Journal of Development Economics 122 (2016) 147–169

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Human capital in the long run

Jong-Wha Lee *, Hanol Lee

Economics Department, Korea University, Sungbuk-Ku, Anam-Ro 145, Seoul 02841, Republic of Korea

ARTICLE INFO

Article history:
Received 25 February 2015
Received in revised form 15 January 2016
Accepted 29 May 2016
Available online 1 June 2016

JEL classification:
120
J24

Keywords:
Education
Enrollment
Human capital

ABSTRACT

This study presents new data sets on long-run enrollment ratios, educational attainment, and human capital stock measures for numerous countries. We construct a complete data set of historical enrollment ratios, subdivided by education level and gender, for 111 countries from 1820 to 1945 (at five-year intervals) by using newly compiled census observations and information on the year of establishment of the oldest school in individual countries. Then, by utilizing these enrollment ratios, as well as available census data from 1945 onward on different age groups' educational attainment, we construct a data set of estimated educational attainment, disaggregated by gender and age group, and aggregate human capital stock that spans from 1870 to 2010. The data show that over the past two centuries, there has been remarkable growth in average educational attainment and human capital stock as well as a narrowing of the gap in average educational attainment between nations.

1. Introduction

Human capital has been emphasized as a critical determinant of economic development. In addition, it exerts a significant influence on social and political issues, such as fertility, the education of children, and democracy (Becker, 2009; Barro, 2012).

Since human capital is multifaceted and includes a complex set of human attributes, it is difficult to precisely and quantitatively measure; as such, many have used educational attainment as a proxy and attempted to measure it across various populations.[1]

This study goes beyond previous ones by presenting new long-run data sets on estimated school enrollment and educational attainment, classified by education level and gender, for a sample of 111 countries. These data sets include measures of enrollment ratios among school-age populations from 1820 onward and educational attainment among those aged 15–64 from 1870 onward.

School enrollment provides valuable information about a society's expansion and investment in education. Enrollment data have been widely used by scholars studying comparative education and politics to investigate the development of national education systems (Benavot and Riddle, 1988; Meyer et al., 1992). It has also been used for cross-national comparative analyses of the relationship between education and economic and social development (Psacharopoulos and Woodhall, 1985; Barro, 1991; Benavot, 1996). Recent studies have also used enrollment ratios to construct estimates of a population's educational attainment (Morrisson and Murtin, 2009; Barro and Lee, 2013, 2015).

UNESCO has collected and published enrollment data for every country in the world; however, this information is limited to the post-World War II era. Before that time, the League of Nations and other international agencies made similar efforts to collect such data. Several compendia (Benavot and Riddle, 1988; Mitchell, 2003a, b, c; Lindert, 2004; Banks and Wilson, 2013) also provide historical enrollment statistics for many independent states and colonies.

Yet, available studies do not provide comprehensive data on historical enrollment rates for each education level (primary, secondary, and tertiary). Moreover, no study has attempted to collect extensive data on female enrollment. Our new data set fills this gap by providing more detailed information regarding the enrollment ratios of 111 countries from 1820 to 1945.

Using a variety of sources, we compiled figures for primary, secondary, and tertiary enrollment among females and the population as a whole. Our data set expands the existing ones by using a significant number of new census observations. We also utilize the information on the years of establishment of the oldest schools at different education levels in individual countries. We then used these figures to calculate enrollment ratio estimates. We present the resulting data set on enrollment, disaggregated by education level (primary, secondary, or tertiary) and gender, for 111 nations and former colonies at five-year intervals from 1820 to 1945. By combing historical enrollment ratios with information from post-1950 UNESCO publications, we are able to construct a complete time series of

* Corresponding author.
E-mail addresses: jongwha@korea.ac.kr (J.-W. Lee), hanollee@korea.ac.kr (H. Lee).

[1] See Barro and Lee (2001) and Barro and Lee (2015) for detailed discussions.

http://dx.doi.org/10.1016/j.jdeveco.2016.05.006
0304-3878/

enrollment ratios up to the year 2010. We then analyze this data set in order to study the trends of school enrollment ratios.

The adult population's aggregate stock of education, which has a significant influence on production and social and political life, cannot be gauged by enrollment ratios, though. Attainment of schooling, as reflected in the enrollment of the school-aged population, is indicative of past inflows to schooling. These flows accumulate to create future stocks of educational attainment, but because the education process takes many years, there is a long lag between the flows and stocks.

Many researchers have attempted to construct direct measures of the stock of educational attainment (De La Fuente and Doménech, 2006; Cohen and Soto, 2007). The most widely used of these measures is the number of years that citizens spend in school. For instance, Barro and Lee (1993, 2001, 2013) constructed a data set on educational attainment for a large number of countries by using census and survey observations on attainment as benchmark stocks, and filled in missing observations by using enrollment-ratio and population structure data. The most recent data set provides estimates of educational attainment for the populations of 146 countries, disaggregated by gender and five-year age group, from 1950 to 2010 at five-year intervals. The data set distinguishes between seven different levels of education: no formal education, incomplete primary, complete primary, lower secondary, upper secondary, incomplete tertiary, and complete tertiary. These data were used to calculate the average years of schooling among the adult population both as a whole and at the primary, secondary, and tertiary levels.

The present study constructs estimates of educational attainment before 1950 by extrapolating later data on various age groups' educational attainment and utilizing estimated historical enrollment data. The newly constructed data set presents estimates of educational attainment, classified by age group (15–24, 25–64, and 15–64) and by gender, for 111 countries from 1870 to 1945 at five-year intervals. Combining this data set with existing data from Barro and Lee (2013), we highlight the major trends in educational attainment from 1870 to 2010 both throughout the world and in major regions.[2]

Our new estimates of historical educational attainment improve upon those of previous studies, such as Morrisson and Murtin (2009), Baier et al. (2006), Tamura et al. (2012), and Van Leeuwen and Van Leeuwen-Li (2014). Morrisson and Murtin (2009) presented a historical database on educational attainment in 74 countries from 1870 to 1960 at 10-year intervals. In order to do so, they used historical enrollment data to generate educational stock measures. Baier et al. (2006) utilized a similar method but only extended their study to 34 countries. Tamura et al. (2012) constructed data set on real output, physical capital, and human capital for 168 countries, 69 of which have human capital stock data from 1890 or from an earlier year onward.

Van Leeuwen and Van Leeuwen-Li (2014) presented a data set on average years of education that covers a wider range of countries since 1870.[3] They combined the estimates of Morrisson and Murtin (2009) with the estimates, for the additional countries, from the researches by Van Leeuwen, Van Leeuwen-Li, and Foldvari (e.g., Földvári and Van Leeuwen, 2014). Our estimates make use of a greater number of original sources on enrollment. More importantly, our estimation method uses actual census data on educational attainment by age group as a benchmark. Thus, this new data set contributes to the existing literature by improving the accuracy of educational attainment estimates for the 1870–1945 time periods. Moreover, our data set subdivides education into seven different levels and presents the distribution of educational attainment among the female, male, and total populations.

We also analyze the evolution and distribution of human capital stock worldwide and across nations based on a human capital measure (as defined by Mincer (1974)) and average years of schooling.

2. Estimates of enrollment ratios by education level and gender from 1820 to 1945

2.1. Data and estimation method

Earlier primary data sources present information on total enrollment. In order to make meaningful comparisons between different countries and time periods, we construct enrollment ratios for each education level. Enrollment ratios are calculated by taking the number of students at each education level and dividing it by the number of persons in the relevant school-age population.

There are several means by which to measure enrollment ratios. An "unadjusted enrollment" ratio" is defined as the ratio of all persons enrolled in a given level of schooling to the population within a specified age bracket, which is consistent across all countries. For example, the earlier UNESCO statistics in the 1960s use constant age categories for all countries: 5–14 for the primary level, 15–19 for the secondary level, and 20–24 for the tertiary level. This measure can be biased, though, because it does not consider the national variations in the duration of schooling. For example, if a country requires all children aged 6 to 11 to attend a six-year cycle of primary school, then the unadjusted enrollment ratio, which uses the 5–14 year old population as a denominator, underestimates the true enrollment ratio.

A "gross enrollment ratio" takes into account the differences in school-age populations that stem from disparate national educational systems. It is constructed by dividing the number of all persons enrolled in a given level of schooling by the population of the age group that should be enrolled at that level according to national regulations or customs. For example, in Korea, where primary education begins at the age of 6 and lasts for 6 years, the total number of registered students in primary school would be compared with the population of children aged 6 to 11.

The gross enrollment ratio is widely used in many official publications by UNESCO and national statistical agencies for cross-national comparisons. Though this statistic is more often available for developing countries, it tends to be skewed when a significant number of students repeat grades, which is a typical occurrence in the developing world.[4] In fact, for many advanced countries the gross enrollment ratio often exceeds 100%.

A "net enrollment ratio" is defined as the ratio of students in a designated age group, at a given level of schooling to the total population of that age group. For example, registered primary school students aged 6–11 are compared with the total population of children between the ages of 6 and 11. As such, this statistic does not count the students in primary school who are younger than 6 or older than 11. If there are a great number of such under- or over-aged children at each level of education, then the net enrollment ratio will suffer from measurement errors. In addition, data on net enrollment ratios are not widely available.

We construct "adjusted enrollment ratios," which are appropriate for comparing expansions in primary and secondary school education across nations (Barro and Lee, 2001). These adjusted ratios are a modification of the gross enrollment ratios in that they account for the repetition of grades in primary and secondary schools. The gross enrollment ratio, whether adjusted or unadjusted, is a standard flow measurement of educational investment in a society's new entrants, and it is widely

[2] Barro and Lee (2015) use a similar method to determine educational attainment from 1870 to 2010 for 89 countries. The study does not use information on the establishment years of the oldest schools to construct historical estimates of enrollment ratios.

[3] The data are available at the Clio-Infra (2014) website (https://www.clio-infra.eu).

[4] Fredriksen (1983) estimated that the average gross enrollment ratio at the primary level was 86% for developing countries in 1980; however, when adjusted for repeaters, the figure fell to 73%. The data on repetition ratios are discussed later in this section.

Table 1
Correlations between sources for the primary, secondary, and tertiary enrollment ratios.

Primary total	Mitchell	Banks	UNESCO	Lindert
Mitchell (obs.781)	–			
Banks (obs.779)	0.9528 (472)	–		
UNESCO (obs.306)	0.9854 (219)	0.9443 (177)	–	
Lindert (obs.341)	0.9164 (240)	0.9501 (235)	0.9804 (40)	–
Benavot and Riddle (obs.577)	0.9642 (314)	0.9587 (288)	0.9630 (131)	0.9798 (277)

Secondary total	Mitchell	Banks	UNESCO
Mitchell (obs. 611)	–		
Banks (obs. 741)	0.9207 (383)	–	
UNESCO (obs. 183)	0.815 (131)	0.7656 (112)	–
Lindert (obs. 221)	0.7674 (163)	0.8077 (162)	0.6187 (26)

Tertiary total	Mitchell	Banks
Mitchell (obs. 608)	–	
Banks (obs. 926)	0.9403 (419)	–
UNESCO (obs. 165)	0.9652 (131)	0.8729 (137)

Notes: The figures are pairwise correlation coefficients. The numbers of observations in the common samples are in parentheses. The raw enrollment data from the original sources are converted into enrollment ratios by using our estimates of the country-specific, school-age populations.

used for cross-country comparisons. However, as compared to the unadjusted ratio, the adjusted ratio is a superior reflection of the inflow of the youth population.[5] Due to limitations in the data, we use unadjusted gross enrollment ratios for tertiary education.

2.1.1. Historical enrollment statistics

The historical data on school enrollments are compiled using a variety of sources.[6] These data on primary, secondary, and tertiary enrollments among the total and female populations are used to construct a complete time series of enrollment ratio estimates.

In its annual statistical yearbooks, UNESCO has published comprehensive and reliable enrollment data for every country in the world, but its information is limited to the post-World War II era. Before then, the League of Nations and other international agencies made similar efforts to collect and record data. The most comprehensive sources of total enrollment data come from Mitchell (2003a, b, c); Benavot and Riddle (1988); Banks and Wilson (2013); Lindert (2004), UNESCO's World Survey of Education, Volumes II, III, IV (1958, 1961, 1966), and the U.S Bureau of Education's *Report of the Commissioner of Education* (various years). We also compiled data from other sources, including national statistical publications.[7]

Mitchell's (2003a, b, c) data set is, undoubtedly, our most important data source. It provides an extensive collection of unadjusted census and survey figures without estimations for missing years. It also notes the important changes in enrollment series over time. However, this data set only includes figures for total enrollment, and furthermore, it does not include all countries.

Banks and Wilson (2013) also provide a comprehensive time series of total enrollment. However, it includes interpolated enrollment figures, which are not always accurate, and reports little data on former colonies.

UNESCO's *World Survey of Education* reports both total and female population enrollment figures for independent states and former colonies; yet, it only goes as far back as the 1930s.

The U.S. Bureau of Education's *Report of the Commissioner of Education* reports enrollment for a multitude of countries during the late 19th and early 20th centuries.

Benavot and Riddle (1988) compile data from all of the existing sources, such as Mitchell's data set and UNESCO's *World Survey of Education*, in order to construct primary enrollment ratios for 126 countries from 1870 to 1940. They estimate "unadjusted enrollment ratios" by using each country's population of 5–14 year-olds as the comparative base. However, they do not provide estimates for secondary and tertiary enrollment.

For total enrollment in secondary and tertiary schools, our principal sources are Mitchell (2003a, b, c); Banks and Wilson (2013); Lindert (2004) (who only provides data on secondary enrollment), and UNESCO's World Survey of Education (1958, 1961, 1966).

Unfortunately, there is very limited data on the female population. Historical statistics on female enrollment are mainly gathered from the U.S. Bureau of Education's *Report of the Commissioner of Education*, UNESCO's *World Survey of Education*, Barnard (1854) and Monroe (1911). In addition, we also rely upon national statistics and other such publications for data on countries such as China, India, Indonesia, and Japan.

We checked all of the data carefully and selected observations that we judge to be accurate and suitable for both cross-national and intertemporal comparisons. Although a wide variety of sources are used for our benchmark figures of actual enrollment, correlations among the different sources are quite high (see Table 1). For instance, there are relatively high correlations among the estimates of Mitchell, Banks and Wilson, and UNESCO.[8] Thus, in many cases, we are able to combine them. For example, we can combine the actual figures from UNESCO's *World Survey of Education* for 1930–1945 with Mitchell's data for the years prior to 1930.

It should also be noted that we carefully addressed several problems posed by the original data. First, the division of education into three principal levels for cross-national comparability involves a number of arbitrary decisions concerning the classification of different types of schools. In principle, we rely on the classification set forth by UNESCO, wherein higher primary or middle schools are generally categorized as secondary-level educational institutions. However, the education level is not always clearly defined in the data. For example, in some cases the only enrollment data available combine first and secondary level education. The classification of vocational and teacher training schools is often complicated as well; for the most part, they are classified as secondary schools, but some technical schools and teacher training colleges (which require the completion of secondary-level education) are classified as tertiary institutions. Some countries do not include vocational or teacher training schools in their survey and census statistics. For

[5] A proper accounting of net educational investment also deducts dropouts from gross enrollment. Note that the actual enrollment numbers (such as the total number of registered students or average yearly enrollment) reported in censuses and surveys exclude those who have dropped out of school at the time of the survey. However, because students drop out while the school year is still in progress, the actual number of children who finish the school year would be lower than the number recorded as "enrolled." For this reason, when constructing a "stock" measure of educational attainment, we use "completion ratios" to account for all of the students who dropout throughout the entire cycle of primary and secondary school.

[6] Benavot and Riddle (1988) discuss the brief history and the quality of primary education enrollment statistics.

[7] The exact sources of each school's actual enrollment figures and estimates for individual countries are available as an Appendix table online (http://barrolee.com/Lee_Lee_LRdata.htm).

[8] The high correlations must reflect that to a great extent these existing databases relied on the same original sources. The pairwise regressions show that the estimated constant terms are statistically not different from zero for all the pairs in primary, secondary, and tertiary education, except for the two pairs in primary between Banks and Lindert and between UNESCO and Benavot. The estimated slope coefficients are close to one.

these cases, we used all other available information (such as data from the next closest year) in order to make the necessary adjustments.

Second, school enrollments do not include education at home and informal schools, which can cause underestimation of true educational attainment especially in the earlier periods. In Western Europe, for example, families and religious institutions had played an important role in the education of children before the expansion of public schools, resulted from compulsory schooling laws and strong state administrations in the late 18th and early 19 centuries.[9] Although it would be better to take account of education outside formal schooling, it is difficult to get an accurate measure of enrollments at informal schools such as Sunday schools in church as well as home and private tutoring.

For developing countries in Asia, Africa, and South America, the exclusion of indigenous, traditional religious, and non-Western schools from the enrollment data poses a particular challenge in the estimation of educational attainment. In these societies, formal education was often well established before Western-style education was introduced by Christian missionaries or colonial authorities (Benavot and Riddle, 1988). In China, for example, compulsory mass education and the imitation of Western models occurred in the late 19th and early 20th centuries. However, there were countrywide establishments of provincial schools and prevalence of the Confucianism tradition of education. For most sources, we lack detailed information on whether the original census includes traditional school enrollment. For many developing countries, the exclusion of these informal schools would yield underestimates of the true enrollment ratio, particularly at primary schools. However, we suspect that this bias would be fairly insignificant for secondary and tertiary education estimates.

