International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Validation of the Wittgenstein Centre Back-projections for Populations by Age, Sex, and Six Levels of Education from 2010 to 1970

Speringer, M., Goujon, A., K.C., S., Eder, J., Potancokova, M. and Bauer, R.

IIASA Interim Report
April 2015

International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report **IR-15-008**

Validation of the Wittgenstein Centre Back-projections for Populations by Age, Sex, and Six Levels of Education from 2010 to 1970

Markus Speringer (speringe@iiasa.ac.at)
Anne Goujon (goujon@iiasa.ac.at)
Samir K.C. (kc@iiasa.ac.at)
Jakob Eder (jakob.eder@oeaw.ac.at)
Michaela Potančoková (potancok@iiasa.ac.at)
Ramon Bauer (ramon.bauer@oeaw.ac.at)

Approved by

Wolfgang Lutz
Program Director, World Population Program

April 1, 2015 – Revised April 23, 2015

Contents

Abstract

There have been few attempts at creating data series on levels of educational attainment of the adult population consistent across time and space by age and sex. They would be needed to estimate the role played by education and human capital in economic, technological, environmental models as correctly as possible. In 2007, Lutz et al developed a methodology to reconstruct (and project) levels of educational attainment based on the information contained in the base-year source of choice for the most recent period (Lutz et al. 2007a). The methodology was applied again in the framework of a new round of population projections published in 2014 online (www.wittgensteincentre.org/dataexplorer) and in the Oxford University book "World Population and Human Capital in the Twenty-First Century" edited by Lutz, Butz and KC. There, the coverage increased to 171 countries and the number of education categories to six. The back-projection methodology was applied to the updated base-year sample in 2010 to arrive at the reconstruction of levels of educational attainment by age and sex for the period 1970-2005. The purpose of this paper is to compare the reconstructed datasets to other existing sources of historical data on education, including the former reconstruction from 2007, collection and other reconstruction exercises.

The validation of the Wittgenstein Centre back-projection model outcomes with available empirical data source enables the evaluation of our back-projection method for the establishment of harmonized and consistent time series on the educational composition of 171 countries in the world. In comparison, the most other available datasets suffer from severe flaws, hampering any valid trend and regression analysis on levels of educational attainment.

The back-projection methodology is explained in Section 2 and Section 3 describes the collection of empirical data for the validation of the WIC 2015 dataset and associated challenges. The validation methodology and results are developed in Section 4. Detailed documentation about the country-specific validation is available from the Appendices.

About the Authors

Markus Speringer is Research Scientist at the Vienna Institute of Demography (VID) of the Austrian Academy of Sciences, and Research Assistant in the World Population Program at the International Institute for Applied Systems Analysis (IIASA), Wittgenstein Centre for Demography and Global Human Capital.

Anne Goujon is Leader of the Research Group *Human capital data lab* at the Vienna Institute of Demography (VID) of the Austrian Academy of Sciences, and Senior Research Scholar in the World Population Program at the International Institute for Applied Systems Analysis (IIASA), Wittgenstein Centre for Demography and Global Human Capital.

Samir KC is Leader of the *Modelling Human Capital Formation* project in the World Population Program at the International Institute for Applied Systems Analysis (IIASA), Wittgenstein Centre for Demography and Global Human Capital.

Jakob Eder is a Research Assistant at the Vienna Institute of Demography (VID) of the Austrian Academy of Sciences, Wittgenstein Centre for Demography and Global Human Capital.

Michaela Potančoková is Research Scientist at the Vienna Institute of Demography (VID) of the Austrian Academy of Sciences, and Research Scholar in the World Population Program at the International Institute for Applied Systems Analysis (IIASA), Wittgenstein Centre for Demography and Global Human Capital.

Ramon Bauer is Research Scientist at the Vienna Institute of Demography (VID) of the Austrian Academy of Sciences, Wittgenstein Centre for Demography and Global Human Capital.

Preface

This is a revision of the original IR-15-008 published April 1, 2015. This version includes one main change in addition to some language revisions.

In the primarily published version of this Interim Report we stated in Section 3.1, with 411, a wrong total number of data-points for the period 1970 up to the base-year. That number included data-points before 1970 that have not been used in the validation. In fact, we can report for the validation 339 unique data-points. This correction does not affect the validation results, illustrated in Section 4.

Finally, the conclusion, including Figure 15, has been modified according to the changes in the main text.

Validation of the Wittgenstein Centre Back-Projections of Populations by Age, Sex, and Six Levels of Education from 2010 to 1970

Markus Speringer
Anne Goujon
Samir K.C.
Jakob Eder
Michaela Potančoková
Ramon Bauer

1 Introduction

In 2007, the International Institute for Applied Systems Analysis (IIASA) and the Vienna Institute of Demography (VID) published their first reconstruction of past levels of education at country level – also called back-projections – from 2000 back to 1970, followed by population projections from 2000 to 2050 (Lutz et al. 2007a; KC et al. 2008; KC et al. 2010). Both the back-projections and projections required the same base-year data on population disaggregated by levels of educational attainment by age and sex. The base-year for the IIASA/VID back-projection exercise (Lutz et al. 2007a) was set to 2000, and the dataset included data on four levels of education by age and sex for 120 countries. A similar exercise – population projections and back-projections by level of educational attainment – was conducted between 2011 and 2014. This new round of projections is documented in Lutz et al. (2014). The full dataset including the population projections and back-projections is available online in the Wittgenstein Data Explorer (WIC 2015).

The recent back-projection dataset – thereafter referred to as WIC 2015 – differs from the 2007 dataset in three main ways: **(1)** The base-year was changed from 2000 to 2010 and updated with more recent census and survey data (see Bauer et al. 2012); **(2)** the number of education categories was increased from four to six (see Table 1 in Section 3.1) to encompass more differentials in levels of educational attainment across the world; and **(3)** the number of countries in the dataset was extended from 120 to 171 to enlarge the representativeness of the dataset.

The aim of this paper is to present the validation exercise that was conducted to evaluate the WIC 2015 back-projections, as Riosmena et al. (2008) accomplished for the Lutz et al. (2007) dataset. (Riosmena et al. 2008; Lutz et al. 2007b) In Section 2 we describe the back-projection methodology. In Section 3 we explain our efforts to collect empirical data for the validation of the WIC 2015 dataset, including the description of data sources, the data management, and the data harmonization procedure. In Section 4 we introduce the back-projection validation procedure, the resulting amendments to the reconstructed WIC 2015 data, and the results of the validation with empirical historical data and other reconstructed datasets. The concluding section summarizes the main features and envisaged applications of the new WIC back-projection dataset and discusses the next steps to improve this data collection on global human capital. Detailed documentation about the country-specific validation is available in Appendix I.

2 The WIC dataset: Back-projection methodology

The back-projections rely on the principle that education is overwhelmingly acquired at young ages and is therefore a fixed attribute later in life. Hence, the educational attainment composition of a population encompasses the educational development of the past, which can be reconstructed along age groups. If we know the proportion of 50-year old with tertiary education in 2010 in country A, their share is a valid estimate of the proportion of 40-year old in 2000 in the same country. However, there are two phenomena that can upset the full equivalence through time: Differential mortality and migration. For instance, if tertiary educated people are less affected by mortality than lesser educated people – which has been demonstrated in the literature (Huisman et al. 2005; Hummer and Lariscy 2011), then it is possible that the share of 40-year old with tertiary education will be less than the share of the 50-year old, their higher share at the age of 50 being the result of their lower mortality in the 10-year period. The mortality differentials are more important at old ages when mortality rates are higher. Another disturbance that can affect the back-projections and has to be taken into account in the reconstruction is migration which rarely follows the education distribution of the host-country in terms of in-flows and out-flows. Lutz et al. (2007a; 2007b) provide a summary of the back-projection methodology and how migration and mortality differentials were dealt with in the previous round of back-projection. The same principles were used in the WIC 2015 round with some amendments.

One amendment was, since schooling occurs mostly in earlier ages of life, that all transitions between the six levels of educational attainment occur by age 30-34 (K.C. et al. 2015). In the earlier reconstruction (Lutz et al. 2007a; Lutz et al. 2007b) with four levels of education (no education, primary, secondary, and tertiary) the transition from no education to primary was possible until age 15-19, from primary to secondary until age 20-24, and from secondary to tertiary until age 30-34. In the WIC 2015 dataset, the findings from the collection and harmonization effort (Bauer et al. 2012) were used, which revealed that the age at transition to the next level of education was often happening at higher ages or younger ages rather than at the standard age at graduation, describing the theoretical age at which a person is supposed to graduate from a particular level of education e.g. in Austria, pupils classically enter upper secondary education at the age of 14 and graduate four years later at the age of 18 years.

Based on these findings, we estimated country- and age-specific transition rates for education in the base-year to reflect the reality of educational transitions. It was, however, not possible to estimate this transition matrix in the past due to the unavailability of consistent data. Therefore, country-specific transition matrices were assumed to be constant over the back-projection period. This assumption could introduce a bias at both ends of the educational scale. Countries which had an elitist education system in the past might have had higher transition rates at earlier ages (compared to the standard graduation age) than now when the education has become more egalitarian and less advantaged pupils tend to transit to the next level later or spend more time to finish levels. Reversely, in some countries where the average school entry occurred at older ages than the standard graduation age, improvements in the access to modern education may decrease the age at school entrance and increase the speed of attaining the next school level (K.C. et al. 2015).

The WIC 2015 back-projection methodology also differs from the earlier one (Lutz et al. 2007a; Lutz et al. 2007b; K.C. et al. 2015) in terms of the differential mortality by education. Both datasets rely on a standard schedule of mortality differentials. However, Lutz et al. (2007b) assumed that the mortality differential expressed in terms of life expectancy at age 15 (e_{15}) between the highest and the lowest education categories was 5 years for both sexes with a 1-2-2 year-pattern between the no education, primary, secondary and tertiary

education categories. In WIC 2015 the differentials were revised in two ways. Firstly, since the education differentials were found to be larger among men than women (Lutz, Butz, and KC 2014), the difference in e_{15} between the highest and the lowest education categories was set to 6 years for men and 4 years for women. Secondly, we assume the education differential in e_{15} to have a 1-1-2-1-1 year-pattern between the no education, some primary, completed primary, lower, upper, and post-secondary education levels respectively – and the same proportionally for women (K.C. et al. 2015).

At each step of the back-projection, the United Nations (UN) estimates of life tables for the population (United Nations 2011) are disaggregated into education specific life tables using the sex specific educational differentials in mortality as described in Lutz et al (2007b). The education, age and sex specific survival ratios are applied to back-project the population in five-year steps. The total age-sex specific populations are then adjusted proportionally to match the UN age-sex distribution in the 1970-2005 period.

Finally, the education transition matrices are applied to back-project the populations in the 15-34 age groups. By virtue of going into the past, the whole process does not have to consider fertility. The education distribution in the highest age group 100 years plus is assigned according to a logistic model fit derived from the base-year data (K.C. et al. 2015).

3 Data collection and validation methodology

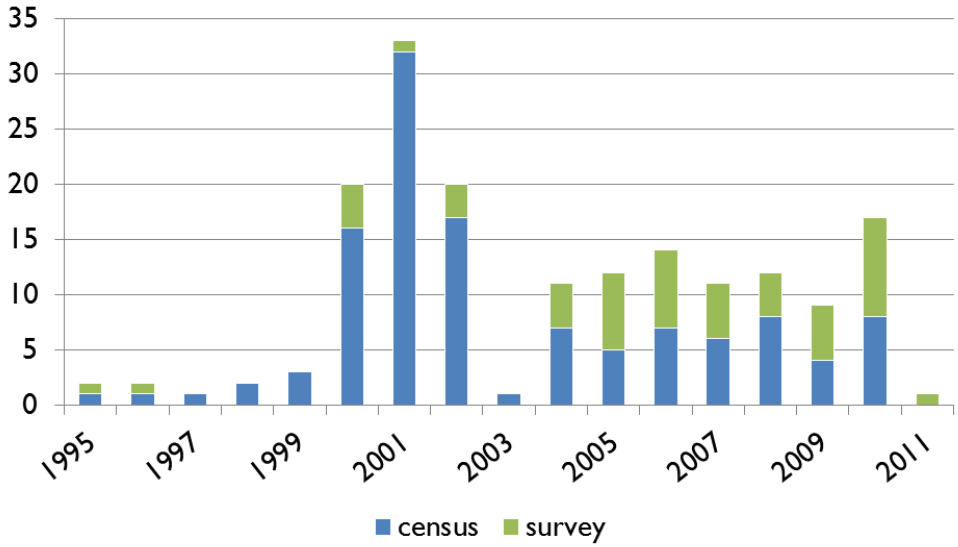
3.1 Data sources for the base-year data

The WIC 2015 back-projection dataset is based on harmonized baseline data on educational attainment from two types of sources (Bauer et al. 2012): **(1)** *census data* collected from major international institutions, primarily from IPUMS [1] (Minnesota Population Center 2014), and also from CELADE [2] (CELADE/CEPAL 2014), EUROSTAT (EUROSTAT 2014), UNESCO Institute for Statistics[3] (UIS 2014b), or directly from National Statistical Offices (NSO); **(2)** *survey data* derived from Demographic Health Surveys[4] (Measure DHS 2015), Labour Force Surveys (LFS), or Multiple Indicator Cluster Survey (MICS). Beside the problems of sampling (size and representativeness), surveys tend to be designed for the national context and are not always immediately comparable. Survey data were used if the census was of poor quality, not available or outdated, whereby the data quality is deeply related to the data source.

The base-year data for most of the 171 countries originated from the period 2000-2002, with a majority of census datasets (119 census or registers/52 surveys). The timespan of the data collected to estimate the base-year reaches from 1995 (Central African Republic [DHS] and Turkmenistan [UIS]) to 2011 with Ethiopia DHS data (see Figure 1).

[1] IPUMS – Integrated Public Use Microdata Series (https://international.ipums.org/)
[2] CELADE – Division de Poblacion, CEPAL (http://www.cepal.org/redatam/default.asp?idioma=IN)
[3] UIS – UNESCO Institute for Statistics (http://www.uis.unesco.org/Pages/default.aspx)
[4] DHS – Demographic Health Surveys (http://www.dhsprogram.com/)

Figure 1. WIC 2015 data sources on educational attainment (effective February, 2014)



The base-year data were harmonized and aggregated into the following six categories based on the ISCED 1997 classification (see Table 1): *no education*, *incomplete primary* **[incomplete ISCED 1]**, *primary* **[ISCED 1]**, *lower secondary* **[ISCED 2]**, *upper secondary* **[ISCED 3]**, and *post-secondary education* **[ISCED 4+5+6]**.

Table 1. Categories of educational attainment (Bauer et al. 2012)

WIC 2015	**ISCED 1997**
No education	No level or ISCED 0 Grade 1 of ISCED 1 not completed
Incomplete primary	Incomplete ISCED 1
Completed primary	Completed ISCED 1 Incomplete ISCED 2
Completed lower secondary	Completed ISCED 2 Incomplete ISCED 3
Completed upper secondary	Completed ISCED 3 Incomplete ISCED 4-5A & 5B
Post-secondary	ISCED 4 & 5B [first diploma, shorter post-secondary courses] ISCED 5A & 6 [longer post-secondary courses, post-graduate level]

Historical datasets are thereby defined as all datasets that are older than the country-specific base-years. For the validation, we aimed to collect as many historical data points as possible, going back from the country-specific base-year to 1970 for the 171 countries. Ideally, this would be about 1148 data points for the period 1970 up to the country-specific base-year (excluding the base-year). However, and not surprisingly, it was not possible to recover all the data-points.

Firstly, the data used originates typically from censuses which are mainly collected every 10 years, and surveys like DHS do not take place at regular intervals. Moreover, the data are not readily available for all the countries in our sample. In total we managed to collect one or more data-points on educational attainment since 1970 by age and sex for 138 countries (81 percent of all 171 countries), that summed up to 339 data-points (30 percent of the potential 1148 data points in the period 1970 up to the base-year). It was not possible to

include 33 countries, because historical data on educational attainment were either not available because it had not been collected (e.g. Ethiopia, Nigeria, see Table 2). Other data at hand were not at a sufficient level of detail or quality but still included into the validation procedure (e.g. Ireland, Netherlands) despite the fact that educational attainment became a standard element of most censuses in the second half of the 20th century.

Additionally, the source of the baseline data puts limitations on the validation exercise. Due to already documented discrepancies between census and survey data e.g. between census and LFS data (Cohen and Soto 2001; Cohen and Soto 2007), or between census and DHS data (Bauer et al. 2012), we prioritize census data as the most reliable source in the collection of historical data. Therefore, we first collected past census data on education from the IPUMS and supplemented these datasets with census data from NSOs, or if not available in UIS datasets (see Section 4.2). In case the source used for the validation did not meet the levels of details required in the WIC database in terms of education or age, we validated the data at a higher level of aggregation, for instance by using population 15 or 25 years and older, which is more often available (e.g. from UIS).

