# OUT OF SCHOOL CHILDREN INITIATIVE

## DEVELOPMENT OF ALTERNATIVES FOR ESTIMATING POPULATION DEMAND ON THE EDUCATIONAL SYSTEM BY SIMPLE AGE

UNESCO
Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
INSTITUTO de ESTADÍSTICA de la UNESCO
unicef

# DEVELOPMENT OF ALTERNATIVES FOR ESTIMATING POPULATION DEMAND ON THE EDUCATIONAL SYSTEM BY SIMPLE AGE

## EXECUTIVE SUMMARY

During the last decades of the twentieth century most of the countries in Latin America and the Caribbean (LAC) have approached -through diverse policies- the goal of universal primary education. Based on these achievements, they have progressively emphasized policies towards the extension of compulsory education (Initial and Secondary levels). These goals are framed in international commitments subscribed by countries. The quantitative goals set by the Millennium Development Goals (MDGs) embody these agreements -among them, the universalization of primary education- and enable the recognition of the signs of progress towards the universalization goals.

Despite these efforts, the coverage estimates of the UNESCO Institute for Statistics realize a visible decrease in the levels of school coverage mainly in the group of primary-school age population. According to the statistics released by the organization, the number of out-of-school boys and girls for primary level would have increased from 2.7 to 4.1 million between 2002 and 2013[1].

These trends would represent an important alert in the region: elapsed the time for achieving the MDGs, and despite the efforts made by countries to extend the right to education, a significant drop in coverage in primary education is estimated. LAC countries could be in an unprecedented process of regression in terms of reaching their educational goals.

[1] Estimations from the UIS UNESCO Data Centre (http://www.uis.unesco.org/datacentre/pages/default.aspx). Retrieved by August 20, 2015

Considering the importance of the situation, this document aims to shed light on this phenomenon, identifying the causes of this declination within certain distortions from the method of calculation of coverage that is regularly used by the countries of the region, in particular, the use of childhood population estimates drawn from population projections.

The measurement of coverage is based on the relation between the total population of a five age (potential demand) and the population attending to the educational system (effective demand). The official statistics of coverage use for the estimation of population of the age to attend compulsory education the data from population projections. The hypothesis of this document is that these estimations of childhood population could be overestimating the total of boys and girls, which would impact on the decline of the coverage of primary level.

This hypothesis is based on the analysis of the assumptions of childbirth, used for the calculation of projections. There are evidences which lead to think that in many countries, by the moment of formulating the projections, the estimation of the number of birth that occurred in the decade of 2000 was higher than the ones which really occurred. This has an impact on the projections of childhood population: the higher the estimation of school age children, the higher is the expected potential demand. If these amounts are overestimated, there will be a number of the population out of school.

The document develops the hypothesis and its components, analyzing the regional situation and taking some countries as examples. As well, other data sources for the calculation of coverage will be analyzed, which are normally used in a less often (household surveys and population census). They serve as a contrast to the decreasing coverage trend.

In parallel, a model is introduced for a complimentary approach to the diagnosis of the coverage, based in the statistical information available at the national systems of education data.

In short, some core ideas are developed throughout the document:

- There is no indicator which can individually reflect with precision the social processes linked to schooling. If the effect of the demographic change of the last decades in LAC is considered, it is necessary to strategically link the different information sources available in order to reach a good approach to the measurement of coverage.
- The use of updated and precise information about out-of-school population represents not only a statistic issue but it also has serious political implications for the education systems management, with regards to the right to education. The selection of a method

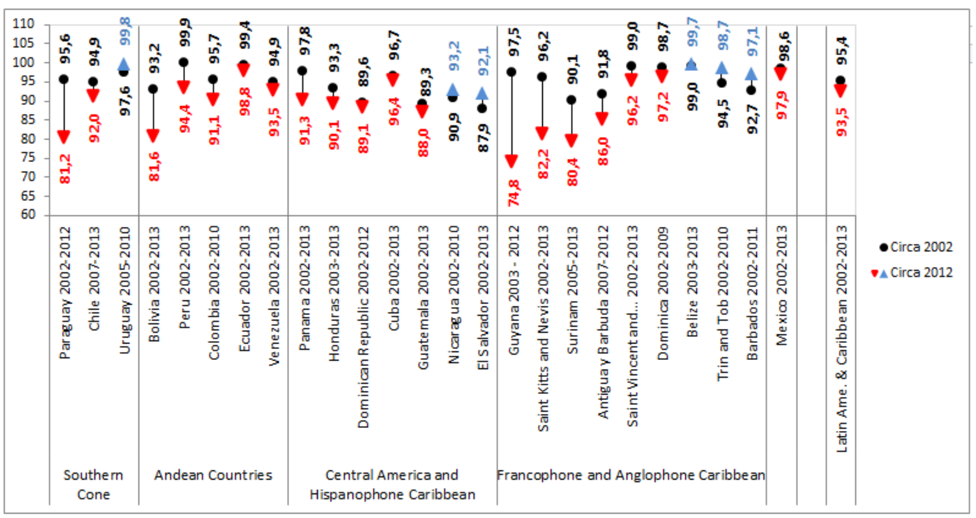
for a diagnosis of coverage must be set in order to identify the political challenges of a country, or of a given context.

- The questions included in this document about the imprecision drawn by the projections for the estimation of childhood population transcend the specific issue of education. It is urgent to build a regional discussion where the specialists, advisors and politicians of the countries involved, as well as those from the cooperation agencies, can analyze, debate and find answers to these questions.

## IS EDUCATIONAL EXCLUSION INCREASING IN LATINAMERICA AND THE CARIBBEAN?

The primary school Adjusted Net Enrolment Ratio (ANER)[2] – the coverage indicator most widely used in the region- outlines for LAC a slight but steady downward trend. Out from the 25 countries with comparative information available from the period 2002-2013, in 19 the ANER shows a declining trend, dropping from 95.4% to 93.5%. This decline in coverage rates represents an increase of the out-of-school population in the period of 1.4 million boys and girls.

### Figure 1.Primary school adjusted net enrolment ratio by country. Selected countries[3] in Latin America and the Caribbean. *Circa* 2002-2013.




*Source: data processing based on UIS Data Centre (http://www.uis.unesco.org/datacentre/pages/default.aspx). Retrieved by August 20, 2015.*

---

[2]Adjusted net enrolment ratio (ANER) is a derivation of the net enrolment ratio (NER). The UNESCO Institute for Statistics defines NER as the "number of students within an age group, theoretically corresponding to a given level of education expressed as a percentage of the total population of that age group."As some of these children could be enrolled at other levels and would be out of the measurement, ANER has been incorporated, which also includes in the numerator the enrolment attending higher levels (UIS, 2009).

[3] By the time of preparing this report, the UIS Data Centre platform failed to return data for the analyzed period from the following countries: Argentina, Brazil, Costa Rica, Granada, Haiti and St. Lucia. In some cases, the lack of available data is due to the disagreement of the country concerning the trend represented by the coverage indicators calculated by the organization.

This drop of the coverage statistics brings the following question:  Why have the coverage rates of primary education decreased over the last decade? Why in recent years the total estimated excluded population of those ages would be increasing? Are these statistical trends a reflection of social processes of educational exclusion that actually occurred in the last decade, as some documents suggest (UIS, 2015)? What evidence is available to support each of these approaches?

**There are clear signs that enable to identify the cause of the problem within distortions associated to the calculation, and not to a decline in the access to education.** Therefore, it is required to explore the method of the calculation of coverage and the sources involved.

The measurement of coverage is based in the relation between the total population and the population that attends to the education system. This indicator identifies the extent of the progress towards the goal of universality, and the proportion of the school-age population that is still out of school. The calculation of the coverage mostly used in the region takes as the numerator (population in school), the information of enrolled students taken from national school records, while the denominator (total population) considers the demographic estimates derived from population projections.

The starting point of the hypothesis developed by this document lays in the data sources which feed the coverage diagnosis: the drop in the primary level coverage in LAC would have as one of the main causes **the distortions introduced by the use of estimates of the population of school age** based on these projections, which would tend to **overestimate the total of boys and girls** who should potentially access compulsory education. All of this linked to a context of demographic greater decline in the level of fertility than it was originally expected for the region.

