A New Data Set of Educational Attainment in the World, 1950–2010[*]

Robert J. Barro[+]

Harvard University

and

Jong-Wha Lee

Korea University

Revised: April 2012

[+]*Send correspondence*: Robert J. Barro, Department of Economics, Harvard University, Cambridge MA 02138-3001. Tel: 617-4953203, Fax: 617-4968629, E-mail: rbarro@harvard.edu

Abstract

Our panel data set on educational attainment has been updated for 146 countries from 1950 to 2010. The data are disaggregated by sex and by 5-year age intervals. We have improved the accuracy of estimation by using information from consistent census data, disaggregated by age group, along with new estimates of mortality rates and completion rates by age and education level. We compare the estimates with our previous ones (Barro and Lee, 2001) and alternative measures (Cohen and Soto, 2007). Our estimates of educational attainment provide a reasonable proxy for the stock of human capital for a broad group of countries and should be useful for a variety of empirical work. ,

1.   Introduction

Many observers have emphasized the crucial importance of human capital, particularly as attained through education, to economic progress (Lucas, 1988 and Mankiw, Romer and Weil, 1992). An abundance of well-educated people goes along with a high level of labor productivity. It also implies larger numbers of more skilled workers and greater ability to absorb advanced technology from developed countries. The level and distribution of educational attainment also have impact on social outcomes, such as child mortality, fertility, education of children, and income distribution (see for example Barro and Lee, 1994, de Gregorio and Lee, 2002, Breierova and Duflo, 2004, Cutler et al., 2006).

There have been a number of attempts to measure educational attainment across countries to quantify the relationship between it and economic and social outcome variables. Earlier empirical studies used school enrollment ratios or literacy rates (Romer, 1990, Barro, 1991, and Mankiw, Romer and Weil, 1992). But although widely available, these data do not adequately measure the aggregate stock of human capital available contemporaneously as an input to production.

Our earlier studies (1993, 1996, and 2001) filled this data gap by constructing measures of educational attainment for a broad group of countries. The figures were constructed at 5-year intervals from 1960 to 2000. The data showed the distribution of educational attainment of the adult population over age 15 and over age 25 by sex at seven levels of schooling. We also constructed measures of average years of schooling at all levels—primary, secondary, and tertiary—for each country and for regions in the world.

In this paper, we update and expand the data set on educational attainment. We extend our previous estimates from 1950 to 2010, and provide more, improved data disaggregated by sex and age. The data are broken down into 5-year age intervals, and the coverage has now expanded to 146 countries by adding 41, including 11 former Soviet republics. The accuracy of estimation has also improved by incorporating recently available census/survey observations.

The new data set improves on the earlier by using more information and better methodology.

We construct new estimates by using information from survey/census data, disaggregated by age group. Previously, we adopted a perpetual inventory method, using the census/survey observations on the educational attainment of the adult population group over age 15 or over age 25 as benchmark stocks and new school entrants as flows that added to the stocks with an appropriate time lag. The flow estimates were estimated using information on school-enrollment ratios and population structure over time. But this method is subject to bias due to inaccuracy in estimated enrollment ratios and in benchmark censuses. In the current estimation, we reduce measurement error by using observations in 5-year age intervals for the previous or subsequent 5-year periods, as in Cohen and Soto (2007). We also construct new estimates of (a) survival/mortality rates by age and by education; and (b) completion ratios by educational attainment and by age group. These measures help improve the accuracy of the backward- and forward-estimation procedure.

The data set improvements address most of the concerns raised by critics, including De La Fuente and Doménech (2006) and Cohen and Soto (2007). They noted that the previous data set of Barro and Lee (1993, 2001) shows implausible time-series profiles of educational attainment for some countries. The new procedures have resolved these problems.

In the next section, we summarize the data and the methodology for constructing the estimates of educational attainment and discuss the modifications that have been made in the present update. In section 3, we highlight the main features of the new data set. In section 4, we compare the estimates with our previous ones (Barro and Lee, 2001) and alternative measures by Cohen and Soto (2007). Section 5 presents our conclusions.

2.   Data and Estimation Methodology

A. The Census data

The benchmark figures on school attainment (621 census/survey observations) are collected from census/survey information, as compiled by UNESCO, Eurostat, national statistic agencies, and other sources.[1] The census/survey figures report the distribution of educational

---

[1] Most of our original data are from full censuses (603 of total 621 observations). Country Notes available online at: http://www.barrolee.com provide information on data sources for individual countries. There are additional data available from OECD sources for a group of OECD countries since 1990s. We have decided not

attainment in the population over age 15 by sex and by 5-year age group, for most cases, in six categories: no formal education (*hu*), incomplete primary (*hpi*), complete primary (*hpc*), lower secondary (*hsi*), upper secondary (*hsc*), and tertiary (*hh*). These classification follows UNESCO's 'International Standard Classification of Education (ISCED)' that facilitates comparisons of education statistics and indicators across countries on the basis of uniform and internationally-agreed definitions.[2]

Table 1 presents the distribution of countries by the number of available census/survey observations since 1950.[3] For total population aged 15 and over, 191 countries have at least 1 observation, and 112 countries have 3 or more observations. Table 2 shows the distribution of countries by census/survey year since 1950 (where the underlying figures are applied to the nearest 5-year value). For total population over age 15, for example, 67 observations are available for 1960, 89 for 1970, 93 for 1980, 84 for 1990, and 79 for 2000. These data points are used as benchmark figures on educational attainment.

When a census provides only numbers for a combination of several categories, such as no formal education, incomplete primary, and complete primary, we use decomposition methods to separate into categories. In many OECD and non-OECD countries, available census data do not report data according to these four broad categories.[4] Some census data do not report *hu* or the proportion of those who have no formal education, and do report *hp*, *hs*, and *hh* among the educated members of the population only. To avoid overestimation of average years of schooling, for census years with missing *hu*, we use the illiteracy rate, primary enrollment ratio, or *hu* from other census years to estimate *hu*. We then adjust *hp*, *hs*, and *hh* to reflect both the educated and uneducated members in the total population.

In some instances, data on *hu* is not missing but overlapped with other category(ies) or subcategory(ies). Some census data report the proportion of those who have reached primary level together with those who have no formal education (*hu+hp*). A number of countries also

---

to use these additional observations for reasons described in section 4.
[2] See http://www.uis.unesco.org/Education/Pages/default.aspx for the definition of ISECD and individual ISCED mappings for each country.
[3] These census/survey observations include the countries/territories for which we could not construct the complete estimates of educational attainment because of other missing information. Appendix Table shows the census/survey information for the 146 countries for which we have constructed complete estimates.
[4] Country Notes explain the censuses for which a decomposition method was applied.

report the combined proportion of those who have reached secondary schooling or less (*hu+hp+hs*). Also, some census data report the combination of those who have reached primary or secondary levels (*hp+hs*). To decompose these overlapping census observations we use enrollment data. Specifically, for census years where *hu* is combined with other category(ies) or subcategory(ies), we use adjusted primary and/or secondary enrollment ratio by age group from earlier or later years and the age distribution profile to decompose the overlapping observations.[5]

B. Estimation of missing observations at the four broad levels

We calculate from 1950 to 2010 at the five year intervals the educational attainment of the population by 5-year age groups. First, we calculate the distribution of educational attainment at four broad categories—no formal education (*hu*), primary (*hp*), secondary (*hs*) and tertiary education (*hh*). Primary includes both incomplete primary (*hpi*) and complete primary (*hpc*), and secondary (*hs*) includes lower secondary (*hsi*) and upper secondary (*hsc*). Tertiary education (*hh*) also includes both junior-level (*hhi*) and higher-level tertiary (*hhc*).

We fill in most of missing observations by forward and backward extrapolation of the census/survey observations on attainment. The estimation procedure extrapolates the census/survey observations on attainment by age group to fill in missing observations with an appropriate time lag.

Let's denote $h_{j,t}^a$ as the proportion of persons in age group *a*, for whom *j* is the highest level of schooling attained— *j*=u for no school, p for primary, s for secondary, and h for higher at time *t*. There are 13 5-year age groups ranging from *a* =1 (15–19 years old) to *a* =13 (75 years and over).

The forward extrapolation method assumes that the distribution of educational attainment of age group *a* at time *t* is the same as that of the age group that was five years younger at time *t*-5:

---

[5] The adjusted enrollment ratio is the gross enrollment ratio minus the proportion of repeaters. The age distribution profile is the relative population distribution by age group within an educational attainment at a specific time period. If the gross enrolment ratio is not available, the net enrolment ratio is used as a proxy for the adjusted enrollment ratio.

$$h_{j,t}^a = h_{j,t-5}^{a-1} \qquad (1)$$

where age group a denotes, $a=3$: 25–29 age group, …$a=11$: 65–69 age group. This setting applies to persons who have completed their schooling by time t-5. As explained below, we adjust this formula by considering different mortality rates by education level for the old population aged 65 and over. For younger groups under age 25, we adopt a different method, considering that part of population is still in school during the transition period from t to t+5.

The backward extrapolation is expressed as:

$$h_{j,t}^a = h_{j,t+5}^{a+1} \qquad (1a)$$

where age group a denotes, $a=2$: 20–24 age group, …$a=10$: 60–64 age group.

Thus, a person's educational attainment remains unchanged between age 25 and 64. An assumption here is that, in the same 5-year age group, the survival rate is the same regardless of a person's educational attainment. When we look at information from available censuses stratified by educational attainment and population structure by age group in the previous or subsequent 5-year periods, we find this assumption holds well for the population aged 64 and under, but not for older age groups. In a typical country, the mortality rate is higher for older people who are less-educated. The assumption of uniform mortality can then cause a downward bias in the estimation of the total educational stock. Note that the survival rate is a relative survival rate that measures the survival rate of each educational attainment group as a fraction of the overall survival rate of the relevant age group. The overall survival rates for each five-year age cohort are reflected in the change of population structure over time derived from population census data for individual countries.

If we consider the differences in survival rate by education levels, the forward extrapolation method is expressed by

$$h_{j,t}^a = h_{j,t-5}^{a-1} \cdot (1 - \rho_j^a) \qquad (2)$$

where ρ_j^a is the age-specific mortality rate over the five years for the population in age group a, for whom j is the highest level of schooling. For the population aged 65 and above, we allow for the different mortality rates by education levels.

By utilizing information from available censuses by age group in the previous and/or next 5-year periods, we have estimated the survival rates (1- ρ_j^a) for the old population in the age group, 65–69, 70–74, and 75-79 by education levels. Due to limited availability of observations, we estimate survival rates for two broad groups of education levels, a less-educated population (uneducated and people who have reached the primary level) and a more-educated population (reached at least secondary schooling), and for broad groups of OECD countries and non-OECD countries. The estimation results show that the more educated people have lower mortality (higher survival) rates. Appendix Notes 1 describes more details on the estimation of survival rates.