As shown in Table 2, the highest coverage of historical datasets compared to our base-year countries, could be reached in Northern America and Latin America, where it was possible to collect datasets for 100 and 97 percent respectively. The least available historical data series could be found for Asia (70 percent) and Africa (72 percent) because census reports or datasets are often not publicly available, in case there have been censuses, which is not always the case.

Table 2. Country coverage of historical datasets (1970 up to the base-year) by UN region

UN region	All countries	Countries with available base-year	Countries with historical data	Coverage of all historical data points (%)	Missing countries
Europe	39	39	34	87.2	**Albania**, **Iceland**, **Malta**, **Ukraine**, **United Kingdom**
Asia	50	43	30	69.8	Afghanistan, **Armenia**, **Azerbaijan**, Brunei, **Bhutan**, **Cambodia**, **East Timor**, **Georgia**, **Iran**, **Iraq**, **Kyrgyzstan**, **Laos**, North Korea, Oman, **Palestine**, **Saudi Arabia**, Sri Lanka, **Turkmenistan**, Uzbekistan, Yemen
Africa	55	46	33	71.7	Angola, Botswana, **Chad**, **Comoros**, **Cape Verde**, Djibouti, **Egypt**, **Equatorial Guinea**, Eritrea, **Ethiopia**, **Gabon**, **Guinea-Bissau**, Libya, **Madagascar**, Mauritania, Mayotte, **Nigeria**, **Reunion**, **Sierra Leone**, **Somalia**, Togo, Western Sahara
Northern America	2	2	2	100.0	-
Latin America	37	34	33	97.1	Barbados, Grenada, **Suriname,** Virgin Islands
Oceania	12	7	6	85.7	Fiji, **French Polynesia**, Guam, Micronesia, Papua New Guinea, Solomon Islands
World	***195***	***171***	***138***	***80.7***	***[171 – 138 = 33 countries]***

Note: Countries where no historical datasets could be found, but with base-year data are underlined and in bold.

3.2 Data harmonization

Despite the sheer lack of comprehensive data on education from 1970 onwards, it is increasingly difficult the further we go back in the past to find usable data sources that are comparable to our education categories (see Table 1) based on the UNESCO ISCED 1997 classification (UNESCO 2006).

Educational categories describing the highest level of attainment always follow national education systems. Due to the variety of nationally distinct education systems, many different types of educational levels exist around the globe, which are additionally changing over time. Any change affect the way data are collected in the census. Unlike other datasets (Barro and Lee 2013; Cohen and Soto 2007; Cohen and Leker 2014; de la Fuente and Doménech 2012) our reconstruction is based on a base-year dataset of consistent estimates of educational attainment by age and sex harmonized into ISCED 1997 levels to achieve a better comparability and to avoid flaws in primary data (Bauer et al. 2012; Potančoková, KC, and Goujon 2014)

Seemingly, in order to make the historical education statistics comparable with the harmonized base-year it is necessary to harmonize them according to the ISCED 1997 mapping[5] (UIS 2014a). To create such a harmonized educational database has often been challenging because the data were not organized in line with the ISCED definitions.

3.2.1 Categories of educational attainment

In order to harmonize the historical data with the WIC 2015 categorization of the base-year data, we collected detailed information based on national education categories – optimally categories as surveyed in censuses without being further processed or aggregated – to allocate the various categories to comparable ISCED 1997-levels. Special emphasis has been placed on the differentiation between completed and incomplete levels, using information about the highest school year or grade attended within the level to distinguish between completed and incomplete levels.

While IPUMS often provides harmonized and unharmonized variables on educational attainment with the information on years of schooling, which makes a differentiation between *'incomplete'* and *'complete'* levels possible, NSO and UIS datasets mainly provide aggregated datasets with the national education categorization.

Therefore, in the process of validation, in the WIC 2015 dataset single categories had to be aggregated to match the national education categorization and to enable a comparison between empirical and back-projection data. NSOs often do not provide data on lower education in the detail which would be needed for validation, but combine them as 'basic' or 'elementary education', which includes *incomplete ISCED 1*, *ISCED 1* and *ISCED 2*, like in Norway or France. In such cases the WIC 2015 categories got aggregated corresponding to the national categorization to conduct the validation.

[5] The UNESCO Institute for Statistics also plans to harmonize their database on educational attainment using the ISCED 1997 categories to provide a comparable dataset. (UIS 2014a)

3.2.2 Changes in national educating systems

Another challenge for determining educational attainment by age is the change in national education systems over time. People of different ages often went through different educational programs and/or systems with respect to the number of grades required to reach a certain ISCED level. Although ISCED 1997 mappings are available for most countries, it is not always clear how to identify and allocate surveyed categories, particularly if the categories in the respective national education program varied from those used in the ISCED mappings.

Besides prolongations in the duration of compulsory education, alterations in the duration of schooling might occur just at one particular level or at more levels at once. Some countries did change their education systems fairly frequently. Since 1970, for example, Cambodia had four different systems and Mozambique and Ukraine reformed their education systems three times. Moreover, such changes in national education systems are often poorly documented and are thus difficult to identify (Bauer et al. 2012).

Aiming to take changing education systems into account when allocating original data into the six WIC 2015 categories, we benefited from a compendium of documented changes compiled by UNESCO Institute for Statistics[6] (UIS 2014c). However, UNESCO does not provide any information before the 1970s and detailed information – duration of compulsory education, theoretical duration of ISCED 1997 levels as well as starting age of education at each level – is given only from 1998 onwards. For all persons enrolled prior to 1998, which is almost the entire population of 15 years and older covered by the WIC 2015 dataset, there is only information on the aggregate duration of secondary education, without any differentiation between lower (ISCED 2) and upper (ISCED 3) secondary education. This is a particular problem, if the cumulative duration of lower and upper secondary education remains the same, but the duration of each level has changed – say, from a 3-4 year system to a 4-3 year one. Such a situation is not unusual, because compulsory education is often extended by adding one or more grades of schooling years to lower secondary education by clipping it off from upper secondary. In other cases, there was evidence that the cumulative durations of secondary education changed, but no further information about the levels or grades affected was available. Due to these circumstances, we acknowledge that at least some changes in educational systems, which potentially would have affected the cohort-wise allocation of particular grades and degrees to ISCED 1997 levels, may have remained concealed to us (Bauer et al. 2012; UNESCO Institute for Statistics 2011).

3.2.3 Changes in age groups of educational attainment

Beside changes in national education systems, age categorizations are often not consistent across time and countries. While the population by education is mostly given from the age of 10 or 15 years in single or five-year age groups, some countries have different starting ages. For example the Norwegian censuses from 1970 to 1980 start with the age of 16 years with differing age intervals in the younger age groups (e.g. 16 -19 years) according to the structure of the national education systems (CBS Norway 1986; CBS Norway 1991; CBS Norway 1999), or the Czech census of 1980 starts the documentation of education in the published reports with the age of 20.(CZSO 1980; CZSO 1991)

Especially in older censuses data are available just for one or two broad age groups such as 15 years plus and/or 25 years plus. Hence, for the validation of the WIC 2015 back-projection data we used the aggregated age group *population 25 years plus* as this age group is widely available or can be calculated over most census points and times.

[6] UNESCO collects evidence on past education systems since 1970: http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=143&IF_Language=eng [last visited March 2014]

4 Validation process and results

In the validation process we compare the results of the historical reconstruction of educational attainment with empirical historical datasets that were harmonized beforehand. With the setting up of the base-year dataset and the aim to collect data on educational attainment from various sources in order to choose the most reliable, it became apparent that census data seem to be the most reliable and comprehensive data. As a matter of fact, different data sources may lead to different results, and in this case in different educational compositions. Picking the most reliable source is certainly the correct solution to this problem. This was not always an unambiguous task, as the following comparisons between different data sources demonstrate.

Based on these experiences we decided to focus our ambitions to collect historical educational data on censuses. The three main sources that provide reliable historical time series of census data via databases, aggregated tables, or reports are IPUMS, NSOs and UIS, which show different patterns of fitting accuracy to our back-projection model, as shown in the following sections. Apart from that we compare the WIC 2015 reconstruction with other reconstruction exercises that tried to deal with the lack of consistent data on education.

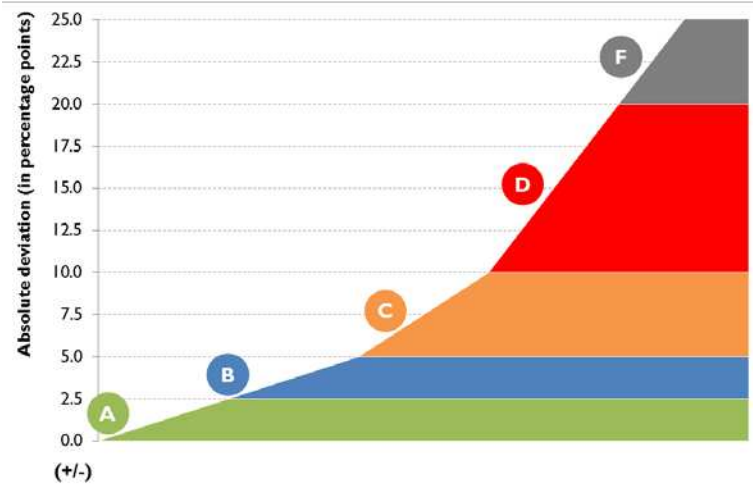
4.1 Data validation categorization

The comparison of empirical historical education data with the WIC back-projection data is based on the lineup of the proportional share of the six WIC education categories of the total population 25 years plus. Hereby we compare the (+/-) deviation between the corresponding educational categories by absolute differences in percentage points (pp). To illustrate this, for Serbia 1971 the WIC 2015 dataset estimates a proportion of 20.1 percent of the total population aged 25 years plus with no education. The collected NSO data for that year show a share of the same educational category of 26.1 percent, which gives a deviation of 6.0 percentage points in this educational category.

The education category with the highest deviation of data point in a country serves as basis for the ***validation categorization***. For instance, in the same example of Serbia 1970 the highest deviation does not occur in the education category *no education*, but with 6.6 pp in the group of *incomplete primary education*. That causes Serbia 1970 to be in category *C*.

The deviations in percentage points are categorized into five groups by predefined thresholds (see Figure 2). The labels of those five groups are referring to the *American school grading scheme* (*A, B, C, D,* and *F*). Therefore deviations beneath 5 pp are labelled as *good* (*A*) or *rather good* (*B*). We made this distinction to highlight the countries with very high accuracy of fit (beneath 2.5 pp). Higher deviations above 5 pp are divided into the categories *rather bad* (*C*) or *bad* (*D*) to get a more differentiated picture of the level of deviations between empirical and model data. Countries and data series with a deviation above 20 pp and no chance of improving the matching accuracy are allocated to the category *not usable* (*F*), since the deviation beyond that point cannot be explained whereas most of the deviations below the 20 pp threshold can be explained.

Figure 2. Validation categorization scheme by (+/-) absolute deviation (in percentage points)



Outliers are, for instance, countries with highly aggregated educational categories, where it is not possible to distinguish educational categories according to the WIC 2015 categories as there are massive overlaps between educational groups e.g. for the Netherlands or Ireland. Another reason could be due to a bias introduced by the educational impact of massive immigration like in Israel where different waves of immigration have affected the education composition.

4.1.1 Validation of historical datasets

The validation is a two-step procedure. First, we compare the harmonized empirical data with the WIC 2015 dataset and attribute to the country- and education specific data point a validation category, depending on the degree of deviation. In the case of a perfect match, like the validation with empirical NSO and IPUMS data of Greece in 1971, 1981 and 1991 (category *A*), we do not apply any further steps. In any other case (categories *B*-*F*), we control the validation output in a second step to find the source of deviation. This second step comprises two parts:

- **Controlling for harmonization mistakes:** In this part we control the gathered empirical data for issues that could have occurred in the harmonization of historical and base-year data, e.g. misallocation of educational categories in the recoding of IPUMS data or we did not account for changes in the country-specific education systems.
- **Controlling for data reliability:** If there are no issues that could be identified and/or solved, we apply an *'age-cohort analysis'* (ACA) described in Section 4.1.1.1 to check for the data reliability. Additionally, we check the reliability of the source data by means of reviewing in detail the data documentation and related literature to identify flaws in the census quality, before checking for historical migration movements that could have biased our validation results.

These corrections are incorporated into our empirical dataset and documentation before finally apply our validation categorization. Examples for this procedure can be found in the following Section 4.1.1.1 and Section 4.2.

4.1.1.1 Age-cohort analysis (ACA)

The *age-cohort analysis* (ACA) was used to evaluate changes in the educational composition of cohorts between several empirical data points. Two or more data points of harmonized educational distributions by age are necessary for this approach e.g. the 40-45 age group in 1980 and the 50-55 age group in 1990. The age groups (single age, five or ten-year age groups) have to correspond to the time intervals between the decennial data points.

By employing ACA we were able to identify outliers, corrupted datasets and unusual changes in educational compositions. The procedure is based on the evaluation of in/consistencies in the development of educational composition of the age cohorts. The ACA allows us to identify unusual age heaping, sudden drops or increases in the educational composition of particular cohorts that points to problems of data quality of the particular source or to other biases such as that of selective migration explaining the sudden changes in shares with post-secondary education in Israel.

The ACA allows to spot issues with the categorization of education. For instance, in the Canadian census of 1971 which is accessible via IPUMS, the post-secondary educated population seems to be highly underrepresented compared to the later census years. According to these datasets the population 25 years and older with post-secondary education would have increased between 1971 and 1981 from 5.4 to 38.4 percent. The ACA confirms a discontinuity in the share of the population with post-secondary education across all cohorts.

In Trinidad and Tobago the available data from IPUMS had an underrepresentation of the no education category in 1980 compared to 1970 and 1990. In such cases where the educational composition of the compared age cohorts in one dataset shows a clear deviation from the other available data points, we discard the dataset in the validation, like in the case of Israel, or look for another dataset, which was the case for Trinidad and Tobago where the data was substituted with available data from the NSO.

4.1.2 Amendments to the reconstruction after the first validation phase

During the validation procedure the comparison unveiled discrepancies between the back-projections and the empirical data. We further identified the datasets that needed to be corrected. We focused on the countries in categories *D* and *F* for which we had collected reliable historical data points from valid data sources, like IPUMS or NSO. In total we could identify twelve countries[7] that fit to the articulated criteria.

For those countries we incorporated the empirical data points into our model and reconstructed from the validated data points the missing data points by applying the WIC 2015 back-projection methodology (see Section 2) to fill the gaps in the empirical data time series. Before going into detail, the used data points had to be processed so that they are applicable to our back-projection procedure.

Firstly, we adjusted the age groups available from the empirical data i.e. split broader age groups (e.g. 10 year age groups) by applying a cubic spline function or extend the last open age group (e.g. age group 65 years plus) up to 100 years plus.

Secondly, missing educational categories in historical data points were estimated. For instance *incomplete* and *completed primary* were aggregated in one educational group in the Czech Republic (1971, 1981, and 1991) or in Hungary (1970, 1980, and 1990). If we use the example of Hungary, in order to get these two educational groups apart, we projected in a first phase back from the base-year 2001 to 1990 and used the educational share between

[7] Countries: Bangladesh, Cameroon, Czech Republic, Finland, Haiti, Hungary, Jamaica, Liberia, Philippines, Spain, Singapore, and Tanzania.

incomplete and completed primary education to split the primary education category in the empirical data for 1990. We further used this adjusted 1990 distribution to project further back to 1985 and 1980, where we applied the same procedure as for 1990. This procedure was repeated until 1970.

The WIC 2015 dataset incorporates the twelve country-specific corrected time series and provides the basis for the validation with the empirical data series from IPUMS, NSOs and UIS as shown in the next section.

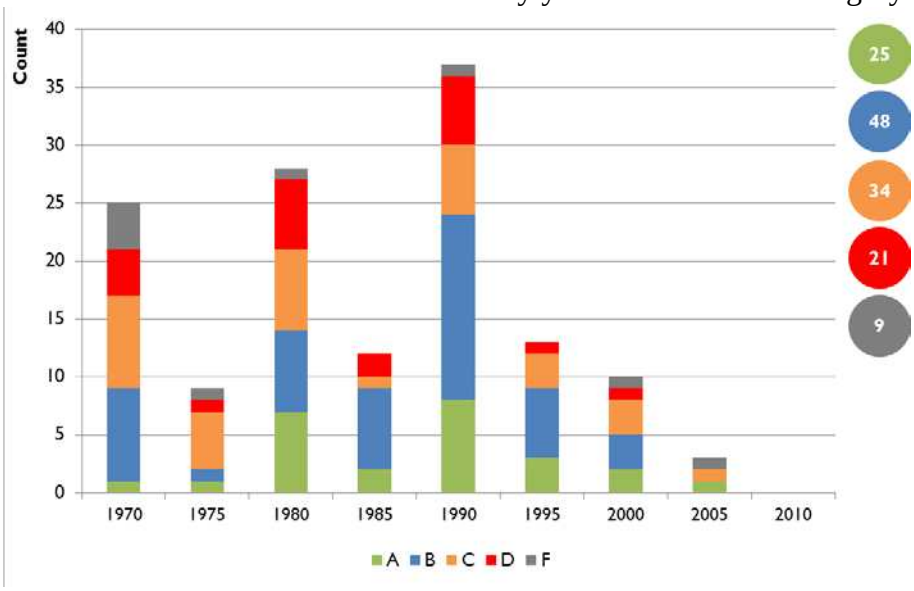
4.2 Validation with empirical datasets

As mentioned earlier, there are three major sources for validating the WIC 2015 dataset: IPUMS, NSOs, and UIS.