## THE DENOMINATOR OF THE COVERAGE: POPULATION ESTIMATES

The most common coverage measurements consider as rate denominator the demographic estimates which are obtained from population projections. These projections are prepared by the countries of the region -with technical support from CELADE or the UNESCO Institute for Statistics- And work as an analytical instrument and statistical base, useful inputs for government planning

The model most widely used to develop population projections by sex and age group, is the "**Model of components**". One of its features is that it  allows the complete and systematic incorporation of the schemes about the evolution of determinant variables of the population dynamics (fertility, mortality and migration) based on the knowledge of the historical evolution in recent times (Mesa Rincón Fajardo and Valenzuela, 2007:29). Now, these hypotheses are based in the results of the last census, which is conducted every 10 years. This period becomes excessively long for the present context, since the population dynamics in LAC usually have short-term changes that are not covered by the projection assumptions (Bay, 2012). In other words, the population projections could show maladjusted estimations whenever there is a distance of several years between the date of the estimation and the elaboration date. The hypotheses about fertility are used to estimate the number of births during a given calendar year. These hypothesis start to build the potential population who will demand educational services through time, particularly at preschool and primary level, in the short time of five years. Therefore, **the quality and accuracy of the hypotheses of fertility are the key assumption for the future educational demand estimates**.

## THE NUMERATOR OF THE COVERAGE INDICATORS: THE ENROLMENT OF THE EDUCATIONAL SYSTEM

The information used for the numerator of the coverage rate comes from the school records (enrolment), that is, the number of students enrolled in the official records. These systems are usually named EMIS, (acronym of Education Management Information Systems). They have started to be developed in the region in the decade of 1980, and they have increasingly improved in the last years reaching higher precision in the collected data, developing in some cases nominal registry systems.

One of the main features of the EMIS is the fact that the data collection is done in an annual base and including all schools. The countries of the region have been regularly supporting these applications for consecutive years, which enables obtaining up-to-date information, comparable throughout time, and with broad levels of disaggregation.

## WHY IS THE COVERAGE IN LATIN AMERICA AND THE CARIBBEAN DECLINING?

It is propose that the way in which coverage is measured is the main cause of the fall of the indicator and the increase of the estimated total out-of-school population. Its cause is identified in the use of childhood population estimates based on population projections which would be tending to overestimate the total of boys and girls who potentially should access compulsory education. In order to advance in this line, a group of core ideas are presented:

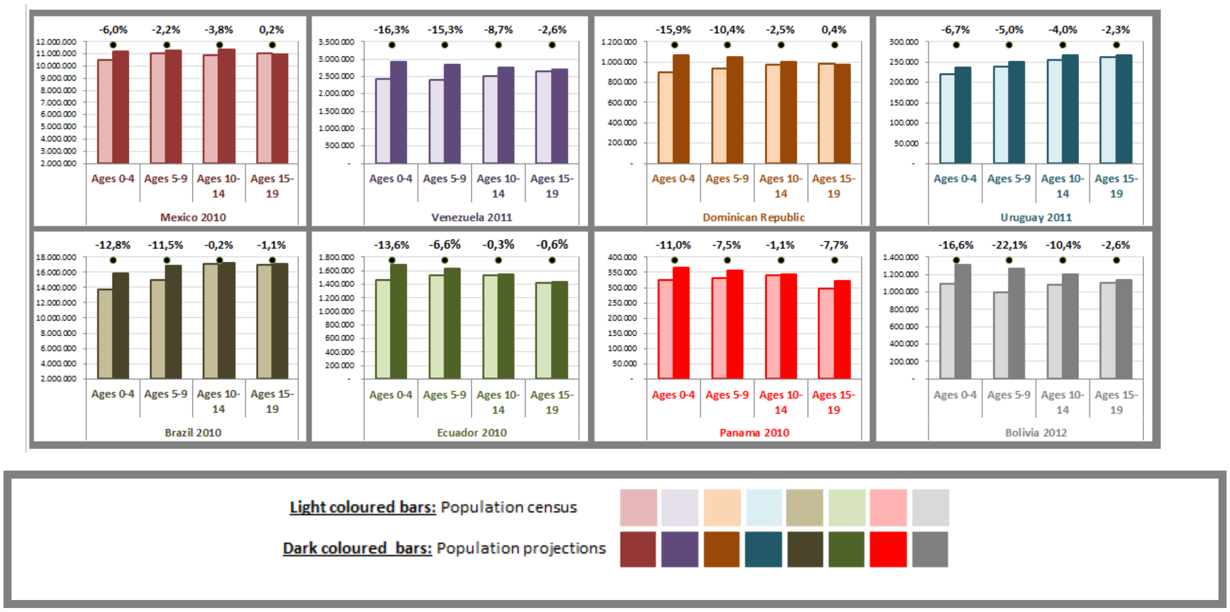
### A systematic difference is observed between the total of childhood population by the census and by the estimates.

A key element is the difference observed between population projections and censuses *circa 2010*. The study presented below was performed on eight countries in the region[4].Seven of the countries under analysis -with the exception of Bolivia- have updated their projections after the population census[5].

[4] Bolivia, Brazil, Ecuador, Mexico, Panama, Dominican Republic, Uruguay and Venezuela. These countries have implemented a census of population between 2010 and 2012. The rest of the countries in the region were excluded for two reasons: (i) the absence of recent census data, or (ii) difficulty in accessing official data disaggregated by simple age.

[5]Other countries in the region still use population projections over a decade old, such as Paraguay and Guatemala. However, they are not included in the sample because of lacking available data of recent census.

**Figure 2.Total Population by five-year age groups by country.0-19 year old. Census Circa 2010 and population projections for the same year. Selected countries.**




*Sources: Data processings based on compiled data from the following institutes: Consejo Nacional de Población (CONAPO), México (http://www.conapo.gob.mx), Instituto Nacional de Estadística (INE), Venezuela (http://www.ine.gov.ve), Oficina Nacional de Estadística (ONE), República Dominicana (http://www.one.gov.do/), Instituto Nacional de Estadística (INE), Uruguay (http://www.ine.gub.uy/), Instituto Nacional de Estadística y Censo (INEC), Ecuador (http://www.ecuadorencifras.gob.ec/), Instituto Brasileiro de Geografía e Estadística (IBGE), Brasil (http://www.ibge.gov.br/home/), Instituto Nacional de Estadística (INE), Bolivia (http://www.ine.gob.bo/), Contraloría General de la República de Panamá, Panamá (https://www.contraloria.gob.pa).*

From the analysis of the graph on figure 2, emerge a number of questions regarding the case under study...Why the population estimates between these two sources are so different only within the population under 10 years? Which of the two is more accurate? Can these differences be related to a drop in coverage?

Two arguments allow finding answers to these questions:

### The projections may overestimate the implicit total births

The demographic components have experienced modifications in the last decades. Specially, Latin American population's fertility has experienced an accelerated decline. It is presented as **a feature an evident overestimation of implicit births, according to fertility assumptions included in the**

**projections of population** for all the considered five-year periods without exception (Bay, 2012: 59. Table 3). It is observed that in the years 2000-2005, 15.6% of implicit births in the population projections would correspond to overestimated births.

**If childhood population is calculated based on fertility rates, the total of school age children would be overestimated, and it would show an "inflated" estimate of the total population who is out of the educational system.**

### The census conciliation and the assumption of sub-reporting of childhood population

For the assessment of the coverage errors –issues of the census omissions- it can be used a methodology named "demographic conciliation"[6], a procedure based on a type of analysis that includes the census information, the other demographic data available (surveys, vital statistics, previous census) and the accumulated experience. That is where the population base for the formulation of projection comes from.

The specialized literature often says that in Latin America there is a proven trend to leave out the children as members of the households (ECLAC and UNFPA, 2014 trend: 35-36 and Chackiel 2009: 19). Therefore, the census results are not often used as a reference base in the estimation of the total of children. The CELADE-Division of Population of CEPAL uses directly the information of implicit birth of the population estimates and projections to estimate the under-registration of births (CEPAL and UNFPA, 2014).

That is, facing the results of each new census, performing the demographic conciliation is based on the idea that it is probable that the total childhood population is greater than the one collected. In these cases, the countries try to build a probable total of the childhood population which allows to estimate this error, and for this the same hypothesis used to project the population are often applied -especially in contexts where vital statistics have proven coverage problems-.

Another of the postulated hypothesis **is that the population projections, even those made based on the demographic conciliation of the Circa 2010 census, may be inaccurate in the adjustment of the population under 10 years of age, because they are based on *a priori* assumptions of a childhood population omission**, which did not allow that the census adjusted fertility assumptions that support the projections.