We have evaluated the accuracy of forward- and backward- extrapolation method by carrying out ex-post simulations as described in Barro and Lee (1993) and found that the method provides reliable estimates for missing observations. An important issue is how to combine forward and backward-flow estimates when both are available for a missing cell. We have carried out a simulation exercise in which we regressed the 'observed' actual census values of the various levels of educational attainment on the estimates generated from forward- and backward-flow estimates (based on both five- or ten-year lead and lagged values from actual censuses). We use the regression results to construct a weighted-average of forward and backward-flow estimates (see Appendix Notes 1 for more details on how to combine forward-flow and backward-flow estimates).

Note that the forward and backward-flow estimates cannot be applicable for the two youngest cohorts between ages 15 and 24 because part of the population is in school during dates *t* and *t*+5. For these age groups (*a* =1: 15–19 age group and *a* =2: 20–24), we construct the estimates by using the estimates of the same age group in *t*–5 (or *t*+5) and the change in (age-specific) enrollment for the corresponding age groups over time (see Appendix Note 1 for more details).

C. Estimation of sub-categories of educational attainment

We have estimated school attainment at four broad levels of schooling: no school, some primary, some secondary, and some higher. We break down the three levels of schooling into incomplete and complete education by using estimates of completion ratios.

First, we describe our procedure for estimating missing observations for the subcategories for the primary schooling category. We filled in the missing cells using information from the available census/survey data. The completion rate at the primary level is expressed as a ratio of people who completed primary schooling but did not enter secondary schooling to people who entered primary school. For the remaining missing cells, we filled them in by forward and backward extrapolation of the census/survey observations on completion ratios with an appropriate time lag. This procedure applies to the age group $a=3$ (25–29) and above.[6] If both forward and backward estimates are available, we combine them by using the results of regression of the 'observed' actual census values of the various levels of completion ratio on the estimates generated from forward- and backward-flow estimates (based on both 5-year or ten-year lead and lagged values from actual censuses). On the other hand, we assume that the completion ratios for aged 15–19 and 20–24 are determined by age specific profile of completion ratios in each country (see Appendix Notes 2).

We applied similar methods to estimate missing observations for the subcategories for secondary and tertiary schooling. Secondary-school enrollees aged 15–19 are treated as incompletely educated at the secondary level, and higher-school enrollees aged 20–24 are treated as incompletely educated at the higher level. Appendix Notes 2 explains more details on how to combine forward-flow and backward-flow estimates of completion ratios. For the countries which do not have actual census observations data for completion rate and thereby no flow estimate is generated, we use the average for a group of OECD countries or non-OECD countries for the same age group and the same period..

D. Average Years of Schooling

The number of years of schooling for the population aged 15 and above, s_t, is constructed as

$$s_t = \sum_{a=1}^{A} l_t^a s_t^a \tag{3}$$

where l_t^a : the population share of group g in population 15 and above and s_t^a : the number of years of schooling of age group a— $(a =1$: 15–19 age group, $a =2$: 20–24 age group, … , a

---

[6] For the countries in which only the completion ratio for total population is available, we break down it into age groups based on the typical age profile of completion ratios constructed using the available data of the countries in the same region.

=13: 75 and above).

The number of years of schooling of age group *a* in time *t* is

$$s_t^a = \sum_j h_{j,t}^a Dur_{j,t}^a$$

(4)

where h_j^a the fraction of group *a* having attained the educational level $j = p, s$ (*incomplete, complete)*, *h (incomplete, complete)*, and *Dur* indicates the corresponding duration in years.

The duration is the typical duration of primary and the two levels of secondary education for each country (available from issues of UNESCO, *Statistical Yearbook*), It is constructed by taking account of changes in the duration system over time in a country. We suppose that changes in the duration of schooling at the primary or secondary level applied to new entrants in primary or secondary school (that is, ages 5–9 or ages 10-14) at the time of change. For higher education, we used a duration of four years for all countries and for all years, and we assigned two years to persons who entered tertiary school but did not complete it.

We use the same sources and methodology to construct a panel data set on educational attainment of females by age group. The data on the distribution of educational attainment among the population, combined with the information for each country on the duration of school at each level, generate the number of years of schooling achieved by the average person at various levels and at all levels of schooling combined.

3.   The Complete Data Set on Educational Attainment, 1950–2010

Tables 3 and 4 summarize the progress in educational attainment of the population aged 15 years and above by region and by income classification from 1950 to 2010 for the 146 countries that have complete information.[7]

---

[7] The additional countries/territories that have complete estimates in the new data set include Albania, Belize, Burundi, Brunei Darussalam, Cambodia, Congo, Cote d'Ivoire, Egypt, Gambia, Gabon, Lao People's Democratic Republic, Luxembourg, Libya, Morocco, Macao Special Administrative Region, Maldives, Malta, Mauritania, Mongolia, Namibia, People's Republic of China, Qatar, Reunion, Rwanda, Saudi Arabia, Tonga, United Arab Emirates, Viet Nam, Yemen. The data set include Croatia, Czech Republic, Serbia, Slovakia, and Slovenia as independent countries, replacing the former Yugoslavia and Czechoslovakia. The former USSR is replaced by the Russian Republic, Armenia, Azerbaijan, Estonia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Republic of Moldova, Tajikistan, and Ukraine.

The table considers two broad groups—24 advanced countries and 122 developing countries. The developing group is further broken down into six regions: Middle East/North Africa (18 countries), Sub-Saharan Africa (33), Latin America/Caribbean (25), East Asia/Pacific (19), South Asia (7), and Europe and Central Asia (20). Regional averages are computed by weighting each country's observation by its share in total population of the region.

Some of the important developments that represent the progress of developing countries in achieving higher educational attainment are summarized as:

- In 2010, the world population aged 15 and above is estimated to have an average of 7.9 years of schooling, increasing steadily from 3.1 years in 1950 and 5.3 years in 1980. The overall population over age 15 in high-income economies is estimated to have 11.3 years of schooling, compared to 7.2 years in developing countries. Both Sub-Saharan African and South Asian countries have the lowest at 5.3 years on average.

- Since 1950, the average years of schooling among the total population aged 15 years and above in developing countries increased significantly from 2.0 years to 7.2 years. In South Asia and Middle East/North Africa regions, average years of schooling have more than doubled since the 1980s. In South Asia, for instance, average years of schooling among the total population aged 15 and over rose from 2.4 years in 1980 to 5.3 years in 2010.

- While higher secondary and tertiary completion and enrollment ratios account for most of the improvements in years of schooling in advanced countries, most of the improvements in developing countries are accounted for by higher primary and secondary completion and enrollment ratios (see Figure 1a).

- Average years of education among the population aged 15–24 years in developing countries rose from 3.1 years in 1950 to 6.83 years in 1990 and to more than 8.90 years in recent years (see Figure 1b). The improvements in completion and enrollment ratios at all levels among the younger cohorts in every generation continually contribute to rising average years of schooling as they mature over time. The biggest improvement in average years of schooling among the younger cohorts was recorded between 1970 and 1990 in advanced countries.

- Figure 2 shows that developing countries have successfully reduced illiteracy rates, especially among the younger cohorts. Specifically, the proportion of the uneducated in the total population over age 15 in developing countries has declined significantly over the past six decades since 1950, from 61.1% in 1950 to 17.4% in 2010. Among 15–24 year olds, this proportion has declined from 45.2% in 1950 to 5.9% in 2010.

- Table 4 summarizes the educational attainment among males and females by region since 1950. It shows that significant progress has been made by developing countries in terms of reducing gender inequality in education among the overall population over age 15. The ratio of female to male average years of schooling increased from around 62.5% in 1950 to 81.2% in 1990 and 85.9% by 2010.

Despite these major developments, many challenges in making education more inclusive remain. Notwithstanding significant improvements, the current level and distribution of educational attainment in developing countries is comparable only to that of advanced countries in the late 1960s (see Figure 1a).

- The gap between developing and advanced countries in average years of schooling among the overall population over age 15 remains high (4.1 years in 2010) as it has narrowed by only less than 1 year in the past 40 years.

  One factor that contributed to the slow reduction in this gap is the continued increase in the proportion of the population in advanced countries reaching higher levels of education.

  Also, the narrowing of the gap in average years of schooling among younger cohorts between developing and developed countries is less than enough to compensate for the huge gap among the older cohorts (see Figure 1b). For example, while the gap between the average years of schooling among 15−24 year olds in developing economies and advanced countries has narrowed since 1970 by around 1.44 years in 2010 (3.73 years in 1970 to 2.29 years in 2010), this gap has even widened by around 1.25 year (from 4.49 years in 1970 to 5.76 years) among those aged 65 years and above (see Figure 1a).

- The challenge of making education more gender inclusive also remains in many developing regions, such as South Asia, Middle East, and sub-Saharan Africa. The ratio of years of schooling among females to males remains below 70% in South Asia.

- The prospects of narrowing the educational attainment gap between developing and developed countries greatly rely on the capability of developing countries to (1) significantly increase enrollment ratios among new entrants (below 15 years old); and (2) catch up with the high rate of survival from primary to secondary level and from secondary to tertiary level (15–24 years old) in developing countries.

4.   Comparison with Alternative Estimates

This section compares our estimates of educational attainment with other estimates. First, we want to check our new estimates with our previous estimates in Barro and Lee (2001). Table 5 shows the means and standard deviation of levels and 10-year differences of the overlapping observations between the new Barro-Lee data set and Barro and Lee (2001) estimates over 1960−2000. The two estimates are highly correlated in levels, with correlation coefficient of 0.96 for all countries and of 0.91 and 0.95 for group of advanced and developing countries respectively. The correlation is lower in 10-year differences, with correlation coefficient of 0.56 for all countries and of 0.29 and 0.63 for group advanced and developing countries respectively. .It also shows that the new Barro-Lee data set displays less dispersion than the previous estimates in both levels and 10-year differences.

Figure 3 shows that, on average, the new Barro-Lee estimates for average years of schooling for advanced countries are higher than the previous Barro-Lee estimates. For developing countries, estimates of average years of schooling until 1990 are slightly lower than the previous estimates. Figure 3 also shows that the new estimates display a smoother increasing trend in average years of schooling, both for developing and advanced countries, than the previous estimates.

Country level estimates are also much smoother over time. Figure 4 compares the new Barro-Lee estimates with Barro-Lee (2001) estimates for selected countries. The new estimates provide smother time profiles of educational attainment in Norway, the United States, Peru,

and Venezuela.