4.2.1 Integrated Public Use Microdata Series (IPUMS)

For our purposes the major source of historical data on education is the IPUMS dataset provided by the Minnesota Population Center at the University of Minnesota. It allows for the validation of 55 countries or 137 census data points with educational attainment. The IPUMS dataset includes census microdata samples (usually 5 or 10 percent samples) and shows relatively high matching with the WIC 2015 dataset. The high degree of details in the microdata allows for the differentiation between incomplete and complete education categories. About 73 datasets show an accuracy of category *A* or *B*, which means an absolute difference between WIC 2015 and the harmonized IPUMS data beneath 5 percent points.

Figure 3. Validation result for IPUMS data source by year and validation category

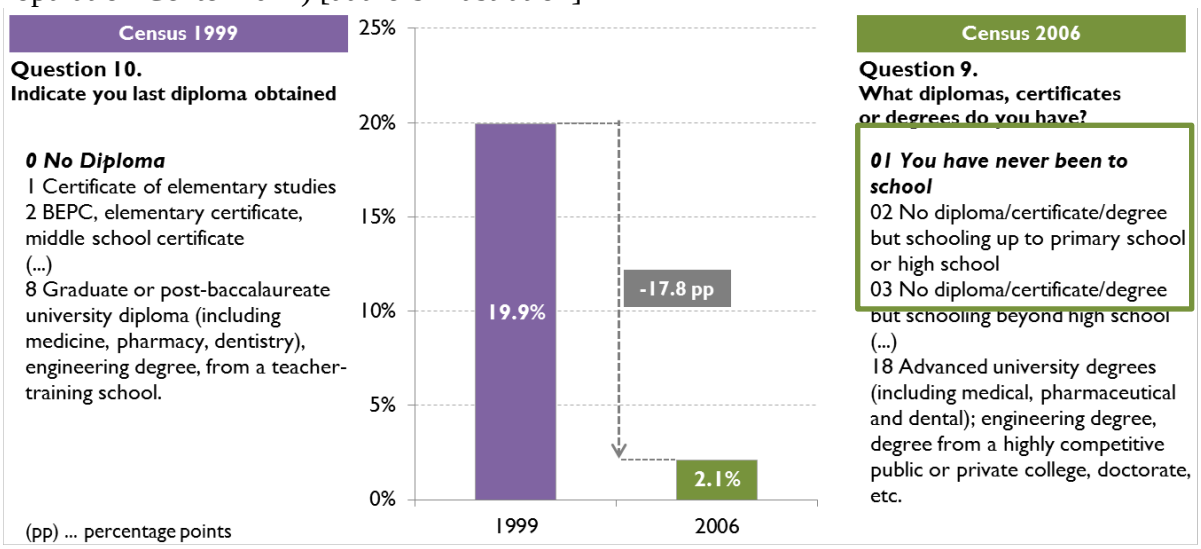


The highest number of data points can be found in the census year 1990 with 37 datasets out of which 24 are categorized as *A* or *B*. The further back in time we compare the WIC 2015 back-projection dataset with IPUMS data the greater the deviation. In total there are 30 datasets that enter categorization *D* and *F* indicating a strong divergence that was evaluated. If the deviation originates from the transition model, we can adjust the back-projection model to the empirical outcomes. If the divergence is due to data reliability, we have to rely on alternative sources, if available. An example for insufficient data reliability is the IPUMS data for France. The IPUMS dataset reveals that *"…'none declared' category encompasses those without schooling/diploma(s) as well as those who did not answer the*

question, depending on the year -- this may explain some variation between years." (Minnesota Population Center 2014)[8]

The share of the total population aged 25 years plus in France in the category *none declared* decreased from 59 percent in the 1962 census to 20 percent in 1999 and 2 percent in 2006. Thereby, the high share of population before 2006 can be explained due to the fact that this category contains both people without schooling/diploma(s) and people who did not declare their education. This may be the result of different approaches in questioning in the different census questionnaires, e.g. 1999 and 2006 (see Figure 4).

Figure 4. Differences in the education question in French censuses, 1999 and 2006 (Minnesota Population Center 2014) [authors illustration]



While the census questionnaires in both years ask for the diploma certificate or degree obtained, the possibilities to answer this question show a different level of graduation. In 1999 it was only possible to answer with *0 No Diploma*, while in 2006 the answer opportunities got notably refined with three choices for those without diploma.

This high shares in the category *none declared* affected our validation result in a way that all validation data points show a high deviation that causes an *F* classification. Therefore, our major source for validating the WIC 2015 dataset was provided by the library of the Institut National D'Etudes Démographiques (INED)[9] and the National Institute of Statistics and Economic Studies (INSEE)[10], which shows a reliable fit with our data (category *A* and *B*).

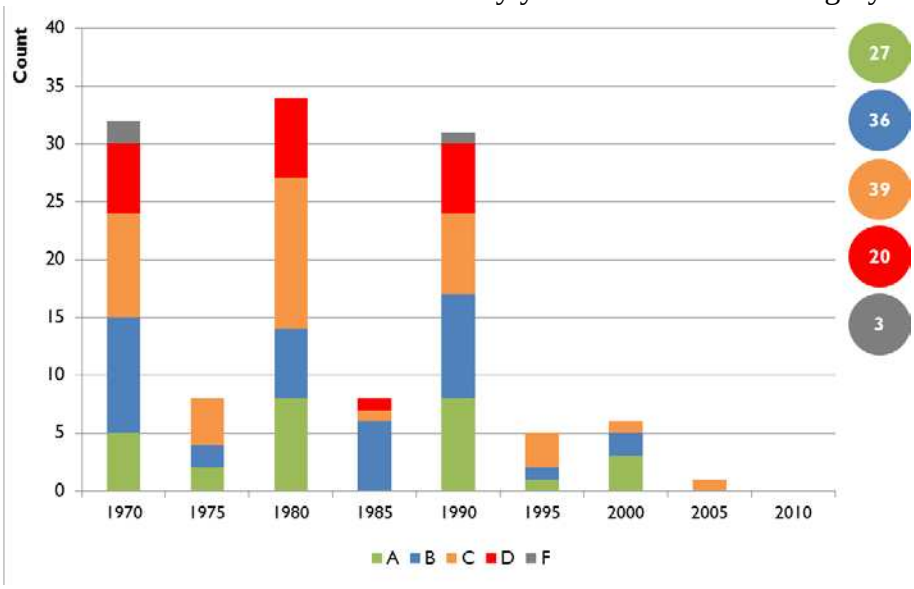
Another issue that can occur in the IPUMS dataset are aggregated educational levels that do not allow for disaggregation and harmonization into WIC 2015 educational categories, such cases are the IPUMS sample for the Netherlands in 1971 or the United Kingdom in 1991 and 2001.

[8] Source: https://international.ipums.org/international-action/variables/EDUCFR#comparability_section (accessed: 01-12-2014)
[9] INED – Institut National D'études Démographiques (http://bibliotheque.web.ined.fr/infos/bibliothequeeng)
[10] INSEE – National Institute of Statistics and Economic Studies (http://www.insee.fr/fr/default.asp)

4.2.2 National Statistical Offices (NSO)

With NSO data, we were able to validate 54 countries and 125 data points out of which almost half exhibit a *good* or *rather good* matching with the WIC 2015 back-projection dataset (see Figure 5).

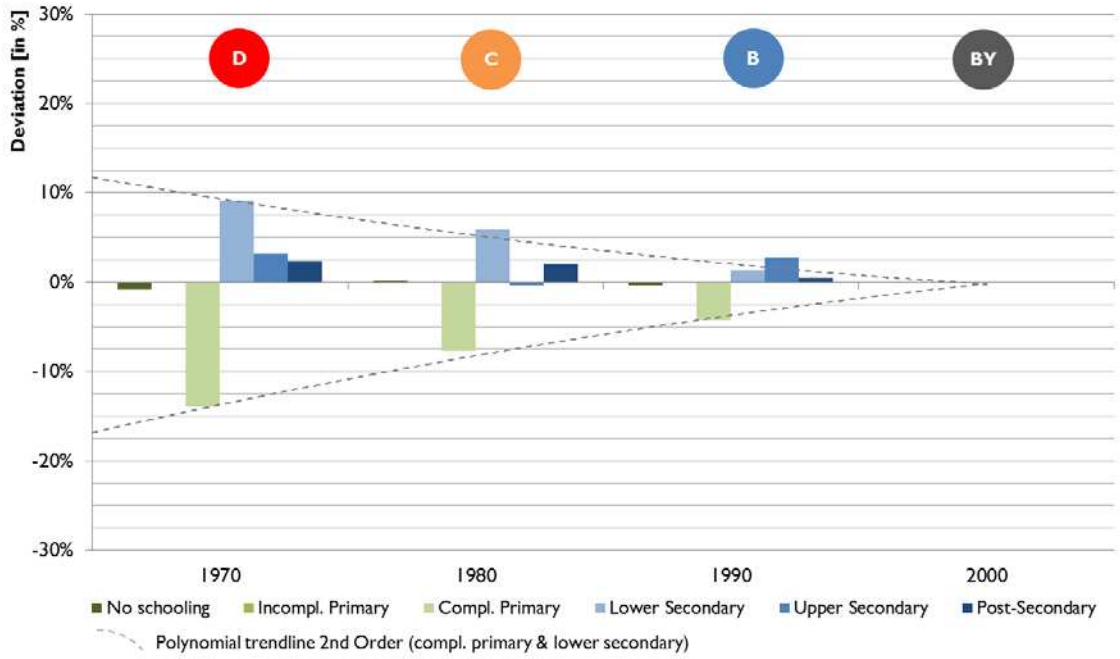
Figure 5. Validation result for NSO data sources by year and validation category



Issues that lead to a deviation of the WIC 2015 dataset from the NSO data are mainly originating from changes in national education systems or educational categories surveyed by the NSOs compared to the applied base-year data, e.g. in the case of Spain. Before 1996 Spain had a 5-3-4 education system corresponding to the number of grades in primary, lower and upper secondary education. In the early 1990s, a reform changed it to a 6-4-2 system. Since we accounted for these changes in the collection and harmonization of historical education dataset (Black and Wilian 2005; Remesal 2007), our reconstruction seems to be valid for Spain. For three other data points, namely Canada 1971, Switzerland 1971, and Latvia 1989, it was not possible to account for the country-specific validation due to issues with data reliability and a clear distinction of national educational categories from the ISCED categories.

The country-specific education transition matrix within the WIC 2015 back-projection model is another source of divergence because it sometimes over- or under-estimates the pace of the transition between educational categories. In the case of Hungary data on educational attainment for the current national borders are available from 1920 onwards by the Hungarian Central Statistical Office (CSO 1992). Additionally, IPUMS provides census microdata from 1970 onwards, which allows us to validate the WIC 2015 data for Hungary with two alternative empirical datasets. Both show a similar pattern of divergence in the educational attainment of the total population of Hungary aged 25 years plus from the base-year 2001 backwards to 1970 (see the case of NSO data in Figure 6) that indicates an overestimation of the transition between *complete primary* and *lower secondary education* in the WIC 2015 back-projection model.

Figure 6. Deviation of the WIC 2015 dataset to NSO data, Hungary, 1970-2000



Note: BY refers to base-year

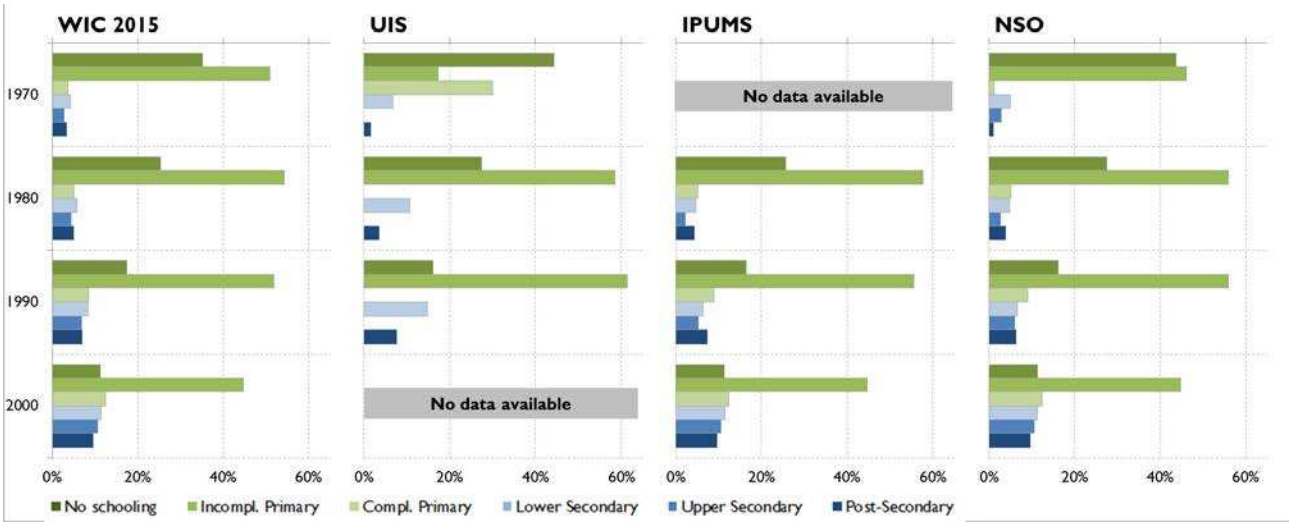
As shown in Figure 6, the WIC 2015 model estimates a share of population aged 25 years plus with completed primary education in 1970 with 40 percent, while the NSO data reveal a share of 52 percent. This divergence in completed primary education is mirrored in the share with lower secondary education, indirectly affecting the shares in upper and post-secondary education.

4.2.3 UNESCO Institute for Statistics (UIS)

UIS has the largest database of data on educational attainment with 257 data points for 127 countries. However the UIS data show the lowest matching accuracy with the WIC 2015 dataset mostly due to the inconsistent coding of the education categories, especially regarding the classification of complete and incomplete education, and the aggregation or overlapping of categories. In general *"…the UIS was not checking the accuracy of the classification into the ISCED categories and consistency across (the) different datasets"* (Potančoková, KC, and Goujon 2014, 18)

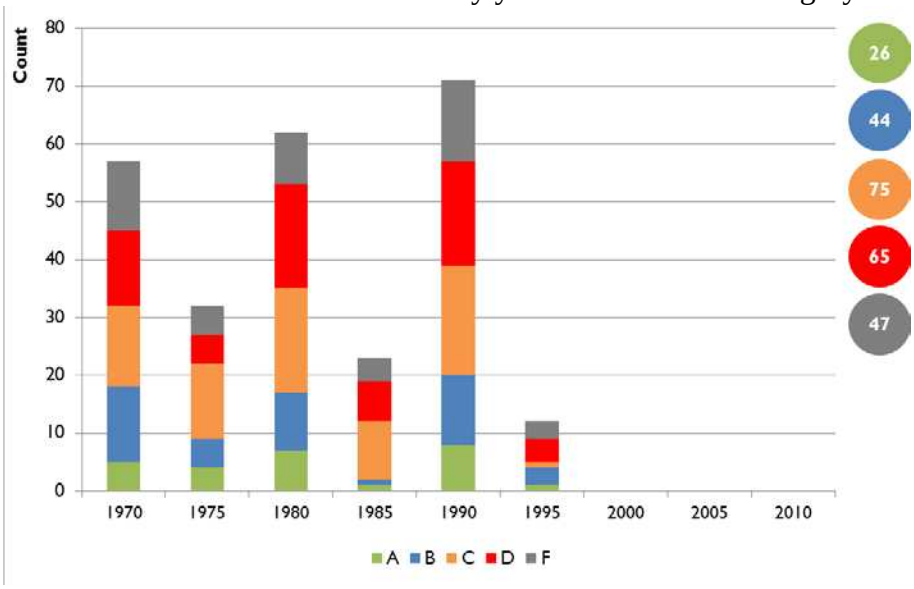
A good example to illustrate these two issues is the case of Portugal where it is possible to compare the four data sources. Figure 7 shows, that the WIC dataset achieves a *good* or *rather good* matching accuracy with empirical data sources i.e. IPUMS and NSO datasets, but not with the UIS. The NSO data (Statistics Portugal 2009a; Statistics Portugal 2009b) conflict with the UIS for instance with the split between incomplete and complete primary education. This is surprising as the UIS dataset represents a compilation of national statistics obtained from the NSO. The absolute deviation of more than 20 percentage points from UIS to other datasets results in the classification of the 1970 data point in the *F* category.