Political and territorial changes are another obstacle that we must face in the construction of our data set. Changes in territorial boundaries or state independence sometimes interfere with the collection of data, but, in most cases, we can reorganize the historical data based on the current national boundaries in order to construct our estimates.

We also construct the "original" enrollment data by making use of the information on years of school establishment (at different education levels) in individual countries. The limited availability of historical enrollment data often indicates that organized, or "modern," schools did not exist. We have found information on the establishment of primary education in 95 countries, the establishment of secondary education in 96 countries, and the establishment of tertiary education in 105 countries. The information is not always very specific in terms of the exact year of foundation of the oldest school. It often explains only the earliest documented contemporary reference to the primary school or the year that the state declared a decree related to secondary education for the first time. Nevertheless, we have compiled data on 24 countries for primary education, 70 countries for secondary education, and 101 countries for tertiary education. Such data provide specific information about the year of establishment of the oldest school. The information is collected from the UNESCO's *World Survey of Education*, *World Education Encyclopedia*, as well as various online and offline sources.[10] The information on the establishment of the oldest school contributes to the generation of a significant number of original enrollment ratios (i.e., zero values for the years prior to school establishment).

Using the aforementioned sources, we compiled all available enrollment ratios for primary, secondary, and tertiary education levels, at five-year intervals from 1820 to 1945 (where the underlying figures are applied to the nearest five-year values).[11]

Table 2
Number of available actual enrollment ratios by year, 1820–1945.

Year	Total population			Female population		
	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary
1820–1845	149	301	422	104	296	369
1850–1870	189	230	365	77	188	259
1875–1895	283	300	391	115	143	210
1900–1920	388	347	412	121	87	175
1925–1945	461	419	448	45	167	256
Total 1820–1945	1470	1597	2038	462	881	1269

Note: The figure refers to the number of available country-year observations of actual enrollment figures in the indicated years at five-year intervals for the sample of 111 countries.

Table 2 summarizes the distribution of the actual enrollment figures (available at five-year intervals since 1820). A great many countries in our sample had actual enrollment data available for the total population, especially for the period after 1870. Actual observations fill 68%, 64%, and 75% of the possible cells for the primary, secondary, and tertiary education levels between 1875 and 1945, respectively. Table 3 shows that 109, 103, and 103 countries respectively have at least one observation of actual primary, secondary and tertiary enrollment data over the period of 1820 to 1945.

Available figures for the female population are even rarer, constituting approximately 16%, 31%, and 44% of the cells for primary, secondary, and tertiary enrollment, respectively. There are 70, 77, and 87 countries that have at least one observation of actual primary, secondary, and tertiary enrollment data over the period 1820–1945, respectively. For a limited sample of cases in which the observations for the female population are missing, when raw census enrollment data for the total population are available, the ratios between total and female enrollments in the nearest year are used to construct the missing values.

2.1.2. Estimation of population age structure

Gross enrollment ratios are calculated by dividing the gross enrollment by the number of children in the relevant school-age population.

The historical demographic data on the national population's age distribution are gathered from Mitchell (2003a, b, c), the United Nations Demographic Yearbook (1955), and the Statistical Yearbook of the League of Nations (various years). Data for Sweden are from the Berkeley Mortality Database (Wilmoth, 1995), and data for Japan and Korea are compiled from national sources.

For 70 countries in our sample, we have at least one actual observation from the historical demographic statistics, and for the advanced countries we can fill about 50% of the cells with actual

[9] See Barro and Lee (2015) for a brief history of the worldwide educational expansion that has occurred over the past two centuries.

[10] The accompanying Appendix Table A, which is available online, contains the details on the data sources of school establishment years for individual countries.

[11] Our data set has many new original pieces of data on raw historical enrollment ratios, even when excluding those generated by the years of school establishment. The historical enrollment ratios used in existing studies, such as Morrison and Murtin (2009) and Baier et al. (2006), are mainly from Mitchell, Lindert, and Banks (with many overlapping observations). Although identifying the exact sources of original data used in these studies is rather difficult, a rough comparison shows that our original data contain 272 additional country-year observations out of 1075 total census-survey observations in primary, 146 out of 818 in secondary, and 158 out of 1019 in tertiary. Without using the information on the years of school establishment, our sample contains a large amount of actual enrollment data available for the total population, especially for the period after 1870. Actual observations fill 57%, 48%, and 48% of the possible cells for the primary, secondary, and tertiary education levels between 1875 and 1945, respectively. There are 64, 43, and 49 countries that have at least one observation of actual primary, secondary, and tertiary enrollment data in the 19th century, respectively. Needless to say, all the observations for raw female enrollment ratios are novel.

Table 3
Number of countries with available actual enrollment ratios by year, 1820–1945.

Year	Total population			Female population		
	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary
1820–1845	33	56	78	22	53	65
1850–1870	56	57	79	25	42	55
1875–1895	78	74	87	48	34	48
1900–1920	97	84	90	51	26	38
1925–1945	109	100	98	11	49	69
Total 1820–1945	109	103	103	70	77	87

Note: The figure refers to the number of countries with available actual enrollment figures in the indicated years at five-year intervals for the sample of 111 countries.

observations.[12] However, there are a significant amount of data missing for many of the developing countries, and as such, we must construct estimates for a large number of missing cells.

We compute the proportions of the total population aged 5–9, 10–14, 15–19, and 20–24 at ten-year intervals from 1820 to 1940.

Fig. 1 shows the changes in the median values of the shares of four specific school-age populations (aged 5–9, 10–14, 15–19 and 20–24) within a subsample of advanced and developing countries.[13] As noted in previous studies (Benavot and Riddle, 1988), the proportion of the population consisting of primary school-age children (those 5–14 years old) has gradually declined in the advanced regions of the world because of declining birth rates. This downward trend is also present, though less noticeable, in the world's developing countries. There is also a reduction in the relative size of advanced countries' secondary school-age populations (those aged 15–19). Due in part to the limited number of observations in the earlier years, the secondary school-age populations of developing countries have exhibited large variations. The share of the tertiary school-age population (20–24) has been relatively stable over time.

When sufficient observations of school-age population shares are available, we fill in the missing observations by using linear interpolation. If a country does not have actual figures for 1820, then we estimate them by using the median value of the subsample to which the country belongs. However, only the advanced countries have enough actual observations for us to calculate these median values. For the developing countries, we use the actual median values from 1870 to 1910 to extrapolate an estimate of the median value for 1820. Once we estimate the school-age population shares for 1820, we can fill in the missing observations by interpolating between these estimates and the actual figures of the next available year.

For 41 developing countries and former colonies, there are no data on the population structure prior to 1950.[14] We fill in these missing observations by means of linear interpolation between the estimated school-age population distribution for 1820 (constructed from the available observations as described above) and the actual census figures recorded in 1950.

The assumptions we must make in order to estimate the missing observations, especially those necessary to extrapolate the figures for 1820, may generate biases; and as such, we run a number of counterfactual simulations. In the first simulation, we use the upper quartile values, instead of the median values, to generate the missing observations; the results of this simulation can be considered as an upper-bound counterfactual estimate. In the second simulation, we use the lower quartile values to get a lower-bound estimate.

We find that the estimates derived from these two simulations are similar to our baseline estimates. Furthermore, the resulting standard deviation estimates are rather small compared to the mean values of the school-age population ratios; therefore, we believe our estimation procedure provides reasonably accurate approximations. Consequently, a different set of assumptions regarding the extrapolation of the school-age populations in 1820 would not cause significant variations in the final primary and secondary enrollment estimates.

Using the estimated proportions of the population aged 5–9, 10–14, 15–19, and 20–24, we estimate the country-specific school-age population shares that correspond to each level of education, which vary according to the duration of schooling as established by the national education system.

The duration refers to the typical length of each country's primary and secondary education program. We construct a complete duration data series by taking the historical changes in each country's educational system into account. For the 1930–1955 duration data, we rely on UNESCO's *World Survey of Education*. We also compiled information about school duration in the early 20th century from Monroe (1911) and the European Commission's Eurybase[15] for a limited number of countries. We combined this data with that provided by Barro and Lee (2013) for the 1950–2010 period. Between 1930 and 1960, 41 countries changed the duration of either their primary or secondary education programs. For countries on which we could find no information regarding changes in the duration of their educational programs, we assumed that no such changes were instituted.

Next, we multiply these estimated shares by the total population (or female population), thereby calculating the number of children enrolled in each country-specific education level. The sample countries' total population data are from Maddison (2010); Klein Goldewijk et al. (2011); Banks and Wilson (2013), and the Wolfram Alpha (2013) database. When no data are available on the female population, we used the total population figure and the average ratio between the total and female populations to construct an estimate of the number of females.

The final enrollment estimates could be biased by our assumptions regarding the duration values in the earlier period. We fill in the missing duration observations from 1820 to 1945 by assuming that they are the same as the earliest available figures from the later period. The direction of changes in duration varies between individual countries. For some countries, the duration of primary or secondary schooling was extended after World War II, whereas it was shortened in others. In most cases, changes in the duration of primary and secondary schooling occurred simultaneously; for example, when primary duration increased, secondary duration decreased, and vice versa. If our assumed duration value is longer than the actual duration value, we would underestimate the enrollment ratio by overestimating the school age population. Nevertheless, this bias would be fairly insignificant for most countries where there was no significant change in schooling duration over time. In fact, there is very little variation in the duration of education across

[12] Throughout the paper, the group of "advanced countries" is defined as including Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.

[13] The structure and change of school-age population shares for the female and male population are similar to those for the total population. We use the procedures for the total population in order to construct estimates for the missing observations of school-age population shares for the female population.

[14] Most of these are former colonies, including Albania, Afghanistan, Bangladesh, Belize, Benin, Bolivia, Cambodia, China, the Democratic Republic of the Congo, Ecuador, Gambia, Ghana, Haiti, Hong Kong, Iceland, Indonesia, Iran, Jordan, Kenya, Kuwait, Lesotho, Liberia, Libya, Mali, Malta, Morocco, Nepal, Pakistan, Paraguay, Reunion, Senegal, Serbia, Sierra Leone, Sudan, Syria, Togo, Tunisia, Uganda, Yemen, Zambia, and Zimbabwe.

[15] Eurybase is a database on Europe's education systems (European Commission, 2014).

countries. For example, around 1930, the most common system of education (utilized by 59 countries in our sample) has six-years for primary duration and between five and seven years for secondary duration (that is, 6 + 5, 6 + 6, or 6 + 7 system). For 80 countries, primary education lasted between six and eight years. For 94 countries, secondary schooling lasted between five and seven years.

2.1.3. Adjustment with repetition ratios

School repeaters are defined as pupils who are enrolled in the same grade as they were in the previous year. We calculate the repetition ratio as the ratio of repeaters to the total number of students enrolled in a given education level. The repetition ratio is then used to adjust the gross enrollment ratio. A higher repetition ratio would generally indicate a lower quality of education or a lack of student effort. UNESCO regularly reports these measures in order to facilitate international comparisons between school systems (UNESCO, 1993). The repetition ratio would, however, also be influenced by variations in national education systems' promotion standards. For example, the low repetition ratios in many East Asian countries, including Japan, Korea, and Malaysia, are due to a policy of automatic promotion.

UNESCO provides data on repeaters for primary schools from 1965 onward and for secondary schools from 1970 onward. Additional data are gathered from Lockheed et al. (1991) and Lee and Barro (2001).

We must fill in the missing observations for the primary and secondary level repetition ratios over the period from 1820 to 1945 by using the earliest available values. Our assumption here is that there was no change in the repetition ratios between 1820 and the next earliest year when actual data are available.

This assumption about the missing repetition ratios could have a significant influence on our adjusted enrollment ratio estimates. The repetition ratios could be higher or lower in the 19th century than in the early 20th century depending on the characteristics of, and changes in the national education systems.

In order to test the influence of our assumption, we do a sensitivity test with two scenarios. For the first simulation, we use the lower quartile value in the distribution of the primary (or secondary) repetition ratios in place of the extrapolated values for 1820–1945 (which are greater than

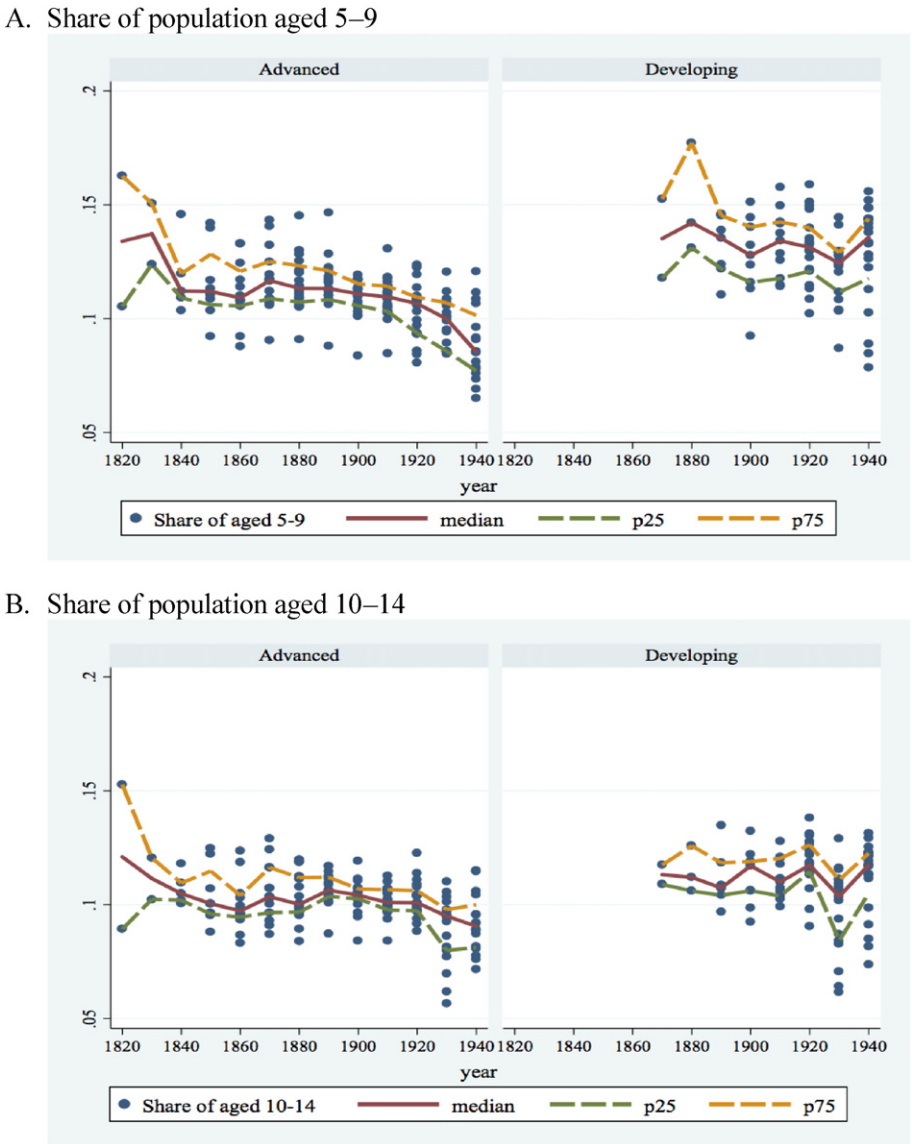
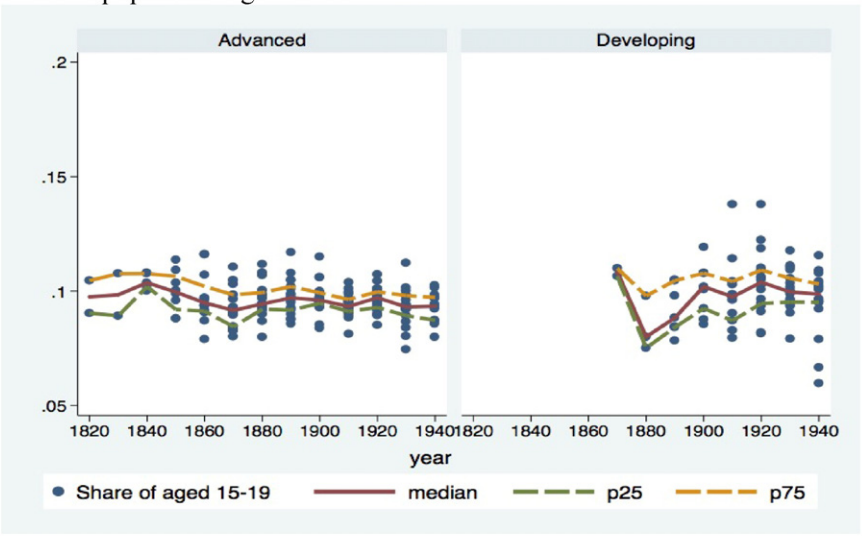


Fig. 1. Share of school-age population within the total population. The dots denote all available country and year observations for the share of population in the particular five-year age group (5–9, 10–14, 15–19, or 20–24 years old). The lines show the changes in the median and upper and lower quintile values of the population shares.