We also compare the new estimates with the estimates by Cohen and Soto (2007). Cohen-Soto constructed a data set for average years of schooling for 95 countries at 10-year intervals for 1960–2010. They adopt data and methodology similar to ours. They use forward-flow and backward-flow methods to fill-in missing observations by extrapolating the census/survey observations on educational attainment by 5-year age group. But there are also significant differences. First, Cohen and Soto use OECD sources for OECD countries and UNESCO sources for non-OECD countries.[8] As discussed in Barro and Lee (2001), there exist significant differences between the OECD data and UNESCO censuses. Most OECD data come from labor-force surveys based on samples of households or individuals, in contrast to the national censuses in the UNESCO database. A part of population, such as persons aged 15-24 years who are in school, are not included in the labor force survey. There are also significant differences in the classification of education systems between the OECD and the UNESCO sources. As a result, Cohen and Soto's procedure tends to over-estimate educational attainment for OECD countries (see Figure 5).

Second, relying on only OECD sources, which are available since the 1990s, underutilizes available information. For example, Cohen and Soto's estimation for the United States relies on only two OECD surveys in 1991 and 1998. The data are then used to estimate missing observations in the earlier years by the backward-flow method, whereas for the age groups for which the backward estimates are not applicable, the estimates are constructed mainly by lagged enrollment rates. In contrast, our estimation for the United States relies on seven UNESCO censuses and one national census from 1950 to 2005, so that the census information on educational attainment by age group is used to fill in missing observations by both forward-flow and backward-flow estimates.

Third, for non-OECD countries, we use substantially more UNESCO censuses than Cohen and Soto. Cohen and Soto's estimation uses only 70 UNESCO censuses for 75 developing countries in their sample, compared to 407 for 122 developing countries in our sample. In fact, Cohen-Soto's estimates for 27 countries (including most sub-Saharan African countries) rely entirely on enrollment data. We believe our estimates based on more censuses must

---

[8] For only a few OECD countries including Finland, Portugal and Turkey, Cohen and Soto combine both OECD and UNESCO sources.

contribute to more accurate estimation of missing observations by forward-flow and backward-flow method by age-group.

Finally, Cohen-Soto do not consider the difference in mortality rates by educational levels nor the change in durations over time.

Table 5 shows means and standard deviations of average schooling years in levels and 10-year differences for the sample of the overlapping observations between the new Barro-Lee data set and Cohen-Soto (2007). The two estimates are highly correlated in levels with correlation coefficient of 0.94., but less in 10-year differences with correlation coefficient of 0.40 for the sample of all overlapping observations. The estimates for advanced countries are less correlated than those for developing countries in both levels and 10-year differences. It also shows that the new Barro-Lee data set displays less dispersion than Cohen-Soto (2007) in levels for all sample groups, except for that of advanced countries in 1960. In contrast, the new Barro-Lee estimates show greater dispersion in 10-year differences for all sample groups. [9]

Figure 5 show that the estimates for advanced countries in the new Barro-Lee data set are on average lower than in Cohen-Soto (2007) for the overall period, 1960-2010. The two estimates are quite diverged in earlier periods. For developing countries, the new Barro-Lee estimates are on average very close to Cohen-Soto (2007) estimates in earlier years but higher for 2010.

Our estimates of educational attainment provide a reasonable proxy for the stock of schooling capital for a broad group of countries. However, the school attainment does not take account of the skills and experience gained after formal education. The measure does not directly measure the skills obtained at schools and, specifically, does not take account for differences in the quality of schooling across countries. Figure 6 compares our concept of educational attainment with Hanushek and Woessmann's (2009) measure of human capital quality, which was constructed by standardizing and combining available international math and science test

---

[9] We estimate reliability ratios for the new Barro-Lee estimates vis-à-vis Cohen-Soto (2007) in levels and 10-year differences. As used by Krueger and Lindahl (2001), the reliability ratio gauges the fraction of the variability of a (unobserved) true variable in the total variability of the variable measured with error. We find the reliability ratio for the new Barro-Lee estimates is greater than that of Cohen Soto (2007) in levels (1.00 vs. 0.89) but lower in 10-year differences (0.31 vs. 0.50).

results, covering 1964–2003 for 50 countries.[10] Educational attainment and human capital quality measures are highly correlated but human capital quality is quite diverse for countries with similar levels of educational attainment.

5. Concluding Remarks

Our new data set on educational attainment applies to 146 countries at five-year intervals from 1950 to 2010. The estimates are disaggregated by sex and by 5-year age intervals. These estimates improve on our previous, widely used data set by utilizing more information and better estimation methodology.

This improved data set on educational attainment should be helpful for a variety of empirical work.[11] Our estimates of educational attainment provide a reasonable proxy for the stock of human capital for a broad group of countries. The data set has been useful for studying the linkages across countries between education and important economic and social variables, such as economic growth, export competitiveness, fertility, income inequality, democracy, institutions, and political freedom. We expect that this new data set will help to improve the reliability of these types of analyses.

---

[10] See Hanushek and Woessmann (2009) for more details about their methodology for estimating human capital quality.

[11] Our earlier estimates of educational attainment have been used in many studies. Up to March 2012, our papers on educational attainment data published in 1993, 1996, and 2001 have been cited in journals over 950 times, according to the *Social Science Citations Index*. The total number of citations by all journal articles, books, and working papers amounts to over 6,800, according to *Google Scholar*.

References

Barro, R.J. 1991. "Economic Growth in a Cross Section of Countries," The Quarterly Journal of Economics, 106(2): 407-43.

Barro, R.J. and J.W. Lee. 1993. "International Comparisons of Educational Attainment," Journal of Monetary Economics, 32, 363-94.

Barro, R.J. and J.W. Lee. 1996. "International Measures of Schooling Years and Schooling Quality," American Economic Review, 86, 218-23.

Barro, R. and J.W. Lee. 1994, "Sources of Economic Growth," Carnegie Conference Series on Public Policy, 40: 1.

Barro, R. and J.W. Lee. 2001. "International Data on Educational Attainment: Updates and Implications," Oxford Economic Papers 53(3).

Breierova, L. and E. Duflo. 2004."The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less than Mothers?" NBER Working Paper No. 10513.

Cohen, D. and M. Soto. 2007 "Growth and Human Capital: Good Data, Good Results," Journal of Economic Growth, 12:51–76.

Cutler, D., A. Deaton, A. Lleras-Muney. 2006. "The Determinants of Mortality," The Journal of Economic Perspectives, 20(3): 97-120.

De Gregorio J. and J.W. Lee. 2002. "Education and Income Inequality: New Evidence from cross-country data," Review of Income and Wealth, 48(3): 395–416.

De La Fuente, A. and R. Doménech. 2006. "Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?," Journal of the European Economic Association, 4(1): 1-36.

Hanushek E.A. and L. Woessmann. 2009. "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation," NBER Working Paper No. 14633.

Krueger, A.B. and M. Lindahl. 2001. "Education for Growth: Why and For Whom?," Journal of Economic Literature, 39 (December): 1101–1136.

Lucas R.E. 1988. "On the Mechanics of Economic Development," Journal of Monetary Economics, July, 3-42.

Mankiw, G., D. Romer, and D. Weil. 1992. "A Contribution to the Empirics of Economic Growth," Quarterly Journal of Economics, 107.

Romer, Paul. 1990. "Endogenous Technological Change," Journal of Political Economy, October, S71-S102.

UNESCO. various years. Statistical Yearbook. Paris.

Table 1.   Breakdown of Number of Countries by Number of Census-Survey Observations

Number of observations (1950–2005)	All		Number of countries Advanced		Developing	
	MF	F	MF	F	MF	F
1	42	41	1	1	41	40
2	37	40	1	1	36	39
3	32	31	3	3	29	28
4	30	31	4	4	26	27
5	28	27	7	7	21	20
6	13	10	3	3	10	7
7	3	4	1	2	2	2
8	4	3	3	2	1	1
9	1	1	1	1	0	0
10	1	1	0	0	1	1
Total	191	189	24	24	167	165

Note: The data refer to census-survey observations for educational attainment for the total (MF) and female (F) populations in each age category.

Table 2.   Breakdown of Number of Countries by Number of Census-survey Year

Census-survey year (to the nearest 5-year value)	All		Number of countries Advanced		Developing	
	MF	F	MF	F	MF	F
1950	25	25	8	8	17	17
1955	16	14	1	1	15	13
1960	67	67	15	15	52	52
1965	30	29	4	4	26	25
1970	89	85	17	17	72	68
1975	44	43	7	7	37	36
1980	93	90	18	18	75	72
1985	26	24	5	5	21	19
1990	84	80	14	13	70	67
1995	25	24	4	4	21	20
2000	79	79	12	12	67	67
2005	43	43	16	16	27	27
Total	621	603	121	120	500	483

Note: The data refer to census-survey observations for educational attainment for the total (MF) and female (F) populations in each age category.

Table 3.   Trends of Educational Attainment of the Total Population Aged 15 and Over by Region