Figure 7. Proportion of educational attainment, different sources, population 25 plus, Portugal 1970 to 2000



The issue with disentangling or aggregation of incomplete and completed education categories is occurring frequently in different variations and intensities in the UIS dataset. Moreover, in the UIS dataset educational categories are very often aggregated into one category, such as incomplete and complete primary education, or lower and upper secondary education, which does not allow for a comparison at the level of the six categories available in the WIC 2015 database. The comparison of the WIC 2015 dataset with the UIS is therefore quite difficult and translates into only 70 datasets (21.7 percent) with a *good* or *rather good* matching accuracy. In 47 cases, the difference is above 20 percentage points (pp) equivalent to category *F*. In such cases it is difficult to check the UIS dataset for correctness as there is no consistent structure or approach to harmonize the national educational categories.

Figure 8. Validation result for UIS data source by year and validation category



4.3 Comparison with other reconstruction exercises

There are only few reconstruction exercises. The most widely used estimates of historical education attainment are the datasets from Barro & Lee (2013 and earlier versions) and De la Fuente & Doménech (2006; 2012). Both datasets rely on UIS and NSO data. Additionally, the WIC 2015 back-projections build upon an earlier reconstruction effort published in Lutz et.al. (2007). In the following sections, we will compare the WIC 2015 back-projections with these three datasets. Further details on the different reconstruction exercises can be found in Appendix II.

4.3.1 Comparison with the Lutz et al (2007) dataset

The main differences between the 2007 Lutz et al. dataset and the WIC 2015 dataset lie in the methodology (see Section 2). Lutz et al. used 2000 as the base-year, and had only four educational categories, namely *no schooling*, *primary*, *secondary*, and *tertiary education* (see Table 3).

Table 3. Categories of educational attainment (Bauer et.al 2012: p.7) [adapted by authors]

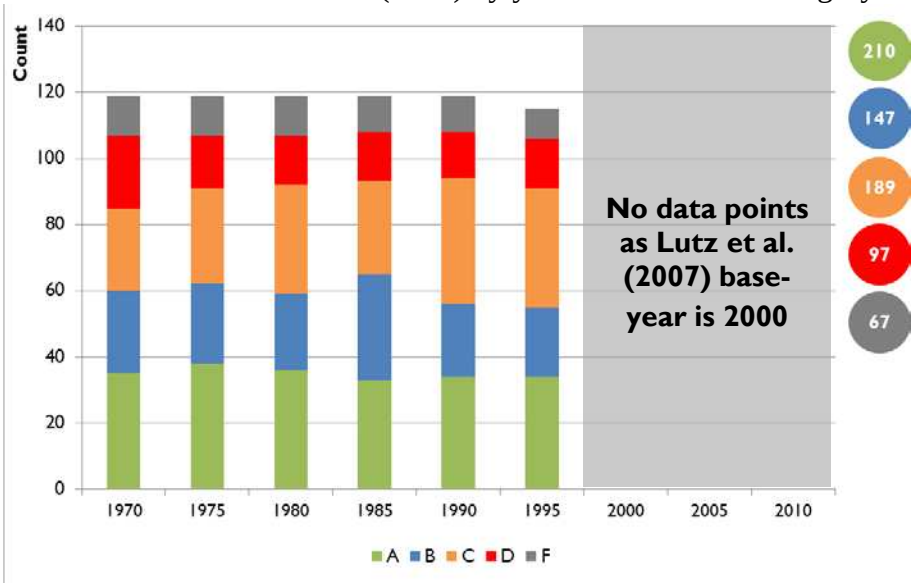
Lutz et.al. (2007)	**ISCED 1997**	**WIC (2015)**
No schooling	No level or ISCED 0 Grade 1 of ISCED 1 not completed	No schooling
Primary	Incomplete ISCED 1	Incomplete primary
	Completed ISCED 1 Incomplete ISCED 2	Completed primary
Secondary	Completed ISCED 2 Incomplete ISCED 3	Completed lower secondary
	Completed ISCED 3 Incomplete ISCED 4	Completed upper secondary
	Completed ISCED 4	Post-secondary
Tertiary	Completed ISCED 5 & 6	

In total it is possible to compare 710 data points for 119 countries[11] from 1970 onwards for the two datasets. From those data points, about 210 (30 percent) are fitting *good* and 147 (21 percent) *rather good*.

The source for deviations in all other cases originates mainly from the different base-year data, data sources and education classification between the two datasets. As shown in Table 3, the 4 categories used in Lutz et al. are not one-to-one comparable to the categories used in the WIC 2015 dataset at the level of secondary and tertiary education. In the 2007 dataset the education category *secondary* includes ISCED 4 (post-secondary non-tertiary) that is in the WIC 2015 dataset part of *post-secondary* education.

[11] The Lutz et al. (2007) dataset includes 120 countries, but Eritrea was disregarded due to data reliability issues for the base-year. (Bauer et al. 2012)

Figure 9. Validation result for Lutz et.al. (2007) by year and validation category



This affects the comparability of the two datasets and makes it necessary to aggregate the ISCED categories 2 to 6 from the original Lutz et al. (2007) dataset to allow a direct lineup of both (see Figure 10 in the case of the Czech Republic).

Figure 10. Educational attainment lineup of Lutz et al. (2007) and WIC 2015, Czech Republic, 1975

Education categories (Lutz et.al. 2007)						Education categories (WIC 2015)	
Lutz et.al. (2007)	/	**ISCED '97**	**original**	/	**aggregate**	/	
No schooling		None	0.7%		0.7%	0.9%	**No schooling**
Primary		ISCED 0	39.1%		39.1%	0.2%	**Incomplete primary**
		ISCED 1				0.8%	**Completed primary**
Secondary		ISCED 2	53.9%		60.2%	37.1%	**Completed lower secondary**
		ISCED 3				51.9%	**Completed upper secondary**
		ISCED 4				9.1%	**Post-secondary**
Tertiary		ISCED 5 & 6	6.3%				

Some other differences result from the allocation between *completed primary* and *completed lower secondary*, which in many cases represents *basic* or *compulsory education*. In Lutz et al. (2007b), when the data could not be disaggregated between the two categories, it was allocated as a rule to the lower education category *primary education.* Altogether, we observe a significant difference between the two datasets in the case of 12 countries[12].

[12] Countries: Bahamas, Czech Republic, Italy, Japan, Jordan, Moldova, Mozambique, Namibia, Nicaragua, Poland, Spain, and Slovakia

4.3.2 Comparison with the Barro and Lee (2013) dataset

The Barro & Lee datasets (1993, 2001, 2010 and 2013) are the most widely used reconstructed datasets on past levels of education. (Barro and Lee 1993; Barro and Lee 2001; Barro and Lee 2010; Barro and Lee 2013) They have addressed the issue of reconstructing time series of the complete age, sex, and levels of educational attainment matrix to the year 2013 by using the *Perpetual Inventory Method (PIM)* to translate enrolment data into educational attainment to fill time gaps in attainment data for the second half of the 20th century.

In their latest revision, Barro & Lee (2013) updated their methodology to one resembling our approach of using the stability of education along cohort lines, and assuming differential mortality by education for the population aged 65 years plus. Thereby the authors estimated survival ratios by distinguishing between two broad educational groups, a less-educated population (uneducated and people who have reached the primary level) and a more-educated population (reached at least secondary schooling), as well as for broad groups of OECD and non-OECD countries.

Their approach is based on the collection of empirical data points, mainly from UIS and national censuses for recent years. They deal with the occurring issue of aggregated and overlapping educational categories by applying decomposition methods on the basis of enrolment data to split up the broader groups of educational attainment into four classes: no formal education, primary, secondary, and tertiary education. By the mean of a completion ratio, they disaggregate these 4 categories into 7.

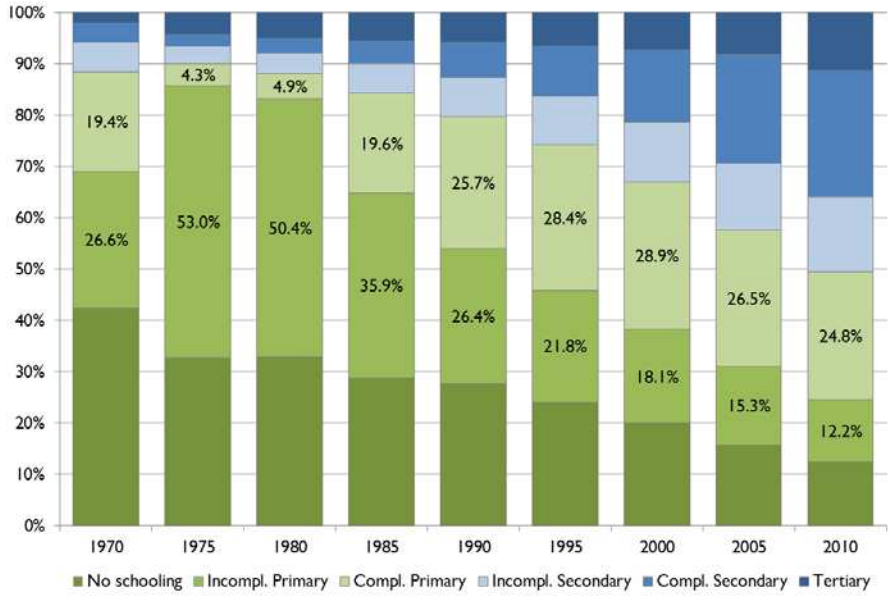
Based on their data points, Barro & Lee (2013) interpolate/extrapolate the population by age and education between/from empirical data points for- and backward. Their dataset includes estimates for 146 countries, including 12 countries[13] that are not listed in the WIC 2015 dataset, mainly for the reason that those datasets are either hardly available from NSOs, the countries have a population beneath 100,000 or the data is of limited reliability.

Beside the usage of unharmonized educational data from UIS, another major drawback of the Barro & Lee approach is the decomposition method applied to separate incomplete from completed education levels that causes several oddities in the time series e.g. for Belize, Bolivia, Brazil, Mozambique, Senegal, South Africa, Swaziland. For instance, in the case of Brazil, a country that provides detailed time series via NSO and IPUMS, the Barro & Lee dataset shows a doubling of *incomplete primary* education from 1970 to 1975 and in the same period a decrease of *complete primary* education from 19 percent to 4 percent. From 1980 to 1985 the share of people aged 25 years plus with completed primary education suddenly quadruples again to almost 20 percent (see Figure 11).

One major drawback of this exercise is that the authors incorporated the data as collected and reported to UIS by national organizations without questioning them. Nevertheless, Barro and Lee show awareness for many problems, so for example for the well-documented tendency of over-reporting in school enrolment data in developing countries, a result of financial advantages for reporting schools or school districts.

[13] Countries: Afghanistan, Barbados, Brunei, Botswana, Fiji, Libya, Mauritania, Papua New Guinea, Sri Lanka, Taiwan, Togo, and Yemen

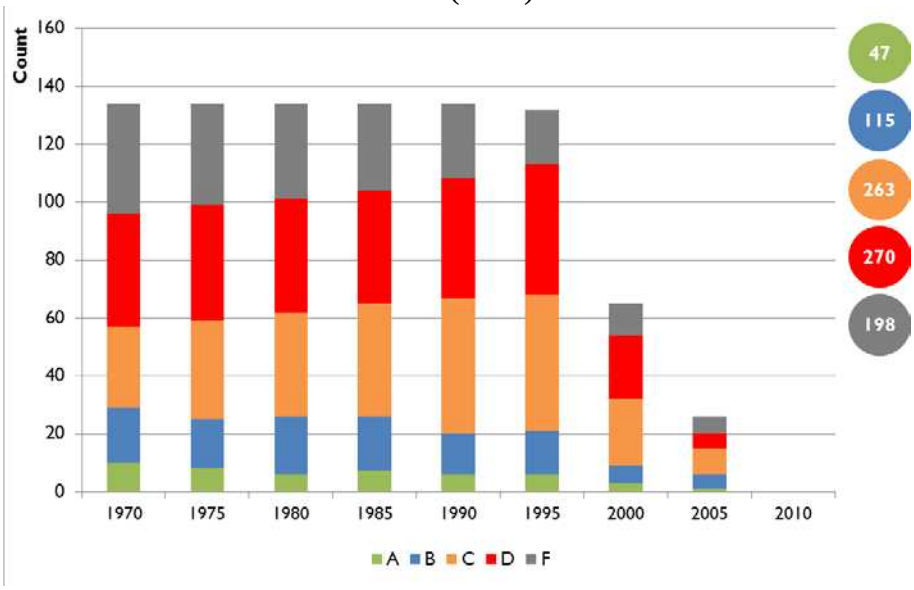
Figure 11. Total population aged 25 years plus by education, Brazil, 1970-2010 (Barro and Lee 2013) [authors illustration]



The example of Brazil is not an isolated case but rather an indication of a more general problem in the Barro & Lee estimates and problem lies in their confidence in the accuracy of the UIS dataset. This becomes especially an issue for countries for which the reconstruction is based on just one or two data points, which is the case for 79 countries out of 146. Additionally, the method to decompose incomplete and completed levels extend the flaws to the source data (Bauer et al. 2012; Potančoková, KC, and Goujon 2014).

These, among other factors, cause the discrepancies between the WIC 2015 and Barro & Lee (2013) datasets. Summarizing, out of the 134 countries and 893 data points available for comparison, only 162 data points or 18 percent show an absolute difference of less than 5 pp (categories *A* [47] and *B* [115]). As shown in Figure 12, the vast majority of the 468 comparable data points deviates by more than 10 pp in one or more educational categories (categories *D* [270] and *F* [198]).

Figure 12. Validation result for Barro and Lee (2013)



4.3.3 Comparison with the De La Fuente and Doménech (2012) dataset

De la Fuente & Doménech (2000; 2012) adapt the methods from Cohen & Soto (2007) and Barro & Lee (2010) to interpolate/extrapolate backward and forward by adding miscellaneous information and their professional judgment to create a smooth time series of educational attainment for 6 education categories[14] for the population 25-years plus in some 21 OECD countries for the period 1960-2010. Thereby, the authors state themselves that:

> *"… the construction of our series involves a fair amount of guesswork. (…) Hence, we have found it preferable to rely on judgment to try to piece together the available information in a coherent manner than to take for granted the accuracy of the primary data."* (de la Fuente and Doménech 2012, 3)

The authors revised and extended their already published data set (de la Fuente and Doménech 2000; de la Fuente and Doménech 2006) in 2012, which will further be named as DF2012.

In general, de la Fuente & Doménech collected data on educational attainment, years of schooling and qualification levels from censuses, surveys (mainly LFS), registers and statistical yearbooks to convert the given data to their educational categories. In the case of missing categories the authors applied the shares of other available data points or proportional thresholds based on their expert opinion.

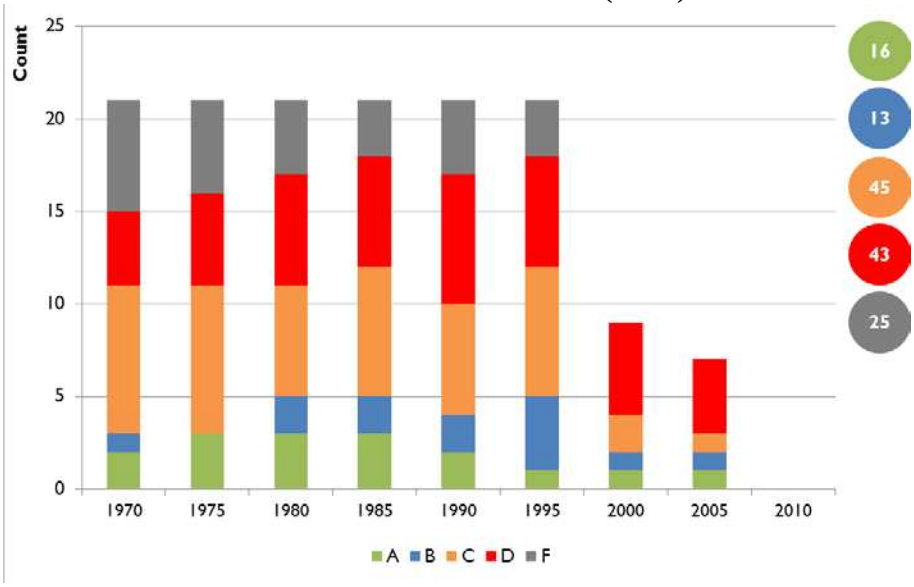
For earlier periods the they used a back-projection method described in Cohen and Soto (2007) that assumes *"… that individual school attainment does not change over time once agents reach the age of 25 (which is probably a rather good approximation), that there are no migration flows (or that migrants have the same educational level as the rest of the population) and that survival probabilities are independent of educational attainment, then the mean educational level of a given 25+ cohort remains constant over time."* (de la Fuente and Doménech 2012, 5f)

These assumptions are the base of their back-projection method, which is different from the WIC 2015 method as they are estimating the education structure of the population aged 25 years plus for an early data point in their time series. The other missing data points are basically resulting from a basic linear interpolation and extrapolation technique to estimate the educational shares for the population 25 years plus.