---

[6] The demographic conciliation is defined as the process where the adjustments to the level and structure of mortality, fertility and migration are made, aiming to achieve coherence with the census figures of population. (Bay, no date.)

## AN ALTERNATIVE TO THE MEASUREMENT OF COVERAGE: ESTIMATES BASED ON SURVEYS

Facing these obvious difficulties in measuring the coverage from school records and population projections, one of the alternatives to improve the accuracy is to use information based on **surveys applied within households**[7], in the form of censuses and household surveys.

Population censuses have universal coverage and perform a thorough accounting of the number and profile of all the population, so its data allows broad levels of geographical disaggregation. They are generally conducted every ten years. While household surveys are a representative sample of households and population, they take place regularly, but the data obtained are representative only for the most aggregated units.

**Even if these instruments have some limitations, they present an advantage regarding its elaboration from the combination of sources of registry and population estimates.** An evidence for this is that estimates based on household surveys do not outline a drop in coverage at the primary level[8]. However, these information resources are not regularly used to calculate the official indicators of coverage of the educational system.

## THE MODEL FOR ESTIMATING THE POPULATION DEMAND FROM THE FLOW OF ENROLMENT

This document also advances in the presentation of a proposal for approaching the problem of measuring the coverage from an unconventional use of the statistical information available in national educational information systems. A model for **estimating the excluded population from the different stages of the educational system** is proposed.

The model is structured around the concept of **inter-annual follow-up of a cohort of students by simple age** and allows determining the potential uncovered demand at certain ages, as a direct result of late entry and drop-out.

[7]In this case, the term refers to the particular homes, for example the person or group of people living under the same roof and sharing the costs of food and/or services, rent. Thus, collective households are excluded, such as those in which people live under anon-family system, such as hospitals, prisons, nursing homes, among others.

[8] On the website of SITEAL - System of Information of Educational Trends in Latin America-it is possible to consult the evolution of the primary school net enrolment ratio for most Latin American countries. This ratio is calculated from processing of national household surveys (http://siteal.iipe-oei.org/). A similar resource is available on the SEDLAC initiative -Socio-Economic Database for Latin America and the Caribbean-from the Center for Distributional, Labor and Social Studies (CEDLAS) and The World Bank (http://sedlac.econo.unlp.edu.ar/eng/).

This model is a complementary approach for the analysis of the coverage. Even though it does not allow dimensioning the total out-of-school population, it is robust to estimate the total of boys and girls who start school but do not have access to certain educational levels.

The use of a model of inter-annual follow-up of an age cohort outlines the school entry and drop-out by age with greater accuracy and potential disaggregation, both closely linked to the coverage measurement. It also allows detecting other phenomena such as estimates of drop-out by the ages to attend secondary, and inequalities that occur in the access to pre-school and primary school.

## CONCLUSION

The analysis of the coverage needs accurate indicators with at least two fundamental characteristics: accuracy in the measurement of coverage according to school age simple ages, in order to detect situations of exclusion associated with educational background of the population, and secondly, precision in the territorial disaggregation in order to recognize those populations where exclusion is generated and the areas where access to school is restricted.

The drop in coverage rates over the last decade for the entire LAC region calls to reflect about the technical aspects of the processes involved in the processes used for its measurement.

That is why this document performs a double interpellation. On the one hand, it aims to emphasize the importance of understanding the scope and limitations of the various data sources, and to recognizing the distorting effects of methodological decisions. On the other hand, it aims to considering in the analysis of indicators the articulation of technical and political criteria involved in the diagnosis of coverage, in order to understand the information needs demanded by educational policy actions and to warn about the consequences of the use of not so accurate sources, with regards to their key role in the decision-making in educational matters.

# DEVELOPMENT OF ALTERNATIVES FOR ESTIMATING POPULATION DEMAND ON THE EDUCATIONAL SYSTEM BY SIMPLE AGE

## METHODOLOGICAL ANNEX

### INTRODUCTION

The measurement of the educational systems' coverage is a complex task, which involves a series of definitions and assumptions regarding childhood population, school enrolment and the educational system. The content proposed in the document "Development of alternatives for estimating population demand on the educational system by simple age" has been organized facing some of these key aspects related to the measurement of school coverage.
Chapter 1 has presented some worrying trends which lead to questioning current data. Chapters 2 and 3 have proposed a critical revision of the indicators and data sources used for the measurement of coverage, making progress towards outlining some assumptions that would contribute to explain the technical-methodological origin of the behavior of the primary level coverage rate, which are currently drawing a de-schooling scenario in LAC.

It is necessary to problematize the data processing that, with different levels of analysis, is done to produce the coverage statistics used in the countries of the region. The development proposed in the document has embodied a series of dilemmas: the information available needs to be revised with a critical eye in order to support the enhancement of the educational indicators needed for planning purposes.

Now, the problem of the measurement of coverage is not fully covered with the discussion about the selection of the most accurate indicator or the revision of the assumptions underlying the construction of every indicator. It is part of a more complex and broader challenge referred to the generation of useful information for decision making in policy. The educational actions for inclusion need information tools that can move beyond the generation of a national indicator for childhood population grouped in age brackets.

For this reason, this section presents a development of an alternative measurement, sustained in the use of the information collected by the school records. The proposal is a model of an **estimate of the population excluded from the different stages of the educational system.**

This model is a complementary approximation to the coverage analysis. Even if it does not enable dimensioning the total out-of-school population, it is robust enough to estimate the total of boys and girls who enter school but do not have access to certain educational offerings. This methodology presents some relevant characteristics for its use:

1. <u>It is precise for conducting disaggregated analysis:</u> it is good enough for estimating the specific potential demand of given ages. With the provision of the possible impact driven from migrations or students' relocation, it can be very useful for the identification of regions with remarkable gaps of pre-school coverage, or for detecting early drop-out.

2. <u>It provides the opportunity of having up-to-date information for annual monitoring:</u> given the fact that the educational information systems of the countries produce information with an annual frequency, it enables a permanently updated monitoring system.

3. <u>It enables the disaggregation of the non-covered demand based on educational variables:</u> the information produced can be disaggregated based on the characteristics of the students, of the groups and of the schools, contributing with key information for educational policy monitoring.

4. <u>It is consistent with regulations of the educational system:</u> by building the measurement of the age based on the educational system's criteria, it is more accurate than other sources (such as census or household surveys) for the diagnosis about pre-school and primary school entry. It enables to account for the situation of the population at the right age for a specific stage (the last year of pre-school, the penultimate year, the first year of primary…).

To sum up, the use of the model of inter-annual follow up of a students' cohort by age intends to complement the analysis of the coverage measurement, describing with precision and potential disaggregation the school entry and drop-out by age, both closely linked to the coverage measurement. Besides, it enables to detect other phenomena, such as estimations of the drop-out levels at the age to attend secondary and the inequalities produced in the pre-school and primary entry.

## A MODEL FOR ESTIMATING POPULATION DEMAND BASED ON THE ENROLMENT FLOW

### Key concepts of the model

The following pages present a model of estimation of the excluded population in specific stages of the educational system, from a non-conventional use of the statistics from the national systems of educational information.

> *Counting with information about* ***students by age,*** *collected in a* ***consistent and comparable*** *manner across different years, through a device that reaches* ***full coverage*** *of its units is a basic requirement of this model.*

The model of inter-annual follow-up of a students' age cohort is based on two core ideas:

IDEA 1: relating time and age in order to reconstruct a cohort

For a group of students enrolled at a certain age in a certain year, in the *subsequent* years only two eventualities can happen:

a) Either they are enrolled in some educational offering with a proportionally higher age, or

b) They have dropped-out of the system.

Complementarily to the previous premise, when seeking this group of students in a *previous* year, it can only happen that:

a) Either they are enrolled with a proportionally lower age, or

b) They are not attending the educational system, taking into account that in this second case they would enroll (or re-enroll) in subsequent years.

On the basis of these ideas, an operational definition of the cohort is formulated, based on the information of enrolment by age:

*The cohort is defined as the total* **T** *of students that attending school at the age* **A** *in a year* **Y,** *regardless of the level they in which they are enrolled. A year that works as a point of comparison in time is selected, defined as* **Y + n**. *The cohort is identified from the total of students at an age that is proportionally similar to the time interval; that is, enrolled students of* **A + n** *years of age. In the comparison between both totals it is obtained, for the year* **Y + n,** a difference $\Delta$ *(delta).*

The formula would be expressed as follows:

$$\Delta_A^{Y+n} = T_A^Y - T_{A+n}^{Y+n}$$

As it is observed, in the inter-annual follow-up of the age cohorts it is necessary that, between one point and another, the time interval and the difference of ages are the same (n value in the equation).