Region (no. of countries) and Year	Population Aged 15 and over (Million)	No Schooling	Highest level attained						Average years of schooling
			Primary		Secondary		Tertiary		
			Total	Completed	Total	Completed	Total	Completed	
			(% of population aged 15 and over						
World (146)									
1950	1588	47.1	38.1	17.1	12.6	5.2	2.2	1.1	3.12
1960	1831	42.4	38.5	17.7	16.3	7.0	2.7	1.5	3.60
1970	2221	35.5	38.3	19.6	22.4	9.5	3.8	2.1	4.39
1980	2761	30.1	32.8	17.4	31.0	13.0	6.1	3.1	5.34
1990	3413	25.7	30.6	18.2	34.7	16.8	9.0	4.7	6.14
2000	4064	19.3	27.7	17.7	41.3	21.5	11.7	6.4	7.10
2010	4759	14.8	24.6	17.3	46.3	25.9	14.2	7.8	7.89
Advanced (24)									
1950	428	9.2	60.1	35.1	25.0	11.7	5.7	2.8	6.10
1960	476	7.7	54.2	32.1	31.2	16.0	7.0	3.8	6.72
1970	541	6.2	45.8	30.2	38.2	21.0	9.8	5.3	7.64
1980	614	5.4	34.6	23.9	44.5	26.3	15.5	8.1	8.74
1990	683	5.4	27.6	19.8	44.4	26.1	22.6	12.3	9.55
2000	746	3.4	19.4	14.6	49.2	28.1	28.0	16.0	10.52
2010	805	2.4	13.7	10.6	51.7	33.6	32.2	17.9	11.30
Developing (122)									
1950	1160	61.1	29.9	10.5	8.1	2.8	0.9	0.5	2.02
1960	1355	54.6	32.9	12.7	11.1	3.8	1.3	0.7	2.50
1970	1681	45.0	35.9	16.1	17.2	5.7	1.9	1.0	3.35
1980	2146	37.2	32.3	15.5	27.1	9.2	3.4	1.7	4.37
1990	2730	30.8	31.3	17.7	32.3	14.4	5.6	2.9	5.28
2000	3318	22.9	29.5	18.4	39.6	20.1	8.0	4.3	6.33
2010	3954	17.4	26.9	18.6	45.2	24.4	10.5	5.7	7.20
By Region									
Middle East and North Africa (18)									
1950	48	88.1	8.5	3.5	2.6	1.1	0.9	0.5	0.76
1960	58	84.3	10.2	4.5	4.2	1.8	1.2	0.7	1.07
1970	75	75.3	14.2	6.3	8.4	3.9	2.0	1.1	1.81
1980	102	61.5	18.9	8.8	16.1	8.3	3.5	1.7	3.07
1990	142	45.6	23.1	11.7	25.8	14.4	5.5	3.1	4.64
2000	196	31.7	25.5	13.8	34.3	19.2	8.4	4.8	6.10
2010	256	23.8	22.9	13.5	41.2	23.3	12.0	7.1	7.25
Sub-Saharan Africa (33)									
1950	61	77.1	17.8	5.8	4.6	1.1	0.5	0.1	1.30
1960	76	72.3	22.1	7.2	5.1	1.3	0.6	0.2	1.54
1970	97	64.6	26.6	7.7	8.1	2.1	0.7	0.2	2.04
1980	129	55.1	32.1	12.5	11.9	3.8	0.8	0.3	2.82
1990	175	44.4	36.1	17.9	17.9	6.0	1.6	0.5	3.89
2000	233	38.1	36.7	20.7	22.7	7.4	2.4	0.8	4.64
2010	295	32.2	38.9	24.3	26.2	8.6	2.7	1.0	5.23
Latin America and the Caribbean (25)									
1950	98	45.9	46.6	15.7	6.4	3.0	1.0	0.6	2.57
1960	124	38.9	49.9	18.1	9.6	4.2	1.6	1.0	3.09
1970	161	30.1	52.3	20.3	15.1	5.9	2.5	1.6	3.84

Region (no. of countries) and Year	Population Aged 15 and over (Million)	No Schooling	Primary		Secondary		Tertiary		Average years of schooling
			Total	Completed	Total	Completed	Total	Completed	
			(% of population aged 15 and over						
1980	215	22.4	52.4	15.7	19.8	8.2	5.3	3.1	4.63
1990	278	17.1	48.9	23.6	26.1	11.7	7.9	4.4	5.90
2000	351	12.6	41.3	23.3	36.3	18.4	9.6	5.7	7.13
2010	425	8.2	34.5	22.3	44.9	25.2	12.2	6.8	8.20
East Asia and the Pacific (19)									
1950	496	67.2	24.7	7.9	7.7	2.0	0.4	0.2	1.71
1960	556	56.6	30.9	11.9	11.6	3.1	0.9	0.5	2.42
1970	695	40.4	39.3	17.7	19.0	5.0	1.3	0.7	3.54
1980	900	26.4	40.4	19.6	31.3	10.0	1.8	0.9	4.84
1990	1168	22.9	37.0	20.0	36.8	19.7	3.3	1.8	5.59
2000	1377	12.1	35.8	20.8	46.0	30.4	6.1	3.5	6.91
2010	1593	7.9	30.1	19.1	51.7	38.3	10.3	5.9	7.95
South Asia (7)									
1950	282	76.0	20.4	5.4	3.0	0.6	0.6	0.3	1.00
1960	341	73.5	22.4	6.8	3.6	0.8	0.6	0.3	1.14
1970	423	68.3	23.8	10.7	6.8	1.2	1.1	0.6	1.62
1980	543	66.6	13.3	8.4	17.9	1.9	2.1	1.1	2.37
1990	694	53.1	18.5	13.8	24.7	2.9	3.8	2.0	3.43
2000	879	44.7	19.1	15.8	31.5	4.5	4.8	2.8	4.26
2010	1100	33.3	21.1	18.8	39.9	6.6	5.7	3.1	5.29
Europe and Central Asia (20)									
1950	174	15.4	60.8	26.8	21.0	9.3	2.7	1.4	4.82
1960	199	11.4	56.9	26.0	28.1	12.6	3.7	1.8	5.49
1970	229	7.7	47.1	25.2	39.7	18.5	5.6	2.7	6.64
1980	257	6.7	32.5	20.0	49.7	25.9	11.1	5.1	8.02
1990	272	5.3	23.1	15.1	51.6	29.4	20.0	9.3	9.30
2000	283	1.8	14.0	9.7	54.8	31.2	29.3	13.7	10.46
2010	284	1.1	10.7	7.8	53.0	31.2	35.2	16.7	10.91

Table 4. Educational Attainment by Sex, 1950–2010

Region (no. of countries) and Year	Average years of schooling (population 15 age and over)		Gender Ratio
	Females (A)	Males (B)	(A/B, %)
World (146)			
1950	2.74	3.50	78.3
1960	3.18	4.03	79.0
1970	3.92	4.87	80.5
1980	4.78	5.91	80.9
1990	5.68	6.59	86.2
2000	6.56	7.63	86.0
2010	7.44	8.35	89.0
Advanced (24)			
1950	5.85	6.36	92.0
1960	6.43	7.04	91.3
1970	7.36	7.95	92.5
1980	8.43	9.08	92.8
1990	9.34	9.76	95.7
2000	10.29	10.78	95.5
2010	11.18	11.43	97.8
Developing (122)			
1950	1.55	2.48	62.5
1960	2.00	3.01	66.5
1970	2.77	3.92	70.8
1980	3.69	5.04	73.3
1990	4.73	5.83	81.2
2000	5.70	6.95	82.0
2010	6.65	7.74	85.9
By Region			
Middle East and North Africa (18)			
1950	0.44	1.08	40.6
1960	0.63	1.51	41.8
1970	1.10	2.53	43.4
1980	2.10	4.02	52.2
1990	3.50	5.72	61.3
2000	5.10	7.06	72.2
2010	6.45	8.02	80.4
Sub-Saharan Africa (33)			
1950	0.97	1.65	58.8
1960	1.12	1.97	56.9
1970	1.49	2.62	57.0
1980	2.09	3.58	58.4
1990	3.14	4.67	67.2
2000	3.97	5.34	74.4
2010	4.65	5.82	80.0
Latin America and the Caribbean (25)			
1950	2.36	2.79	84.4
1960	2.87	3.31	86.8
1970	3.60	4.09	88.1
1980	4.43	4.84	91.6
1990	5.82	5.99	97.2
2000	7.04	7.22	97.5

Region (no. of countries) and Year	Average years of schooling (population 15 age and over)		Gender Ratio
	Females (A)	Males (B)	(A/B, %)
2010	8.13	8.27	98.4
East Asia and the Pacific (19)			
1950	1.12	2.27	49.4
1960	1.72	3.09	55.8
1970	2.87	4.19	68.4
1980	4.11	5.54	74.2
1990	5.24	5.93	88.3
2000	6.31	7.50	84.2
2010	7.46	8.42	88.5
South Asia (7)			
1950	0.41	1.54	26.6
1960	0.52	1.71	30.4
1970	0.88	2.32	37.7
1980	1.38	3.29	42.1
1990	2.28	4.51	50.7
2000	3.16	5.31	59.5
2010	4.29	6.25	68.6
Europe and Central Asia (20)			
1950	4.20	5.65	74.4
1960	5.04	6.06	83.2
1970	6.08	7.31	83.2
1980	7.49	8.65	86.6
1990	8.92	9.75	91.4
2000	10.30	10.65	96.7
2010	10.82	11.01	98.3

Table 5. Comparison of Average Years of Schooling (Over Age 15) Between Series

A. Barro-Lee 2012 and Barro-Lee 2001	Obs	Correlation	Barro-Lee 2012		Barro-Lee 2001	
			1960	2000	1960	2000
World						
Levels	984	0.96	3.95	7.10	4.30	6.58
			2.52	*2.76*	*2.54*	*2.79*
10-year differences	427	0.56	0.97		0.75	
			0.58		*0.64*	
Advanced countries						
Levels	201	0.91	6.94	10.53	6.96	9.77
			2.18	*1.55*	*2.27*	*1.85*
10-year differences	89	0.29	0.92		0.67	
			0.61		*0.66*	
Developing countries						
Levels	783	0.95	2.44	6.28	2.96	5.82
			1.96	*2.61*	*1.99*	*2.43*
10-year differences	338	0.63	0.98		0.77	
			0.58		*0.64*	

B. Barro-Lee 2012 and Cohen-Soto 2007	Obs	Correlation	Barro-Lee 2012		Cohen-Soto 2007	
			1960	2010	1960	2010
World						
Levels	540	0.94	3.48	7.85	4.03	7.40
			2.53	*2.68*	*2.85*	*3.05*
10-year differences	450	0.40	0.96		0.82	
			0.54		*0.42*	
Advanced countries						
Levels	132	0.83	6.72	11.30	8.14	11.66
			2.20	*1.48*	*2.10*	*1.77*
10-year differences	110	0.15	0.86		0.71	
			0.60		*0.26*	
Developing countries						
Levels	408	0.95	2.06	7.01	2.23	6.37
			1.76	*2.41*	*1.91*	*2.53*
10-year differences	340	0.46	0.99		0.85	
			0.51		*0.45*	

Obs = overlapping observations

Notes: Figures presented in this table represent overlapping observations only. The new Barro-Lee data set consists of a total of 1,898 observations on average years of schooling at 5-year intervals for 146 countries (1950–2010); Barro-Lee data set (2001): 930 observations for 107 countries (1960-2000); Cohen-Soto (2007): 570 observations at 10-year intervals for 95 countries (1960–2010). Of these 95 countries, 5 countries are not in Barro-Lee (2010). Numbers in italics are standard deviations.

Source: Authors' calculations based on Barro-Lee (2001), Cohen-Soto (2007) data sets and own data.