Despite the similarities in the number and characteristics of the education categories, the WIC 2015 dataset and the DF 2012 dataset hardly match because 68 of the 142 data points fall in the category *D* or *F* (see Figure 13). Since De la Fuente and Doménech (2012) provide an exhaustive documentation on country specific datasets, sources and estimation methods, it enabled a detailed comparison with the WIC 2015 dataset. We compared 142 data points for the 21 countries that are in the WIC 2015 dataset.

[14] Categories: Illiterates, Primary schooling, Lower secondary schooling, Upper secondary schooling, Higher education/first cycle or short post-secondary courses, Higher education/second cycle or full-length courses (de la Fuente and Doménech 2012, 3)

Figure 13. Validation result for de la Fuente & Doménech (2012)



The major reasons for the deviations between the two datasets are based in the processing and harmonization of the available educational data as basis for filling the data gaps. There are several examples where different data sources, like surveys, are used and/or the given educational classifications are not consistently transposed into the DF2012 dataset.

In the case of *Australia* and *New Zealand* de la Fuente & Doménech are using census information on post-school qualifications and data on school leaving ages, both indicators by age and sex, to estimate the educational attainment structure for the available census years. Depending on the indicated age at leaving school and the information about the school duration in the country specific education system, the authors allocate the people to the educational categories primary schooling, lower secondary and upper secondary. With the information on the qualification level of the population by age and sex it is possible to estimate the amount of people with apprenticeships, short vocational training and higher education.

The obvious risk of this approach is to misallocate school repeaters; however this concerns a small share of the population and results in minor error. A much greater issue is the treatment of people in different age groups with *unknown* or *not stated* qualification (up to 37 percent) or year/age at leaving school (up to 11 percent) (ABS 1986).

Another source for deviations in the share of educational attainment between the two datasets is apparently different education harmonization approaches. Thereby DF2012 does not always use the ISCED classification, like in the case of the *Netherlands* (see Figure 14). This can lead to a mismatch in the allocation of educational categories into the DF2012 categories. For instance for 2001 DF2012 uses the LFS 2001 for the Netherlands and the national educational categories, namely the SOI categories, that compile different ISCED categories that belong to upper secondary and post-secondary groups. The SOI classification does not allow a clear distinction between single ISCED categories (Schaart, Bernelot Moens, and Westerman 2008) and is therefore hardly comparable to the harmonized WIC 2015 categories (Bauer et al. 2012).

Figure 14. The population by level of educational achievement in the Netherlands 2001 (DF2012 vs WIC2015) [authors illustration]

Education categories (LFS 2001)		SOI	ISCED '97	DF2012		e1	(e2)	e3	e4	e5	e6	unk	cat
Primary education		SOI 1 & 2	ISCED 0 & 1	L1	11.4%	4.0%		13.8%					e1 e3
Vmbo, mbo1, avo onderbouw		SOI 3	ISCED 2 & 3C	L2.1	24.2%				23.2%				e4
Havo, vwo, mbo	*Mbo 2 en 3*	SOI 4.1 & 4.2	ISCED 3A & 3C	L2.2	40.4%					35.7%	23.4%		e5
	Mbo 4	SOI 4.3	ISCED 3A, 4B, 4C, 5A										
	Havo, vwo	SOI 4.2 & 4.3	ISCED 3A, 3C, 4B, 4C, 5A										
Hbo, wo bachelor		SOI 5	ISCED 5A & 5B	L3	23.5%								e6
WO masters, doctor		SOI 6 & 7	ISCED 5A & 6										
Unknown level of education		Unknown		unk	0.5%							-	unk

Note: WIC2015 - (e1) no education, (e1) incomplete primary education, (e2) completed primary education, (e3) lower secondary education, (e5) upper secondary education, (e6) post-secondary education, (unk) unknown | DF2012 - (L0) Illiterates, (L1) Primary schooling, (L2.1) Lower secondary schooling, (L2.2) Upper secondary schooling, (L3.1) Higher education, first cycle or short post-secondary courses, (L3.2) Higher education, second cycle or full-length courses (Source: de la Fuente and Doménech 2012, 3)

Apart from the different approaches in compiling, harmonizing and processing of the empirical data sets used, the differences get extended due to the use of a linear interpolation method in the DF2012 dataset to estimate missing data points and the smoothing of the time series due to country-specific correction factors.

5 Conclusion

The measurement of educational attainment on a globally comparable scale has always been a problem due to internationally inconsistent classification and diverse national education systems. Despite isolated attempts to standardize levels of educational attainment e.g. ISCED 1997, the discrepancies brought by differences in categorization across countries and times have persisted, particularly in earlier years. The WIC 2015 back-projection exercise, as other reconstruction works, attempts at overcoming those issues and creating consistent time-series of educational attainment by age and sex. All problems have not been surmounted, but the validation shows that our effort certainly addressed the main issues and adopts clear and systematic measures to overcome them.

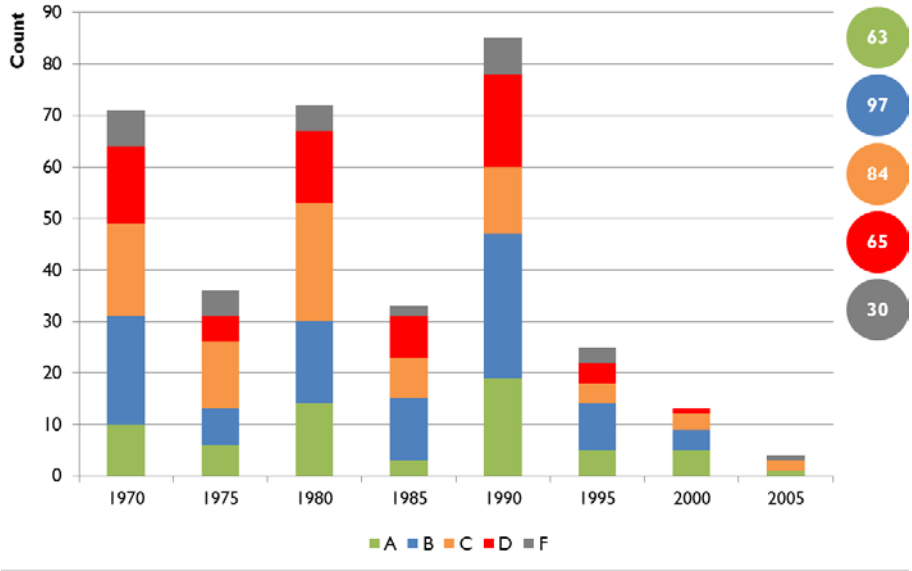
These measures unite a comprehensive approach to harmonize historical and base-year data in terms of educational attainment, and a methodology to reconstruct the educational attainment for 171 countries and validate and evaluate the outcome with empirical data. This approach makes this dataset unique and hardly comparable with other approaches. By validating the WIC 2015 data series with globally collected and harmonized empirical data we can show the accuracy but also the insufficiencies of this dataset.

This paper contains the validation of the WIC 2015 dataset on the estimated educational composition by age and sex for 171 countries from 1970 up to the country-specific base-year with 339 empirical historical datasets (excluding duplicates from other sources)[15] for 138 countries (81 percent of all 171 countries). This corresponds to a coverage

[15] In total it was possible to collect and harmonize 519 data points. After excluding duplicates, which could occur due the availability of educational data for one country in a certain period from different data sources, we could identify 339 empirical data points with high data reliability for the validation of the WIC 2015 dataset.

of 30 percent of the overall potential 1148 data points in the period 1970 up to the base-year. In total, about 160 data points or 47 percent show a *good* or *rather good* fitting accuracy with empirical data, while with 30 data points about 9 percent show a very high deviation and were therefore classified in category *F* (see Figure 15).

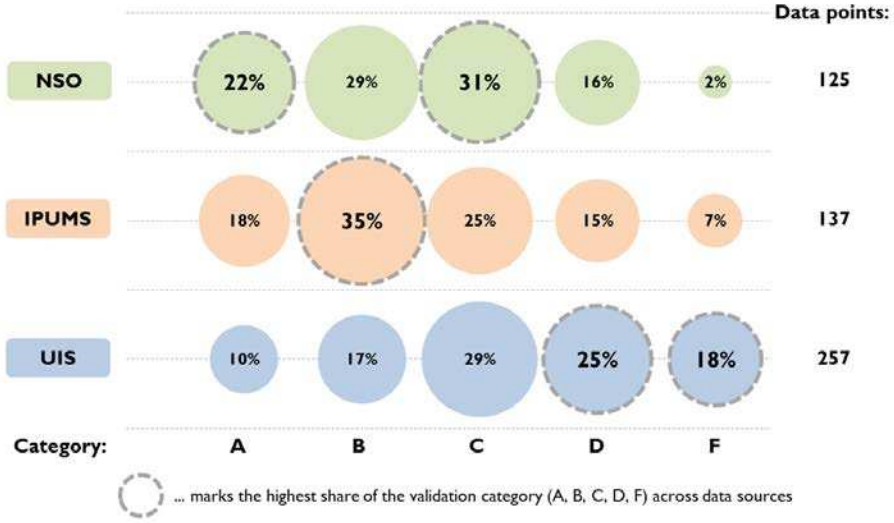
Figure 15. Validation Result for all data sources by year and validation category



The fitting accuracy of the WIC 2015 dataset with empirical datasets is thereby highly influenced by the data origin. While for NSO and IPUMS data the concordance to the WIC 2015 is with respectively 50 and 53 percent of the data points in the categories *good* (*A*) or *rather good* (*B*) is very high, the UIS data shows a lower accuracy of about 27 percent in those categories.

At the same time just about 2 or 7 percent of NSO or IPUMS data had to be classified as category *F*, which makes up a very small proportion compared to UIS data, where about 18 percent had to be assigned to this category (see Figure 16). Again, this highlights the unsatisfactory data quality of the UIS dataset and shows that it should be used with caution.

Figure 16. Validation Result by data sources and proportion data points by validation category



In general it was possible to achieve a very high matching accuracy of the WIC 2015 dataset with empirical datasets from NSO, IPUMS and UIS, which brings us one step closer to the harmonization of levels of educational attainment of the global population. What remains to be done is to enhance the data collection and classification efforts, especially beyond the 1970s to draw a picture of the global educational development for the 20th century in order to fill the gaps in the availability of data.

Education is a key indicator for appraising the level of socio-economic development of the population in a country. In turn, its measurement can indicate economic capabilities and adaptability of societies for instance to climate change related disasters. Therefore, the creation of a comprehensive harmonized dataset on levels of educational attainment by age and sex can have an important additional value either for policy-makers, scientists and therefore for the wider public. In this study we did not further decompose the reasons for the discrepancies between the reconstructed data and other sources of valid data, which could have been due to irregular education-specific migration or mortality patterns of unusual patterns of age-specific education progressions. This will be the topic of a subsequent study.

At the time of finishing this report, the WIC2015 dataset will be available online in the Wittgenstein Data Explorer[16]. We plan to regularly update the historical dataset and the online WIC 2015 back-projection database.

6 References

ABS. 1986. Census of Population and Housing, 30 June 1986. CENSUS 86 — Cross-Classified Characteristics of Persons and Dwellings. Australia. 2498.0. Canberra: Australian Bureau of Statistics.

Barro, Robert J., and Jong Wha Lee. 1993. “International Comparison of Educational Attainment.” Journal of Monetary Economics 32 (3): 363–94.

———. 2001. “International Data on Educational Attainment: Updates and Implications.” Oxford Economic Papers 53 (3): 541–63. doi:10.1093/oep/53.3.541.

———. 2010. A New Data Set of Educational Attainment in the World, 1950-2010. NBER Working Paper No.15902. Cambridge, Massachusetts: National Bureau of Economic Research. http://www.nber.org/papers/w15902.

———. 2013. “A New Data Set of Educational Attainment in the World, 1950–2010.” Journal of Development Economics 104 (September): 184–98. doi:10.1016/j.jdeveco.2012.10.001.

Bauer, Ramon, Michaela Potančoková, Anne Goujon, and Samir KC. 2012. Populations for 171 Countries by Age, Sex, and Level of Education around 2010: Harmonized Estimates of the Baseline Data for the Wittgenstein Centre Projections. Interim Report IR-12-016. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://www.iiasa.ac.at/publication/more_IR-12-016.php.

Black, Paul, and Dylan Wilian. 2005. “Lessons from around the World: How Policies, Politics and Cultures Constrain and Afford Assessment Practices.,” The Curriculum Journal, 16 (2): p.249–61.

CBS Norway. 1986. Population and Housing Census 1980. Volume IV Main Results of the Censuses 1960, 1970 and 1980. Oslo, Kongsvinger: Statistik Sentralbyra - Central Bureau of Statistics of Norway.

———. 1991. Folke- og boligtelling 1990. Forelopige hovedtall. Oslo, Kongsvinger: Statistik Sentralbyra - Central Bureau of Statistics of Norway.

———. 1999. Populatoin and Housing Census 1990. Documentation and Main Figures. Oslo, Kongsvinger: Statistik Sentralbyra - Central Bureau of Statistics of Norway.

[16] Wittgenstein Data Explorer (http://www.wittgensteincentre.org/dataexplorer)

CELADE/CEPAL. 2014. “Redata Informa. Software para procesar y mapear datose censos y encuestas para analisis local y regional.” Redatam Informa. http://www.cepal.org/cgi-bin/getprod.asp?xml=/redatam/noticias/paginas/8/14188/P14188.xml&xsl=/redatam/tpl/p18f.xsl&base=/redatam/tpl-i/top-bottom.xsl.

Cohen, Daniel, and Laura Leker. 2014. “Health and Education: Another Look with the Proper Data.” Paris. http://www.parisschoolofeconomics.eu/docs/cohen-daniel/cohen-leker-health-and-education-2014.pdf.

Cohen, Daniel, and Marcelo Soto. 2001. Growth and Human Capital: Good Data, Good Results. OECD Development Centre Technical Papers 179. Paris: Organisation for Economic Co-operation and Development.

———. 2007. “Growth and Human Capital: Good Data, Good Results.” Journal of Economic Growth 12 (1): 51–76. doi:10.1007/s10887-007-9011-5.

CSO. 1992. Time Series of Historical Statistics 1867-1992, Volume 1, Population - Vital Statistics. Vol.1. Budapest, Hungary: Hungarian Central Statistical Office.

CZSO. 1980. 1980 Population and Housing Census – Czech Socialist Republic. Prague: Czech Statistical Office (CZSO).

———. 1991. 1991 Population and Housing Census. Prague: Czech Statistical Office (CZSO).

De la Fuente, Angel, and Rafael Doménech. 2000. Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?. OECD Economics Department Working Papers 262. Paris: Organisation for Economic Co-operation and Development.

———. 2006. “Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?” Journal of the European Economic Association 4 (1): 1–36. doi:10.1162/jeea.2006.4.1.1.

———. 2012. Educational Attainment in the OECD, 1960-2010.

EUROSTAT. 2014. “EUROSTAT. Your Key to European Statistics. Database.” EUROSTAT Database. http://ec.europa.eu/eurostat/data/database.

Huisman, Martijn, Anton E. Kunst, Matthias Bopp, Jens-Kristian Borgan, Carme Borrell, Giuseppe Costa, Patrick Deboosere, et al. 2005. “Educational Inequalities in Cause-Specific Mortality in Middle-Aged and Older Men and Women in Eight Western European Populations.” Lancet 365 (9458): 493–500. doi:10.1016/S0140-6736(05)17867-2.

Hummer, Robert A., and Joseph T. Lariscy. 2011. “Educational Attainment and Adult Mortality.” In International Handbook of Adult Mortality, Vol. 2:pp.241–61. International Handbooks of Population. Rotterdam, Netherlands: Springer Netherlands.

KC, Samir, Bilal Barakat, Anne Goujon, Vegard Skirbekk, and Wolfgang Lutz. 2008. Projection of Populations by Level of Educational Attainment, Age and Sex for 120 Countries for 2005– 2050. Interim Report IR-08-038. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://www.demographic-research.org/volumes/vol22/15/22-15.pdf.

KC, Samir, Bilal Barakat, Anne Goujon, Vegard Skirbekk, Warren C. Sanderson, and Wolfgang Lutz. 2010. “Projection of Populations by Level of Educational Attainment, Age, and Sex for 120 Countries for 2005-2050.” Demographic Research 22 (Article 15): 383–472. doi:10.4054/DemRes.2010.22.15.