C. Share of population aged 15–19



D. Share of population aged 20–24

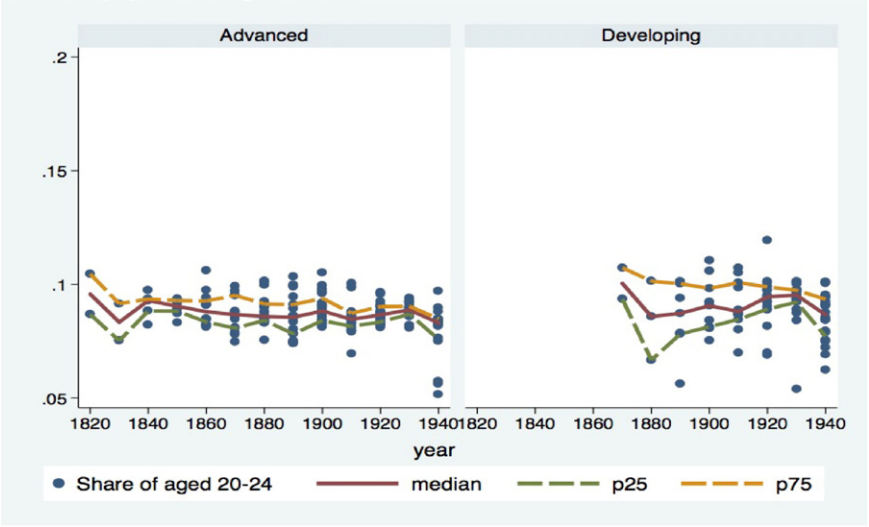


Fig. 1 (*continued*).

the lower quartile value).[16] Assuming that the repetition ratios are lower, this simulation should overestimate the enrollment ratios. The second simulation uses the upper quartile values, and should consequently underestimate the enrollment ratios.[17] The results of the two simulations could be considered as upper-bound and lower-bound counterfactual estimates.

Fig. 2 shows the final primary and secondary enrollment estimates for the 1820–1945 period, which are averaged (un-weighted) for each group of countries and compared to the simulated lower- and upper-bound estimates. The changes (as measured by the average standard deviations) in the secondary enrollment ratios are smaller because the mean values are smaller. Broadly speaking, our assumptions regarding the missing repetition ratios would not cause sizable biases in the enrollment ratio estimates. They would not result in significant variations over time either, unless the repetition ratios fluctuated over time.

2.1.4. Construction of missing enrollment data

Because complete enrollment data are not available for most countries from 1820 to 1945, we must construct estimates for a significant number of missing cells, especially for developing countries and former colonies. We use interpolation and extrapolation techniques to fill in the missing adjusted enrollment ratio observations. If the missing observation is located between two actual enrollment figures, we use linear interpolation to estimate the value. However, we employ interpolation based on a logistic trend between the first actual enrollment ratio observation and the next observation. When interpolation is not feasible, we also utilize a logistic trend to extrapolate the estimate. The logistic growth model assumes that each country's enrollment ratio ($enroll_j$) for education level j (primary, secondary, or tertiary) grows logistically over time (t), until it approaches the maximum ratio ($enroll_j^{max}$). That is, the enrollment ratios follow a logistic growth time trend such that:

$$enroll_{j,t} = enroll_j^{max} / (1 + \exp(-\alpha_j - \beta_j time)) \quad (1)$$

Hence, in order to estimate the enrollment ratios for 1820–1945, we fit the actual enrollment data from 1820 to 2010 to the following

[16] The lower quartile value of the primary repetition ratio distribution is 0.5 for advanced countries and 6 for developing countries. The value for secondary repetition is 5 for advanced countries and 3 for developing countries.

[17] The upper quartile value of the primary repetition ratio distribution is 23.5 for advanced countries and 21 for developing countries; for secondary repetition, it is 11.8 for advanced countries and 16.7 for developing countries.

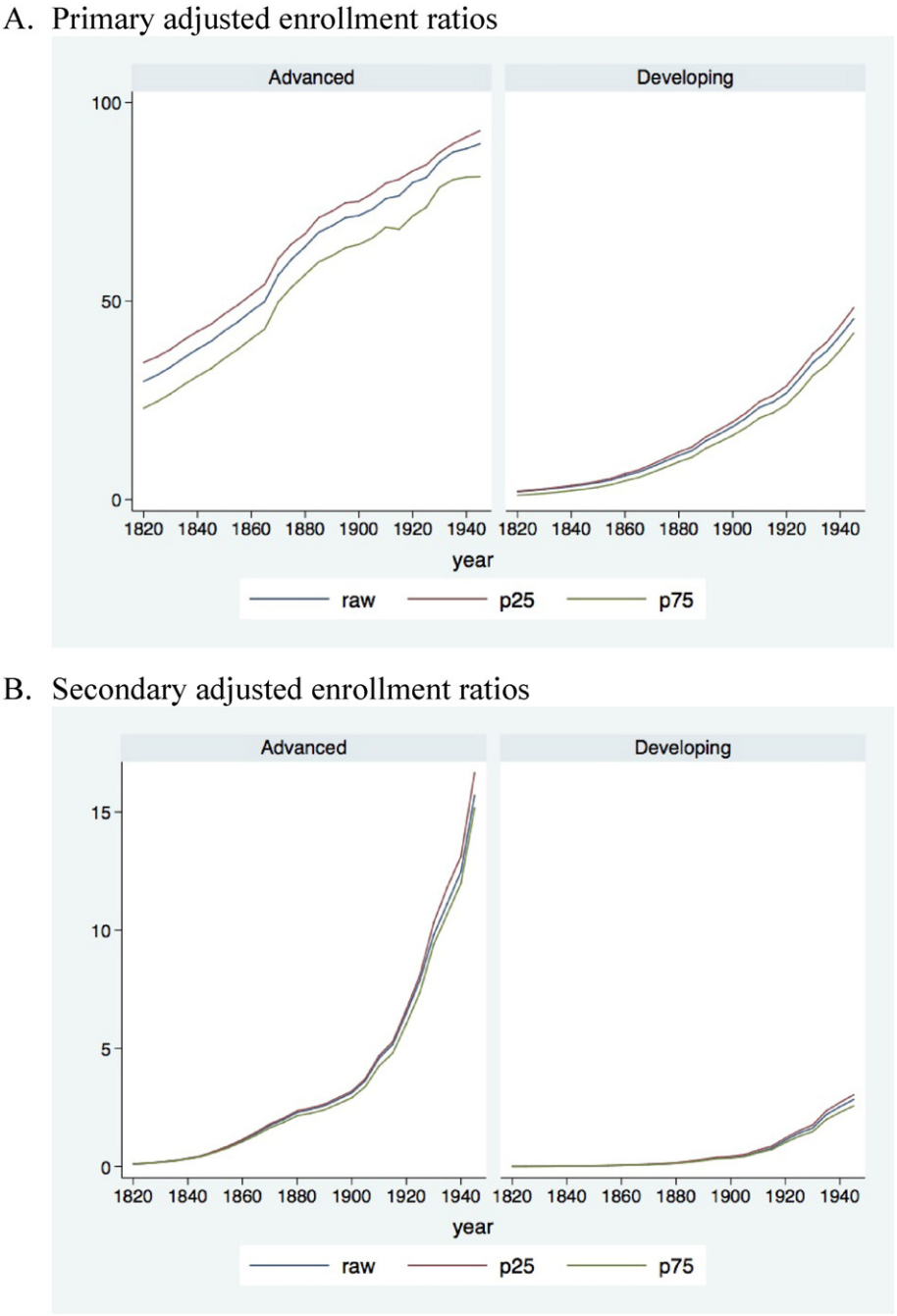


Fig. 2. Sensitivity analysis with primary and secondary repetition ratios. The raw estimates of adjusted enrollment ratios are calculated by assuming the same values for missing repetition ratios as the earliest available values from the later period. The simulations use the lower (p25) and upper quartile (p75) values from the distribution of the actual primary (or secondary) repetition ratios in place of the extrapolated repetition ratios. The values are the (un-weighted) averages of the primary and secondary enrollment ratios.

equation:

$$\ln\left(enroll_{j,t}/\left(enroll_j^{max}-enroll_{j,t}\right)\right)=\alpha_j+\beta_j time+\mu_{j,t} \quad (2)$$

where α_j is the constant term, β_j is the slope coefficient, and $\mu_{j,t}$ is the disturbance term. It should be noted that the slope of the logistic curve varies over time, changing in proportion to the size of the slope coefficient, β_j.[18] The estimation uses a panel data set on the different education level enrollment ratios for 111 countries (divided into advanced and developing countries) from 1820 to 2010 (recorded at five-year intervals). The dependent variable is the enrollment ratio at each education level for either the total population or the female population. As the unobserved, persistent characteristics unique to each country can influence the growth of enrollment ratios, we estimate the regression with country fixed-effects so that the intercept varies from nation to nation. Furthermore, the slope coefficient is also allowed to differ by country as well.

Fig. 3 presents the results of our regressions. It displays the estimates of the slope coefficients on individual countries' enrollment ratios for each education level. For both the total and female populations the mean and median values of the estimates are lower for primary enrollment than for secondary or tertiary enrollments. This implies that from

[18] In the logistic growth model, the growth rate is time-varying and equals to $\beta_j \cdot enroll_{j,t}(1-enroll_{j,t}/enroll_j^{max})$.

1820 to 2010, secondary and tertiary enrollment ratios have grown more rapidly than primary enrollment ratios. The total population estimates are higher in developing countries than in advanced countries at all education levels, implying that school enrollment ratios have grown more rapidly in developing countries. The estimates also indicate that at each education level, female enrollment ratios have grown more rapidly than total population (male) enrollment ratios.

Fig. 3 confirms that the estimated slope coefficients do not vary greatly between developed and developing countries. This, in turn, indicates that the standard deviations of the estimated coefficients are small compared to the mean and median values of the estimates at all education levels, for both the total and female populations. These results seem to indicate that the logistic growth model is effective in explaining the behaviors of individual countries' enrollment ratios at each education level, for both the total and female populations.

2.1.5. Estimation of enrollment ratios for the United States and Zimbabwe

Using the United States and Zimbabwe as illustrative examples, we explain the data and methodology used to estimate enrollment ratios. Table 4 summarizes the raw data on enrollment, population, and

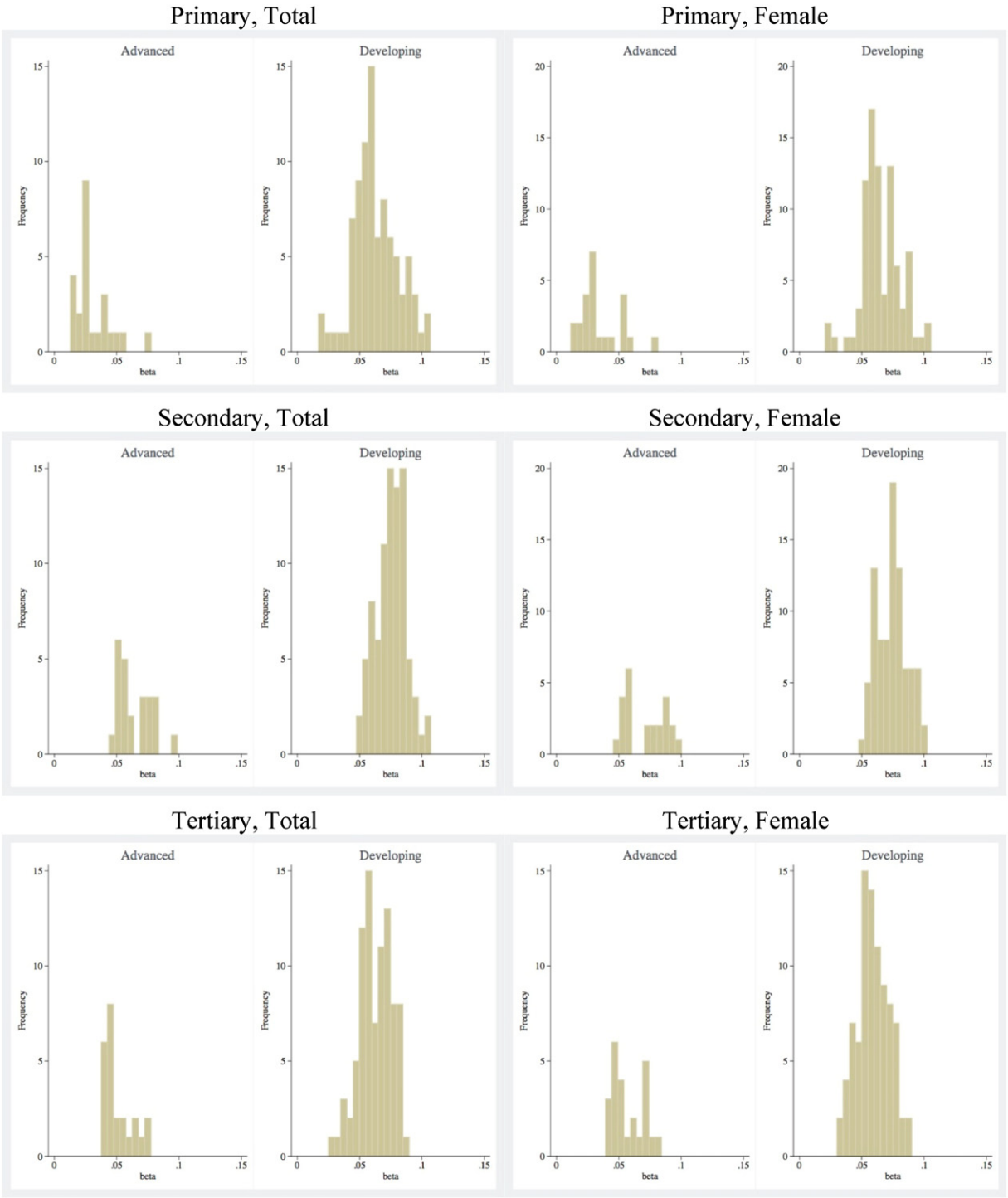


Fig. 3. Distribution of the estimated slope coefficients from the logistic trend regressions on the enrollment ratios Notes: The figures present the estimates of the slope coefficients (β) from the regressions on the enrollment ratios for each education level among the total and female population. The specification used in these regressions is the logistic time trend model (Eq. (1) in the text), which allows the coefficients to vary between countries.

Table 4
Enrollment, population, and enrollment ratios, 1820–1950.