Figure 1. Educational Attainment of the Total Population over Age 15

a. Average years of schooling, by educational level
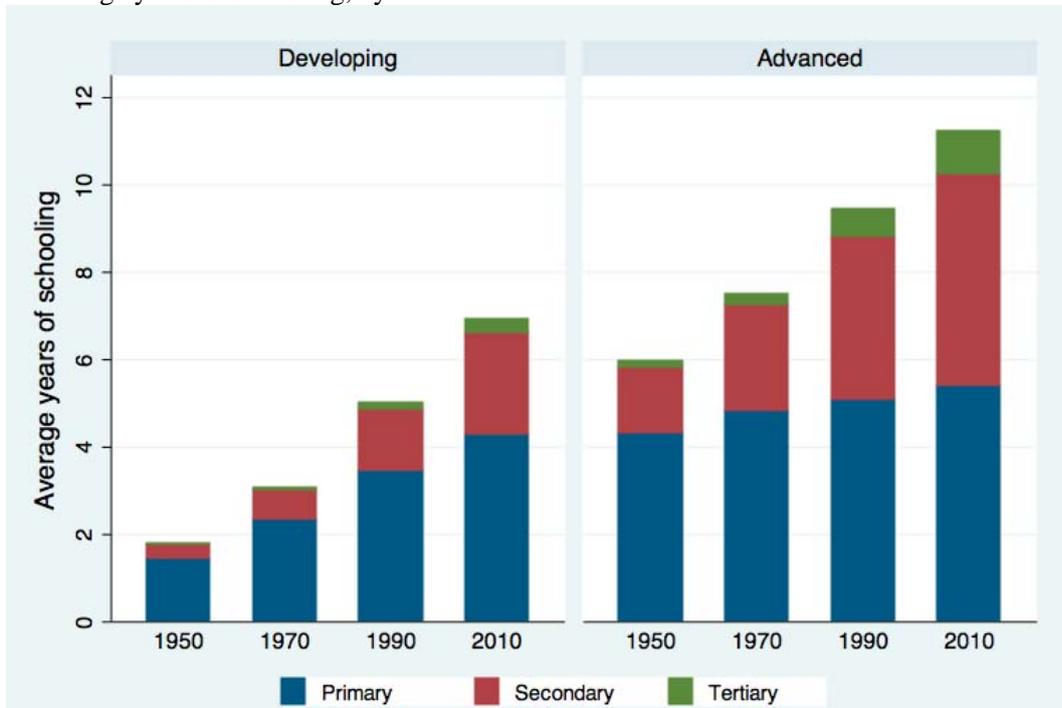
Note: Advanced countries = Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, USA, United Kingdom.

b. Average years of schooling, by age group

.
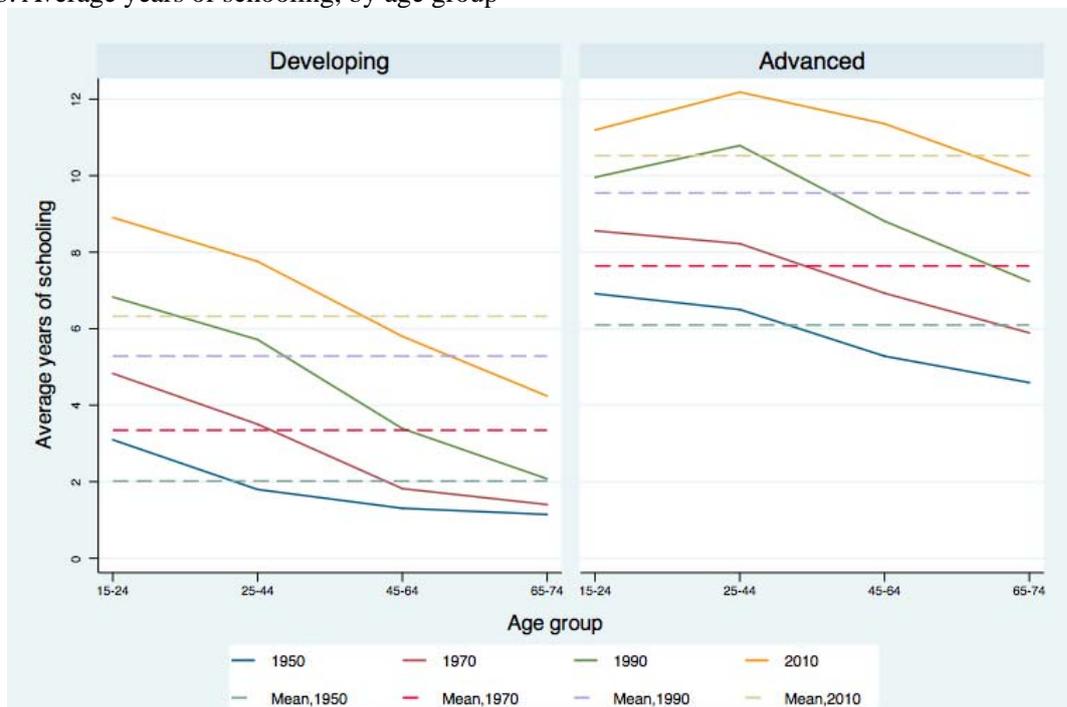
Figure 2. Proportion of Population (15 Years Old and Above) with No schooling, by Age Group
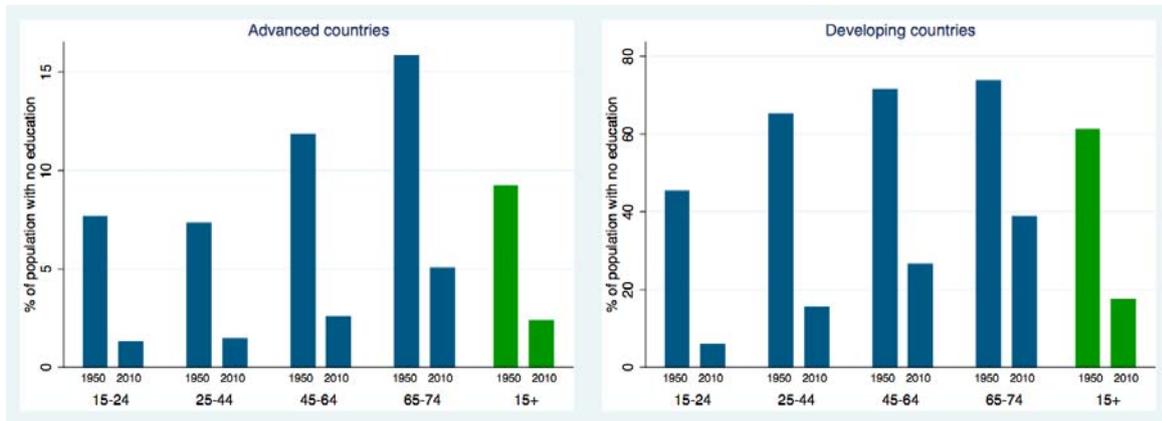
Figure 3. Comparison of Barro-Lee (2010) and Barro-Lee (2001) Estimates
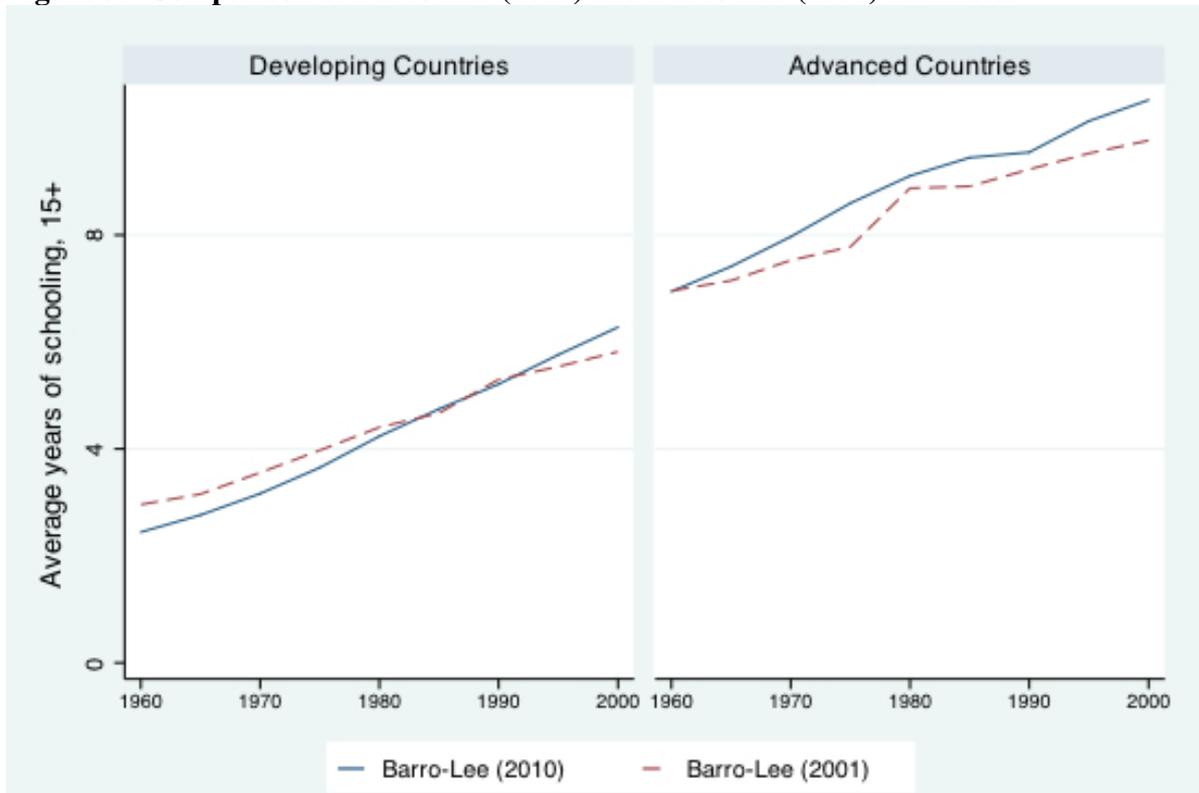
Figure 4. Average Years of Schooling, Barro-Lee (2010) and Barro-Lee (2001) Estimates, Selected Countries
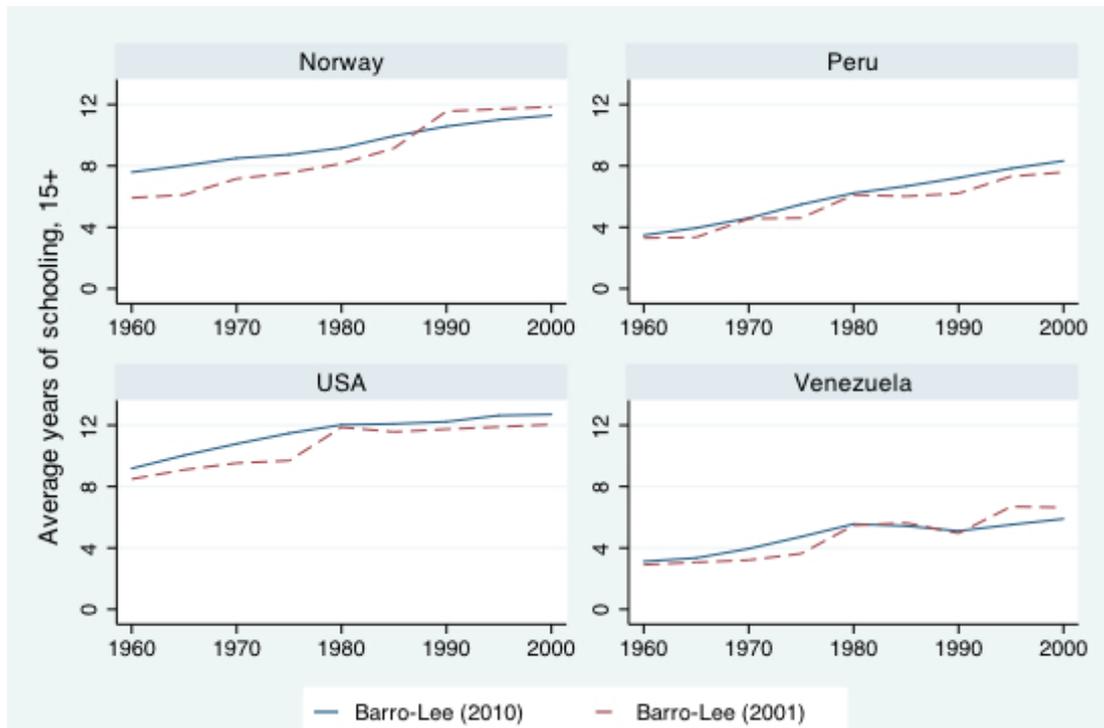
Figure 5. Comparison of Barro-Lee (2010) and Cohen-Soto (2007) Estimates
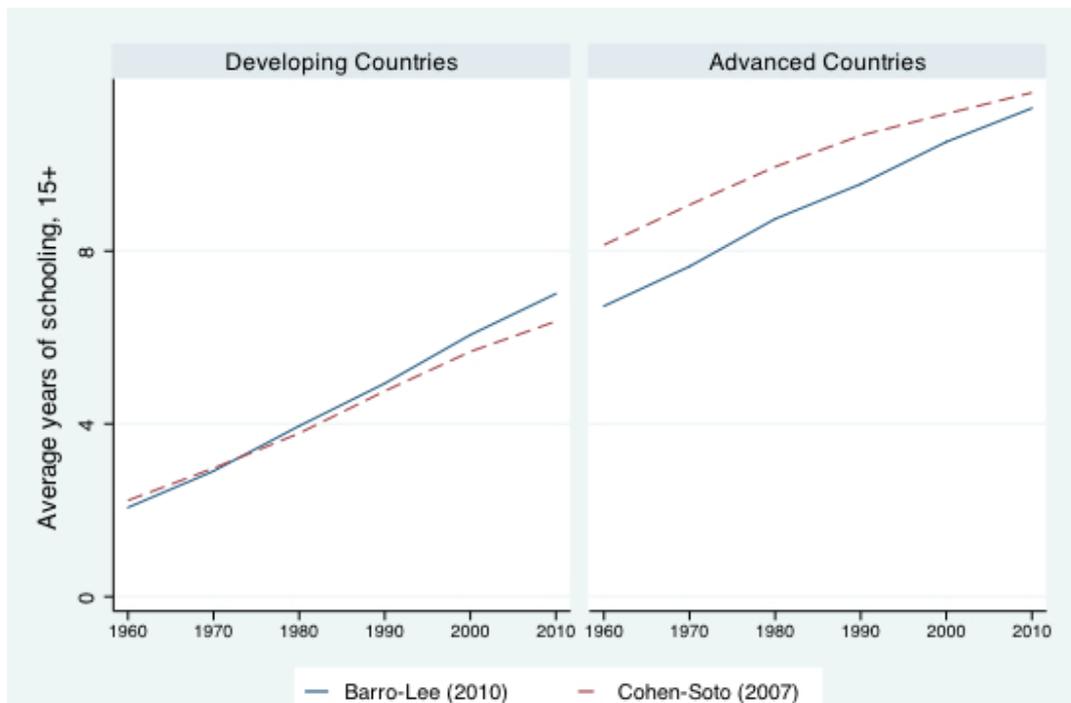
Figure 6. Comparison of Barro-Lee (2010) Years of Schooling Estimates and Hanushek-Woessmann (2009) Labor Force Quality Estimates
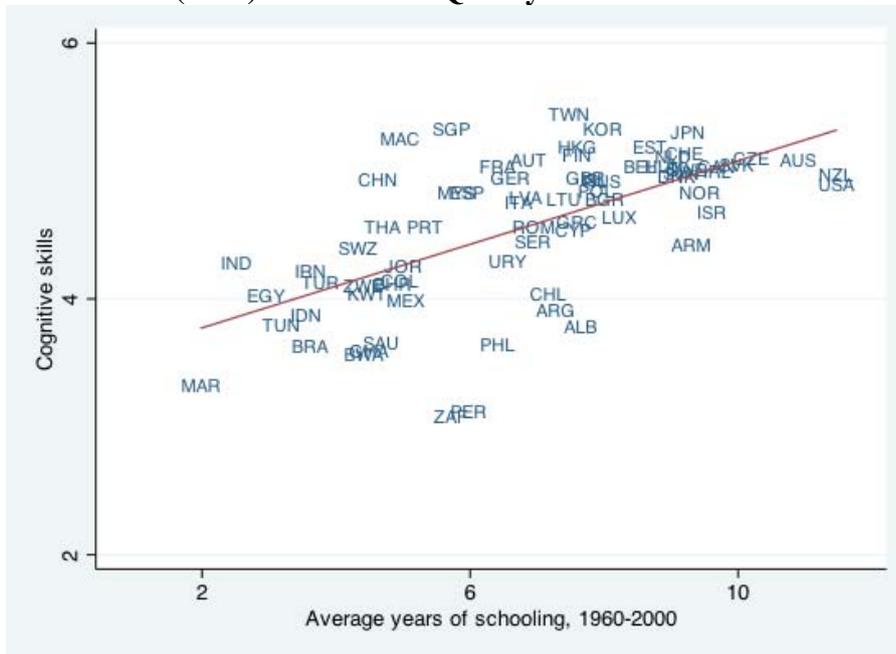
Note: Figures on years of schooling are simple average years of schooling for 1960–2000.