K.C., Samir, Erich Striessnig, Bilal Barakat, and Markus Speringer. 2015 (forthcoming). Wittgenstein Centre Back-Projections Methodology for Populations by Age, Sex, and Six Levels of Education. Interim Report IR-15-xxx. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Lutz, Wolfgang, William P. Butz, and Samir KC, eds. 2014. World Population and Human Capital in the 21st Century. Oxford University Press. http://ukcatalogue.oup.com/product/9780198703167.do.
Lutz, Wolfgang, Anne Goujon, Samir KC, and Warren C. Sanderson. 2007a. "Reconstruction of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 1970-2000." Vienna Yearbook of Population Research 2007, 193–235.
———. 2007b. Reconstruction of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 1970-2000. Interim Report IR-07-002. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Measure DHS. 2015. "The DHS Program. Demographic and Health Surveys." The DHS Program. www.measuredhs.com.
Minnesota Population Center. 2014. "Integrated Public Use Microdata Series, International: Version 6.3 [Machine-Readable Database]." IPUMS International. https://international.ipums.org/international/index.shtml.
Morrisson, Christian, and Fabrice Murtin. 2009. "The Century of Education." Journal of Human Capital 3 (1): 1–42.
Potančoková, Michaela, Samir KC, and Anne Goujon. 2014. Global Estimates of Mean Years of Schooling: A New Methodology. Interim Report IR-14-005. Laxenburg, Austria: International Institute for Applied Systems Analysis (IIASA). http://www.iiasa.ac.at/publication/more_IR-14-005.php.
Remesal, Ana. 2007. "Educational Reform and Primary and Secondary Teachers' Conceptions of Assessment. The Spanish Instance, Building upon Black & Wiliam (2005)," The Curriculum Journal, 18 (1): p.27–38. doi:10.1080/09585170701292133.
Riosmena, Fernando, Isolde Prommer, Anne Goujon, Samir KC, and Wolfgang Lutz. 2008. An Evaluation of the IIASA/VID Education-Specific Back Projections. Interim Report IR-08-019. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Schaart, Roel, Mies Bernelot Moens, and Sue Westerman. 2008. The Dutch Standard Classification of Education, SOI 2006. Voorburg, Netherlands: Statistics Netherlands. http://www.cbs.nl/NR/rdonlyres/2D265545-433C-484B-849C-5BBCCAAF6229/0/2008thedutchstandardclassificationofeducationsoiart.pdf.
Statistics Portugal. 2009a. 50 Years of Education Statistics, Volume 1. Vol. 1. Lisbon, Portugal. http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=82895775&PUBLICACOESmodo=2.
———. 2009b. 50 Years of Education Statistics, Volume 2. Vol. 2. Lisbon, Portugal. http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=82895775&PUBLICACOESmodo=2.
UIS. 2014a. "ISCED 1997 Mappings." http://www.uis.unesco.org/Education/ISCEDMappings/Pages/default.aspx.
———. 2014b. "UN Data. A World of Information. Population 15 Years of Age and Over, by Educational Attainment, Age and Sex." Data.un.org. http://data.un.org/Data.aspx?d=POP&f=tableCode%3a30.
———. 2014c. "UNESCO Institute for Statistics. Data Centre." UIS Data Centre. http://www.uis.unesco.org/datacentre/pages/default.aspx?SPSLanguage=EN.
UNESCO. 2006. International Standard Classification of Education: ISCED 1997 (Reprint). Montreal, Canada: UNESCO Institute for Statistics. http://www.uis.unesco.org/Library/Documents/isced97-en.pdf.
UNESCO Institute for Statistics. 2011. "ISCED: International Standard Classification of Education." http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx.

United Nations. 2011. World Population Prospects: The 2010 Revision. New York: Department of Economic and Social Affairs, Population Division.
WIC. 2015. “Wittgenstein Centre Data Explorer Version 1.2.” www.wittgensteincentre.org/dataexplorer.

7 Appendix I - Country Data Documentation

7.1 Appendix Ia – Country-Specific Documentation by Year

Table 4. Availability of educational attainment from empirical census/survey data by country and year

IPUMS NSO UIS Others

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
						Africa						
Angola	(A)	-										
Burundi	(B, D)	1					***1990***					
	(B, D)	1									***2010***	Others: DHS
Benin	(B, C, D)	2			***1979***		***1992***					
	(B, C, D)	1								***2006***		Others: DHS
Burkina Faso	--	2				***1985***		***1996***				
	--	1								***2006***		
Botswana	(A)	-										
Central African Republic	(B, C, D)	2		***1975***			***1988***					
	(B, C, D)	1						***1995***				Others: DHS
Cote d'Ivoire	(B, C, D)	1					***1988***					
	(B, C, D)	1								***2005***		Others: DHS
Cameroon	(B, D)	2		***1976***		***1987***						
	(B, D)	1		***1976***								
	(B, D)	1								***2004***		Others: DHS
Congo DR	(B, C, D)	1				***1984***						
	(B, C, D)	1								***2007***		Others: DHS
Congo	(B, C, D)	1				***1984***						

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
	(B, C, D)	1								***2005***		Others: DHS
Comoros	(G)	1						***1996***				Others: DHS
Cape Verde	(G)	1							***2000***			
Algeria	(B, D)	1				***1987***						
	(B, D)	1	***1971***									
	(B, D)	1							***2002***			Others: PAPFAM
Egypt	(G)	1								***2006***		
Ethiopia	(G)	1									***2011***	Others: DHS
Gabon	(G)	1							***2000***			Others: DHS
Ghana	(B, C, D)	1							***2000***			
	(B, C, D)	1	***1970***									
Guinea	(F)	2				***1983***		***1996***				
Gambia	(B, C, D)	1		***1973***								
	(B, C, D)	1							***2000***			Others: UNICEF
Guinea-Bissau	(G)	1							***2000***			Others: UNICEF
Equatorial Guinea	(G)	1							***2000***			Others: UNICEF
Kenya	(F)	4	***1969***		***1979***		***1989***		***1999***			
	(F)	2	***1969***		***1979***							
Liberia	(B, C, D, F)	1		***1974***								
	(B, C, D, F)	1								***2007***		Others: DHS
Libya	(A)	-										
Lesotho	(B, F)	1		***1976***								
	(B, F)	1									***2009***	Others: DHS
Morocco	--	2						***1994***		***2004***		
	--	1	***1971***									
Madagascar	(G)	1									***2008***	Others: DHS

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Mali	--	2				***1987***			***1998***			
	--	1		***1976***								
Mozambique	(B, C, D)	1								***2007***		
	(B, C, D)	1			***1980***							
Mauritania	(A)	-										
Mauritius	(B, D)	1							***2000***			
	(B, D)	3	***1972***			***1983***	***1990***					
Malawi	(B, D)	2				***1987***					***2008***	
	(B, D)	2		***1977***		***1987***						
Namibia	(B, C, D)	1					***1991***					
	(B, C, D)	1								***2007***		Others: DHS
Niger	(B, C, D)	1		***1977***								
	(B, C, D)	1								***2006***		Others: DHS
Nigeria	(G)	1									***2008***	Others: DHS
Reunion	(G)	1									***2008***	
Rwanda	(B, F)	2					***1991***		***2002***			
	(B, F)	1			***1978***							
Sudan	(B, F)	1									***2008***	
	(B, F)	1				***1983***						
Senegal	--	2					***1988***		***2002***			
Sierra Leone	(G)	1								***2004***		
Somalia	(G)	1								***2006***		Others: UNICEF
Sao Tome & Principe	(B, C, D)	1			***1981***							
	(B, C, D)	1									***2009***	Others: DHS
Swaziland	(B, D)	2		***1976***		***1986***						
	(B, D)	1								***2006***		Others: DHS
Chad	(G)	1								***2004***		Others: DHS

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Togo	(A)	-										
Tunisia	(B, D)	1									***2010***	
	(B, D)	3		***1975***	***1980***	***1984***						
Tanzania	(B, F)	2					***1988***		***2002***			
	(B, F)	1					***1988***					
Uganda	(D)	2					***1991***		***2002***			
	(D)	1					***1991***					
South Africa	(B, C, D)	2						***1996***	***2001***			
	(B, C, D)	1								***2007***		
	(B, C, D)	3	***1970***		***1980***	***1985***						
Zambia	(B, C, D, F)	3	***1969***		***1980***		***1990***					
	(B, C, D, F)	1							***2002***			Others: DHS
Zimbabwe	--	1					***1992***					
	--	1								***2005***		Others: DHS
						Asia						
Afghanistan	(A)	-										
United Arab Emirates	(B, C, D)	1								***2005***		
	(B, C, D)	1		***1975***								
Armenia	(G)	1							***2001***			
Azerbaijan	(G)	1								***2006***		Others: DHS
Bangladesh	(E, F)	2					***1991***		***2001***			
	(E, F)	2		***1974***	***1981***							
	(E, F)	1								***2004***		Others: DHS
Bahrain	(B, F)	1							***2001***			
	(B, F)	2	***1971***				***1991***					

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Brunei	(A)	-										
Bhutan	(G)	1								*2005*		
China	(B, D)	2			*1982*		*1990*					
	(B, D)	2			*1982*		*1990*					
	(B, D)	1								*2005*		Others: NPS
Cyprus	(G)	1					*1992*					
	(G)	1							*2001*			Others: EUROSTAT
Georgia	(G)	1							*2002*			
Hong Kong	(B, D)	2							*2001*	*2006*		
	(B, D)	6	*1971*	*1976*	*1981*	*1986*	*1991*	*1996*				
Indonesia	(D)	8	*1971*	*1976*	*1980*	*1985*	*1990*	*1995*	*2000*	*2005*		
	(D)	1									*2010*	
	(D)	3	*1971*		*1980*		*1990*					
India	(D)	3			*1983*	*1987*		*1993*				
	(D)	2	*1971*						*2001*			
	(D)	2	*1971*		*1981*							
Iran	(G)	1								*2006*		
Iraq	(G)	1						*1997*				
Israel	(B, D, E)	3	*1972*			*1983*		*1995*				
	(B, D, E)	4	*1972*		*1982*	*1983*				*2004*		
Jordan	(G)	1								*2004*		
Japan	(F)	4	*1970*		*1980*		*1990*				*2010*	
	(F)	3	*1970*		*1980*		*1990*					
Kazakhstan	(G)	1									*2009*	
	(G)	1					*1989*					
Kyrgyzstan	(G)	1							*1999*			
Cambodia	(G)	1									*2008*	

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
South Korea	(F)	7	*1970*	*1975*	*1980*	*1985*	*1990*	*1995*			*2010*	
	(F)	6	*1970*	*1975*	*1980*	*1985*	*1990*	*1995*				
Kuwait	(B, D)	1								*2005*		
	(B, D)	5	*1970*	*1975*	*1980*	*1985*	*1988*					
Laos	(G)	1								*2005*		Others: UNFPA
Lebanon	(B, C, D, F)	1								*2007*		
	(B, C, D, F)	1	*1970*									
Sri Lanka	(A)	-										
Macau	(B, C, D)	1								*2006*		
	(B, C, D)	2	*1970*				*1991*					
Maldives	(G)	1								*2006*		
	(G)	1					*1990*					
Myanmar	(B, C, D)	2		*1973*		*1983*						
	(B, C, D)	1								*2007*		Others: UNFPA
Mongolia	(F)	2					*1989*		*2000*			
Malaysia	(F)	3	*1970*		*1980*		*1991*					
	(F)	4	*1970*		*1980*		*1991*		*2000*			
	(F)	3			*1980*		*1991*	*1996*				
Nepal	(F)	1							*2001*			
	(F)	2	*1971*		*1981*							
Oman	(A)	-										
Pakistan	(D)	3	*1973*		*1981*				*1998*			
	(D)	3	*1972*		*1981*		*1990*					
Philippines	(B, D)	3					*1990*	*1995*	*2000*			
	(B, D)	4	*1970*	*1975*	*1980*		*1990*					
Palestine	(G)	1								*2007*		

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Qatar	(B, C, D)	1									***2010***	
	(B, C, D)	1				***1986***						
Saudi Arabia	(G)	1								***2004***		
Singapore	(B, D, F)	1									***2010***	
	(B, D, F)	3			***1980***		***1990***	***1995***				
Syria	(D)	4	***1970***		***1981***			***1994***		***2004***		
	(D)	1	***1970***									
Thailand	--	3	***1970***		***1980***		***1990***					
	--	1							***2000***			
	--	2	***1970***		***1980***							
Tajikistan	(B, C, D)	1					***1989***					
	(B, C, D)	1									***2009***	Others: WB
Turkmenistan	(G)	1						***1995***				
Timor-Leste	(G)	1									***2009***	Others: DHS
Turkey	(B)	2				***1985***	***1990***					
	(B)	4		***1975***	***1980***	***1985***		***1993***				
	(B)	1							***2000***			Others: EUROSTAT
Uzbekistan	(A)	-										
Viet Nam	(B, C, D)	3					***1989***		***1999***		***2009***	
	(B, C, D)	2			***1979***		***1989***					
						Europe						
Albania	(G)	1							***2002***			Others: WORLD BANK
Austria	(B, D)	4	***1971***		***1981***		***1991***		***2001***			
	(B, D)	5	***1971***		***1981***		***1991***		***2001***		***2008***	
	(B, D)	3	***1971***		***1981***		***1991***					

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Belgium	(B, C, D)	3			*1981*		*1991*		*2001*			
	(B, C, D)	1	*1970*									
Bulgaria	(B, D)	1							*2001*			
	(B, D)	2		*1975*			*1992*					
Bosnia-Herzegovina	(D)	2	*1971*								*2010*	
Belarus	(B, C, D)	1							*1999*			
	(B, C, D)	1					*1989*					
Switzerland	(B)	3	*1970*		*1980*		*1990*					
	(B)	4	*1970*		*1980*		*1990*		*2000*			
	(B)	2	*1970*		*1980*							
Czech Republic	(F)	2			*1980*		*1991*					
	(F)	1					*1991*					
	(F)	1							*2001*			Others: EUROSTAT
Germany	(B, D)	1	*1971*									
	(B, D)	1									*2010*	
Denmark	--	1					*1991*					
	--	1							*2001*			Others: EUROSTAT
Spain	(F)	2			*1981*		*1991*					
	(F)	3	*1970*		*1981*		*1991*					
	(F)	4	*1970*		*1981*	*1986*	*1991*					
	(F)	1							*2001*			Others: EUROSTAT
Estonia	(B, D)	4	*1970*		*1979*		*1989*		*2000*			
	(B, D)	1					*1989*					
Finland	(B)	7	*1970*	*1975*	*1980*	*1985*	*1990*	*1995*			*2009*	
	(B)	2				*1985*	*1990*					
France	--	6	*1968*	*1975*	*1982*		*1990*		*1999*	*2006*		
	--	6	*1968*	*1975*	*1982*		*1990*		*1999*		*2008*	

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
	--	1					***1990***					
United Kingdom	(G)	1							***2001***			
Greece	--	3	***1971***		***1981***		***1991***					
	--	2			***1981***		***1991***					
	--	1							***2001***			Others: EUROSTAT
Croatia	(B, D)	2	***1971***						***2001***			
	(B, D)	2					***1991***	***1997***				
Hungary	(F)	4	***1970***		***1980***		***1990***		***2001***			
	(F)	3	***1970***		***1980***		***1990***					
	(F)	3	***1970***		***1980***		***1990***					
Ireland	(B, D)	4	***1971***		***1979***		***1991***	***1996***				
	(B, D)	4	***1971***		***1981***		***1991***	***1996***				
	(B, D)	3	***1971***		***1981***		***1991***					
	(B, D)	1							***2002***			Others: EUROSTAT
Iceland	(G)	1									***2010***	
Italy	--	3	***1971***		***1981***		***1991***					
	--	2	***1971***		***1981***							
	--	1							***2001***			Others: EUROSTAT
Liechtenstein	(A)	-										
Lithuania	(B, D)	1							***2001***			
	(B, D)	1					***1989***					
Luxembourg	(B, D)	1					***1991***					
	(B, D)	1							***2001***			Others: EUROSTAT
Latvia	(B, C, D)	1					***1988***					
	(B, C, D)	1							***2000***			Others: EUROSTAT
Moldova	(B, D, F)	1								***2004***		
	(B, D, F)	1					***1989***					

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Macedonia	(B, D, E, F)	1									*2008*	
	(B, D, E, F)	1						*1994*				
Malta	(G)	1									*2010*	
Montenegro	(G)	2	*1971*							*2003*		
Netherlands	(B)	1	*1971*									
	(B)	1	*1971*									
	(B)	1							*2001*			Others: EUROSTAT
Norway	(F)	5	*1970*		*1980*		*1990*		*2000*		*2010*	
	(F)	4	*1970*	*1975*	*1980*		*1990*					
Poland	(B, D)	3	*1970*		*1978*		*1988*					
	(B, D)	1							*2002*			Others: EUROSTAT
Portugal	(F)	3			*1981*		*1991*		*2001*			
	(F)	3	*1970*		*1981*		*1991*					
Romania	(F)	2		*1977*			*1992*					
	(F)	2		*1977*			*1992*					
	(F)	1							*2002*			Others: EUROSTAT
Russia	(B, D)	1							*2002*			
	(B, D)	1					*1989*					
Serbia	(G)	2	*1971*						*2002*			
Slovakia	(B, D)	3	*1970*		*1980*				*2001*			
	(B, D)	1					*1991*					
Slovenia	(B, D)	3	*1971*		*1981*				*2002*			
	(B, D)	1					*1991*					
Sweden	(F)	6				*1985*	*1991*	*1995*	*2001*	*2005*	*2010*	
Ukraine	(G)	1							*2001*			