	1820	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	1900	05	10	15	20	25	30	35	40	45	50
A. United States																											
(A) Enrollment (thousands)																											
Primary (duration: 8)		1087	1400	1712	2024	2833	3642	4667	5693	6505	7481	8693	9757	11,243	12,520	13,894	14,984	15,789	16,899	18,375	18,897	20,311	22,803	22,062	19,673	19,168	21,707
Secondary (duration: 4)							*38*				80	96	110	135	203	350	519	680	915	1329	2200	3757	5593	6436	6933	6327	6596
Tertiary			15	15	16	17	17	25	34	43	52	83	116	137	157	198	238	264	355	404	598	941	1154	1208	1404	1677	2302
(B) Enrollment (thousands)																											
Primary (duration: 6)		*935*	*1204*	*1472*	*1741*	*2436*	*3132*	*4014*	*4896*	*5594*	*6434*	*7476*	*8391*	*9669*	*10,767*	*11,949*	*12,886*	*13,579*	*14,533*	*15,582*	*15,949*	*16,899*	*18,698*	*17,738*	*15,562*	*15,219*	*17,344*
Secondary (duration: 6)							*548*				*1127*	*1310*	*1476*	*1729*	*1956*	*2295*	*2617*	*2890*	*3281*	*4122*	*5148*	*7169*	*9697*	*10,760*	*11,045*	*10,276*	*10,960*
(C) Population structure	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes
(D) Total population (millions)	9.6	*11.6*	12.8	*15.3*	17.1	*20.5*	23.2	*27.7*	31.3	*36.0*	38.5	*45.3*	43.4	*56.9*	62.5	*69.9*	75.5	*84.2*	91.4	*100.9*	105.1	*116.3*	122.1	*127.9*	131.0	*140.5*	157.8
(E) School-age population (thousands)																											
Primary (6–11 years old)	*1842*	*2088*	*2164*	*2550*	*2797*	*3339*	*3750*	*4373*	*4818*	*5471*	*5762*	*6675*	*6306*	*8165*	*8863*	*9713*	*10,292*	*10,986*	*11,406*	*12,696*	*13,325*	*14,419*	*14,803*	*14,180*	*13,172*	*14,031*	*15,657*
Secondary (12–17 years old)	*2092*	*2372*	*2461*	*2938*	*3266*	*3969*	*4537*	*5345*	*5951*	*6950*	*7528*	*8691*	*8181*	*10,753*	*11,851*	*12,965*	*13,712*	*15,131*	*16,271*	*17,409*	*17,546*	*19,493*	*20,557*	*21,185*	*21,353*	*20,806*	*21,024*
Tertiary (18–21 years old)	*1490*	*1903*	*2224*	*2675*	*2996*	*3606*	*4085*	*4845*	*5432*	*6168*	*6495*	*7665*	*7378*	*9730*	*10,760*	*11,925*	*12,782*	*14,354*	*15,710*	*16,651*	*16,613*	*18,359*	*19,258*	*20,320*	*20,973*	*20,949*	*21,803*
(F) Adjusted enrollment ratio																											
Primary	*41.4*	*42.9*	*53.4*	*55.4*	*59.7*	*70.0*	*80.1*	*88.0*	*97.5*	*98.1*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*
Secondary	*3.6*	*4.3*	*5.2*	*6.3*	*7.6*	*9.1*	*10.9*	*11.5*	*12.2*	*12.8*	*13.5*	*13.6*	*16.2*	*14.5*	*14.9*	*15.9*	*17.2*	*17.2*	*18.1*	*21.3*	*26.4*	*33.1*	*42.5*	*45.7*	*46.6*	*44.4*	*28.8*
Tertiary	*0.5*	*0.6*	*0.7*	*0.6*	*0.5*	*0.4*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*1.1*	*1.6*	*1.4*	*1.5*	*1.7*	*1.9*	*1.8*	*2.3*	*2.4*	*3.6*	*5.1*	*6.0*	*5.9*	*6.7*	*8.0*	*10.5*
B. Zimbabwe																											
(A) Enrollment (thousands)																											
Primary (duration: 5)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				15		54		127		112		236
Secondary (duration: 6)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0												3
Tertiary	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	*0*	0	0	0	0	0	0	0
(B) Population structure																											Yes
(C) Total population (millions)	*0.7*	*0.8*	*0.8*	*0.8*	*0.9*	*0.9*	*1.0*	*1.0*	*1.1*	*1.1*	*1.2*	*1.2*	*1.3*	*1.4*	*1.5*	*1.5*	*1.6*	*1.7*	*1.8*	*1.9*	*2.0*	*2.1*	*2.2*	*2.3*	*2.5*	*2.6*	2.7
(D) School-age population (thousands)																											
Primary (7–11 years old)	*92*	*96*	*101*	*107*	*112*	*118*	*124*	*130*	*137*	*145*	*152*	*160*	*169*	*178*	*187*	*198*	*209*	*220*	*232*	*246*	*259*	*274*	*290*	*307*	*324*	*343*	*363*
Secondary (12–17 years old)	*152*	*158*	*165*	*173*	*180*	*189*	*197*	*206*	*216*	*226*	*236*	*247*	*259*	*271*	*284*	*298*	*312*	*327*	*343*	*360*	*377*	*397*	*416*	*437*	*459*	*483*	*507*
Tertiary (18–21 years old)	*152*	*159*	*165*	*171*	*178*	*185*	*192*	*200*	*208*	*217*	*225*	*234*	*244*	*254*	*264*	*275*	*286*	*299*	*311*	*324*	*338*	*352*	*367*	*383*	*399*	*417*	*434*
(E) Adjusted enrollment ratio																											
Primary	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.8*	*2.4*	*5.1*	*9.6*	*20.5*	*31.4*	*48.6*	*65.9*	*58.9*	*51.9*	*58.9*	*66.0*
Secondary	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.1*	*0.1*	*0.1*	*0.2*	*0.4*	*0.5*
Tertiary	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

Note: The figures that are in italics are estimates using the methodology described in the text.

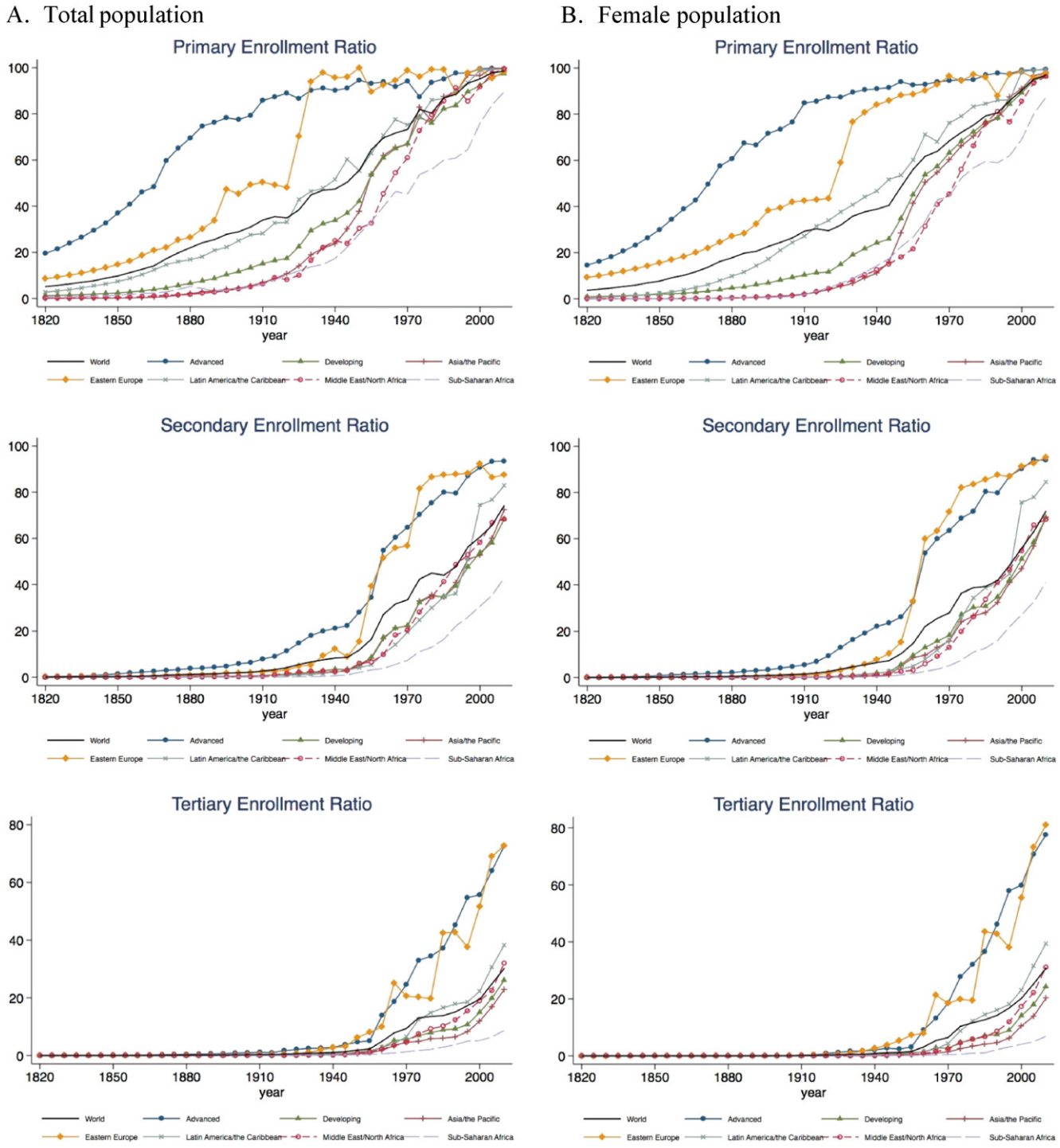


Fig. 4. Trends in (weighted-) average enrollment ratios by region, 1820–2010. The values are averages weighted by each country's school-age population at each education level.

population structure for the total populations from 1820 and 1945 (with missing observations estimated and observations recorded at five-year intervals).

Note: The figures that are in italics are estimates using the methodology described in the text.

Table 4A shows the information on the total population in the United States. Row (A) of Table 4A shows the availability of raw historical enrollment data for primary, secondary, and tertiary education. The data are compiled from various sources: For primary, Banks and Wilson (2013) are used for 1825–1865, 1875, and 1885; Mitchell (2003a, b, c) is used for 1870, 1880, and 1890–1925; and UNESCO's *World Survey of Education* provides 1930–1945. For secondary, Mitchell provides data for 1870, 1880, and 1890–1925; Banks and Wilson are used for 1875 and 1885, and UNESCO is used for 1930–1945. The U.S. Bureau of the Census (1975) provides data for 1850.[19] For tertiary, data for 1830–1865, 1875, 1885, and 1895 come from Banks and Wilson; Mitchell is used for 1870, 1880, 1890, and 1900–1925; and UNESCO is used for 1930–1945.

[19] The United States Bureau of the Census (1975, Series H 433–441) provides historical enrollment statistics. The figures are quite close to those reported here, which are drawn from international sources.

C. Male population

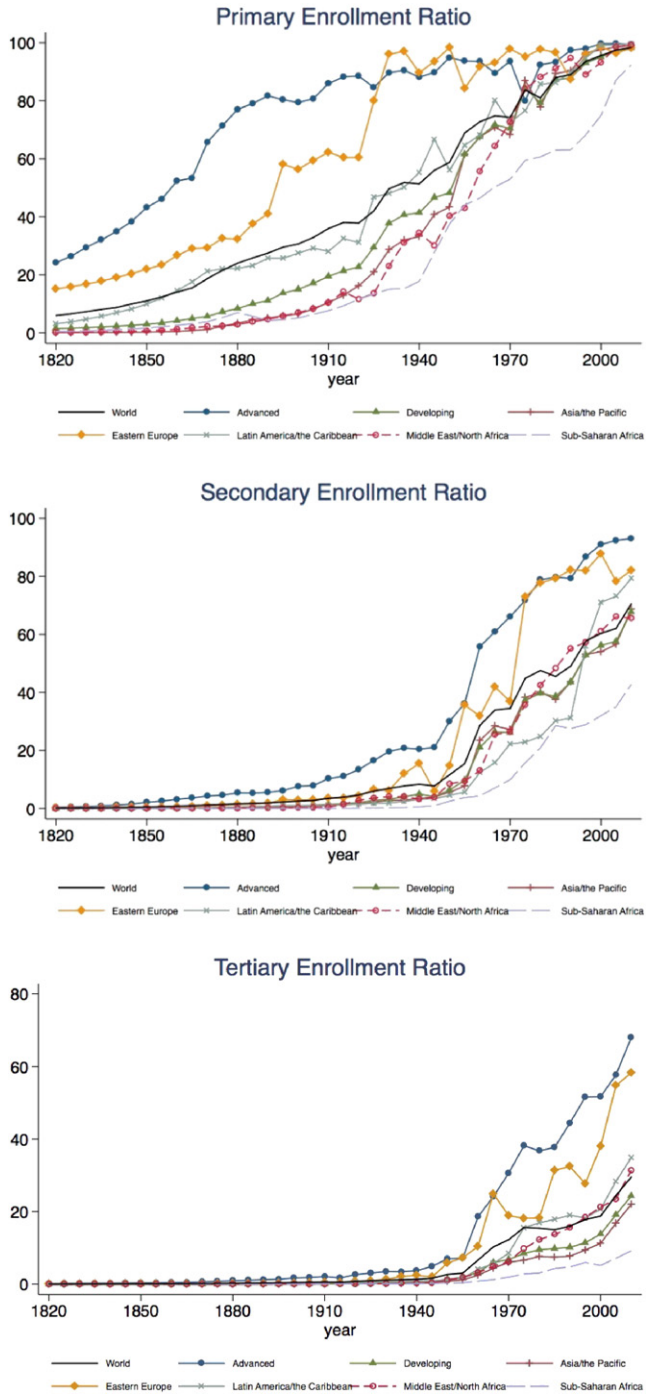


Fig. 4 (*continued*).

In the United States, raw data on historical enrollment in primary and secondary education are classified by the duration system of 8 + 4; that is, 8 years of primary and 4 years of secondary education. However, following UNESCO's international standard that classifies the United States' upper primary-level grades (seventh and eighth) as secondary-level education, we rearrange the enrollment by a duration cycle of 6 + 6.[20] This reallocation is also important for combining the historical enrollment ratio estimates with UNESCO's primary and secondary enrollment statistics from after 1950, which follow the 6 + 6 duration cycle.

In order to obtain seventh and eighth grade enrollment figures, we compare the enrollment data from Mitchell (2003a, b, c) with that from the U.S. Department of Education (1993). The U.S. Department of Education (1993) presents enrollment statistics by grade under the 8 + 4 duration system. Accordingly, we discover that the total enrollment figures reported therein are very close to those presented by Mitchell. We calculate the proportion of seventh and eighth grade enrollments to first through eighth grade enrollments and use this fraction to reclassify Mitchell's primary and secondary enrollment figures. Because the U.S. Department of Education (1993) has only reported annual data on the number of pupils per grade since 1910, we assume no change in the proportion between 1820 and 1910. The new primary and secondary enrollment estimates are shown in Row (B) of Table 4A.

We need the number of the school-age population enrolled in each education level in order to calculate the gross enrollment ratios. Row (C) shows the availability of population structure data (Mitchell, 2003a, b, c). The missing observations, as explained in Section 2, are linearly interpolated from the available figures. Row (D) shows the total population numbers (Mitchell, 2003a, b, c; Maddison, 2010). Using the total population and age structure data, we can calculate the school-age population for each education level. These estimates are shown in Row (E). Finally, the gross primary, secondary, and tertiary enrollment ratios are obtained by dividing the enrollment figures by the relevant school-age population. We then adjust the primary and secondary enrollment ratios by the estimated repetition ratios, which are 4.1% and 10.0% for the United States' primary and secondary schools, respectively. Consequently, we obtain the final adjusted enrollment ratio estimates, which are reported in Row (F). We assumed that the values that came in at over 100% are simply equal to 100%.

Table 4B shows the raw data availability and calculation of missing observations for Zimbabwe. The amount of historical data is very limited. The first census observation of primary enrollments is available from 1910. There are no observations for actual secondary and tertiary enrollments until 1945. However, by using information on the school establishment, we add zero actual enrollments for primary and secondary until 1885 and for tertiary until 1950. Notably, the secondary enrollment in the first observation in 1950 is very small, such that the estimation using a long interpolation between 1890 and 1950 would not be subject to a significant measurement error. Indeed, the enrollments in the first observations, especially at secondary and tertiary levels in developing countries, are very small for most cases. Thus, the interpolation or extrapolation for a long time span should be reasonable.

We can repeat the above procedures for the female population in the United States and Zimbabwe to construct estimates of female enrollment ratios, while the raw enrollment data are less available.

2.2. Trends in school enrollment, 1820–2010

Our complete data set includes comparable data on the total and female populations of 111 countries from 1820 to 1945 (reported at five-year intervals). We combine our historical estimates for this period with the actual figures from 1950 to 2010 in order to show the long-run trends in enrollment ratios.[21] Most of the world's major countries are included in our sample; such that, together, they accounted for approximately 96% of the world's total population in 1945.

Fig. 4 summarizes this data and shows the trends in the various adjusted enrollment ratios (among the different education levels and for

[20] UNESCO's *World Survey of Education* (Volume II, pp.1250–1251) points out that a 6-year duration is common for primary school, particularly in urban areas; additionally, grades 7 and 8, which are upper division or upper primary, can be considered as secondary.

[21] The actual enrollment ratios can be found in UNESCO's *Statistical Yearbook* (various years). The data for 1950 and 1955 are from the 1967 *Yearbook*, which reports "unadjusted enrollment ratios." We converted these figures to adjusted enrollment ratios by using UN data on the age structure of each country's population.

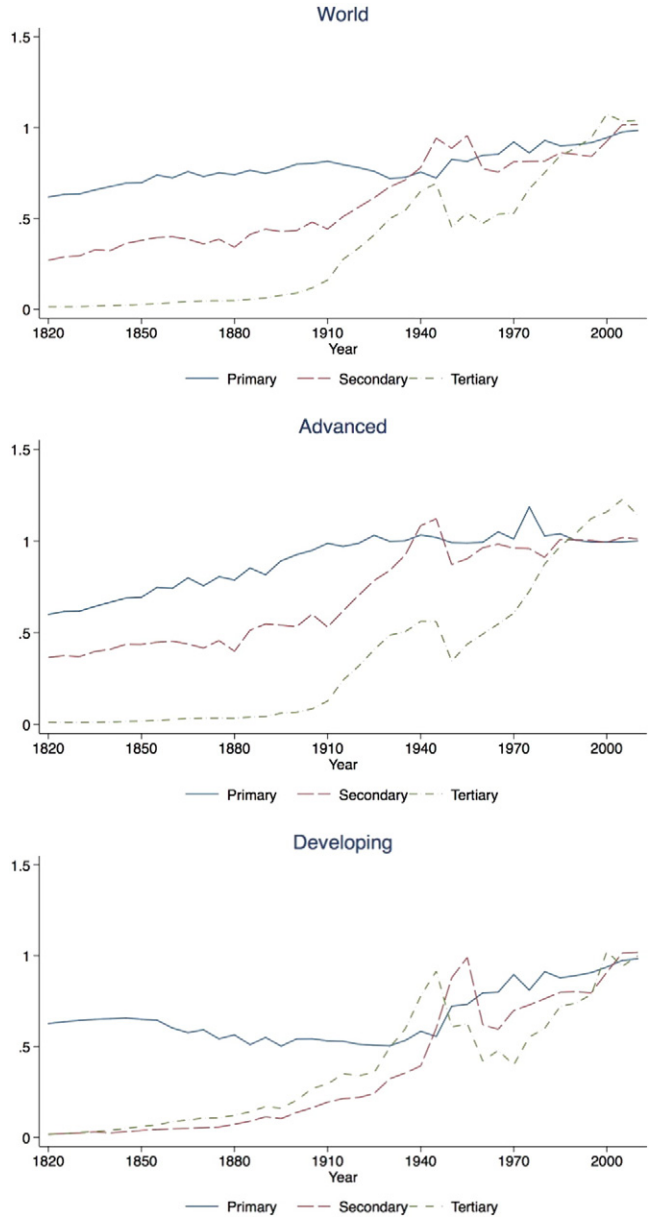


Fig. 5. Trends in gender enrollment ratios, 1820–2010, by economic development and for the world. The values are averages weighted by each country's school-age population at each education level.

the total, male, and female populations) from 1820 to 2010.[22] The figure presents information on two broad groups: "advanced countries" (of which there are 24) and "developing countries" (of which there are 87). The developing countries are further broken down into five regions: Asia/Pacific (17 countries), Eastern Europe (8), Latin America/Caribbean (25), the Middle East/North Africa (13), and Sub-Saharan Africa (24). The regional averages are weighted by each country's school-age population at each education level.[23]

Fig. 4A shows that the proportion of the school-age population attending school increased significantly after 1820. For the world as a whole, primary enrollment ratios increased from about 10% in 1850 to 47% in 1940. Furthermore, secondary enrollment ratios rose from 0.3% to about 8.3%, and tertiary ratios rose from nearly 0% to 1.1% over the same period.