Appendix Notes: Estimation Procedures

We use census/survey information compiled by UNESCO, Eurostat, and others as benchmark figures to estimate average years of schooling at 5-year intervals from 1950 to 2010. As discussed in the main text, these census figures report the distribution of educational attainment in the population over age 15 by sex and by 5-year age groups, at 5-year intervals. In most cases, the distribution of educational attainment is classified according to the following four broad categories: no formal education (*hu*), primary (*hp*), secondary (*hs*), and tertiary (*hh*). It is further classified in many cases into subcategories: incomplete primary (*hpi*), incomplete secondary (*hsi*), and incomplete tertiary (*hhi*).

1. *Estimation of missing attainment data in four broad categories*

We fill in most of the missing census observations by forward and backward extrapolation of the census/survey observations on attainment by age group, with an appropriate time lag. Table A.1 below summarizes the backward and forward estimation procedure by age group.

Table A.1. General Rules for Estimating Missing Observations through Forward and Backward Extrapolation

Age group (*a*)	Backward extrapolation	Forward extrapolation
15–19	$h_{j,t}^{15-19} = h_{j,t+5}^{15-19} - \Delta enroll_{j,t}^{15-19}$	$h_{j,t}^{15-19} = h_{j,t-5}^{15-19} + \Delta enroll_{j,t}^{15-19}$
20–24	$h_{j,t}^{20-24} = h_{j,t+5}^{20-24}$	$h_{j,t}^{20-24} = h_{j,t-5}^{20-24} + \Delta enroll_{j,t}^{20-24}$
25–29, 30–34,…, 55-59, 60–64	$h_{j,t}^{a} = h_{j,t+5}^{a+1}$	$h_{j,t}^{a} = h_{j,t-5}^{a-1}$
65-69	$h_{j,t}^{a} = h_{j,t+5}^{a+1} \cdot (1-\rho_j^a)^{-1}$	$h_{j,t}^{a} = h_{j,t-5}^{a-1}$
70–74, 75–79	$h_{j,t}^{a} = h_{j,t+5}^{a+1} \cdot (1-\rho_j^a)^{-1}$	$h_{j,t}^{a} = h_{j,t-5}^{a-1} \cdot (1-\rho_j^a)$

Note: $h_{j,t}^a$ is the proportion of people in age group *a*, for whom *j* is the highest level of schooling attained at time *t*, $\Delta enroll_{j,t}^a$ is the enrollment adjustment factor for age group *a* in level *j* at time *t*, and $(1-\rho_j^a)$ is the survival ratio for age group *a* in education level *j* over the five year at time *t*.

We perform either backward or forward extrapolation when at least one benchmark figure is available from either an earlier or later period. If more than one benchmark figure is available, we use the figure from the closest period as the benchmark figure.

Aged 25–64. We assume that an individual's educational attainment remains unchanged from age 25 to 64 and that mortality is uniform across all individuals, regardless of educational attainment. Hence, for age groups between 25 and 64, we fill the missing attainment data using the attainment of the younger age group from the previous period (forward) as benchmark or the attainment of the older age group from the succeeding period (backward).

Aged 15–19 and 20–24. Since direct forward extrapolation is not applicable for these two youngest age groups, we use attainment and enrollment data to estimate missing attainment data. We assume that the change in enrollment leads to a proportional change in attainment over time with time lag. Hence, for these age groups, we use estimates for the same age group from the previous (or in the next) period as benchmark and adjust this benchmark figure by the change in enrollment over time or the enrollment adjustment factor. Table A.2 summarizes how the age-specific enrollment adjustment factors are derived in case of forward extrapolation. When using backward extrapolation for the 20-24 age group, we use the estimate for the 25-29 age group from a later period without adjustment (see Table A.1)

Table A.2. Enrollment Adjustment Factor for Age Groups, 15–19 and 20–24

Level	Forward extrapolation
No education	$-(enroll^a_{pri,t} - enroll^a_{pri,t-5})$
Primary	$(enroll^a_{pri,t} - enroll^a_{pri,t-5}) - (enroll^a_{sec,t} - enroll^a_{sec,t-5})$
Secondary	$(enroll^a_{sec,t} - enroll^a_{sec,t-5}) - (enroll^a_{ter,t} - enroll^a_{ter,t-5})$
Tertiary	$(enroll^a_{ter,t} - enroll^a_{ter,t-5})$

Note: $enroll^a_{j,t}$ is the enrollment rate for age group a in level j at time t.