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Latin America and the Caribbean												
Aruba	(B, C, D)	1									***2010***	
	(B, C, D)	1					***1991***					
Netherlands Antilles	(G)	1							***2001***			
	(G)	6	***1971***	***1975***	***1981***	***1987***	***1988***	***1995***				
Argentina	(B, D)	3	***1970***		***1980***		***1991***					
	(B, D)	3	***1970***		***1980***		***1991***					
	(B, D)	1							***2001***			Others: CELADE
Bahamas	(B, C, D, E)	1					***1990***					
	(B, C, D, E)	1							***2000***			Others: CARICOM
Belize	(B, D, E, F)	3	***1970***		***1980***		***1991***					
	(B, D, E, F)	1							***2000***			Others: CELADE
Bolivia	(F)	3		***1976***			***1992***		***2001***			
	(F)	2		***1976***			***1992***					
Brazil	(B, D)	4	***1970***		***1980***		***1991***		***2000***			
	(B, D)	1									***2010***	
	(B, D)	3	***1970***	***1976***	***1980***							
Barbados	(A)	-										
Chile	(B, D, F)	3	***1970***		***1982***		***1992***					
	(B, D, F)	2	***1970***				***1992***					
	(B, D, F)	3	***1970***		***1982***		***1992***					
	(B, D, F)	1							***2002***			Others: CELADE
Colombia	(F)	3		***1973***		***1985***		***1993***				
	(F)	1						***1993***				
	(F)	1								***2005***		Others: CELADE

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Costa Rica	(F)	2		*1973*		*1984*						
	(F)	2	*1968*	*1973*								
	(F)	1							*2000*			Others: CELADE
Cuba	(G)	1							*2002*			
Dominica	(A)	-										
Dominican Republic	(B, C, D)	1	*1970*									
	(B, C, D)	1							*2002*			Others: CELADE
Ecuador	(F)	3		*1974*	*1982*		*1990*					
	(F)	3		*1974*	*1982*		*1990*					
	(F)	1							*2001*			Others: CELADE
Guadeloupe	(B, C, D)	1									*2008*	
	(B, C, D)	1			*1982*							
Grenada	(A)	-										
Guatemala	--	2		*1973*	*1981*							
	--	1							*2002*			Others: OTHER
French Guiana	(B, C, D)	1									*2008*	
	(B, C, D)	1			*1982*							
Guyana	(B, D)	1							*2002*			
	(B, D)	2	*1970*		*1980*							
Honduras	(B, F)	2		*1974*		*1983*						
	(B, F)	1							*2001*			Others: CELADE
Haiti	(B, D)	2	*1971*		*1982*							
	(B, D)	3	*1971*		*1982*	*1986*						
	(B, D)	1								*2005*		Others: DHS
Jamaica	(B, D)	2			*1982*		*1991*					
	(B, D)	3	*1970*		*1982*		*1991*					
	(B, D)	1							*2001*			Others: CARICOM

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Saint Lucia	(B, D, F)	2			*1980*		*1991*					
	(B, D, F)	3	*1970*		*1980*		*1991*					
	(B, D, F)	1							*2001*			Others: CARICOM
Mexico	(F)	6	*1970*		*1980*		*1990*	*1995*	*2000*	*2005*		
	(F)	1									*2010*	
	(F)	2			*1980*		*1990*					
Martinique	(G)	1									*2008*	
	(G)	1			*1982*							
Nicaragua	(F)	2	*1971*					*1995*				
	(F)	1	*1971*									
	(F)	1								*2005*		Others: CELADE
Panama	(F)	4	*1970*		*1980*		*1990*		*2000*			
	(F)	3	*1970*		*1980*		*1990*					
	(F)	1									*2010*	Others: CELADE
Peru	(B, D)	2						*1993*		*2007*		
	(B, D)	3	*1972*		*1981*			*1993*				
Puerto Rico	(F)	4	*1970*		*1980*		*1990*		*2000*			
	(F)	3	*1970*		*1980*		*1990*					
Paraguay	(B, D)	2	*1972*		*1982*							
	(B, D)	1							*2002*			Others: CELADE
El Salvador	--	1					*1992*					
	--	2	*1971*				*1992*					
	--	1								*2007*		Others: CELADE
Suriname	(G)	1								*2004*		Others: CARICOM
Trinidad & Tobago	(B, C, D)	3	*1970*		*1980*		*1990*					
	(B, C, D)	1							*2000*			Others: CARICOM
Uruguay	(B, D)	4		*1975*		*1985*		*1996*		*2006*		

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
	(B, D)	3		*1975*		*1985*		*1996*				
Saint Vincent	(B, D)	2	*1970*		*1980*							
	(B, D)	1							*2001*			Others: CARICOM
Venezuela	(B, D, F)	4	*1971*		*1981*		*1990*		*2001*			
	(B, D, F)	3	*1971*		*1981*		*1990*					
United States Virgin Is	(A)	-										
Northern America												
Bermuda	(A)	-										
Canada	(B, D, F)	3	*1971*		*1981*		*1991*					
	(B, D, F)	4	*1971*		*1981*		*1991*		*2001*			
	(B, D, F)	5	*1970*	*1976*	*1981*	*1986*	*1991*					
United States	(E, F)	5	*1970*		*1980*		*1990*		*2000*	*2005*		
	(E, F)	5	*1970*	*1975*	*1979*		*1990*	*1994*				
Oceania												
American Samoa	(A)	-										
Australia	(B, D)	1								*2006*		
	(B, D)	1	*1971*									
Fiji	(A)	-										
Guam	(A)	-										
Kiribati	(A)	-										
New Caledonia	(B, C, D)	1									*2009*	
	(B, C, D)	1					*1989*					
New Zealand	(B, C, D)	1							*2001*			
	(B, C, D)	2			*1981*		*1991*					
French Polynesia	(G)	1								*2007*		

Region/country	Notes	#	Empirical census/survey year									Source
			1970	1975	1980	1985	1990	1995	2000	2005	2010	
Tonga	(B, C, D)	1								***2006***		
	(B, C, D)	1				***1986***						
Vanuatu	(B, C, D)	1									***2009***	
	(B, C, D)	1			***1979***							
Samoa	(B, C, D)	1							***2001***			
	(B, C, D)	3	***1971***	***1976***	***1981***							

***Note*:** (A) No base-year data, (B) Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups, (C) ACA is not possible due to missing time series and/or aggregated age groups, (D) Data reliability, (E) Migration impact on the country-specific education distribution, (F) Transition model issue, and (G) No historical datasets found for validation process;

7.2 Appendix Ib – Country-Specific Documentation with Comments

Table 5. Documentation on validation of WIC 2015 to empirical census data by country

Region/country	From…	to…	Notes	Comments
			Africa	
Angola	-	-	(A)	No base-year data.
Burundi	2010	1990	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Benin	2006	1979	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. Comparison of UIS to WIC indicates an overestimation of educational transition from 'no education' to 'incomplete ISCED 1'. But this cannot be proven with an ACA due the lack of comparable age groups in the UIS time series.
Burkina Faso	2006	1985	--	--
Botswana	-	-	(A)	No base-year data.
Central African Republic	1995	1975	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Cote d'Ivoire	2005	1988	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Cameroon	2004	1976	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The back-projection is based on DHS base-year data which differ from the IPUMS dataset for Cameroon Census 2005 (published in 2013). It is recommended to change the base-year data from DHS to the IPUMS dataset in the next round of the back-projection.

Region/country	From…	to…	Notes	Comments
Congo DR	2007	1984	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. The back-projection is based on DHS base-year data which shows overweighed shares of the population in the higher educational groups.
Congo	2005	1984	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. The back-projection is based on DHS base-year data which shows overweighed shares of the population in the higher educational groups.
Comoros	1996	1996	(G)	No historical datasets found for validation process.
Cape Verde	2000	2000	(G)	No historical datasets found for validation process.
Algeria	2002	1971	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The ACA on the UIS data for 1966 and 1971 indicates the correctness of the UIS data, although UIS aggregates *incomplete* and *complete primary* education into primary education, and *lower* and *upper secondary* education into secondary education. From the ACA it looks like the WIC model overestimates the transition from "no education" into "primary education".
Egypt	2006	2006	(G)	No historical datasets found for validation process.
Ethiopia	2011	2011	(G)	No historical datasets found for validation process.
Gabon	2000	2000	(G)	No historical datasets found for validation process.
Ghana	2000	1970	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Guinea	1996	1983	(F)	Transition model issue. The ACA indicates a slight underestimation of educational transition from "no education" to "completed primary" education

Region/country	From…	to…	Notes	Comments
Gambia	2000	1973	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. UIS categories show a higher share in "no education" to the disadvantage of "incomplete primary" education, which is most likely due to the insufficient split of incomplete and complete educational levels.
Guinea-Bissau	2000	2000	(G)	No historical datasets found for validation process.
Equatorial Guinea	2000	2000	(G)	No historical datasets found for validation process.
Kenya	1999	1969	(F)	Transition model issue. The ACA indicates that the WIC model overestimates the pace of educational transition from "no education" to "lower secondary" education from 1969 to 2009.
Liberia	2007	1974	(B, C, D, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. Transition model issue. UIS categorization for 1970 is not clear and deviates from back-projection model, especially in the categories below 'ISCED 1'. As there is no ACA possible the deviation could result from a lack of data accuracy from UIS or – more realistically as UIS provides here 6 categories – that the transition model underestimates the transition since 1974 from 'no education' to 'ISCED 1'. DHS used for base-year 2007 vastly deviates from NSO census 2008 data, e.g. for 'no education' DHS states for population 25 years plus around 80.8 percent and NSO around 55.3 percent.
Libya	-	-	(A)	No base-year data.
Lesotho	2009	1976	(B, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Transition model issue. The ACA is difficult to implement due to the time lag between the data points, but it indicates that the model slightly underestimates the pace of transition from lower to higher education.
Morocco	2004	1971	--	--
Madagascar	2008	2008	(G)	No historical datasets found for validation process.
Mali	1998	1976	--	--

Region/country	From…	to…	Notes	Comments
Mozambique	2007	1980	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. UIS categorization is not clear and deviates from back-projection model. ACA gives no results as UIS provides not plausible age specific educational data for 1980 with for instance 89 percent of all 25 to 34 year old having ‘no education’, but with just 65.6 percent in the broader age group 35 to 54 years.
Mauritania	-	-	(A)	No base-year data.
Mauritius	2000	1972	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Malawi	2008	1977	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. IPUMS data fits perfectly with WIC model for 1987. UIS categorization is not clear for 1977 and deviates from back-projection model.
Namibia	2007	1991	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Niger	2006	1977	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Nigeria	2008	2008	(G)	No historical datasets found for validation process.
Reunion	2008	2008	(G)	No historical datasets found for validation process.
Rwanda	2002	1978	(B, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Transition model issue. The ACA indicates a slight underestimation of educational transition from "no education" to "completed primary" education from 1978 to 2002.

Region/country	From…	to…	Notes	Comments
Sudan	2008	1983	(B, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Transition model issue. UIS categorization is not clear and deviates from back-projection model. ACA is not possible as UIS does not provide education by age, but just for the category '25 years plus'. But a too fast educational transition looks plausible, even if not provable, due the educational shares of 1956, which show 91.2 percent of population 25 years and older with 'no education'. Our model shows this share for 1973.
Senegal	2002	1988	--	--
Sierra Leone	2004	2004	(G)	No historical datasets found for validation process.
Somalia	2006	2006	(G)	No historical datasets found for validation process.
Sao Tome & Principe	2009	1981	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Swaziland	2006	1976	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. ACA indicates incorrect base-year data from DHS 2006 or underestimation of the educational transition from ISCED 1 to ISCED 2.
Chad	2004	2004	(G)	No historical datasets found for validation process.
Togo	-	-	(A)	No base-year data.
Tunisia	2010	1975	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. UIS data report throughout the years a higher share of the population with "completed primary" education to the disadvantage of those with "lower secondary" education. This seems to indicate a systematic mismatch of educational categories in UIS data.
Tanzania	2002	1988	(B, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Transition model issue. ACA indicates an underestimation of the pace of educational transition from "lower" to "higher" education.
Uganda	2002	1991	(D)	Data reliability. The empirical IPUMS data on "post-secondary" education in Uganda 1991 report a much lower share in this group that cannot be supported with the ACA.

Region/country	From…	to…	Notes	Comments
South Africa	2007	1970	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. The ACA for UIS 1980 and IPUMS 1996 fits until the base-year in 2007. The UIS 1970 and 1985 deviate largely from the other census points and even the ACA does not work out for those two years.
Zambia	2002	1969	(B, C, D, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. Transition model issue. UIS categorization is not clear and deviates from back-projection model. Especially UIS 1990 deviates strongly from WIC model. The ACA analysis is not possible due not comparable age groups in the UIS time series. But for 1969 to 1980 we see an overestimated transition from ‘no education’ to ‘incomplete ISCED 1’, which seems plausible.
Zimbabwe	2005	1992	--	--
			Asia	
Afghanistan	-	1968	(A)	No base-year data.
United Arab Emirates	2005	1975	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Armenia	2001	2001	(G)	No historical datasets found for validation process.
Azerbaijan	2006	2006	(G)	No historical datasets found for validation process.
Bangladesh	2004	1974	(E, F)	Migration impact on the country-specific education distribution. Transition model issue. Bangladesh recorded high migration movements which cannot be taken into account in the transition model that additionally underestimates the pace of educational transition from 'no to 'completed primary' education.

Region/country	From…	to…	Notes	Comments
Bahrain	2001	1971	(B, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Transition model issue. The ACA of the UIS data indicate that the model slightly underestimates the education transition from lower to higher education groups between 1971 and 2001.
Brunei	-	-	(A)	No base-year data.
Bhutan	2005	2005	(G)	No historical datasets found for validation process.
China	2005	1982	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. There is no 'incomplete ISCED 1' in the WIC 2012 dataset while IPUMS data for 1982 cover this category.
Cyprus	2001	1992	(G)	No historical datasets found for validation process.
Georgia	2002	2002	(G)	No historical datasets found for validation process.
Hong Kong	2006	1971	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Indonesia	2010	1971	(D)	Data reliability. Indonesia's census results before 2000 are affected by political motivations and therefore overestimate the national educational structure
India	2001	1971	(D)	Data reliability. India's data via IPUMS do not refer to the census but to an 'Employment Survey' with smaller sample.
Iran	2006	2006	(G)	No historical datasets found for validation process.
Iraq	1997	1997	(G)	No historical datasets found for validation process.
Israel	2004	1972	(B, D, E)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Migration impact on the country-specific education distribution. The ACA of the IPUMS data indicate strong shifts in the age and education structure that are mainly caused by massive migration streams of Jewish population from USA, European countries, Russia, etc., which make the comparison impossible.
Jordan	2004	1979	(G)	No historical datasets found for validation process.
Japan	2010	1970	(F)	Transition model issue. Minor deviations in the transition from 'lower' to 'upper secondary' education.
Kazakhstan	2009	1989	(G)	No historical datasets found for validation process.

Region/country	From…	to…	Notes	Comments
Kyrgyzstan	1999	1999	(G)	No historical datasets found for validation process.
Cambodia	2008	2008	(G)	No historical datasets found for validation process.
South Korea	2010	1970	(F)	Transition model issue. The transition model underestimates the pace of educational transition that has happened in South Korea in the second half of 20th Century due to drastic educational reforms.
Kuwait	2005	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Laos	2005	2005	(G)	No historical datasets found for validation process.
Lebanon	2007	1970	(B, C, D, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. Transition model issue. UIS categorization for 1970 is not clear and deviates from back-projection model, especially in the categories below 'ISCED 1'. As there is no ACA possible the deviation could result from a lack of accuracy in UIS data or – more realistically as UIS reports 6 categories – that the transition model underestimates the transition since 1970 from 'no education' to 'ISCED 1'.
Sri Lanka	-	-	(A)	No base-year data.
Macau	2006	1970	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Maldives	2006	1990	(G)	No historical datasets found for validation process.
Myanmar	2007	1973	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. Missing educational categories, like *completed ISCED 1* in 1973 and 1983, or missing *no education* in 1983 cause deviations. 1983 does not cover the whole country due to an unstable political situation.
Mongolia	2000	1989	(F)	Transition model issue. The ACA indicates a slight overestimation of education transition from *lower* to *upper secondary* education from 1989 to 2000.