Enrollment rates expanded significantly after World War II. In 1950, the world's adjusted enrollment ratios were approximately 56%, 12%, and 2.2% for primary, secondary, and tertiary education levels, respectively. By 2010, the figures jumped to almost 99% for primary, 74% for secondary, and 30% for tertiary.

Fig. 4A further shows the regional variations in the long-run growth of school enrollment ratios. In advanced regions of the world, primary education expanded rapidly in the 19th century followed by secondary and tertiary education in the 20th century. For most developing countries, modern primary education began to expand in the latter half of the 19th century or the early 20th century, while secondary and tertiary education grew steadily after World War II.

In the 24 advanced countries, primary education expanded rapidly with the spread of compulsory primary schooling, which occurred in the late 18th and early 19th centuries. The average primary enrollment ratio increased dramatically from about 37% in 1850 to about 90% in 1940. Secondary enrollment ratios also demonstrated strong growth, rising from 1.5% to roughly 21% over the same period. In contrast, the average tertiary enrollment ratio remained relatively low (around 3%) even through 1940. Fig. 4A shows that the growth in secondary and tertiary enrollments accelerated after World War II, reaching as high as 93% and 73%, respectively, in 2010.

Enrollments also showed significant growth (at all levels) in the 87 developing countries. The average primary enrollment ratio increased from a mere 3% in 1850 to 34% in 1940 and reached 97% in 2010. Secondary enrollment ratios, starting from a level of almost 0% in 1850, increased to 4% in 1940 and then jumped to 68% by 2010. The increase in primary and secondary enrollment ratios during the post-World War II period reflects global educational expansion, especially in large countries such as Brazil, China, and India. As of 1940, tertiary enrollment remained quite low and did not, in fact, exceed 1%. However, throughout the past half century, it has expanded rapidly, rising to 26% in 2010.

Among the developing regions, the expansion in school enrollments was most prominent during the late 19th and early 20th centuries in the Eastern European region, which includes Albania, Bulgaria, Hungary, Poland, and Russia. By 1940, the average enrollment ratios reached 96% for the primary level, 12% for the secondary level, and 2.8% for the tertiary level. This evolution reflects the relatively advanced economic and social systems present within these countries, as well as the enactment of strong, state-led educational development initiatives.

For most developing countries and former colonies in Asia, Africa, and Latin America, the expansion in education occurred later than it did in the Eastern Europe; although, Latin America had relatively high enrollment ratios for all three levels of education before World War II. Since World War II, there has been significant growth in primary, secondary, and tertiary enrollment ratios in all of the developing regions. Compared to other developing countries, though, those in Sub-Saharan Africa enjoyed lower growth in secondary and tertiary enrollment ratios, which still remain quite low.

Fig. 4A further shows that, although the developing countries made great progress in education over the past century and a half, their average enrollment ratios (at all education levels) lagged far behind those of advanced regions. The strong educational progress made by developing countries over the past 50 years has significantly narrowed the gap in primary enrollment ratios, almost bringing the developing world to the same level as the advanced. Yet, there are still substantial gaps in secondary and tertiary enrollment ratios.

Fig. 4B and C display the changes in the average primary, secondary, and tertiary enrolment ratios for the female and the male populations, respectively, in advanced and developing countries and different regions. They show that since 1850, significant progress has been made in both male and female school enrollment ratios in all regions, particularly in the advanced countries and five of the developing regions. For

[22] The complete data set, which is available as an Appendix online, presents the estimates for each country's total, female, and male populations at five-year intervals. The accompanying Appendix Fig. A also shows the figures for individual countries.

[23] The unweighted averages of enrollment ratios and gender gap show very similar patterns with the population-weighted averages in Fig. 4A, B, and C and Fig. 5.

the world as a whole, female primary enrollment ratios increased from about 8% in 1850 to 39% in 1940 and then jumped to 97% in 2010. As with the total population, secondary and tertiary enrollment ratios for the female population grew rapidly after World War II. The average female secondary enrollment ratio increased significantly from about 7% in 1940 to about 72% in 2010, while the ratio for tertiary education increased from 1% to 31%. Fig. 4C shows that the trends in male enrollment ratios over the past two centuries are generally similar to those in the female population.

Fig. 5 compares the enrollment ratios for females with those for males in advanced and developing countries. The values listed are the "gender ratios," which we define as the proportion of female-to-male enrollment ratio for each education level. In the advanced countries, the gender disparity in primary education was fairly low even in the 19th century, with a gender ratio of 0.7 in 1850. This ratio, along with those for both secondary and tertiary education, has improved steadily over the past century. The gender ratio rose steadily from 0.44 in 1850 to 1.01 in 2010 for secondary education, and from 0.02 to nearly 1.14 for tertiary education.

In both advanced and developing regions, gender inequality has declined steadily in all levels of education over the past 150 years. Notably, in the developing world, the gender gap in secondary and tertiary education has declined rapidly over the past few decades, resulting in almost equal enrollment ratios for both genders.

3. Estimating historical educational attainment

3.1. Data and estimation method

The estimated enrollment data demonstrate that the world has observed a great expansion in education over the past two centuries.

An interesting topic is the long-term evolution of educational capital stocks since the beginning of educational expansion in the 19th century. In order to investigate this pattern, we construct estimates of historical educational attainment for the total and female populations from 1870 to 1945. By utilizing the historical enrollment ratio estimates from 1820 to 1945, as well as the available census data from the later period on educational attainment by age, we are able to construct a complete data set of educational attainment estimates for the total and female populations of 111 countries from 1870 to 1945 (with observations at five-year intervals).

We divide educational attainment into seven categories and fill in observations for the total population aged 15–64 (subdivided into five-year age groups) from 1870 to 1945. We construct ten five-year age groups, ranging from 15–19 years old ($a = 1$) to 60–64 years old ($a = 10$).

First, we use actual census and survey observations of different age groups' educational attainment from 1950 onward (available from Barro and Lee (2013)'s data set) as benchmarks for backward extrapolation.[24]

Let $h^a_{j,t}$ denote the proportion of people in age group a for whom j is the highest level of schooling that they have attained: $j = 0$ for no school, 1 for primary, 2 for secondary, and 3 for higher education at time t. The process of backward extrapolation can then be expressed as:

$$h^a_{j,t} = h^{a+1}_{j,t+5} \tag{3}$$

where age group a can represent one of the 10 five-year age groups, such that $a = 2$ equates to the 20–24 age group and $a = 10$ equates to the 60–64 age group. This procedure assumes that the survival rate is the same within each five-year age group, regardless of a person's educational attainment. The overall survival rates for each five-year age cohort are reflected in the changes in each county's population structure over time, which we derive from population census data. When we extrapolate the educational attainment estimates for the over-65 age group from 1950 onward, we adjust the formula by considering differing mortality rates among the various education levels.[25]

For those aged 15 to 19 ($a = 1$), we use Eq. (3) to estimate tertiary educational attainment. For primary and secondary education, we adopt a different method that accounts for the fact that part of population is still in school during the transition period from t to $t + 5$. In order to do so, we use both estimates for that age group in $t + 5$ and the change in their (age specific) enrollment ratios over time.[26] Then, following the procedure set forth by Barro and Lee (2013), we divide the broad levels of schooling into completion and non-completion subcategories by using age-specific profile information on completion ratios in the nearest year.[27]

Next, we fill in any observations that are still missing by using information on previous school enrollment rates and the population's age structure. It should be noted that, following the "backward-extrapolation method," the estimate for the 20–24 age group in 1910 is filled in with the distribution of educational attainment for the 60–64 age group in 1950; as such, we cannot use the backward extrapolation method to construct estimates before 1910.

This estimation procedure (which we call "enrollment-based flow estimation") assumes that the enrollments for various levels of schooling among the 5–24 year-old population can produce (with the application of appropriate time lags) estimates of the current flows of attainment in the 15–19 and 20–24 year-old populations. We use the adjusted primary enrollment ratios at time t–5 and t–10, with the information on country-specific schooling duration and starting age, to estimate the share of the 15–19 year-old population that completed primary school. Similarly, the adjusted secondary enrollment ratios at times t and t–5 (along with country-specific secondary schooling duration data) are used to estimate the share of those with a secondary education among the population aged 15–19. The share of those with a tertiary education among this age group is calculated by multiplying the adjusted tertiary enrollment ratio at time t by the fraction of years that correspond to the duration of tertiary school (as established by the national education system's standards). Similar procedures are used for the 20–24 age groups, with appropriate durations and time lags taken into account.

We are then able to fill in the missing observations for the older age groups by extrapolating forward the "enrollment-based flow" estimates. We assume that the distribution of educational attainment among the 25–29 age group at time t is the same as that of the 20–24 age group at time t–5, and then apply the same procedure to the later age groups.

When both backward and forward extrapolations were feasible, we elected to fill the missing observations with backward extrapolation estimates; this is because they are derived from the original census and survey figures.

The data on the distribution of educational attainment among the population, combined with information on each country's population

[24] We use four additional censuses in the 1940s from the United States, Argentina, Bulgaria, and Cyprus as benchmarks for those countries. To the best of our knowledge, there are no other census data available for the years before 1945.

[25] The varying mortality rates by education level for the older age groups (65–69, 70–74, and 75–79) are from Barro and Lee (2013), who estimated different age groups' survival rates by utilizing information from available censuses from both before and after the time period in question. The estimation results show that more-educated individuals have lower mortality (higher survival) rates. See Barro and Lee (2013, Appendix Notes 1) for details.

[26] This procedure is based on the work of Barro and Lee (2013). Based on the *ex-post* simulation, we use in the estimation of the 1950–2010 data set, the backward extrapolations for the 15–19 age groups are repeated up to eight times using the actual census and survey observations from 1950 and thereafter for those groups.

[27] Dividing each broad period of schooling into completion and non-completion is required to construct a measure of average years of schooling in a more accurate and consistent manner over the whole time period. The estimates of "completion ratios" in Barro and Lee (2013) are based on drop-out ratios. Note that the completion ratios up to 1910 at the youngest age-cohort (e.g., age 20–24) are based on the backward extrapolation of the actual observation (e.g., age 60–64 in 1950) from the Barro–Lee data set.

Table 5
Trends in educational attainment of the population aged 15–64 for selected years.

Region (no. of countries) and year	Total					Female					Male				
	No schooling	Primary	Secondary	Tertiary	Average years of schooling	No schooling	Primary	Secondary	Tertiary	Average years of schooling	No schooling	Primary	Secondary	Tertiary	Average years of schooling
	(% of population aged 15–64)														
World (111)															
1870	90.0	9.0	0.9	0.1	0.49	91.6	7.8	0.6	0.02	0.42	88.5	10.2	1.2	0.2	0.55
1910	73.5	22.2	3.8	0.5	1.37	77.5	19.1	3.1	0.3	1.19	69.7	25.2	4.4	0.7	1.54
1950	47.9	36.8	13.1	2.2	3.20	52.8	33.5	12.0	1.6	2.88	43.1	40.0	14.2	2.8	3.54
2010	12.9	21.0	51.5	14.6	8.40	17.0	20.9	47.7	14.4	7.98	8.8	21.1	55.3	14.8	8.81
Advanced (24)															
1870	65.3	31.1	3.4	0.3	1.70	70.9	26.8	2.3	0.03	1.44	59.6	35.4	4.5	0.6	1.94
1910	28.5	57.9	11.8	1.7	3.94	34.5	53.8	10.7	1.1	3.61	22.5	62.1	13.0	2.4	4.27
1950	9.1	58.2	26.9	5.8	6.58	10.1	58.3	27.0	4.6	6.40	8.0	58.0	26.8	7.1	6.79
2010	1.2	10.0	53.1	35.8	11.94	1.5	10.2	51.2	37.2	11.94	0.9	9.7	54.9	34.5	11.94
Developing (87)															
1870	97.5	2.4	0.1	0.03	0.12	98.2	1.8	0.03	0.02	0.09	96.9	2.9	0.2	0.03	0.15
1910	89.3	9.7	0.9	0.1	0.47	92.9	6.7	0.4	0.1	0.33	85.8	12.6	1.5	0.2	0.61
1950	62.6	28.7	7.9	0.9	1.93	69.6	23.8	6.1	0.5	1.50	55.7	33.5	9.6	1.2	2.36
2010	15.1	23.1	51.2	10.5	7.72	20.1	23.0	47.0	10.0	7.21	10.4	23.2	55.4	11.1	8.22
Asia/the Pacific (17)															
1870	99.8	0.1	0.1	0.02	0.01	100.0	0.05	0.02	0.02	0.01	99.7	0.2	0.2	0.02	0.02
1910	95.7	3.6	0.6	0.1	0.18	98.7	1.1	0.1	0.02	0.05	93.0	6.0	1.0	0.07	0.30
1950	70.3	23.0	6.2	0.5	1.51	80.1	16.1	3.7	0.2	0.95	61.3	29.3	8.6	0.8	2.04
2010	15.4	22.1	54.5	8.1	7.51	21.3	22.5	49.4	6.8	6.85	9.7	21.6	59.4	9.3	8.15
Eastern Europe (8)															
1870	85.7	13.9	0.4	0.06	0.70	88.9	11.1	0.06	0.02	0.56	82.4	16.7	0.8	0.1	0.82
1910	67.0	30.6	2.1	0.3	1.47	75.3	23.7	0.9	0.2	1.19	58.0	38.1	3.4	0.5	1.76
1950	13.5	61.7	22.0	2.9	4.58	16.6	60.7	20.6	2.1	4.17	9.5	62.8	23.9	3.9	5.13
2010	0.5	3.5	52.1	44.0	11.84	0.5	3.3	48.1	48.2	11.91	0.5	3.6	56.3	39.7	11.75
Latin America/the Caribbean (25)															
1870	93.7	6.2	0.07	0.04	0.23	97.6	2.4	0.02	0.02	0.09	90.0	9.9	0.1	0.05	0.35
1910	73.7	23.8	2.2	0.4	1.16	76.9	21.0	2.0	0.1	1.02	70.5	26.6	2.3	0.6	1.29
1950	45.3	47.0	6.7	1.1	2.73	49.2	44.1	6.3	0.5	2.51	41.4	49.9	7.0	1.7	2.94
2010	5.6	32.3	48.6	13.4	8.68	6.0	31.2	48.5	14.3	8.73	5.3	33.5	48.8	12.5	8.62
Middle East/North Africa (13)															
1870	99.5	0.4	0.05	0.02	0.02	99.9	0.1	0.02	0.02	0.01	99.2	0.7	0.07	0.02	0.04
1910	96.6	3.1	0.3	0.06	0.15	98.1	1.7	0.1	0.03	0.08	95.1	4.4	0.4	0.1	0.21
1950	90.4	7.0	2.1	0.6	0.58	94.9	3.9	1.1	0.2	0.29	85.9	10.1	3.0	1.0	0.85
2010	22.9	22.8	40.7	13.7	7.43	29.4	19.9	38.3	12.5	6.93	16.5	25.7	43.0	14.8	7.92
Sub-Saharan Africa (24)															
1870	99.2	0.9	0.02	0.02	0.05	100.0	0.06	0.02	0.02	0.01	98.4	1.7	0.02	0.02	0.09
1910	95.7	4.2	0.1	0.04	0.22	98.4	1.5	0.1	0.04	0.09	92.9	6.8	0.2	0.05	0.36
1950	76.8	17.5	5.1	0.6	1.35	83.7	11.2	4.6	0.5	1.02	69.6	24.0	5.6	0.8	1.69
2010	32.0	33.6	31.6	2.8	5.50	37.6	32.1	28.0	2.4	4.95	26.5	35.1	35.2	3.3	6.03

Note: Regional averages are weighted by each country's population aged 15–64.

structure and the duration of each level of schooling, allow us to approximate the number of years of schooling completed by the average citizen.

The average years of schooling for the population aged 15–64 (S) is defined as

$$S = \sum_a \sum_j dur_j^a l_j^a,\ a = 15\text{–}19, \ldots, 60\text{–}64 \quad (4)$$

where dur_j^a is the duration of schooling level j (no formal education, incomplete primary, complete primary, lower secondary, upper secondary, incomplete tertiary, or complete tertiary) for population group a, and l_j^a is the fraction of population group a that has attained education level j.