Aged 65 and over. For older age groups, however, we distinguish between a less-educated population (uneducated and people who have reached the primary level) and a more-educated population (reached at least secondary schooling). We assume mortality is higher for the less-educated and lower for the more- educated. We estimate the survival rate for less-educated ($1-\rho_L^R$) and for more-educated ($1-\rho_U^R$) individuals, for advanced countries (R = OECD) and for developing countries (R = non-OECD) using a weighted least squares procedure with the available census information and the following equations.

$$h_{j,t}^a = (1-\rho_k^R) \cdot h_{j,t-5}^{a-1}, \; a = 70\text{-}74, \; 75\text{-}79; \; R = OECD, \; NOECD; \; k = U, L \qquad (1a)$$

$$h_{j,t}^a = (1-\rho_k^R)^2 h_{j,t-10}^{a-2}, \; a = 75\text{-}79; \; R = OECD, \; NOECD; \; k = U, L \qquad (1b)$$

We have obtained estimates $1-\hat{\rho}_L^{OECD} = 0.966$ (s.e. = 0.01, t-stat =87.9) and $1-\hat{\rho}_U^{OECD} = 1.065$ (s.e. = 0.02, t-stat = 65.7) for advanced countries, and $1-\hat{\rho}_L^{NOECD} = 0.969$ (s.e. = 0.01, t-stat =132.8) and $1-\hat{\rho}_U^{NOOECD} = 1.068$ (s.e. = 0.03, t-stat = 38.1) for developing countries. We then apply the estimated survival rate to adjust the backward or forward estimate for mortality rate differences between less-educated and more-educated individuals (see Table A.1.). Note that the estimated survival rate is a relative survival rate that measures the survival rate of each educational attainment group as a fraction of the overall survival rate of the relevant age group. The overall survival rates for each five-year age cohort are reflected in the change of population structure over time derived from population census data for individual countries.

If two or more benchmark figures are available from both earlier and later periods, a weighted average of backward and forward estimates is used as the benchmark. We derive the weights for combing the backward and forward estimates for OECD countries and for non-OECD countries, by estimating the following system of simultaneous equations through a weighted least squares estimation procedure. The estimation uses the sample of available actual censuses. We apply a same weight to all education categories to keep the sum of all education categories equal to 1,

$$h_{j,t}^{a} = \beta_{11}^{R} h_{j,t-5}^{a} + \beta_{21}^{R} h_{j,t+5}^{a} \text{ , where } \beta_{11}^{R} + \beta_{21}^{R} = 1 \tag{2a}$$

$$h_{j,t}^{a} = \beta_{12}^{R} h_{j,t-5}^{a} + \beta_{22}^{R} h_{j,t+10}^{a}, \text{ where } \beta_{12}^{R} + \beta_{22}^{R} = 1 \text{ and} \tag{2b}$$

$$\beta_{12}^{R} = \beta_{11}^{R} / (1 - \beta_{11}^{R} + \beta_{11}^{R\,2})$$

$$h_{j,t}^{a} = \beta_{13}^{R} h_{j,t-10}^{a} + \beta_{23}^{R} h_{j,t+5}^{a} \text{ , where } \beta_{13}^{R} + \beta_{23}^{R} = 1 \text{ and} \tag{2c}$$

$$\beta_{13}^{R} = \beta_{11}^{R\,2} / (1 - \beta_{11}^{R} + \beta_{11}^{R\,2})$$

$$h_{j,t}^{a} = \beta_{14}^{R} h_{j,t-10}^{a} + \beta_{24}^{R} h_{j,t+10}^{a} \text{ , where } \beta_{14}^{R} + \beta_{24}^{R} = 1 \text{ and} \tag{2d}$$

$$\beta_{14}^{R} = \beta_{11}^{R\,2} / (1 - 2\beta_{11}^{R} + 2\beta_{11}^{R\,2})$$

We have obtained $\widehat{\beta}_{11}^{OECD} = 0.461$ (se = 0.01, t-stat = 82.5) for OECD countries. For non-OECD countries, $\widehat{\beta}_{11}^{NOECD} = 0.549$ (se = 0.01, t-stat = 102.7).

2. *Estimation of missing data in subcategories*

To decompose the estimates of the three broad educated categories (i.e., *hp, hs, hh*) by subcategory (i.e., *hpi, hsi, hhi, hpc, hsc, hhc*), we estimate missing observations of completion ratios. For countries with complete and available completion ratio data (i.e., for all age groups at either the primary, secondary, or tertiary level) for at least one year, we use a backward or forward estimation procedure to estimate the completion ratio for earlier and later years, respectively. The following describes the procedure for estimating missing data on the completion ratio in more detail.

Estimating the primary and secondary completion ratio. Table A.3 below presents the rules for extrapolating from earlier or later years through a backward or forward extrapolation procedure for missing primary and secondary completion ratio data. If available, we use country-specific

completion ratio data to perform either or both backward or forward extrapolation of missing completion ratio data. Otherwise, we use the average estimates for group of OECD or non-OECD countries for the same age group and the same period.

Table A.3. Rules for Extrapolating Primary and Secondary Completion Ratio

Age group (a)	Backward extrapolation	Forward extrapolation
15–19	$c_{j,t}^{15-19} = c_{j,t+5}^{25-29} \cdot (c_{j,t+5}^{15-19} / c_{j,t+5}^{20-24})$	$c_{j,t}^{15-19} = c_{j,t-5}^{20-24} \cdot (c_{j,t-5}^{15-19} / c_{j,t-5}^{25-29})$
20–24	$c_{j,t}^{20-24} = c_{j,t+5}^{25-29}$	$c_{j,t}^{20-24} = c_{j,t-5}^{20-24} \cdot (c_{j,t-5}^{20-24} / c_{j,t-5}^{25-29})$
25–29, 30–34,…, 70–74	$c_{j,t}^{a} = c_{j,t+5}^{a+1}$	$c_{j,t}^{a} = c_{j,t-5}^{a-1}$
75–79	$c_{j,t}^{75-79} = c_{j,t+5}^{75-79} \cdot (c_{j,t+5}^{75-79} / c_{j,t+5}^{70-74})$	$c_{j,t}^{75-79} = sh_{t-5}^{70-74} \cdot c_{j,t-5}^{70-74} + sh_{t-5}^{75-79} \cdot c_{j,t-5}^{75-79}$

Note: $c_{j,t}^{a}$ is the completion ratio or the proportion of people in age group a, for whom j is the highest level of schooling attained at time t who have completed j. $sh_t^a = pop_t^a / pop_t^{70-79}$ or the share of the population in age group a (70-74, 75-79) in the population aged 70-79 years at time t.

Aged 15–19 and 20–24. As with attainment data, we cannot directly estimate the completion ratio for those aged 15–19 and 20–24. We assume that the distribution of completion between two age groups will be stable across time. Hence, as shown in the table below, to estimate the completion ratio for 15–19 year olds through forward estimation, we use the completion ratio for the older age group from a later period as the benchmark figure and adjust this by multiplying with the ratio between the completion ratio among 15–19 year olds and the completion ratio among 20–25 year olds during the benchmark period. For 20-24 year olds, we use the completion ratio for the same age group from the earlier period and adjust it by the ratio between the completion ratio of 20–24 and 25–29 year olds.

When using backward estimation, we use the completion ratio among 25–29 year olds from the later period for the completion ratio for 20–24 age group. For 15-19 year olds, we use the completion ratio among 25–29 year olds from the later period as benchmark and adjust it by the

ratio of the completion ratio between 15–19 year olds and 20–24 year olds during the benchmark period.

Aged 75 and over. For 75 years and above, we use the population weighted average of the completion ratios for 70–74 year olds and 75 and above during the reference period. Since direct backward estimation is not applicable for individuals aged 75 years and above, when using backward estimation we use the completion ratio of the same age group and adjust it by the ratio of completion ratio among 75–79 year olds during the benchmark period to that of 70–74 year olds.

If complete country-specific completion ratio data are available from both earlier and later periods, we combine backward and forward estimates using averages primary/secondary completion weights for group of OECD and non-OECD countries. We derive the weights by estimating the following system of simultaneous equations for each education category, using available completion ratio data through a weighted least squares estimation procedure.

$$c_{j,t}^{a} = \beta_{15}^{R} c_{j,t-5}^{a} + \beta_{25}^{R} c_{j,t+5}^{a} \text{ , where } \beta_{15}^{R} + \beta_{25}^{R} = 1 \tag{3a}$$

$$c_{j,t}^{a} = \beta_{16}^{R} c_{j,t-5}^{a} + \beta_{26}^{R} c_{j,t+10}^{a} \text{ , where } \beta_{16}^{R} + \beta_{26}^{R} = 1 \text{ and} \tag{3b}$$

$$\beta_{16}^{R} = \beta_{15}^{R} / (1 - \beta_{15}^{R} + \beta_{15}^{R^{2}})$$

$$c_{j,t}^{a} = \beta_{17}^{R} c_{j,t-10}^{a} + \beta_{27}^{R} c_{j,t+5}^{a} \text{ , } \beta_{17}^{R} + \beta_{27}^{R} = 1 \text{ and} \tag{3c}$$

$$\beta_{17}^{R} = \beta_{15}^{R^{2}} / (1 - \beta_{15}^{R} + \beta_{15}^{R^{2}})$$

$$c_{j,t}^{a} = \beta_{18}^{R} c_{j,t-10}^{a} + \beta_{28}^{R} c_{j,t+10}^{a} \text{ , } \beta_{18}^{R} + \beta_{28}^{R} = 1 \text{ and} \tag{3d}$$

$$\beta_{18}^{R} = \beta_{15}^{R^{2}} / (1 - 2\beta_{15}^{R} + 2\beta_{15}^{R^{2}}) .$$

For primary completion ratio, we have obtained the estimates: $\hat{\beta}_{15}^{OECD} = 0.475$ (se = 0.02, t-stat = 26.50) for OECD countries, and $\hat{\beta}_{15}^{NOECD} = 0.308$ (se = 0.04, t-stat = 8.45) for non-OECD countries, for secondary completion ratio, we have obtained $\hat{\beta}_{15}^{OECD} = 0.250$ (se = 0.05, t-stat = 4.78), for OECD countries, and $\hat{\beta}_{15}^{NOECD} = 0.593$ (se = 0.03, t-stat = 20.7) for non-OECD

countries.

B. Estimating tertiary completion ratio. Since tertiary data is not reported by subcategory for most countries, we use available tertiary completion ratio data reported by the UN Demographic Yearbook (various years) and Kaneko (1986) to derive country-specific and income/region (advanced and developing countries) tertiary completion ratio estimates by age group. Again, we use a backward and forward estimation procedure to estimate completion ratio for years with missing information.