Region/country	From…	to…	Notes	Comments
Malaysia	2000	1970	(F)	Transition model issue. The base-year 2000 aggregates *incomplete ISCED 1* and *ISCED 1* in one category. The ACA indicates an overestimation of educational transition from *no education* to *incomplete* and *completed ISCED 1* education from 1970 to 2000.
Nepal	2001	1971	(F)	Transition model issue. The ACA indicates a slight underestimation of educational transition from *no education* to *completed primary* education from 1971, 1981 to 2001.
Oman	-	-	(A)	No base-year data.
Pakistan	1998	1972	(D)	Data reliability. Data for 1981 excludes federally administered tribal areas.
Philippines	2000	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. This problem occurs with UIS data.
Palestine	2007	2007	(G)	No historical datasets found for validation process.
Qatar	2010	1986	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Saudi Arabia	2004	2004	(G)	No historical datasets found for validation process.
Singapore	2010	1980	(B, D, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Transition model issue. The ACA indicates, apart from issues with categorization and data reliability, the underestimation of the pace of educational transition from lower to higher education from 1980 to 2010.
Syria	2004	1970	(D)	Data reliability. 1979 covers the East Bank region only - excluding about 50% of the population.
Thailand	2000	1970	--	--
Tajikistan	2009	1989	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Turkmenistan	1995	1995	(G)	No historical datasets found for validation process.
Timor-Leste	2009	2009	(G)	No historical datasets found for validation process.

Region/country	From…	to…	Notes	Comments
Turkey	2000	1970	(B)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. UIS categorization for 1975 and 1980 is not clear and shows slight deviations from WIC model. The comparison with the IPUMS data from 1985 and 1990 shows a high comparability with WIC.
Uzbekistan	-	-	(A)	No base-year data.
Viet Nam	2009	1979	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. IPUMS aggregates the 8th and 9th grade, which does not allow for a disaggregation between "ISCED 1" and "ISCED 2" in 1989 and 1999.
Europe				
Albania	2002	2002	(G)	No historical datasets found for validation process.
Austria	2008	1971	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. WIC data shows a 6 percentage points higher share in "post-secondary" education compared to NSO data. This is most likely due to changes in the national classification of education groups.
Belgium	2001	1970	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Bulgaria	2001	1975	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The ACA shows the plausibility of the UIS dataset, which indicates a model underestimation from lower to higher educational groups from 1975 to 2001 in the WIC 2015 model.
Bosnia-Herzegovina	2010	1971	(D)	Data reliability. 1981 just covers the economically active population. Emigration in the context of the Yugoslav war can't be considered in the back-projection model and causes deviations in the educational structure.
Belarus	1999	1989	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.

Region/country	From…	to…	Notes	Comments
Switzerland	2000	1970	(B)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. IPUMS provides the category "no education" for 1970 to 1992 (~ 1.5%) which are not considered in WIC data. IPUMS combines the categories from "incomplete primary" up to "lower secondary" education. Minor underestimation of share with "upper secondary" education to the disadvantage of shares with "post-secondary" education (~4%)
Czech Republic	2001	1980	(F)	Transition model issue.
Germany	2010	1971	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The different educational models in the Federal Republic of Germany (FRG) and the German Democratic Republic (GDR) before the Reunification in 1990 do not allow for comparison in the IPUMS data.
Denmark	2001	1991	--	--
Spain	2001	1970	(F)	Transition model issue. Deviation between WIC and IPUMS data increases the further the validation goes back in time. As the ACA shows a high reliability in the IPUMS data it is most likely that the WIC 2015 model underestimates the pace in transition from lower to higher education.
Estonia	2000	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Finland	2009	1970	(B)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Base-year 2009 vs 2011
France	2008	1968	--	--
United Kingdom	2001	2001	(G)	No historical datasets found for validation process.
Greece	2001	1971	--	--
Croatia	2001	1971	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. ACA analysis shows that UIS 1991 and 1997 report identical educational shares.
Hungary	2001	1970	(F)	Transition WIC 2015 model issue. Deviation between WIC and IPUMS data increases the further the validation goes back in time. As the ACA shows a high reliability in the IPUMS data it is most likely that the model underestimates the pace in transition from lower to higher education.

Region/country	From…	to…	Notes	Comments
Ireland	2002	1971	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Data do not take into consideration the population "still at school/university", and aggregate ISCED 3 and 4. Due to the high share of people with "unknown" education (~10%) the empirical data are hardly comparable with the WIC 2015 dataset.
Iceland	2010	2010	(G)	No historical datasets found for validation process.
Italy	2001	1971	--	--
Liechtenstein	-	-	(A)	No base-year data.
Lithuania	2001	1989	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Luxembourg	2001	1991	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. UIS data report ~9.2% of the population 25 years plus with "unknown" education.
Latvia	2000	1988	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Moldova	2004	1989	(B, D, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Transition model issue. Between 1989 and 2004 the share of the population 25+ with "no education" drops from 12.7% to about 1%. Due to the high share of population 65years plus in 1989 with "no education" (58%) it is most likely that these age cohorts got replaced by younger more educated cohorts until the base-year (Life expectancy 1989 about 67years). That drop in "no education" due to higher mortality in this group, is not considered in the WIC dataset.

Region/country	From…	to…	Notes	Comments
Macedonia	2008	1994	(B, D, E, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Migration impact on the country-specific education distribution. Transition model issue. Between 1994 and 2008 the share of the population 25+ with "no education" dropped from 28.0% to 8.6%. Due to the high share of population 65years plus in 1989 with "no education" (71.8%) it is most likely that these age cohorts got replaced by younger more educated cohorts until the base-year (Life expectancy 1994 about 71.9years). That drop in "no education" due higher mortality in this group, is not considered in the WIC dataset. Additionally between 1998 and 2001 around 230.000 people (~12%) emigrated, which changed the educational structure of a population.
Malta	2010	2010	(G)	No historical datasets found for validation process.
Montenegro	2003	1971	(G)	No historical datasets found for validation process.
Netherlands	2001	1971	(B)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. IPUMS provides the national education categories that are due to their level of aggregation not transferable into the ISCED based WIC classification. Additionally about 19% of the population 25+ are labelled with "unknown" education.
Norway	2010	1970	(F)	Transition model issue. Deviation between WIC and NSO data increases the further the validation goes back in time. As the ACA shows a high reliability in the NSO data it is most likely that the WIC 2015 model underestimates the pace in transition from basic (ISCED 2 and lower) to higher education.
Poland	2002	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Portugal	2001	1970	(F)	Transition model issue. The ACA indicates a slight overestimation of educational transition from "no education" up to "lower secondary" from 1970 to 2001.
Romania	2002	1977	(F)	Transition model issue. The ACA indicates a slight underestimation of educational transition from "no education" to "completed primary" education from 1977 to 2002.
Russia	2002	1989	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Serbia	2002	1971	(G)	No historical datasets found for validation process.

Region/country	From…	to…	Notes	Comments
Slovakia	2001	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The NSO data aggregates the categories "incomplete primary" up to "lower secondary", which causes minor deviations to the WIC dataset.
Slovenia	2002	1971	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Sweden	2010	1985	(F)	Transition model issue. The ACA from NSO data indicates that the WIC dataset slightly overestimates the pace of transition from "lower" to "higher" education.
Ukraine	2001	2001	(G)	No historical datasets found for validation process.
Latin America and the Caribbean				
Aruba	2010	1991	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Netherlands Antilles	2001	1971	(G)	No historical datasets found for validation process.
Argentina	2001	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The time series are hardly comparable due to frequent changes in the national education system. Additionally NSO data differentiate between incomplete and complete grades, but not between lower and upper secondary. Therefore people with incomplete secondary education are allocated to primary education, although many of those are probably in upper secondary grades.
Bahamas	2000	1990	(B, C, D, E)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.

Region/country	From…	to…	Notes	Comments
Belize	2000	1970	(B, D, E, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Migration impact on the country-specific education distribution. Transition model issue. The increase in "no education" in UIS data up to 1970 is not reflected in the WIC data. Especially the increase of "no education" in the base-year 2000 compared to 1980 and 1990 can probably be exclusively explained by emigration of educated population.
Bolivia	2001	1976	(F)	Transition model issue. The ACA shows the plausibility of the IPUMS dataset and indicates an underestimation in the pace in transition from lower to higher education.
Brazil	2010	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The IPUMS data show an overrepresentation of population with "completed primary" education to the disadvantage of the other categories, compared to WIC dataset. It is most likely that incomplete higher categories are allocated to ISCED 1.
Barbados	-	-	(A)	No base-year data.
Chile	2002	1970	(B, D, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Transition model issue. The ACA indicates that the WIC model overestimates the pace of educational transition from "no education" to "lower secondary" education from 1970 to 2002.
Colombia	2005	1973	(F)	Transition model issue. Deviation between WIC and IPUMS data increases the further the validation goes back in time. As the ACA shows a high reliability in the IPUMS data it is most likely that the model slightly underestimates the pace in transition from lower to higher education.
Costa Rica	2000	1968	(F)	Transition model issue. Deviation between WIC and IPUMS data increases the further the validation goes back in time. As the ACA shows a high reliability in the IPUMS data it is most likely that the model underestimates the pace in transition from lower to higher education.
Cuba	2002	1980	(G)	No historical datasets found for validation process.
Dominica	-	-	(A)	No base-year data.

Region/country	From…	to…	Notes	Comments
Dominican Republic	2002	1970	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Ecuador	2001	1974	(F)	Transition model issue. Pace of transition between 1974 and 1982 seems to be underestimated by the WIC 2015 model. As the ACA shows a high reliability in the IPUMS data it is most likely that the model underestimates the pace in transition from lower to higher education before 1982. This impression gets supported by older IPUMS and UIS datasets from 1950 and 1962, which seem consistent across time. In the NSO data incomplete and complete primary education is difficult to disentangle as NSO divides between 1-3 years primary and 4-6 years primary in a system where primary lasts for 6 years.
Guadeloupe	2008	1982	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Grenada	-	-	(A)	No base-year data.
Guatemala	2002	1973	--	--
French Guiana	2008	1982	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Guyana	2002	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Honduras	2001	1974	(B, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Transition model issue. In 1983 UIS data "incomplete primary" education is missing. The ACA indicates an underestimation in the pace of transition from "no education" up to "lower secondary" education in the WIC 2015 model.
Haiti	2005	1971	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.

Region/country	From…	to…	Notes	Comments
Jamaica	2001	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The IPUMS data aggregate "incomplete primary"/"completed primary" and "lower and upper secondary" education.
Saint Lucia	2001	1970	(B, D, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Transition model issue. UIS and IPUMS categorization is not clear plus summing up of 6 categories into 4. ACA of IPUMS data indicates the correctness of the data, but ACA of UIS data is not possible due to shifts between categories. The WIC model seems to underestimate pace of increase from lower to higher educational groups.
Mexico	2010	1970	(F)	Transition model issue. Deviation between WIC and IPUMS data slightly increases the further the validation goes back in time. As the ACA shows a consistent picture, the deviation indicates a slight underestimation of educational transition from ‘no education’ to ‘ISCED 1’.
Martinique	2008	1982	(G)	No historical datasets found for validation process.
Nicaragua	2005	1971	(F)	Transition model issue. Deviation between WIC and IPUMS data increases the further the validation goes back in time. As the ACA shows a consistent picture, the deviation indicates an underestimation of educational transition from ‘no education’ to ‘ISCED 1’.
Panama	2010	1970	(F)	Transition model issue. Deviation between WIC and IPUMS data increases the further the validation goes back in time. As the ACA shows a consistent picture, the deviation indicates an underestimation of education transition from lower to higher educational categories.
Peru	2007	1972	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Puerto Rico	2000	1970	(F)	Transition model issue. Deviations between WIC and IPUMS indicate a slight underestimation the pace of transition from lower to higher educational categories.
Paraguay	2002	1972	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. UIS distinction of "incomplete" and "completed primary" education unclear.
El Salvador	2007	1971	--	--
Suriname	2004	2004	(G)	No historical datasets found for validation process.

Region/country	From…	to…	Notes	Comments
Trinidad & Tobago	2000	1970	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points. The ACA indicates an abnormal educational composition in the UIS dataset for 1980.
Uruguay	2006	1975	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The ACA indicates that the share of the population with "incomplete primary" is overrepresented in the IPUMS data, which causes a validation result of *C*.
Saint Vincent	2001	1970	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points.
Venezuela	2001	1971	(B, D, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Transition model issue. In 1971 and 1991 IPUMS shows about 6.5% and more people with "unknown" education. The ACA for 1981 indicates that people with "no education" got allocated in the IPUMS dataset to "incomplete primary" education, which is the reason why 1981 does not fit into the IPUMS time series and is hardly comparable with the WIC dataset. Additionally the WIC model seems to slightly overestimate the pace of transition from lower to higher education between 1970 and 2001.
United States Virgin Is	-	-	(A)	No base-year data.
Northern America				
Bermuda	-	-	(A)	No base-year data.
Canada	2001	1970	(B, D, F)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. Transition model issue. There is an overestimation of post-secondary education and an underestimation of ‘ISCED 1. The 1970s immigration wave of high skilled population changed the education structure between 1971 and 1981 (see Borjas 1993). Additionally there is no ‘no education’ category in IPUMS ("Below grade 5" that covers category no education and incomplete primary). | There is no ‘ISCED 2’ ("Grades 9-13" that covers lower and upper secondary education). We allocated this category for the validation to upper secondary education.

Region/country	From…	to…	Notes	Comments
United States	2005	1970	(E, F)	Migration impact on the country-specific education distribution. Transition model issue. WIC model overestimates the transition from 'ISCED 1' to secondary education. Additionally migration streams in this period can cause deviations from WIC model.
Oceania				
American Samoa	-	-	(A)	No base-year data.
Australia	2006	1971	(B, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. Data reliability issues in one or more data points. The ACA of 1966 and 1971 shows deviations and indicates problems with UIS data quality.
Fiji	-	-	(A)	No base-year data.
Guam	-	-	(A)	No base-year data.
Kiribati	-	-	(A)	No base-year data.
New Caledonia	2009	1989	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
New Zealand	2001	1981	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
French Polynesia	2007	2007	(G)	No historical datasets found for validation process.
Tonga	2006	1986	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Vanuatu	2009	1979	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.
Samoa	2001	1971	(B, C, D)	Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups. ACA is not possible due to missing time series and/or aggregated age groups. Data reliability issues in one or more data points.

***Note*:** (A) No base-year data, (B) Educational categorization in empirical data or its allocation to ISCED mapping is not clear due to aggregation of education groups, (C) ACA is not possible due to missing time series and/or aggregated age groups, (D) Data reliability, (E) Migration impact on the country-specific education distribution, (F) Transition model issue, and (G) No historical datasets found for validation process;

8 Appendix II - Comparison of selected reconstruction efforts of levels of educational attainment

	Barro & Lee (2013)	Cohen & Soto (2007); Cohen & Leker (2014)	De la Fuente & Doménech (2012)	Morrisson & Murtin (2009)	Lutz et al. (2007)	WIC 2015
Age groups	5-year age groups: 15-19; 20-24; … 75+	5-year age groups: 15-19; 20-24; … 80+	One large age group: 25+	Two large age groups: 15+ & 15-64	5-year age groups: 15-19; 20-24; … 65+	5-year age groups: 15-19; 20-24; … 85+
Sex	male/female	total	total		male/female	male/female
Education indicators	Proportion by highest level attained + MYS[17]	Only MYS	Proportion by highest level attained + MYS	Only MYS	Proportion by highest level attained + MYS	
Time frame	1950 to 2010 (5-year steps)	1960 to 2020 (10-year steps)	1960 to 2010 (5-year steps)	1870 to 2010 (10-year steps)	1970 to 2000 (5-year steps)	1970 to 2010 (5-year steps)
Specific education categories used	**7 categories:** no schooling; first level (total / complete); secondary (total / complete); tertiary (total / complete)	Not mentioned	**6 categories:** illiterates; primary schooling; lower and upper secondary schooling; first and second cycle of higher education	Not mentioned	**4 categories:** no schooling; primary; secondary; tertiary	**6 categories:** no schooling; incomplete primary; complete primary; lower secondary; upper secondary; post-secondary
Spatial coverage	146 countries	95 countries	21 countries (OECD)	74 countries	120 countries	171 countries
Empirical data source	Censuses and enrolment series	OECD, censuses, Mitchell Series	Censuses and surveys	Enrolment series, Cohen and Soto (2007)	Censuses, IPUMS, DHS[18], LFS[19]	
Methodology	Interpolation/ extrapolation, decomposition method	Extrapolate backward-assumption of constant proportions assumed, Net School Intake Rate used in case of no census data	Proceeding backward from 1990, 1995 or 2010 by backward and forward interpolation, or rely on miscellaneous information	Perpetual inventory method	Reconstruct 5-year age groups along cohort lines from **2000** backwards considering mortality/ migration differentials	Reconstruct 5-year age groups along cohort lines from **2010** backwards considering mortality/ migration differentials

Sources: (Barro and Lee 1993; Barro and Lee 2013; Cohen and Leker 2014; Cohen and Soto 2007; de la Fuente and Doménech 2000; de la Fuente and Doménech 2012; Morrisson and Murtin 2009; Lutz et al. 2007a; Bauer et al. 2012; Lutz et al. 2007b)

[17] MYS refers to Mean Years of Schooling
[18] DHS refers to Demographic Health Survey
[19] LFS refers to Labour Force Survey