We take into account the changes that countries have made over the years to the structure and duration of their education programs. As explained in Section 2, we collected information on schooling duration for the earlier period and constructed complete data series for our sample countries. We assume that, at the primary or secondary level, changes in the duration of schooling applied to students who had just entered primary or secondary school (students between the ages of 5 and 9 or 10 and 14) at the time of the change. For higher education, we set duration equal to four years for all countries and years, and we assign two years to persons who entered tertiary school but did not complete it.

Furthermore, by interpolating and extrapolating from the available figures, we also computed the population shares for each country from 1820 to 1950 at five-year intervals. We use the same sources and methodology to construct a panel data set on the educational attainment of females and males (subdivided by age group).

3.2. Trends in educational attainment

We now combine our estimates of educational attainment from 1870 to 1945 with the existing estimates in Barro and Lee's (2013) data set, which includes estimates for 146 countries from 1950 to 2010. Here, we summarize the educational attainment data for each region and for the world from 1870 to 2010. Table 5 shows data on the educational attainment at four broad levels of schooling of the total, female, and male population aged 15–64. The table also reports the average years of schooling achieved within each region.[28] Regional averages are computed by weighting each country's observation by its share of the regional population aged 15–64.

In 1870, global participation in formal education was quite limited; the average number of years spent in school was only 0.5 years for the total population aged 15–64. The proportion of the uneducated among the total population reached 90%. Furthermore, secondary and tertiary education was still undeveloped; only 0.9% of the total population had attended secondary school, and only 0.1% had some tertiary education. In 1870, the advanced countries were almost exclusively the participants in formal education. Among the 24 advanced countries, the population spent an average of 1.7 years in school. In the developing world, the average was only 0.1 years.

Table 5 further shows that education expanded greatly over the next 140 years, especially in the 20th century. For the world as a whole, the average years of schooling for the total population aged 15–64 increased from about 1.4 years in 1910 to 3.2 years in 1950 and 8.4 years in 2010. This dramatic increase in educational attainment reflects increases in school enrollment and completion ratios. Primary education began to spread in the early 20th century, and was followed by an expansion in secondary and tertiary education in the latter half of the century. The proportion of the total population that attained, but did not surpass the primary level of education increased from 22% in 1910 to 37% in 1950 and thereafter declined steadily to about 21% in 2010, as more primary school graduates continued on to secondary-level education. The proportion of those whose highest level of education was the secondary level increased rapidly from 13% in 1950 to 52% in 2010. The proportion

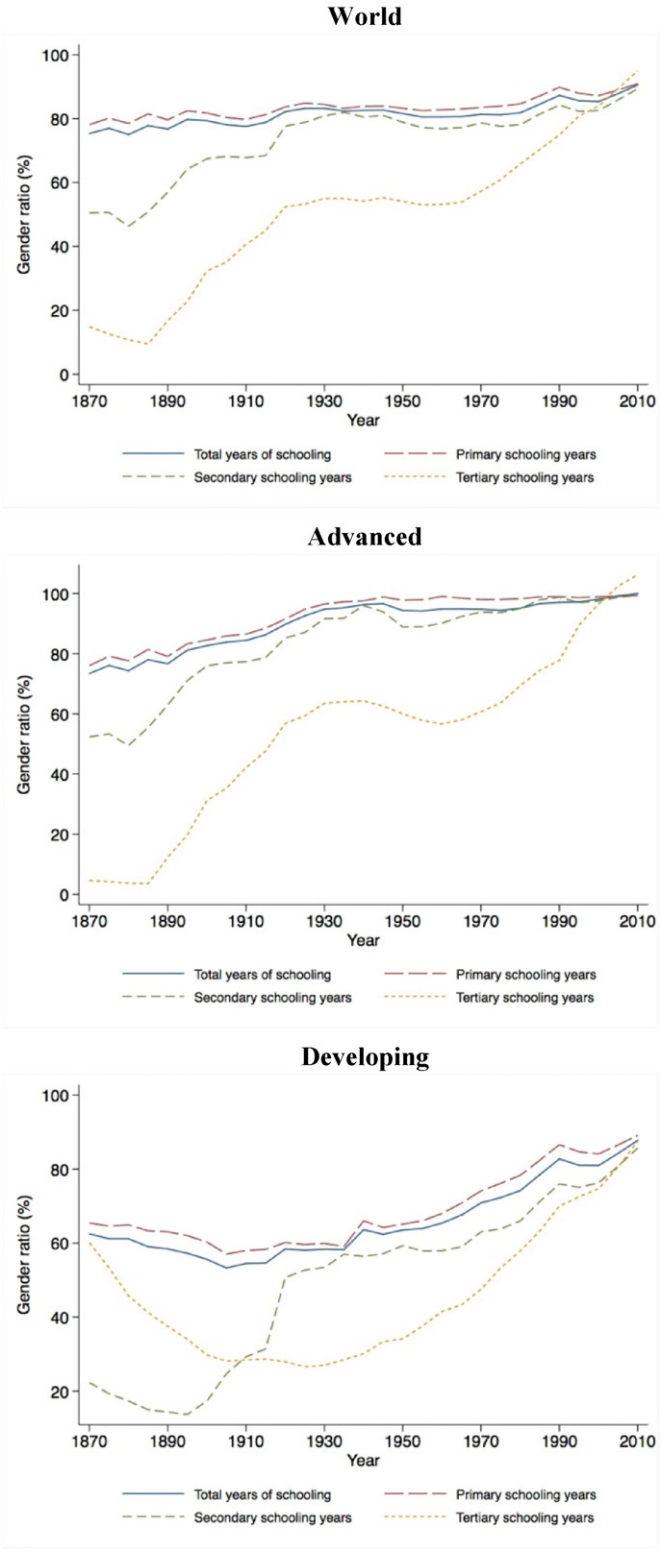


Fig. 6. Trends in gender ratios for educational attainment, by economic development and for the world, for the population aged 15–64. The values are population-weighted averages.

[28] The accompanying Appendix Table B presents the educational attainment data for the individual countries, following the format of Table 5. The accompanying Appendix Fig. B shows the graphs for the average years of schooling (at different education levels) for the total, female, and male population aged 15–64 for individual countries. The full data set presenting educational attainment for 1870–2010 at the individual country level is also available online. More detailed tables are available at a website (http://www.barrolee.com/Lee_Lee_LRdata.htm).

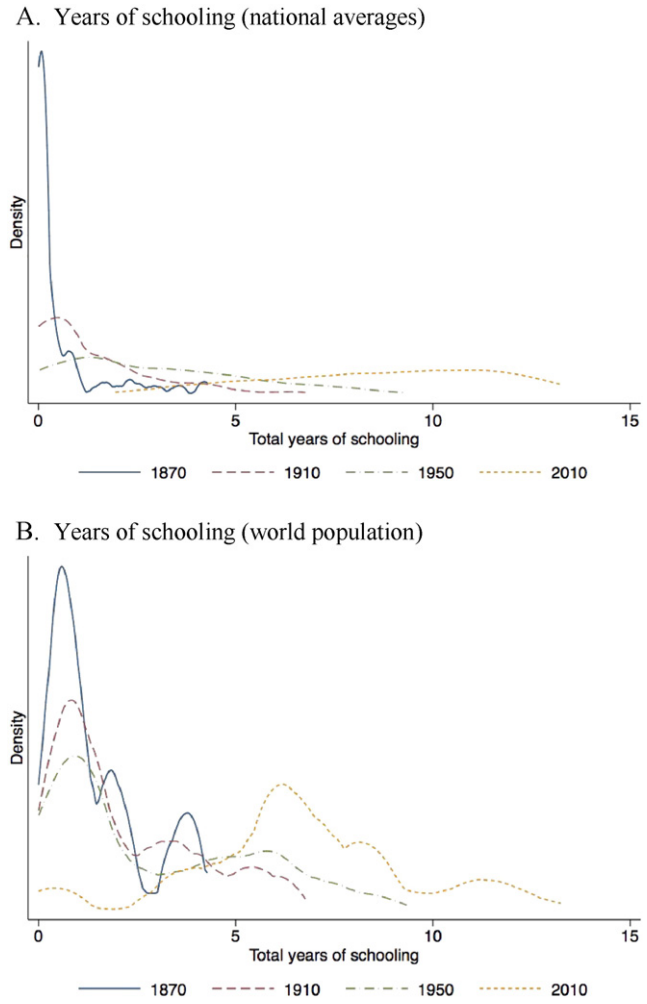


Fig. 7. World distribution of educational attainment for selected years.

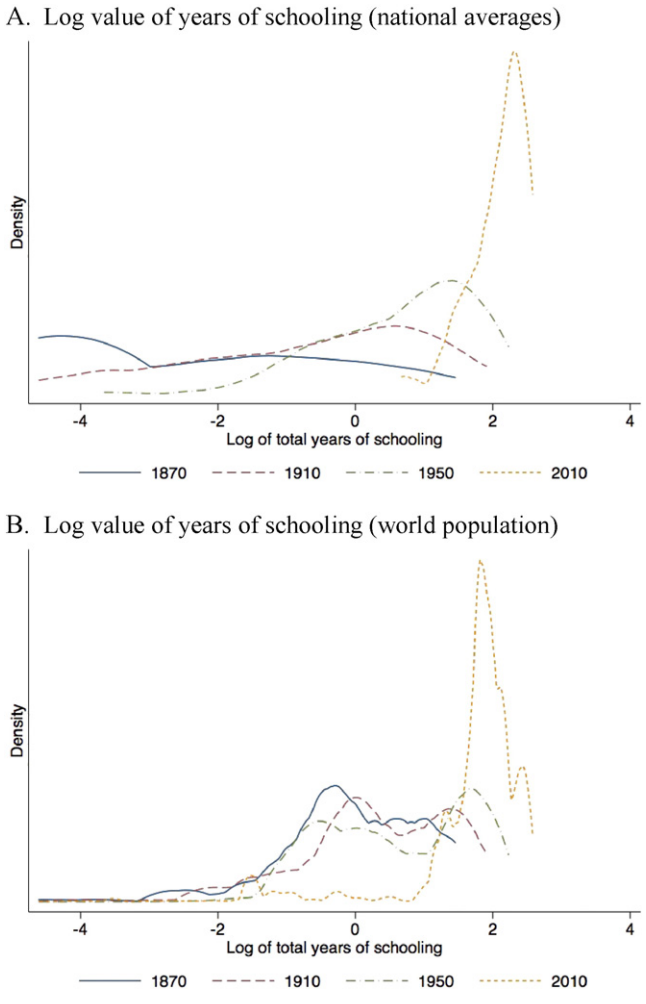


Fig. 8. World distribution of educational attainment (log scale) for selected years.

of the population with a tertiary-level education also rose significantly from 2.2% in 1950 to 15% in 2010.

This significant progress in educational attainment took place in both advanced and developing countries. In 2010, the overall population aged 15–64 in advanced countries was estimated to have an average of 11.9 years of schooling, a significant improvement upon the average of 3.9 years in 1910 and 6.6 years in 1950. In the developing countries, the average years of schooling reached 7.7 years in 2010, increasing from 0.5 years in 1910 and 1.9 years in 1950.

The increases in secondary and tertiary attainment, which occurred in most advanced countries in the post-World War II period, effectively led to the increase in the average years of schooling. The post-World War II increases that occurred in the developing countries, though, can be attributed to increases in primary and secondary educational attainment. Specifically, in developing countries, the proportion of the uneducated among the total population (aged 15–64) declined significantly from 89% in 1910 to 63% in 1950 and 15% in 2010. The share of those with a secondary education (as their highest level) increased dramatically from 1% in 1910 to 8% in 1950 and 51% in 2010. The improvements in the enrollment ratios among the younger cohorts contributed to the rise in the average years of schooling (though the effect was lagged).

Notwithstanding these significant improvements, the gap between developing and advanced countries still remains high, with a difference of 4.2 years in average education as of 2010. Despite this large absolute difference in average years of schooling, there has been a rapid decline in the proportional differences between the two groups. In 1910, the average years of schooling in the developing countries was only 12% of that in the advanced countries; but this ratio has increased rapidly over time to 29% in 1950 and 65% in 2010.

Table 5 shows the variations in the growth of educational attainment among the five developing regions over the past 140 years. Among the developing regions, the expansion of educational attainment was most prominent in Eastern Europe, where the average years of schooling for the total population increased from 1.5 years in 1910 to 4.6 in 1950 and 11.8 in 2010. By contrast, educational attainment in the Sub-Saharan African countries grew at a slower rate, rising from 0.2 years in 1910 to 1.4 years in 1950, and only reaching 5.5 by 2010. Additionally, the enrollment ratios in secondary and tertiary education in these countries are still very low compared to other developing nations (Fig. 4).

We also constructed data on educational attainment among the female population. Table 5 shows that the trend in educational attainment among the female population aged 15–64 is generally similar to that in the total population. In the 19th century, few females were educated; the proportion of uneducated women among the female population was 78% and they attended school for an average of 1.2 years in 1910. Female education expanded greatly over the 20th century; for the world as a whole, the average years of schooling increased to 2.9 years in 1950 and to 8.0 years in 2010.

In addition, there was a significant gap between advanced and developing regions in female education. In 1910, females attended school for an average of 3.6 years in advanced countries, while in developing

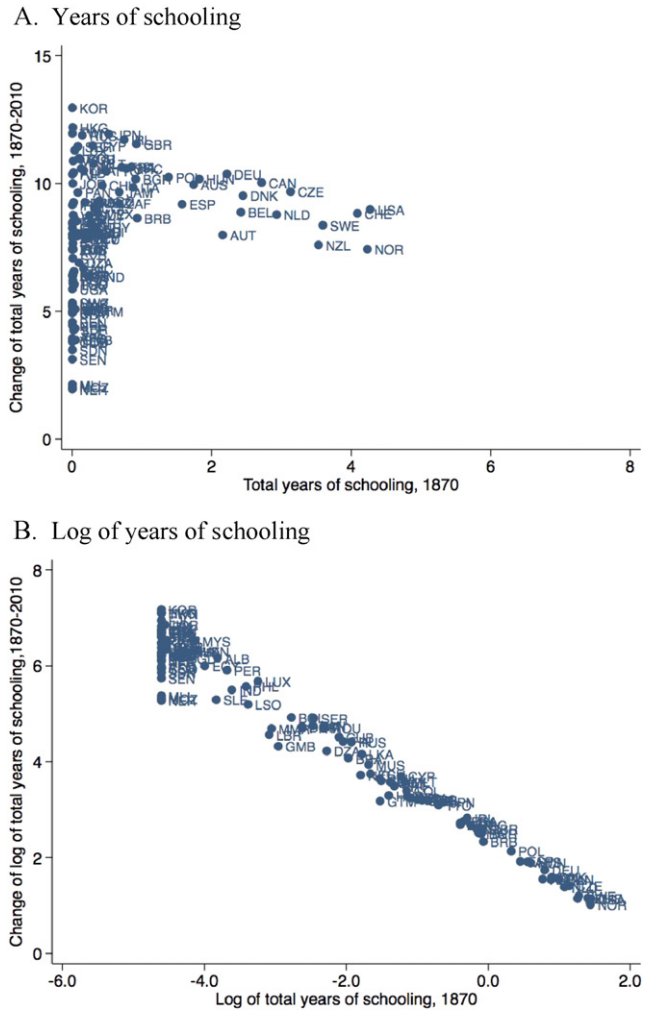


Fig. 9. Growth versus initial years of schooling, 1870–2010.

countries they only attended for 0.3 years on average. Since then, female education has expanded rapidly in both regions. In 2010, the average years of schooling reached 11.9 years in advanced countries, while it increased to 7.2 years in developing countries. It should be noted, though, that the growth in female educational attainment varied widely from region to region. In Eastern Europe, the average years of schooling increased from 1.2 years in 1910 to 11.9 years in 2010, while it increased from a mere 0.1 years to 5 years in the Sub-Saharan African countries over the same period.

Table 5 also shows the trends in educational attainment among the male population, which are similar to those in the total population in terms of improvements and regional variations.

Fig. 6 displays these results, comparing the regional figures (population-weighted averages) for females with those for males, from 1870 to 2010. It shows the "gender ratio," defined as the female-to-male ratio of average years of schooling (expressed as a percentage), at each level of education and overall, both worldwide and by group.

During the late 19th century, there was little improvement in gender equality throughout the world; in fact, in developing regions there was a slight increase in gender inequality. This result stems from the fact that, while the enrollment rates for both genders were initially quite low, male educational expansion began before that of females, and as such, the growth in male enrollment rates far outpaced those of females (Fig. 5). Fig. 6 shows that gender equality has improved across all regions in the 20th century. For the world population aged 15–64, the female-to-male ratio of average years of schooling rose from 78% in 1910 to 82% in 1950 and 91% in 2010. The ratio increased dramatically, from 55% in 1910 to 64% in 1950 and 88% in 2010 in the developing world. Over the past century, the ratios rose most dramatically in Asia/Pacific (17% in 1910 to 84% in 2010), Sub-Saharan Africa (24% to 82%), and the Middle East/North Africa (38% to 88%). In contrast, the ratios in the advanced countries, Eastern Europe, and Latin America remained high throughout the period and currently exceed 100% in some areas. Note that despite the significant improvements over time and toward gender parity in school enrollments (Fig. 5), the gender gap in the average years of schooling still remains in developing countries today.