When performing the difference between totals, the result can be greater, equal or lower than zero. Each one of these three scenarios represents a different situation:

$\Delta$**> 0 (positive difference or greater than zero)**: the resulting difference points out an increase of the cohort's size. It indicates an incorporation of new students that were not registered in the beginning of the cohort.

$\Delta$**~ 0 (difference that is equal or similar to zero)**: the resulting difference points out stability of the cohort size. It indicates a situation of permanence within a cohort of students.

$\Delta$**< 0 (negative difference or lower than zero)**: the resulting difference points out a decrease of the cohort size. It indicates a loss of students of the cohort.

It is important to make two observations, which will be deepened in the development of the models' assumptions. In the first place, it is about characterizing the resulting movement of the total of students. As every other model of cohorts, built on totalized information, it is not possible to give an account of the coexistence of school entry and drop-out trends during the period of time between the two measurements. On the other hand, even if the greatest part of those movements is related to the phenomena of school entry and drop-out, there are other reasons that can explain some of the variations of the cohorts, linked to demographic events such as childhood/adolescent mortality or migrations.

IDEA 2: it is possible to identify the age in which the population will reach its maximum level of enrolment

When school attendance by simple age is analyzed through any source based on data collection (censuses or surveys), it is possible to identify that there is at least one age in which the maximum level of school coverage of population is attained. This maximum level does not necessarily mean the inclusion of the entire population of that age, but it is the highest level of coverage attained: in the higher and lower ages, school attendance tends to diminish.

The model hereby proposed takes that observation as a starting point, assuming that is possible to define a specific age in which a country attains its maximum levels of coverage at a specific moment of time. Given that this coverage never effectively attains 100% of the population (even in countries with the highest school system's entry rates, there is a group of children that in all ages, at times statistically invisible, who do not attend school), it is assumed that there is a group of children –of unknown size- who do not attend school at the age of maximum coverage.

The total of students that attend at the maximum enrolment age is named "*total of population that will enter the educational system".* It is assumed that the population that does not attend at the age of maximum enrolment will never enter the educational system. It is important to clarify that this definition does not involve ignoring or denying the possibility that this excluded population will ever enter the educational system at some point, even during the adolescence ages. On the contrary, it is about performing an operational delimitation that will enable the identification of a cohort of students by age[9].

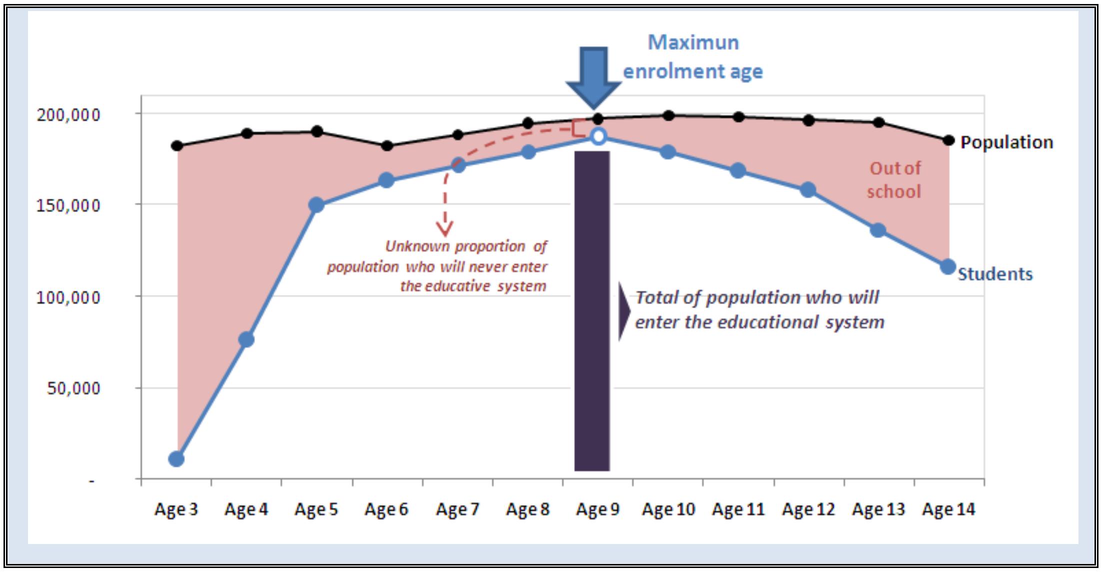
This maximum enrolment age is taken as a base in the model for developing an inter-annual follow-up of cohorts by simple age. This way, it is possible to estimate the total of population that has remained excluded from certain stages of the educational system, but having enrolled at some point. However, given that the entire population is not taken as a base, it is not possible to analyze the coverage since the number of children that do not attend at the age of maximum enrolment is unknown.

Thereon, it is important to point out that the progress made in most of the LAC countries in terms of school coverage acknowledges broad levels of access, close to the total of population. In these cases, the total of students enrolled at this age can reach values that are *very similar* to the total of population. Therefore, in these contexts the implementation of this methodology is a very good approximation to the estimation of total coverage.

*The base age for the analysis of cohorts is defined as the one where the maximum level of enrolment is attained in a specific moment of time. The total of students at this age is named population that will enter the educational system. The population that does not attend at this age is assumed as unknown, and it is named population that never enters the educational system. The following scheme is a synthesis of these operational definitions, from the figure of total of population and students by simple age in a given year:*

**Scheme1. Relation between the total of population and the total of students by age; and operational definitions of the analysis methodology of cohorts by simple age**

[9] On the other hand, the empirical observation of the indicators of the countries from the region shows that those who do not manage to attend school at the ages of the country's maximum level of coverage, will unlikely manage to enter the system during their childhood or adolescence.



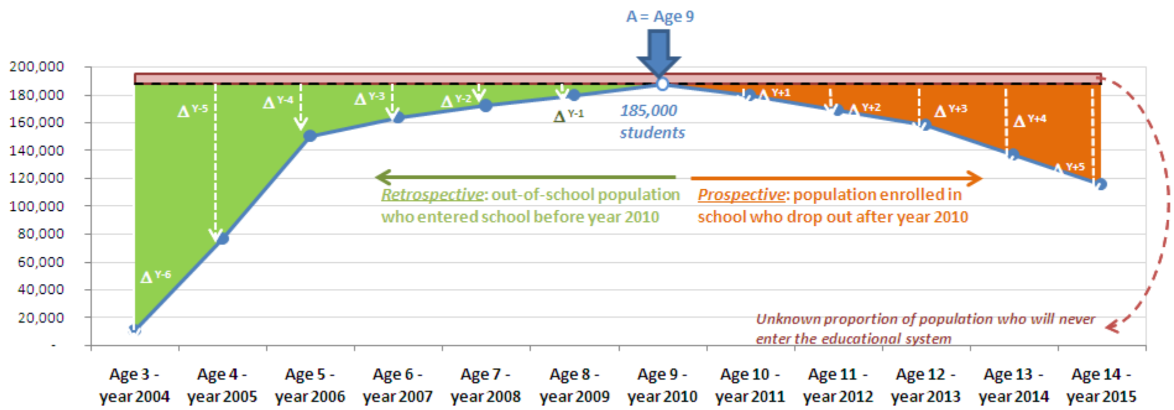

## Presentation of the model

The model for estimating population demand herby developed is based on the application of the methodology of inter-annual follow-up of cohort by simple age (idea 1) and the group of students that attend at the maximum age of enrolment to the educational system (idea 2).

Taking as a base the enrolment that attends at the maximum enrolment age (population that will enter the educational system) the analysis of prospective and retrospective cohort can be built. If the age was correctly identified, in both cases the cohort follow-up will reveal negative results: the cohort cannot grow more because its baseline is equal to the maximum coverage.