Tertiary completion weights for group of OECD and non-OECD countries. If both backward and forward estimates are available, we again combine backward and forward estimates using tertiary completion weights for OECD and non-OECD countries. We also estimate these weights as in primary and secondary completion weights. Using the same method used for primary and secondary completion ratios, we have obtained estimates, $\widehat{\beta}_{h,15}^{OECD} = 0.511$ (se = 0.03, t-stat = 18.9) for OECD countries and $\widehat{\beta}_{h,15}^{NOECD} = 0.468$ (se = 0.06, t-stat = 8.32) for non-OECD countries.

Table A.4. Rules for Extrapolating Tertiary Ratio

Age group (a)	Backward extrapolation	Forward extrapolation
15–19, 20–24	$c_{j,t}^{a} = c_{j,t+5}^{a}$	$c_{j,t}^{a} = c_{j,t-5}^{a}$
25–29	$c_{j,t}^{25-29} = c_{j,t+5}^{30-34}$	$c_{j,t}^{25-29} = c_{j,t-5}^{25-29} \cdot (c_{j,t-5}^{25-29} / c_{j,t-5}^{30-34})$
30–34, 35-39,…, 70–74	$c_{j,t}^{a} = c_{j,t+5}^{a+1}$	$c_{j,t}^{a} = c_{j,t-5}^{a-1}$
75–79	$c_{j,t}^{75-79} = c_{j,t+5}^{75-79} \cdot (c_{j,t+5}^{75-79} / c_{j,t+5}^{70-74})$	$c_{j,t}^{75-79} = sh_{t-5}^{70-74} \cdot c_{j,t-5}^{70-74} + sh_{t-5}^{75-79} \cdot c_{j,t}^{75 \cdot}$

Note: $c_{j,t}^{a}$ is the completion ratio or the proportion of persons in age group a, for whom j is the highest level of schooling attained at time t who have completed j. $sh_{t}^{a} = pop_{t}^{a} / pop_{t}^{70-79}$ or the share of the population in age group a in the total population aged 75-79 years at time t.

Aged 15–19 and 20–24. We assume that tertiary completion is relatively stable for the two younger age groups. Hence, when using either a backward or forward estimation procedure to estimate missing tertiary completion ratio for the 15–19 and 20–24 age groups, we use an unadjusted tertiary completion ratio for the same age group in the benchmark year (see Table A.4).

Aged 25–29. When using forward estimation for ages 25–29, we multiply the completion ratio for the same age group with the ratio of completion ratio between 25–29 and 30–34 in the benchmark period.

Aged 75 and above. When estimating tertiary completion ratio for individuals 75 years and above through backward estimation, we also use the tertiary ratio for the same age group in the next period as the benchmark figure and adjust it by the ratio of the tertiary completion ratio between 75–79 to 70–74 in the same benchmark figure. When estimating through forward estimation, we use both the population weighted average of the tertiary completion ratio of 70–74 and 75 and above during the benchmark period.

Appendix Table. Availability of Educational Attainment Census/Survey Data by Country

Region/Country	No. of censuses	Original census year											
		1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005
Middle East and North Africa													
Algeria	4		1954		1966*	1971						2000	
Bahrain	4				1965#	1971				1991		2001	
Cyprus	5	1946#		1960						1992		2001	2005
Egypt	2						1976		1986				
Iran, Islamic Republic of	4		1956		1966						1996		2006
Iraq	2		1957		1965								
Israel	4			1961		1972		1982					2006
Jordan	2			1961				1979					
Kuwait	6					1970	1975	1980	1985		1995		2006
Libyan Arab Jamahiriya	3				1964#		1973		1984				
Malta	3	1948			1967								2005
Morocco	1					1971							
Qatar	2								1986				2004
Saudi Arabia	1												2004
Syrian Arab Republic	3			1960		1970						2002	
Tunisia	5				1966		1975	1980	1984		1994		
United Arab Emirates	2						1975						2005
Yemen	1						1975#						
Sub-Saharan Africa													
Benin	3							1979#		1992*		2000	
Botswana	4				1964#	1971		1981#		1991			
Burundi	1									1990			
Cameroon	1					1976							
Central Africa	2						1975			1988			
Congo	1							1984					
Cote d'Ivoire	2									1988	1998+		
Democratic Republic of the Congo	1		1955										
Gabon	1										1993		
Gambia	3						1973				1993	2000	
Ghana	2			1960#		1970							
Kenya	3			1962		1969		1979					
Lesotho	2				1966		1976						
Liberia	2			1962#			1974#						
Malawi	4				1966		1977		1987			1998	
Mali	1						1976						
Mauritania	1									1988			
Mauritius	6	1952		1962		1972			1983	1990		2000	
Mozambique	2							1980			1997		
Namibia	3			1960						1991			

Region/Country	No. of censuses	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005
												2001	
Niger	1						1977						
Reunion	2		195#4		1967#								
Rwanda	1							1978					
Senegal	1						1976						
Sierra Leone	1				1963#								
South Africa	6			1960		1970		1980*	1985		1996	2001	
Sudan	2		1956						1983				
Swaziland	3				1966		1976		1986				
Togo	2					1970		1981					
Uganda	4			1959#		1969				1991		2002	
United Republic of Tanzania	1											2000	
Zambia	4					1969		1980#		1990	1993#		
Zimbabwe	2									1992		2002	
Latin America and the Caribbean													
Argentina	6	1947#		1960#		1970		1980#		1991		2001	
Barbados	3					1970		1980				2000	
Belize	4			1960		1970		1980		1991			
Bolivia	3						1976			1992		2001	
Brazil	5	1950				1970	1976	1980					2004
Chile	6	1952		1960#		1970		1982		1992		2002	
Colombia	4	1951					1973#				1993		2006
Costa Rica	5	1950			1963	1968	1973						2007
Cuba	2		1953					1981					
Dominican Republic	2			1960#		1970							
Ecuador	6	1950		1962			1974	1982		1990		2001	
El Salvador	5	1950		1961#		1971*				1992			2006
Guatemala	6	1950			1964#		1973	1981				2002+	2006
Guyana	3					1970		1980				2002	
Haiti	4	1950				1971		1982	1986*				
Honduras	3			1961			1974		1983				
Jamaica	5			1960		1970		1982		1991		2001	
Mexico	6			1960#		1970*#		1980		1990		2000	2006
Nicaragua	2	1950				1971*							
Panama	6	1950		1960		1970*		1980		1990		2000	
Paraguay	7	1950		1962#		1972#		1982		1992		2002	2006
Peru	4			1961		1972		1981			1993		
Trinidad and Tobago	3					1970		1980		1990			
Uruguay	5				1963		1975		1985		1996		2006
Venezuela	5	1950		1961		1971		1981		1990			
East Asia and the Pacific													
Brunei Darussalam	3			1960		1971		1981					
Cambodia	1											1998	

Region/Country	No. of censuses	Original census year											
		1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005
China	3							1982		1990*		2000	
China, Hong Kong SAR	7				1966	1971	1976	1981	1986	1991		2001	
China, Macau SAR	4					1970				1991#		2001+	2006
Fiji	5				1965		1976		1986		1996		2007
Indonesia	5			1961		1971		1980		1990		2000#	
Lao, People's Democratic Republic	1										1995		
Malaysia	4			1957#				1980*		1991		2000	
Mongolia	2									1989		2000	
Myanmar	4		1953#				1973		1983	1991			
Papua New Guinea	3					1971		1980				2000#	
Philippines	8	1948	1956	1960#		1970	1975#	1980#		1990#		2000	
Republic of Korea	10		1955#	1960	1965	1970	1975	1980	1985	1990	1995	2000	
Singapore	5					1970		1980		1990		2000#	2006
Taiwan	5				1965#		1975	1980*				2001	2005
Thailand	4			1960		1970		1980				2000	
Tonga	3								1986#		1996		2006
Viet Nam	2							1979		1989			
South Asia													
Afghanistan	1							1979					
Bangladesh	4			1961#			1974	1981				2001	
India	4			1961		1971#		1981		1991			
Maldives	3								1985			2000	2006
Nepal	5			1961		1971		1981*		1991		2001	
Pakistan	5			1961#				1981		1990		1998	2006
Sri Lanka	4				1963	1969		1981				2001+	
Europe and Central Asia													
Albania	1											2001	
Armenia	1											2001	
Bulgaria	4		1956		1965					1992		2001	
Croatia	2									1991		2001	
Czech Republic	5			1961		1970		1980		1991			2006
Estonia	2									1989		2000	
Hungary	6			1960	1963	1970		1980		1990		2001	
Kazakhstan	2									1989		1999	
Kyrgyzstan	1											1999	
Latvia	2									1989		2000	
Lithuania	3									1989		2001	2007
Republic of Moldova	1									1989			
Poland	5			1960		1970		1978		1988		2002	
Romania	5		1953#		1966		1977*			1992		2002	
Russian Federation	5			1959		1970				1989+	1994**	2002+	
Serbia	5		1953*			1971		1981		1991		2002	
Slovakia	5			1961		1970		1980		1991		2001	
Slovenia	5			1961+		1971+		1981+		1994+		2002+	

Region/Country	No. of censuses	Original census year

Region/Country	No. of censuses	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005
Tajikistan	1									1989			
Ukraine	2					1970						2001	
Advanced Countries													
Australia	3				1966	1971		1981					
Austria	5			1961#		1971		1981		1991			2005
Belgium	3			1961#		1970							2006
Canada	9	1951		1961		1970	1975	1981	1986	1991		2001	2006
Denmark	4							1983		1991	1994	2001	
Finland	8	1950		1960		1970#		1980	1985	1990		2000	2006
France	5		1954#	1962				1982		1990			2004
Germany	6					1970#	1978	1980	1985			2001	2006
Greece	6	1951		1961				1981		1991		2001	2005
Iceland	1			1960									
Ireland	5				1966	1971		1981		1991		2002	
Italy	6	1951		1961		1971		1981				2001+	2005
Japan	4			1960		1970		1980		1990			
Luxembourg	2									1991		2001	
Netherlands	3			1960		1971							2005
New Zealand	5				1966#		1976	1981		1991		2001	
Norway	8	1950#		1960		1970	1975	1980		1990		2001	2006
Portugal	5			1960		1970		1981		1991			2006
Spain	4					1970		1981		1991			2006
Sweden	5					1970#	1974	1979			1995		2005
Switzerland	5			1960		1970		1980				2000	2005
Turkey	7	1950#			1965#		1975	1980	1985		1993		2006
United Kingdom	4	1950#		1961		1971	1976#						
United States	8	1950		1960		1970		1980		1990*	1994	2002	2005+

Notes: * indicates that the census has information for total population only; # indicates that the census has information for a broad age group only; + indicates that census data are from individual countries' national sources.
SAR = Special Administrative Region