	✓	...	✓	✓	X	...	X	X
2009	X	✓	X	✓	X	X	✓	X	...	X	...	✓	X	✓	X	✓	X	X
2010	X	X	X	✓	X	...	X	✓	✓	X	✓	X	X
2011	X	✓	X	✓	X	X	✓	X	...	X	✓	✓	X	...	X	X
2012	X	✓	X	...	X	X	✓	X	...	X	✓	✓	X	X	X	X
2013	X	✓	X	✓	X	X	✓	X	...	X	✓	✓	X	X	X	X
2014	X	✓	✓	...	X	X	✓	X	X	X	✓	X	X	✓	X	X	X	X
2015	...	✓	✓	✓	X	✓	✓	X	...	X	X	✓	X	X	X	...
2016	X	✓	✓	...	X	✓	✓	X	...	X	✓	...	X	✓	X	X	X	...
2017	X	✓	✓	✓	X	✓	✓	X	X	✓	X	X	X	...
2018	X	✓	✓	...	X	✓	✓	✓	...	X	✓	...	X	✓	X	X	✓	...
2019	X	✓	✓	...	X	✓	✓	✓	...	X	X	✓	X	X	✓	...
2020	X	✓	✓	✓	X	✓	✓	✓	✓	✓	X	X	X	...
2021	X	✓	✓	...	X	✓	✓	✓	X	✓	X	...	X	...

This table has the information about types of sampling, where:

✓: upm and strata are available, X: upm and strata are not available, ... : no data