Fig. 7A and B show the evolution of international educational disparities over the 1870–2010 period. Fig. 7A presents the estimated distributions of the average years of schooling among the 111 countries in 1870, 1910, 1950, and 2010. The rightward shift in the density plots shows the continuous increase in average years of schooling worldwide. As the share of the uneducated population fell over time, the concentration at the lower level became less pronounced. The spreading out of the density plots indicates that there has been a considerable increase in the dispersion of the average years of schooling across the countries; this is partly due to the increase in the mean values of the average years of schooling.

Since countries differ greatly in population size and average years of schooling, the distribution of the average years of schooling by nation is not necessarily the same as that by world population. Fig. 7B plots the evolution of educational disparities among the world population in 1870, 1910, 1950, and 2010. In this case, more populous countries, such as Brazil, China, India, and Russia, receive greater weight. The density plots in this figure show that the distributions of average years of schooling were less spread out compared to those in Fig. 7A, which relied on the national averages. The 1870 density plot shows that the majority of the world's population had very little education, though a fair proportion was able to complete primary school. The 2010 density plot shows that the majority of the world's population received an average of five to nine years of education.

Fig. 8A and B presents the distributions of the log of average years of schooling across countries and for the world population, respectively. The logarithmic scale is more appropriate for showing the proportional changes in the average number of years of schooling. The figures show a considerable decline in the proportional gap between 1870 and 2010, indicating that many developing countries have expanded their educational attainment to be more in line with advanced countries.

This pattern of "catching-up" in terms of relative scale is also clearly shown in Fig. 9A and B. Fig. 9A plots the changes in the average years of schooling over the 1870–2010 period against the average years of schooling in 1870 for our sample of 111 countries. There is no evidence of "convergence" of educational attainment across nations, in the sense that there is no negative relationship between the initial level and changes in income (Interestingly, though, this negative relationship is evident in the sample of advanced countries). However, there has been a wide variation in educational growth performance, especially among the countries with low levels of educational attainment in 1870.[29] By contrast, in Fig. 9B, where in the variables are expressed using a logarithmic scale, there is strong evidence of "convergence" of educational attainment across nations.[30]

3.3. Comparison with alternative estimates

We now compare our estimates of educational attainment for the overall population aged 15–64 with those developed by Morrisson and Murtin (2009) and Van Leeuwen and Van Leeuwen-Li (2014),

[29] The disparities in per capita income between the poor and rich regions of the world have long been a topic of interest for economists. For example, the debate over "convergence" focuses on whether the cross-country disparity in per capita income has decreased over time (Barro and Sala-i-Martin, 2004, Chapter 1).

[30] The values in 1870 are truncated so that 0.01 is the minimum.

Table 6
Comparison between the average years of schooling estimates in the new data set and in Morrisson and Murtin (2009) and Van Leeuwen and Van Leeuwen-Li (2014).
Source: Authors' calculations are based on their own data and on the data sets of Morrisson and Murtin's (2009) and Van Leeuwen and Van Leeuwen-Li (2014).

A. This data set and Morrisson and Murtin	Obs	Correlation	This data set			Morrisson and Murtin		
			1870	1910	1940	1870	1910	1940
World								
Levels	560	0.93	0.79	2.08	3.35	1.64	2.93	3.82
			(1.17)	(1.74)	(2.37)	(1.76)	(2.50)	(2.81)
10-year differences	490	0.56		0.36			0.29	
				(0.29)			(0.28)	
Advanced countries								
Levels	176	0.86	1.70	3.94	6.03	3.51	5.86	7.22
			(1.40)	(1.69)	(2.03)	(1.96)	(2.23)	(2.24)
10-year differences	154	0.17		0.57			0.48	
				(0.30)			(0.33)	
Developing countries								
Levels	384	0.91	0.12	0.55	1.22	0.27	0.53	1.12
			(0.32)	(0.88)	(1.42)	(0.47)	(0.97)	(1.45)
10-year differences	336	0.71		0.17			0.13	
				(0.24)			(0.2)	

B. This data set and Van Leeuwen and Van Leeuwen-Li	Obs	Correlation	This data set			Van Leeuwen and Van Leeuwen-Li		
			1870	1910	1940	1870	1910	1940
World								
Levels	678	0.92	0.49	1.37	2.77	1.29	2.22	3.15
			(1.05)	(1.59)	(2.19)	(1.71)	(2.29)	(2.68)
10-year differences	587	0.64		0.33			0.26	
				(0.29)			(0.26)	
Advanced countries								
Levels	179	0.86	1.70	3.94	6.03	3.35	5.58	7.04
			(1.39)	(1.67)	(1.99)	(1.91)	(2.15)	(2.18)
10-year differences	156	0.34		0.57			0.49	
				(0.30)			(0.25)	
Developing countries								
Levels	449	0.89	0.12	0.47	1.62	0.61	0.98	1.73
			(0.46)	(0.92)	(1.48)	(0.54)	(1.22)	(1.65)
10-year differences	432	0.68		0.24			0.17	
				(0.26)			(0.23)	

Notes: Obs indicates overlapping observations. This data set consists of a total of 1665 observations (for 111 countries from 1870 to 1940 at five-year intervals); Morrisson and Murtin's (2009) data set includes 692 observations (for 74 countries from 1870 to 1940 at ten-year intervals), and that of Van Leeuwen and Van Leeuwen-Li (2014) includes 795 observations (for 116 countries from 1920 to 1940 at 10-year intervals). The numbers in parentheses are standard deviations.

who constructed historical data on educational attainment since 1870 at 10-year intervals. The methodology that they used is similar to our "enrollment-based flow estimation"; they compiled historical enrollment data, and then used it to generate their educational stock measures for the total population.

We have compiled a greater number of actual (non-estimated) enrollment ratios than Morrisson and Murtin (2009) and Van Leeuwen and Van Leeuwen-Li (2014). More importantly, our estimates were extrapolated backward from actual census and survey observations on attainment by age group. Furthermore, our estimates are also classified by gender.

Table 6 shows the means and standard deviations of the levels and ten-year differences in average years of schooling for the sample of overlapping observations between our data set and that of Morrisson and Murtin (2009) for the 1870–1940 period. There is a high correlation between the levels of the two estimates, with a correlation coefficient of 0.93, but the ten-year differences are less correlated, with a correlation coefficient of 0.56.

The estimates for advanced countries are less correlated than those for developing countries in both the levels and ten-year differences. The correlation between the two estimates is particularly low (0.17) for the ten-year differences in the sample of advanced countries.

Morrisson and Murtin's estimates for the advanced countries are, on average, higher than ours for the overall period. Fig. 10 plots the samples' common estimates for the years 1870 and 1910 against each other. They are closely correlated, but some differences can be seen.

We also compare our estimates of educational attainment with those of Van Leeuwen and Van Leeuwen-Li (2014). Given that the majority of their estimates are from Morrisson and Murtin, the comparison between our data set and that of Van Leeuwen and Van Leeuwen-Li shows similar correlations in terms of both levels and 10-year differences. In Table 6, the correlations between the two estimates, i.e., ours and Van Leeuwen and Van Leeuwen-Li's, are 0.34 and 0.68 for the 10-year differences in the sample of advanced countries and developing countries, respectively. The samples' common estimates for the years 1870 and 1910 present values that are similar to the scatterplots in Fig. 10.

We suspect that the discrepancies between our estimates and those of Morrisson and Murtin or Van Leeuwen and Van Leeuwen-Li for these advanced countries originate mainly from their use of literacy rates in gauging years of schooling for the earlier years in the 19th century.[31] Literacy, (i.e., the ability to read and write a simple short statement), is an alternative measure of educational attainment. It can take account of education outside formal schools. However, an accurate conversion of literacy into equivalent enrollments or years of formal education is a significant challenge. This is because of the cases where a group of people without formal schooling may be able to read and write a simple sentence, and the one even with formal schooling can remain illiterate. As presented in Van Leeuwen and Van Leeuwen-Li (2014), in the United States in 1947, about 20% of the people who could not receive any formal education were literate and, after 1 year of formal schooling, only 33% of them became literate. Moreover, literacy concepts tend to be

[31] See the Appendix of Morrisson and Murtin (2009).

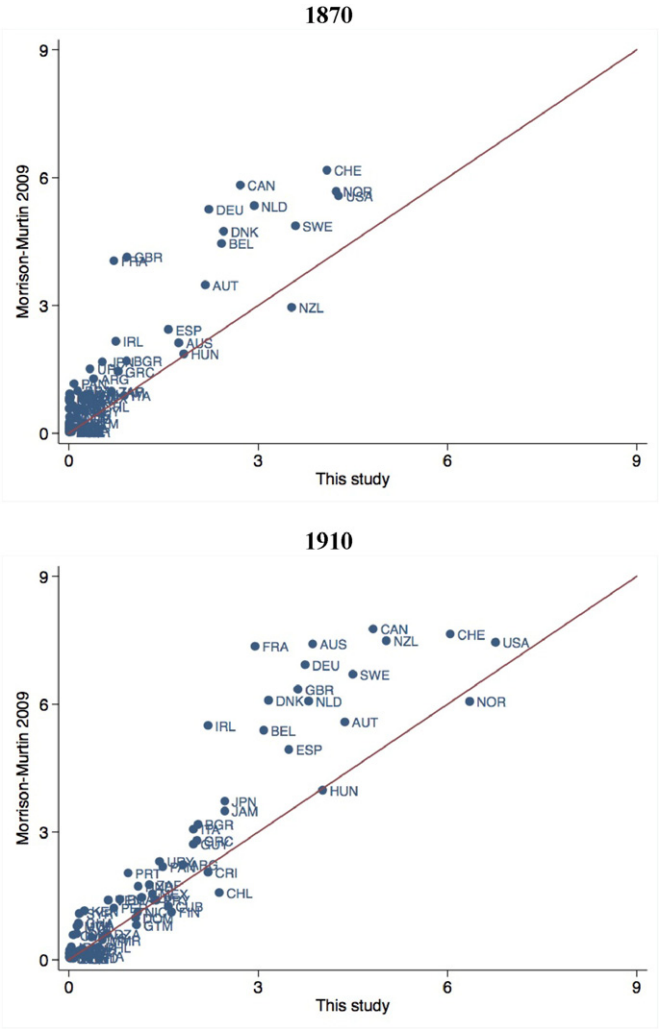


Fig. 10. Comparison between the average years of schooling estimates in the new data set and in Morrisson and Murtin (2009), for 1870 and 1910.

less comparable over time and across countries, as the measure is not based on a common and consistent criterion (Barro and Lee, 1993).

While our estimates do not take account of home and informal schooling and thereby underestimate true educational attainment in the early 19th century, the potential biases must disappear over time as formal schooling becomes the prevailing form of education and comprises different types of traditional and informal education. In contrast, the estimates by Morrisson and Murtin and Van Leeuwen and Van Leeuwen-Li can be subject to upward biases, which may persist.[32]

4. Estimating human capital stock per worker

This section discusses the construction of a measure of human capital stock that reflects the educational attainment of a nation's population.[33] The average years of schooling figure (S) calculated in Equation (4) provides an internationally comparable measure of human capital stock per person. Equation (4) assumes that uneducated workers do not provide any human capital and that human capital stock increases proportionally with the duration of schooling.

An alternative measure of human capital stock (h) assumes a Mincerian log-linear relationship between the number of years of schooling and human capital, such that

$$h = \sum_a \sum_j e^{\theta_j^a \left(dur_j^a\right)} l_j^a, a = 15\text{–}19, \ldots, 60\text{–}64 \quad (5)$$

where dur_j^a is the duration of education level j for population group a, and l_j^a is the fraction of population group a that has attained educational level j (no formal education, incomplete primary, complete primary, lower secondary, upper secondary, incomplete tertiary, and complete tertiary).

In this equation, $\theta_j^a(dur_j^a)$ measures the efficiency of a unit of labor from population group a with education level j, relative to one from a person without any schooling. Consequently, the derivative of $\theta_j^a(dur_j^a)$ can be interpreted as the marginal return to an additional year of schooling at level j. If we assume that this relationship is linear, then Eq. (5) can be rewritten as

$$h = \sum_a \sum_j e^{\theta_j^a dur_j^a} l_j^a \quad (6)$$

This measure indicates that human capital per worker is equal to the sum of the shares of workers, weighted by relative wage rates, across all education categories. If the marginal rate of return to a year of schooling is constant for all education levels and countries, and everyone has the same level of schooling, Eq. (6) can be further simplified to $h = e^{\theta S}$, where S represents the mean years of schooling.[34]

It should be noted that Eq. (6) assumes that human capital stock for uneducated workers is constant across countries, and equal to 1. This assumption of internationally identical productivity among uneducated workers has been challenged by Manuelli and Seshadri (2014). According to them, if early childhood investments (e.g., medical care and nutrition) are important determinants of an individual's early levels of human capital, then uneducated workers' human capital stock would be greater in wealthier countries than in poorer ones. They suggest that a six-year old from a country in the bottom decile would have less than 50% of the human capital of a child from the United States. However, there is no strong evidence that the productivity of uneducated workers varies from country to country or that it is greater in wealthier nations. Jones (2014) presents conflicting evidence regarding this issue. Using the U.S. census from 2000, Jones shows that mean wages for uneducated workers born in the United States are about 17% lower compared to those for immigrants from the very poorest countries. Using Mexican microdata, he further shows that this wage gap is not driven by a bias stemming from immigrants' endogenous decisions.

In Eq. (6), the marginal return to an additional year of schooling at level j can be derived from education/wage profiles It is often assumed that the marginal rate of return to an additional year of education (measured by the wage rate) is constant at 10%, which is the world average of the estimates compiled by Psacharopoulos (1994). However, this specification does not control for the quality of education. If we consider that the marginal rate of return to an additional year of education may be positively correlated with education quality, then the 10% rate of return may be an overestimate.

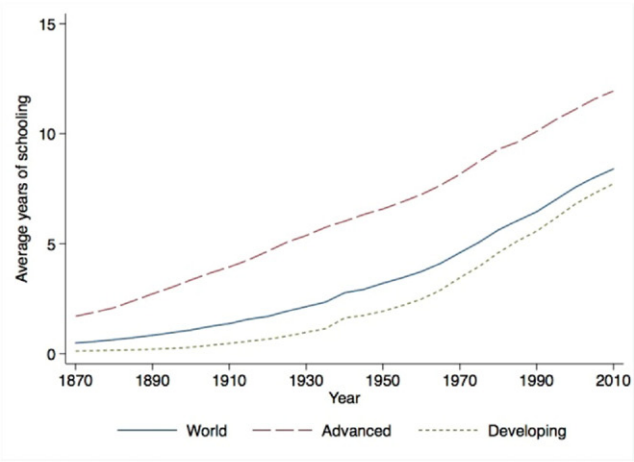
Bratsberg and Terrell (2002) and Schoellman (2012) use the estimated returns to immigrants' schooling in the United States to measure international differences in educational quality. They assume that the country-specific return to schooling that foreign-educated immigrants receive must be associated with the quality of education in the

[32] Morrisson and Murtin acknowledged that their estimates for some advanced countries, including France and Canada, for 1960 are outliers.

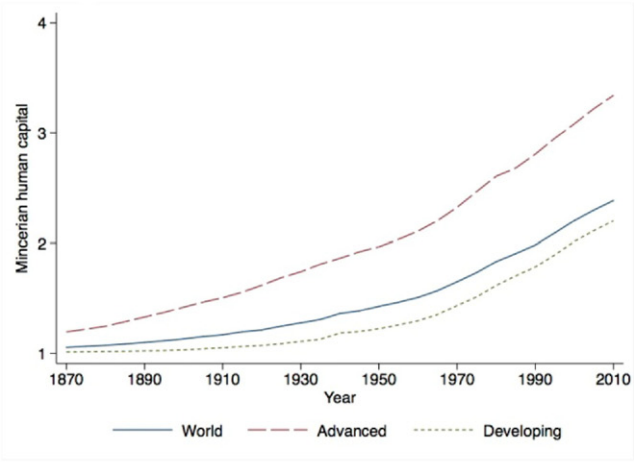
[33] Our measure of human capital stock from schooling is obviously a very limited concept owing to the exclusion of other forms of human capital, such as early childhood, health, and experience.

[34] Klenow and Rodríguez-Clare (1997) and Hall and Jones (1999) use this simple specification to construct measures of human capital stock.