The *prospective* or forward perspective allows acknowledging in the resulting figure $\Delta$the total of out-of-school population of that cohort due to drop-out. Instead, the *retrospective* or backwards perspective is an estimation that represents in the resulting figure $\Delta$the out-of-school population of that cohort due that did was not enrolled in the educational system but who eventually entered during the course of the period. The following scheme presents an example, based on an assumption (arbitrarily selected for this example) that the maximum coverage is achieved at age 9:

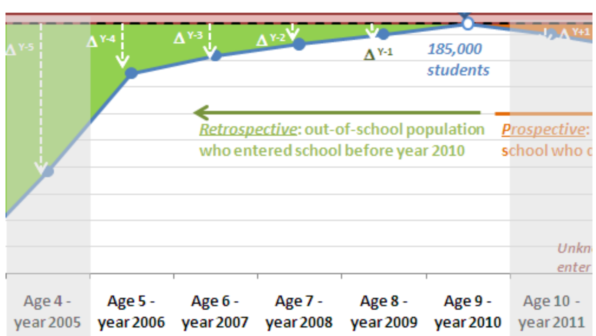
**Scheme 2. Example of the reconstruction of school entry and drop-out based on the use of prospective and retrospective cohorts.**



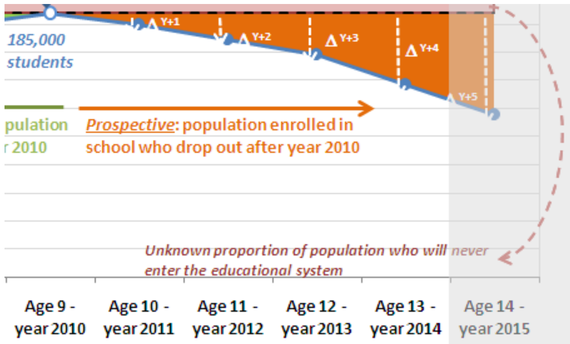

Source: data processing based on information of one country of the region.

In the example within the graph, age 9 is the stage in which it is assumed that coverage reaches the maximum levels in year 2010. The red top area represents the unknown total of population of that age that does not attend. From that starting point, the situation of school entry and drop-out in the previous and subsequent years is reconstructed based on these ages. Two examples will be analyzed.

<u>Example 1: Approach to the out-of-school population of 5 years of age:</u> This approach can be done for the year 2006, based on the calculation of the figure $\Delta$ in a retrospective cohort, that is, obtaining the difference between the enrolment of age 9 in 2010 and the corresponding one of age 5 in 2006. If at age 5 there were 150,000 students attending school, it is possible to estimate that the population of the cohort that was not attending school at age 5 is 35,000 (185,000 enrolled students of age 9 in 2010– 150,000 enrolled students of age 5 in 2006). In other words, by the year 2006 there were 35,000 boys and girls who were out-of-school and who subsequently entered the educational system. Some of them entered at age 6 and others at age 7, 8 or even 9. This difference of simple ages reveals the size of the challenge of inclusion for each specific age. To estimate the total pending coverage, it is necessary to add to this calculation the total of population that never enters the educational system.



Example 2: Approach to the out-of-school population of age 13. Similar to the former example, this approach can be made for the year 2014, based on the calculation of the figure $\Delta$in a prospective cohort, that is, obtaining the difference between the enrolment of age 9 in 2010 and of age 13 in 2014. If at age 13 there were 135,000 students attending school, it is possible to estimate that the population of age 13 who were out-of-school in 2013 is 50,000 boys and girls in the country of the example (185,000 enrolled students of age 9 in 2010 – 135,000 enrolled students of age 13 in 2014). This means that by the year 2014 there were50,000 adolescents who do not attend the educational system, but who used to formerly attend.




This out-of-school population in the year 2014 had dropped out within 2010 and 2014, some at age 10, others at age 11, 12 or 13. In this difference, there is the size of the challenge of reducing drop-out, specific for the population within these ages.

It can be seen from the examples that the calculation is relatively easy to apply, and that its results are intuitive regarding their interpretation. And, as it can be deducted, its implementation is easier when estimating the pending coverage at the ages that are close to the ones that reach the maximum figure; but when the distance between ages is greater (*n* value in the formula) it may lead to some complications.

For example, if there is the intention to estimate the out-of-school population at age 17 in year 2013 in the country of the example, it is necessary to refer to the number of enrolled students of that age in the year 2005 (*n*=8). Is it possible to keep the same age as the maximum coverage assumption for that year? For some countries this is possible, but it may become an aspect to rework for others.

A bigger problem appears in the attempt of estimating the pending coverage of lower ages: if the intention is to estimate how many boys and girls of age 4 do not attend school in 2013 in this country… the information should be awaited until the year 2018!

That is, given the broadness of ages that are crucial to analyze in the inter-annual cohort follow-up, some of the analysis with this methodology may be weak, or delayed for various years. This restriction seems to shatter the model's advantages. It is not clear the advantage of the application of a calculation which not only does not represents the total coverage of the population, but just of the population that is assumed that will enter school at some point, and that additionally builds outdated diagnosis.

However, it is important to clarify that this model of inter-annual follow-up of students' by age is the basic methodological scheme for the development of three alternative proposals for the estimation of the demand, which are built as derivations. These proposals take the shape of specific indicators that enable making some interesting diagnosis for the approach to the specific pending coverage for a population cohort that is expected to enter the educational system.

### Assumptions of the model for estimating population demand

The model for estimating population demand from the inter-annual follow-up of a cohort by simple age – and all its variations presented in this document – is based on a series of basic assumptions and simplifications that enable its application and usefulness. In general terms, these simplifications bring over slight distortions or bias between the indicator's result and the phenomenon that is intended to capture, which do not have an influence in the global results neither in the interpretations. However, in some particular contexts it is convenient to keep these elements in mind, in order to detect a possible imprecision in the measurement.

It is important to highlight that these assumptions are similar to the ones that support the model of reconstructed theoretical cohorts developed by UNESCO. Next, the main assumptions and their possible incidence in the result are listed:

a. The assumption of a closed system. The estimation of school entry and drop-out by age is based on the assumption that the members of the cohort can only adopt two positions: they are counted when they are enrolled in the educational system, and they are not counted when they are not. A third option is not considered: being enrolled in an offering where they are not counted. If this is the case, what is represented as drop-out may include the passage of students from one offering to another which means an overestimation of the specific coverage gaps. Therefore, it is recommended that, under certain conditions, the analysis of the proposed indicators considers estimations or approximations about the magnitude in these passages. This assumption requires taking precautions about two issues:

(i) When calculating the total of enrolled students by age, it must be considered all the educational offerings, or at least those that involve a certification that is equivalent to the one offered at the regular education. If only regular education is considered, it must be understood that a number of the students that drop-out can re-enroll in a different modality, such as special or adults education.

(ii) On the other hand, this assumption establishes some limitations for the disaggregated analysis. When it refers to territorial disaggregation, it must be considered that migrations into or out of the geographical area under analysis will bring some distortions.

They can be slight and not affect the trends if it involves regions of a great size, or areas with scarce influx of population. Instead, in small regions with a frequent daily movement of population across frontiers, the distortions can become bigger. In a similar way, it happens with disaggregations related to the characteristics of the supply: if an inter-annual follow-up of cohorts by simple age is made within institutions of the public sector, it is not possible to distinguish drop-out from the passage to the private sector.

b. <u>Measurement of average trends regarding school entry and drop-out:</u> the measurement of the potential demand based on drop-out is built as a resulting trend between the movements of entry and drop-out from the educational system developing in an intermediate period between measurements. That is, if out a cohort of 1,000 students 200 students drop out from school between one year and the other, but 100 other students enter the system, the drop-out measurement will assume a loss of 10%. This assumption is not exclusive of the model; *all* measurements of school drop-out commonly performed with the school records information consider the resulting trend, with the exception of those that are built through nominal records. Regarding this point, even if many countries have made some progress in the application of student's nominal records, it is not always possible to reconstruct the real cohorts in an inter-annual manner.

### Applications of the population demand estimation

Next, three proposals for the application of the population demand estimation will be introduced; they are useful for an approximation diagnosis of the specific potential demand at specific ages. For each of them, the indicators and derived calculations will be presented:

#### 1) The estimation of *specific* uncovered demand for the ages related to pre-school and primary entry.

This first derivation has a specific context of application: it is useful to estimate the uncovered demand of certain educative services related to the access to pre-school and primary education. It is based upon the following reasoning: in the coverage diagnosis for the population of the age to attend pre-school or to enter primary, it is key to distinguish the access situation of each particular age in order to design sound policies.

For example, in a country where entering primary happens at age 6, it is critical to distinguish:

A) The population that does not attend at age 6 but who enters at age 7, since they represent the issue of late entry to primary. This issue is mostly addressed encouraging families to send their children to school at lower ages.