A. Years of schooling



B. Human capital stock (with the assumption of perfect substitutability)



C. Alternative human capital stock (with the assumption of imperfect substitutability)

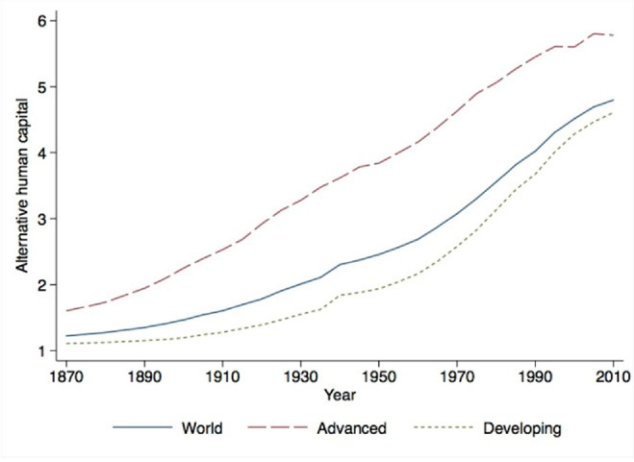


Fig. 11. Trends in average years of schooling and the human capital measures, 1870–2010, worldwide and by group, for the population aged 15–64. The human capital stock measure is constructed by assuming that different types of human capital are perfect substitutes. The alternative human capital stock measure assumes imperfect substitutability between skilled and unskilled labor with the elasticity of substitution, equal to 2. The values are population-weighted averages.

immigrants' birth countries. Using 2000 U.S. census data, Schoellman (2012) measures the average rate of return among 130 countries to be 11.1%, which is close to the world average rate of return. However, the estimates vary by country. Immigrants from countries with higher quality education (which is measured using primary and secondary school students' internationally comparable test scores) tend to enjoy greater returns in the U.S. labor market.[35] While these results highlight the importance of educational quality in determining earnings, these estimates are subject to selection bias due to endogenous immigration decisions.

We assume that the relative wage rate of a worker with schooling S is determined by the constant marginal rate of return to an additional year of schooling, which we set equal to 10% (the world average rate of return).[36]

In regard to the aggregate human capital stock measure (h), we have assumed that different types of human capital are perfect substitutes, which allows easy calculation. However, the perfect substitutability assumption can cause biases, if skilled workers are not easily substituted for unskilled workers. This is because the productivity of unskilled workers will increase over time with the educational expansion which brings the scarcity of the unskilled. This potential bias would lead to an underestimation of the true differences in human capital across countries (Jones, 2014). Empirical evidence also rejects the perfect substitutability between different education levels (Goldin and Katz, 2009, Ciccone and Peri, 2005, and Jones, 2014).

For comparison, we build an alternative human capital stock measure based on the assumption of imperfect substitutability between different skill types. We assume that the aggregate labor input can be thought of as a constant-elasticity-of-substitution (CES) aggregate of two types of labor — unskilled and skilled labor. We classify that everyone who has attained upper-level secondary schooling is skilled, and those who have not are unskilled.[37] Hence, unskilled labor (h_u) is a weighted sum of the first four categories, from no education to some secondary, while h_s is a weighted sum of the other three categories, from upper secondary attained to completed tertiary education.

$$\begin{aligned} h &= \left[h_u^{\rho} + h_s^{\rho}\right]^{1/\rho} \\ &= \left[\left(\sum_a \sum_{j=1}^{4} e^{\theta_j^a dur_j^a} l_j^a\right)^{\rho} + \left(\sum_a \sum_{j=5}^{7} e^{\theta_j^a dur_j^a} l_j^a\right)^{\rho}\right]^{1/\rho}, \end{aligned} \tag{7}$$

where a = 15–19,…, 60–64, j (1 = no formal education, 2 = incomplete primary, 3 = complete primary, 4 = lower secondary, 5 = upper secondary, 6 = incomplete tertiary, and 7 = complete tertiary), ρ is the substitution parameter, and the elasticity of substitution between skilled and unskilled labor (σ) is $1/(1-\rho)$.

In this specification, workers are assumed to be perfectly substitutable within each subgroup. It is practically difficult to estimate the elasticity of substitution between all different skill types, for example, between uneducated workers and primary educated workers.

Micro-evidence analyzing the elasticity of substitution between skilled and unskilled labor suggests the estimate in the range of between 1 and 2 (Ciccone and Peri, 2005; Jones, 2014). We set the

[35] Hanushek and Woessmann (2008) survey various studies that assess the impacts of educational quality on individual earnings both within and across countries.

[36] We can assume that the rates of return to schooling vary by education level. Caselli (2005) and Hall and Jones (1999) suggest that rates of return to schooling (θ) change according to the value of education; for instance, $\theta = 0.13$ for $s \le 4$, $\theta = 0.10$ for $4 < s \le 8$, and $\theta = 0.07$ for $8 < s$. The human capital stock measure based on these values is similar to one based on the uniform 10% rate of return, with a correlation coefficient of 0.98 for the 111 countries from 1870 to 2010. Furthermore, we can also utilize the country-specific rates of return. For example, Caselli and Ciccone (2013) compile Mincerian return rates for 95 countries prior to 1990. The correlation coefficient between the human capital stock measure using the uniform Mincerian rate and that using the country-specific return rates is very high in the common sample over 1870 to 2010.

[37] We can use Caselli and Coleman (2006)'s alternative definition of skilled labor by completed primary schooling. The results discussed below are similar with this classification.

substitution elasticity (σ) equal to 2 (i.e., $\rho = 1/2$) to construct human capital stock for the imperfect substitutability case.[38]

Fig. 11 shows the changes in the human capital stock measures with alternative values for the substitution parameter ($\rho = 1$ or $1/2$) from 1870 to 2010, in the developing and advanced regions as well as the world as a whole, compared to the changes in the average years of schooling.[39] Fig. 11.B and C show that since 1870, human capital accumulation, regardless of the values for the substitution elasticity, has increased dramatically in both the advanced and developing regions of the world. A lower elasticity of substitution, however, tends to amplify the differences in human capital stock across economies as well as over time. Because the log value of the human capital stock measure with perfect substitutability equals the average years of schooling (multiplied by 0.1), the proportional change (i.e., log difference) in the measure should be the same as the change in the absolute difference in the average years of schooling. Likewise, the patterns of change in the distribution of the logged Mincerian human capital stock measures should be the same as those in the average years of schooling (as shown in Fig. 8).

5. Concluding remarks

This study constructed new data sets on enrollment ratios and educational attainment for a large number of countries. The resulting data set on historical enrollment ratios contained information on 111 countries for the 1820–1945 period with observations recorded at five-year intervals. The data refer to adjusted enrollment ratios among school-age populations for both sexes at primary, secondary, and tertiary education levels.

By utilizing the estimates of historical enrollment ratios, as well as the available census data on educational attainment by age, we then constructed a complete data set of educational attainment estimates for the total and female populations of 111 countries from 1870 to 1945 (with observations, once again, recorded at the five-year intervals). These estimates were further disaggregated by gender and age group.

In order to analyze the trends in educational attainment and school enrollment ratios throughout the world, we combined the historical estimates with the available data from 1950 to 2010. Based on estimates of rates of return to schooling, we also presented a measure of human capital stock per worker.

The data show strong growth in average educational attainment and human capital stock throughout the world and across regions over the long run. The data also show that while the absolute differences in average educational attainment (measured by total years of schooling) between nations and among the world's population have been persistent, there has been a rapid decline in the proportional differences over the past two centuries. The data also show that gender equality in education has significantly improved across regions in the 20th century.

The need to construct estimates that were comparable across countries and over such a long period forced us to confront several fundamental issues that stemmed from a lack of accurate information on the classification of different types of schools, country-specific school-age populations, school repeaters and dropouts, mortality rates, and so on. In order to construct our data sets, we had to make several assumptions, some of which were critical to the estimations. There is a possibility that our final estimates may suffer from measurement errors. Our estimates of school enrollments and educational attainment may not appropriately include education at home and non-formal traditional schools and thereby underestimate true educational attainment particularly in the 19th century. Educational attainment and enrollment ratios only measure the quantity of school, but do not capture the quality of students' education. As such, there is still room for us to improve upon our data set.[40]

This study reveals the remarkable achievements in educational expansion throughout the world over the past two centuries. During this period, many countries also experienced significant progress in economic development. The existing long-term data indicate that the rate of income growth accelerated rapidly over the past 200 years. This increased growth was accompanied by rapid capital accumulation and technological progress. Furthermore, it was also associated with decreases in fertility and mortality rates (Galor, 2005). Investment in human capital is also influenced by income growth, as well as changes in infant mortality rates. The interactions among income, human capital accumulation, and fertility rates must have played an important role in many countries' transitions from low-income, uneducated societies to high-income, well-educated societies and their structural change from rural and agricultural to urban and industrial. In addition, the level of educational attainment is considered to have a strong impact on social and political outcomes, such as democracy and the rule of law.

We believe our new data sets on educational attainment and human capital stock can contribute to the exploration of the effects of human capital on economic, social, and political development throughout history.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jdeveco.2016.05.006.

Acknowledgements

We would like to thank Robert J. Barro and anonymous referees for helpful suggestions and Eunbi Song for able assistance in research. The full data set utilized in this study is available at a website (http://www.barrolee.com/Lee_Lee_LRdata.htm).

References

Baier, S.L., Dwyer JR., G.P., Tamura, R., 2006. How important are capital and total factor productivity for economic growth? Econ. Inq. 44, 23–49.
Banks, A.S., Wilson, K.A., 2013. Cross-national Time-series Data Archive. Databanks International, Jerusalem, Israel, Israel.
Barnard, H., 1854. National Education in Europe, Charles B. Norton, New York.
Barro, R.J., 1991. Economic growth in a cross section of countries. Q. J. Econ. 106, 407–443.
Barro, R.J., 2012. Convergence and Modernization Revisited NBER Working Paper No. w18295.
Barro, R.J., Lee, J.W., 1993. International comparisons of educational attainment. J. Monet. Econ. 32, 363–394.
Barro, R.J., Lee, J.W., 2001. International data on educational attainment: Updates and implications. Oxf. Econ. Pap. 53, 541–563.
Barro, R.J., Lee, J.W., 2013. A new data set of educational attainment in the world, 1950-2010. J. Dev. Econ. 104, 184–198.
Barro, R.J., Lee, J.W., 2015. Education Matters: Global Schooling Gains From The 19th to 21st Century. Oxf. Univ. Press, New York.
Barro, R.J., Sala-i-Martin, X.I., 2004. Economic Growth. second ed. MIT Press, Boston.
Becker, G.S., 2009. Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education. Univ. Chic Press, Chicago.
Benavot, A., 1996. Education and political democratization: a cross-national and longitudinal study. Comp. Educ. Rev. 40, 377–403.
Benavot, A., Riddle, P., 1988. The expansion of primary education, 1870–1940: trends and issues. Sociol. Educ. 61, 191–210.
Bratsberg, B., Terrell, D., 2002. School quality and returns to education of US immigrants. Econ. Inq. 40, 177–198.
Caselli, F., 2005. Accounting for cross-country income differences. In: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth Vol. 1. Elsevier Amsterdam, Netherlands, pp. 679–741.
Caselli, F., Ciccone, A., 2013. The contribution of schooling in development accounting: results from a nonparametric upper bound. J. Dev. Econ. 104, 199–211.
Caselli, F., Coleman, W.J., 2006. The world technology frontier. Am. Econ. Rev. 96, 499–522.
Ciccone, A., Peri, G., 2005. Long-run substitutability between more and less educated workers: evidence from US states, 1950–1990. Rev. Econ. Stat. 87, 652–663.
Clio-Infra, 2014. https://www.clio-infra.eu (access February 16, 2015).

[38] An important caveat is that the estimates of the elasticity of substitution in the existing literature are based on microdata that are from advanced countries, and identify skilled workers with the tertiary educated.

[39] The complete data on human capital stock measures for individual countries are available online.

[40] Quality-adjusted human capital stock can be constructed by taking into account the differences in the quality of education that individuals received in each year of schooling. Barro and Lee (2015) present a data set on educational quality (as measured by primary and secondary schools' aggregate math and science test scores) that includes 134 countries and spans from 1965 to 2010 (at five-year intervals).

Cohen, D., Soto, M., 2007. Growth and human capital: good data, good results. J. Econ. Growth 12, 51–76.
De La Fuente, A., Doménech, R., 2006. Human capital in growth regressions: how much difference does data quality make? J. Eur. Econ. Assoc. 4 (1), 1–36.
European Commission, 2014. Eurybase: The European Community Database on Education. Eurydice, Sweden.
Földvári, P., van Leeuwen, B., 2014. Educational and income inequality in Europe, ca. 1870–2000. Cliometrica 8 (3), 271–300.
Fredriksen, B., 1983. On the collection of education statistics in developing countries: purpose, principles, and problems. Int. J. Educ. Dev. 3, 291–304.
Galor, E., 2005. The transition from stagnation to growth: unified growth theory. In: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth Vol. 1. Elsevier, Amsterdam Netherlands, pp. 171–293.
Goldin, C.D., Katz, L.F., 2009. The Race Between Education and Technology. Harvard University Press, Cambridge, MA.
Hall, R.E., Jones, C.I., 1999. Why do some countries produce so much more output per worker than others? Q. J. Econ. 114, 83–116.
Hanushek, E.A., Woessmann, L., 2008. The role of cognitive skill in economic development. J. Econ. Lit. 46, 607–668.
Jones, B.F., 2014. The human capital stock: A generalized approach. Am. Econ. Rev. 104 (11), 3257–3777.
Klein Goldewijk, L., Beusen, A., van Drecht, G., de Vos, M., 2011. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. Glob. Ecol. Biogeogr. 20, 73–86.
Klenow, P., Rodríguez-Clare, A., 1997. The neoclassical revival in growth economics: has it gone too far? In: Bernake, B.S., Rotemberg, J.J. (Eds.), National Bureau of Economic Research Macroeconomics Annual. MIT Press, Boston, pp. 73–114
League of Nations, various years, Statistical Yearbook of the League of Nations, League of Nations, Geneva, Switzerland.
Lee, J.W., Barro, R.J., 2001. Schooling quality in a cross–section of countries. Economica 68, 465–488.
Lindert, P., 2004. Growing Public. Cambridge University Press, Cambridge.
Lockheed, M.E., Verspoor, A.M., Associates, 1991. Improving Primary Education in Developing Countries. Oxford University Press for the World Bank, New York.
Maddison, A., 2010. Historical Statistics of the World Economy: 1–2008 CE. http://www.ggdc.net/maddison/oriindex.htm (access July 1, 2013).
Manuelli, R.E., Seshadri, A., 2014. Human capital and the wealth of nations. Am. Econ. Rev. 104 (9), 2736–2762.
Meyer, J.W., Ramirez, F.O., Soysal, Y.N., 1992. World expansion of mass education, 1870–1980. Sociol. Educ. 65, 128–149.
Mincer, J.A., 1974. Schooling, Experience, and Earnings. National Bureau of Economic Research, Cambridge.
Mitchell, B.R., 2003a. International Historical Statistics: Africa, Asia and Oceania 1750–2000. fourth ed. Palgrave Macmillan, London, UK.
Mitchell, B.R., 2003b. International Historical Statistics: The Americas 1750–2000. fifth ed. Palgrave Macmillan, London, UK.
Mitchell, B.R., 2003c. International Historical Statistics: Europe 1750–2000. fifth ed. Palgrave Macmillan, London, UK.
Monroe, P., 1911. A Cyclopedia of Education. Gale Research Company, Detroit.
Morrisson, C., Murtin, F., 2009. The century of education. J. Hum. Cap. 3, 1–42.
Psacharopoulos, G., 1994. Returns to investment in education: a global update. World Dev. 22, 1325–1343.
Psacharopoulos, G., Woodhall, B., 1985. Education for Development: An Analysis of Investment Choices. Oxford University Press for the World Bank, New York.
Schoellman, T., 2012. Education quality and development accounting. Rev. Econ. Stud. 79, 388–417.
Tamura, R., Dwyer, J., Devereux, J., Baier, S.L., 2012. Economic growth in the long run MPRA Paper no. 41324.
U.S. Bureau of Education, Report of the Commissioner of Education, U.S. Government Printing Office, Various years, Washington, DC.
U.S. Bureau of the Census, 1975. Historical Statistics of the United States, Colonial Times to 1970. U.S. Government Printing Office, Washington, DC.
U.S. Department of Education, 1993. National Center for Education Statistics, 120 Years of American Education: A Statistical Portrait. U.S. Government Printing Office, Washington, DC.
United Nations Demographic Yearbook, 1955, United Nations, New York.
United Nations Educational, Scientific and Cultural Organization (UNESCO), 1958z. World Survey of Education II: Primary Education. UNESCO, Paris, France.
United Nations Educational, Scientific and Cultural Organization (UNESCO), 1961z. World Survey of Education III: Secondary Education. UNESCO, Paris, France.
United Nations Educational, Scientific and Cultural Organization (UNESCO), 1966z. World Survey of Education IV: Tertiary Education. UNESCO, Paris, France.
United Nations Educational, Scientific and Cultural Organization (UNESCO), 1993z. World Education Report 1993. UNESCO, Paris, France.
United Nations Educational, Scientific and Cultural Organization (UNESCO), various years, Statistical Yearbook, UNESCO; Paris, France.
Van Leeuwen, B., van Leeuwen-Li, J., 2014. Education since 1820. In: van Zanden, J.L., Baten, J., d'Ercole, M.M., Rijpma, A., Smith, C., Timmer, M. (Eds.), How Was Life? Global Well-Being Since 1820. MIT Press, Boston, pp. 87–100.
Wilmoth, J.R., 1995. Berkeley Mortality Database. University of California, Berkeley. WolframAlpha, http://www.wolframalpha.com 2013, (access August 3, 2013).
Wolfram, Alpha, 2013. http://www.wolframalpha.com (access August 3, 2013).