B) The population that does not attend at age 5 but enters at age 6. They represent, broadly speaking, the dilemma of those who enter in a timely manner to primary school but remaining excluded from pre-school or pre-primary education. Here, the factor under analysis is the group of associated causes to non-access to pre-school, which can be very different from those that produce late entry. An example can be a shortage of supply of the level.

C) The population that does not attend at age 4 but who enters at age 5. They represent, broadly speaking, those who will enter primary with the experience of only one year of pre-school. Dimensioning this pending coverage will acquires greater relevance whenever the country includes broadening the years of school experience in the level within its policy goals.

The analysis can continue in lower ages, keeping the same logic. As it is observed, this methodology is based on dimensioning the uncovered demand that is specific on certain ages, given that they have a direct link to different stages of schooling.

With regards to the general model, the formula is similar to the one formerly presented but it does not use the assumption of the maximum coverage age as a base of the cohort. A retrospective cohort is used, between simple ages in two consecutive years, taking the higher age as a base of the cohort. This way, it is possible to display the general formulation through a group of complementary indicators:

- Specific coverage gap by the age of entering primary education ($\Delta_{PR}$):

$$\Delta_{PR}^{Y-1} = T_{Apr+1}^{Y} - T_{Apr}^{Y-1}$$

Where $\mathbf{T^{Y}_{Apr+1}}$ is the total of students in the year Y that attend with an age one year older than the official primary entry age, and $\mathbf{T^{Y-1}_{Apr}}$ is the total of students who in the year Y-1 with the expected age to attend primary. That is to say, in a country where entry age is set at age 6, it represents to those who did not attend at age 6 but who did by the following year at age 7. This way, the estimation of late entry to primary is built.

In a complementary way, it is possible to calculate the *Rate of specific coverage at the age to entering primary* ($RSC_{PR}$):

$$RSC_{PR}^{Y-1} = \frac{T_{A_{pr}}^{Y-1}}{T_{A_{pr}+1}^{Y}} x\ 100$$

It expresses the percentage of students of 7 years of age (or its equivalent of lagging behind 1 year regarding the official primary entry age) who attended the previous year at age 6.

The indicator that complements the specific coverage rate in order to reach 100% represents the percentage of students of age 6 –or the equivalent one- that did not attend school the previous year, and could be named *Rate of 1 year late entry* ($RLE_{PR+1}$) regarding the age stipulated for starting primary:

$$RLE_{PR+1}^{Y-1} = 1 - SCR_{PR}^{Y-1}$$

This indicator can become a very good approximation regarding late entry to primary school from those students who did not attend the educational system the previous year (pure late entry)[10],even though it must be considered that this indicator includes those who enter pre-school with two years of school lag.

- Specific coverage gap by the expected age for attending the last year of pre-school or pre-primary ($\Delta_{PP}$):

$$\Delta_{PP}^{Y-1} = T_{Apr}^{Y} - T_{App}^{Y-1}$$

Where $\mathbf{T^{Y}_{Apr}}$ is the total of students in the year Y who attend by the age corresponding to primary entryand$\mathbf{T^{Y-1}_{App}}$ is the total students who by the year Y-1 have attained the age to attend the last year of pre-school or pre-primary. That is, in a country with primary entry stipulated by age 6, it represents those who did not attend at age 5, but who did attend the following year at age 6. This way, it is obtained an estimation of the population with timely entry to primary but no pre-school experience.

In a complementary manner, the *Rate of specific coverage at the age of attending the last year of pre-school*($RSC_{PP}$) can be calculated:

$$RSC_{PP}^{Y-1} = \frac{T_{App}^{Y-1}}{T_{Apr}^{Y}} x\ 100$$

It expresses the percentage of students of age 6 (or its equivalent for the official age for primary entry) who attended the previous year with age 5.

The indicator that complements the specific coverage rate in order to reach 100% is represented by the percentage of students of age 6 -in line with the previous example- who did not attend school the previous year, which can be called *Rate of school entry at the age to start primary* ($REA_{PR}$):

$$REA_{PR}^{Y-1} = 1 - SCR_{PP}^{Y-1}$$

This indicator can become a good approximation to the phenomenon of primary entry with no pre-school experience, even though it must be considered that this indicator includes those who attend pre-school with age-grade gap.

[10]With regards to this concept, Chapter IV, section D, of the document *Developing alternatives for estimating population demand on the educational system by simple age* explains the proxy indicators to calculate the paths of pre-school and primary entry, including the definitions of the late entry indicators.

### 2) The estimation of drop-out by age

This second derivation is about the implementation of the model to the calculation of the school drop-out between two consecutive years, therefore it is usually applied to higher ages regarding the maximum coverage ages.

In this case it is not intended to build an approximation to total coverage, but to estimate school drop-out at specific ages. This condition has some implications: it does not use data from the educational system (such as level or grade) to identify the population of the cohort to be analyzed. Therefore, it does not allow relating drop-out with the schooling pathways or with the grade structure. Instead, it uses a demographic characteristic of the population: age. For this reason, it is easier to link the dynamic of drop-out with the coverage profile: it is expected that in all ages where drop-out is higher, the educational system's coverage drops in similar proportions.

In this case, the assumption of maximum coverage is also not used to select the age of start of the cohort. Through a prospective cohort, it is intended to estimate how many students of a specific age A have dropped-out. The use of simple ages is not a necessary condition; the model can be applied to an age group. At the same time, the estimation of drop-out by age can be realized with periods larger than a calendar year, as long as there are a proportional time gap and a proportional difference between ages.

It is important to highlight that the main restriction to this model lays in the fact that there cannot be an extension of the analysis beyond the official age for the last grade of secondary (or equivalent to ISCED 3). On the contrary, it is possible that it includes students who graduated from secondary education as "drop-outs".

The formulation to calculate drop-out is the same that is used in the general model:

$$\Delta_{A\rightarrow A+n}^{Y\rightarrow Y+n} = T_A^Y - T_{A+n}^{Y+n}$$

Where $\mathbf{T^Y_A}$is the total of students who attained an age (or age group) in a certain year, $\mathbf{T^{Y+n}_{A+n}}$ is the total of students of this same cohort in a subsequent year A+n, and $\Delta_{A\rightarrow A+n}^{Y\rightarrow Y+n}$ is the total of students who dropped-out in the period between years Y and Y+n.

Two indicators can be built from this calculation. There are the *Rate of inter-annual drop-out by age* (RIDA) and the *Rate of students' conservation by age* (RSCA).

In the first case, the drop-out is expressed as a percentage of the initial cohort of age A in a period of time *n:*

$$RIDA_{A}^{Y\to Y+n} = \frac{\Delta_{A\to A+n}^{Y\to Y+n}}{T_{A}^{Y}} x\ 100$$

The second one, the *Rate of students' conservation by age* expresses the capability of the system to retain the students of a cohort of age A along a period of time *n*:

$$RSCA_{A}^{Y\to Y+n} = \frac{T_{A+n}^{Y+n}}{T_{A}^{Y}} x\ 100$$

Given the following matrix of students by age / grade (expressed in thousands of children), some examples of the calculation of drop-out by age are presented:

| | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|
| **Age 12** | 677 | 670 | 686 | 683 |
| **Age 13** | 640 | 628 | 634 | 642 |
| **Age 14** | 580 | 579 | 579 | 579 |
| **Age 15** | 501 | 507 | 514 | 510 |

- *Rate of inter-annual drop-out of students of 12 years of age between 2012 and 2013:*

$$RIDA_{12}^{2012\to 2013} = \frac{686\ thousand - 642\ thousand}{686\ thousand} x\ 100 = 6.4\%$$

Interpretation: 6.4% of students of 12 years of age dropped out between 2012 and 2013.

- *Rate of inter-annual drop-out of students of ages 12 to 14 between 2012 and 2013:*

$RIDA_{12 to 14}^{2012 \rightarrow 2013}$

$$= \frac{(686\ thousand + 634 thousand + 579\ thousand) - (642\ thousand + 579\ thousand + 510\ thousand)}{(686\ thousand + 634\ thousand + 579\ thousand)} x\ 100$$

$= 8.8\%$

Interpretation: 8.8% of students of ages 12 to 14 dropped-out between 2012 and 2013.

*- Rate of conservation of students of age 12 between 2010 and 2013:*

$$RSCA_{12}^{2010 \rightarrow 2013} = \frac{510\ thousand}{677\ thousand} x\ 100 = 75.3\%$$

Interpretation: 75.3% of the students who attended with 10 years of age in the year 2010 remain in the educational system by 2013.

### 3) Theoretical simulation of a reconstructed cohort by age

This third derivation is inspired in the follow-up methodology of reconstructed cohorts, detailed by UNESCO aiming to reconstruct the flow of students of a cohort who enter a specific level in a specific moment of time (UNESCO 1999, Klein 1998, Taccari 2001).Its design is very similar and it keeps some of the assumptions of this methodology.

It is about a model of a higher level of complexity when compared to the ones that were previously introduced. It aims to reconstruct the flow of entry / drop-out of students from a cohort, applying the inter-annual rates of entry and drop-out by age for a specific range of ages to a group of enrolled students of a specific age. That is, it is intended to simulate what would happen with the school entry and drop-out of a cohort of students if the same levels of a specific year remained constant.

The goal of this model is to predict the future behavior of a cohort based on the current values, since it is not possible to establish what would happen in the subsequent years: the pathways will be influenced by a group of forces that are modified year after year, and which are subject to noticeable variation as a consequence of the implementation of certain policies. On the contrary, projecting for the future the current situation is a way of characterizing the present, turning dynamic the present static scenario. In other words, asking what would happen with the school entry and drop-out if the current trends remained constant represents a strategy to understand the present situation.

The following steps must be taken for its calculation, with its corresponding example:

*a. Calculation of the $\Delta$ value for two consecutive years*: taking as a reference the last two consecutive years with complete and consistent information available regarding the total of students by age, the value of $\Delta$ must be calculated for each one of the inter-annual cohorts. It is important to take the precaution of considering a range of ages whose maximum value is equal to the one that corresponds to one year prior to finishing secondary education. Next, the case of Bolivia for the years 2009 and 2010 is presented as an example:

## Table1.Enrolment by age, Bolivia, years 2009 and 2010, and the application of the calculation of the $\Delta$value

| | Age 3 | Age 4 | Age 5 | Age 6 | Age 7 | Age 8 | Age 9 |
|---|---|---|---|---|---|---|---|
| Year 2009 | 13,404 | 83,732 | 178,452 | 213,247 | 218,707 | 228,487 | 234,782 |
| Year 2010 | 10,895 | 82,752 | 175,031 | 203,920 | 213,662 | 216,356 | 225,910 |

| | Age 10 | Age 11 | Age 12 | Age 13 | Age 14 | Age 15 | Age 16 | Age 17 |
|---|---|---|---|---|---|---|---|---|
| Year 2009 | 223,930 | 221,492 | 213,735 | 205,010 | 191,944 | 171,761 | 153,029 | 108,633 |
| Year 2010 | 231,651 | 219,053 | 215,446 | 204,872 | 191,819 | 177,636 | 154,183 | 111,379 |

Source: data processing based on the information provided by the UNESCO Institute of Statistics (UIS). International Database of Education, October of 2013.

| | | | | | | $\Delta_A^{Y+n}$ |
|---|---|---|---|---|---|---|
| Age 3 - 2009: | 13,404 | --> | Age 4 - 2010: | 82,752 | --> | *69,348* |
| Age 4 - 2009: | 83,732 | --> | Age 5 - 2010: | 175,031 | --> | *91,299* |
| Age 5 - 2009: | 178,452 | --> | Age 6 - 2010: | 203,920 | --> | *25,468* |
| Age 6 - 2009: | 213,247 | --> | Age 7 - 2010: | 213,662 | --> | *415* |
| Age 7 - 2009: | 218,707 | --> | Age 8 - 2010: | 216,356 | --> | *-2,351* |
| Age 8 - 2009: | 228,487 | --> | Age9 - 2010: | 225,910 | --> | *-2,577* |
| Age 9 - 2009: | 234,782 | --> | Age10 - 2010: | 231,651 | --> | *-3,131* |
| Age10 - 2009: | 223,930 | --> | Age11 - 2010: | 219,053 | --> | *-4,877* |
| Age11 - 2009: | 221,492 | --> | Age12 - 2010: | 215,446 | --> | *-6,046* |
| Age12 - 2009: | 213,735 | --> | Age13 - 2010: | 204,872 | --> | *-8,863* |
| Age13 - 2009: | 205,010 | --> | Age14 - 2010: | 191,819 | --> | *-13,191* |
| Age14 - 2009: | 191,944 | --> | Age15 - 2010: | 177,636 | --> | *-14,308* |
| Age15 - 2009: | 171,761 | --> | Age16 - 2010: | 154,183 | --> | *-17,578* |
| Age16 - 2009: | 153,029 | --> | Age17 - 2010: | 111,379 | --> | *-41,650* |

Source: data processing based on the information provided by the UNESCO Institute of Statistics (UIS). International Database of Education, October of 2013.

As it can be observed in the table, the value of $\Delta$ is positive in the ages when the entry to the educational system begins, while it turns into a negative value when reaching the ages in which drop-out is present.

*<u>b. The identification of the maximum age of school enrolment</u>*: this model requires taking as a base the age that reach the highest levels of enrolment to the educational system. In general terms, it corresponds to the age in which the value of $\Delta$ turns from a positive into a negative sign. That is to say, in the example of Bolivia, the maximum age of school enrolment would correspond to age 7.

In the countries where the change of sign is registered in more than one age, it will be necessary to select one of them to use as the year base, selecting the one where higher levels of school enrolment are most probably reached. In the case that this probability cannot be determined, any of the two options can be selected indistinctly. Such is the case of El Salvador for the years 2010 and 2011:

**Table2.Enrolment by age, El Salvador, years 2010 and 2011, and $\Delta$ value**

| | Age 5 | Age 6 | Age 7 | Age 8 | Age 9 | Age 10 | Age 11 | Age 12 |
|---|---|---|---|---|---|---|---|---|
| **Year 2010** | 70,363 | 103,384 | 117,072 | 127,972 | 133,991 | 146,512 | 144,584 | 146,239 |
| **Year 2011** | 67,203 | 101,381 | 114,923 | 121,317 | 127,128 | 134,194 | 145,475 | 142,657 |
| $\Delta_A^{Y+n}$ | | 31,018 | 11,539 | 4,245 | -844 | 203 | -1,037 | -1,927 |

Source: data processing based on information from the Regional Series of Education Indicators about School Failure (SRIEFE, in its Spanish acronym – CECC/SICA).

*In the case of El Salvador, presented as an example, it is observed the presence of two ages where the* $\Delta$ *value changes from a positive value into a negative one: at age 8 and age 10. In this case, the results of the Household Survey on Multiple Purposes (EHPM, in its Spanish acronym) are used in order to establish the convenience of the information of age 8 as a maximum level of school enrolment*[11]*.*

This age is considered as the base age of the cohort, and it is defined as the "*total of the population of an age cohort that will enter the educational system".*

*c. Calculation of the percentage variation of the* $\Delta$ *value around the maximum age of school enrolment*: once that the base age is identified and the maximum level of school enrolment is assumed, the calculation of the percentage variation of $\%\Delta$ is calculated, addressing a specific feature:

- For the higher ages regarding the base age, the denominator of the percentage variation must be the base year of the cohort. In other words:

$$\%\Delta_A^{Y+1} = \frac{\Delta_A^{Y+1}}{T_A^Y} x\ 100$$

Applied to the case of Bolivia, the relative variation between ages 8 and 9 would be:

$$\%\Delta_8^{2011} = \frac{225{,}910 - 228{,}847}{228{,}847} x\ 100 = -1.1\%$$

- Instead, for the lower ages regarding the base age, the denominator of the percentage variation must be the end year of the cohort. That is to say:

$$\%\Delta'^{Y+1}_A = \frac{\Delta_A^{Y+1}}{T_{A+1}^{Y+1}} x\ 100$$

Applied to the case of Bolivia, the relative variation between ages 5 and 6 would be:

[11] According to the data presented in the report "Household Survey on Multiple Purposes 2013", the percentage of attendance of the population of 7 to 9 years of age is higher than such of the group of 10 to 12 years of age (DIGESTYC 2014).

$$\%\Delta'^{2011}_{5} = \frac{203{,}920 - 178{,}452}{203{,}920} x\ 100 = \ 12.5\%$$

This way, applying the formula to each of the $\Delta$ values, the following table is obtained:

**Table3.Application of the calculation of the $\Delta$and %$\Delta$ values. Bolivia, years 2009 and 2010.**

| | | | | | | | $\Delta_A^{Y+n}$ | $\%\Delta'^{Y+1}_A$ | $\%\Delta^{Y+1}_A$ |
|---|---|---|---|---|---|---|---|---|---|
| Age 3 - 2009: | 13,404 | --> | Age 4 - 2010: | 82,752 | --> | | *69,348* | 83.8% | |
| Age 4 - 2009: | 83,732 | --> | Age 5 - 2010: | 175,031 | --> | | *91,299* | 52.2% | |
| Age 5 - 2009: | 178,452 | --> | Age 6 - 2010: | 203,920 | --> | | *25,468* | 12.5% | |
| Age 6 - 2009: | 213,247 | --> | Age 7 - 2010: | 213,662 | --> | | *415* | 0.2% | |
| Age 7 - 2009: | 218,707 | --> | Age 8 - 2010: | 216,356 | --> | | *-2,351* | | -1.1% |
| Age 8 - 2009: | 228,487 | --> | Age 9 - 2010: | 225,910 | --> | | *-2,577* | | -1.1% |
| Age 9 - 2009: | 234,782 | --> | Age 10 - 2010: | 231,651 | --> | | *-3,131* | | -1.3% |
| Age 10 - 2009: | 223,930 | --> | Age 11 - 2010: | 219,053 | --> | | *-4,877* | | -2.2% |
| Age 11 - 2009: | 221,492 | --> | Age 12 - 2010: | 215,446 | --> | | *-6,046* | | -2.7% |
| Age 12 - 2009: | 213,735 | --> | Age 13 - 2010: | 204,872 | --> | | *-8,863* | | -4.1% |
| Age 13 - 2009: | 205,010 | --> | Age 14 - 2010: | 191,819 | --> | | *-13,191* | | -6.4% |
| Age 14 - 2009: | 191,944 | --> | Age 15 - 2010: | 177,636 | --> | | *-14,308* | | -7.5% |
| Age 15 - 2009: | 171,761 | --> | Age 16 - 2010: | 154,183 | --> | | *-17,578* | | -10.2% |
| Age 16 - 2009: | 153,029 | --> | Age 17 - 2010: | 111,379 | --> | | *-41,650* | | -27.2% |

Source: data processing based on the information provided by the UNESCO Institute of Statistics (UIS). International Database of Education, October of 2013.

*d. Reconstruction of a theoretical cohort by simple age*: it is based on the assumption that the total of enrolled students in the base age is the 100% of population that will enter the educational system. Then, the percentage variation established by %$\Delta$is applied to each higher age, in a consecutive and iterative manner. The relative values of negative sign mean a reduction of the cohort, while the positive signs must express an increment. Instead, the variation of %$\Delta'$ must be applied to each lower age with the inverse criterion: the positive sign values mean a reduction of the cohort, and the negative ones an increment. The following table presents the procedure for the calculation.

**Table 4.Reconstruction of school entry and drop-out based on the use of prospective and retrospective cohorts from the calculation of the values of $\Delta$and %$\Delta$.Bolivia, 2009 and 2010.**

| | | | | | | $\Delta_A^{Y+n}$ | $\%\Delta'^{Y+1}_A$ | $\%\Delta^{Y+1}_A$ | *Reconstructed cohort* | |
|---|---|---|---|---|---|---|---|---|---|---|
| Age 3 - 2009: | 13,404 | --> | Age 4 - 2010: | 82,752 | --> | *69,348* | 83.8% | | 41.8% | Age 4 |
| Age 4 - 2009: | 83,732 | --> | Age 5 - 2010: | 175,031 | --> | *91,299* | 52.2% | | 87.3% | Age 5 |
| Age 5 - 2009: | 178,452 | --> | Age 6 - 2010: | 203,920 | --> | *25,468* | 12.5% | | 99.8% | Age 6 |
| Age 6 - 2009: | 213,247 | --> | Age 7 - 2010: | 213,662 | --> | *415* | 0.2% | | **100.0%** | **Age 7** |
| Age 7 - 2009: | 218,707 | --> | Age 8 - 2010: | 216,356 | --> | *-2,351* | | -1.1% | 98.9% | Age 8 |
| Age 8 - 2009: | 228,487 | --> | Age 9 - 2010: | 225,910 | --> | *-2,577* | | -1.1% | 97.8% | Age 9 |
| Age 9 - 2009: | 234,782 | --> | Age 10 - 2010: | 231,651 | --> | *-3,131* | | -1.3% | 96.5% | Age 10 |
| Age 10 - 2009: | 223,930 | --> | Age 11 - 2010: | 219,053 | --> | *-4,877* | | -2.2% | 94.4% | Age 11 |
| Age 11 - 2009: | 221,492 | --> | Age 12 - 2010: | 215,446 | --> | *-6,046* | | -2.7% | 91.8% | Age 12 |
| Age 12 - 2009: | 213,735 | --> | Age 13 - 2010: | 204,872 | --> | *-8,863* | | -4.1% | 88.0% | Age 13 |
| Age 13 - 2009: | 205,010 | --> | Age 14 - 2010: | 191,819 | --> | *-13,191* | | -6.4% | 82.4% | Age 14 |
| Age 14 - 2009: | 191,944 | --> | Age 15 - 2010: | 177,636 | --> | *-14,308* | | -7.5% | 76.2% | Age 15 |
| Age 15 - 2009: | 171,761 | --> | Age 16 - 2010: | 154,183 | --> | *-17,578* | | -10.2% | 68.4% | Age 16 |
| Age 16 - 2009: | 153,029 | --> | Age 17 - 2010: | 111,379 | --> | *-41,650* | | -27.2% | 49.8% | Age 17 |

Source: data processing based on the information provided by the UNESCO Institute of Statistics (UIS). International Database of Education, October of 2013.

The value of the right column represents the percentage of the cohort that corresponds to each age, according to the trends outlined between the two last years of available information. The percentages must be read in the terms of a prospective cohort for the higher ages, and they indicate the probability of survival in school of enrolled students of 7 years of age.

Considering the example of Bolivia, it is possible to say that if the rates of school entry and drop-out remain constant in the passage from 2009 to 2010:

- 41.8% of the population that manages to enter the educational system attends from 4 years of age.

- The specific coverage at age 5 reaches 87.3% of the population that manage to enter the educational system.

- Instead, at age 12, 91.8% of the cohort will continue attending, meaning that 8.2% will drop out before attaining that age.

- Slightly less than half of the cohort (49.8%) will continue attending the educational system by age 17.

As it can be observed, once the calculations are applied for the cohort reconstruction, the interpretation is rather easy. It should be recalled that, when considering the population who manages to attend the educational system as a theoretical base, the analysis excludes the children that remain completely out of the system.

These indicators may be useful to identify bottlenecks at the stage of access or drop-out of the educational system and, compared along time, they allow dimensioning the improvement impact in retention projected in the medium term.

## ADVANTAGES AND CONTRIBUTIONS OF THE MODEL FOR ESTIMATING POPULATION DEMAND

Having developed the methodology for the application of the follow-up model of cohorts by age, noticing the assumptions in which it is based and its possible bias, and bearing in mind that it is not a methodology for the measurement of coverage but for the approximation to the uncovered demand of specific levels or stages of the educational system, it is possible to identify the advantages of these indicators:

a. Precision for disaggregated analysis: While the assumption of a closed system can be supported, the model of inter-annual follow-up of cohorts by age is strong enough to estimate the specific potential demand at certain ages. For this reason, prior to its application it is necessary to assess the possible impact of students' migrations or relocations. With this provision, it may become very useful for the identification of regions with remarkable gaps of pre-school coverage, or to detect early drop-out.

b. Opportunity given by annual information and monitoring: the educational information systems of the countries of the region produce information with an annual frequency, which enables establishing a permanently updated monitoring system.

c. Disaggregation from educational variables: the information produced through this methodology can be disaggregated based on the characteristics of students, of the groups and of the schools, enabling the focus on the analysis of specific dimensions of intervention on the educational system. In particular, performing some cross-analysis of inter-annual cohorts by age and attendance with theoretical age or age-gap lead to some interesting diagnosis about the combination of pathways and school drop-out.

d. Consistency with regulations of the educational system: the methodology of the follow-up of inter-annual cohorts by age, when building the measurement of the age with the educational system's criteria, is much more precise than other sources (such as census or household surveys) for the diagnosis of entry to pre-school and primary. It enables to give an account of the situation of the population *with the expected* age to attend a specific stage (the last year of pre-school, to the penultimate year, to the first year of primary…).

.