Political Concepts
Committee on Concepts and Methods
Working Paper Series

57

November 2012

Gauging Cross-National Differences in Education Attainment

Peter Nardulli (nardulli@illinois.edu)

Buddy Peyton

Joe Bajjalieh

Cline Center for Democracy,

University of Illinois at Urbana-Champaign

C&M
The Committee on Concepts and Methods
www.concepts-methods.org

IPSA
International Political Science Association
www.ipsa.org

CIDE
Research and Teaching in the Social Sciences
www.cide.edu

The C&M working paper series are published by the Committee on Concepts and Methods (C&M), the Research Committee No. 1 of the International Political Science Association (IPSA), hosted at CIDE in Mexico City. C&M working papers are meant to share work in progress in a timely way before formal publication. Authors bear full responsibility for the content of their contributions.

The Committee on Concepts and Methods (C&M) promotes conceptual and methodological discussion in political science. It provides a forum of debate between methodological schools who otherwise tend to conduct their deliberations at separate tables. It publishes two series of working papers: "Political Concepts" and "Political Methodology."

Political Concepts contains work of excellence on political concepts and political language. It seeks to include innovative contributions to concept analysis, language usage, concept operationalization, and measurement.

Political Methodology contains work of excellence on methods and methodology in the study of politics. It invites innovative work on fundamental questions of research design, the construction and evaluation of empirical evidence, theory building and theory testing. The series welcomes, and hopes to foster, contributions that cut across conventional methodological divides, as between quantitative and qualitative methods, or between interpretative and observational approaches.

Submissions. All papers are subject to review by either a member of the Editorial Board or an external reviewer. Only English-language papers can be admitted. Authors interested in including their work in the C&M Series may send their paper to wp[at]concepts-methods.org. They may also seek initial endorsement by one editorial board member.

The C&M webpage offers full access to past working papers.

www.concepts-methods.org

Table of Contents

Gauging Cross-national Differences in Education Attainment

Introduction

Human capital is an important concept within the Societal Infrastructures and Development project (SID); it is a welfare indicator, a factor that shapes developmental processes, and a driver of other welfare indicators, such as economic growth. Educational attainment is a widely used and accepted indicator of the stock of human capital in a country (for example, see Psacharopoulos and Arriagada, 1986, 1992; Lau, Jamison, and Louat, 1991; Barro and Lee, 1993, 2001; Nehru, et. al., 1995). The main source of cross-national data on educational attainment is the United Nations Educational, Scientific and Cultural Organization (UNESCO).

UNESCO collects education data by surveying its member countries; its methodology is relatively straight-forward. At five-year intervals it sends questionnaires to individual countries. The first year with truly comprehensive educational data is 1960; the series continues at five-year intervals through 2005. UNESCO's questionnaire is sent to the governmental unit in each country that is responsible for handling educational data. These units are designated by the member countries and are normally an education-related ministry or an organization that collects national statistical information. While UNESCO personnel work with the designated organizations within member countries to generate accurate information, the quality of the data undoubtedly varies across countries. Moreover, there is some "play" in what is acceptable to submit to UNESCO. For example, while the UNESCO questionnaire explicitly solicits information for a given year, its instruction manual says that if the data solicited are not available a country may report the most recent data available. Countries are also encouraged by UNESCO to provide estimates for missing or incomplete data.

Data quality issues in the UNESCO archive are compounded by the fact that countries are not obligated to complete the questionnaire. Consequently, there is a considerable amount of missing data, though it varies by year, country, and region of the world. This is troubling because some of the missing data are systematic (i.e., non-random); they disproportionately affect poor countries or politically isolated entities that lack either the resources or the will to complete the survey. There is also a geographical bias; most of the countries lacking data are African, Asian, Middle Eastern or Southeast Asian; after the break-up of the Soviet Union much data are missing for the post-Soviet States. Because of this missing data problem various researchers have developed estimation procedures to address it. The importance of educational attainment data to the SID project is such that a significant investment was made to build on these efforts.

Our principal objective was to generate as complete an educational attainment dataset as possible for the 175 countries in the SID project for as much of the post-WWII era (to 2005) as possible. The philosophy underlying these efforts rests on two premises. The first is the fact that the non-random occurrence of missing data on factors such as educational attainment can bias statistical analyses. Thus, having missing data for a disproportionate share of exceptionally poor countries (e.g., African nations) can generate

misleading inferences about the developmental role of education. The second factor is that the existence of "good data" for almost two-thirds of the SID country-years means that we have a solid information base for estimating missing data – at least for selected countries and time frames. The value of the existing data is enhanced by the fact, for most countries and most time frames, educational attainment changes slowly and is highly structured – over time and across cognate countries.

Within a given country the proportion of residents who complete secondary education does not fluctuate greatly over time; that is, countries do not routinely experience marked annual jumps in attainment. Moving large segments of the population through educational institutions requires time and resources – and is constrained by communal norms and individual lifestyles. All of these factors militate against large variations and distinguish educational attainment data from such data as inflation and crop yields, which can fluctuate markedly from year to year. The structured nature of educational data also means that, even if we totally lack data for a country, we are not necessarily without information to generate refined estimates. For example, the available data on African educational attainment rates suggests that they are far below global averages. This provides the basis for making some informed estimates about education attainment rates in African countries by using cognate groups of African countries with available data. The error introduced by using such estimates may be less than the bias introduced by failing to incorporation a sizeable and distinct subset of the global community of nations.

To achieve our objective we examined the structure of the missing educational attainment data and decomposed the problem into distinct sources of missing data. Whenever methodologically sound solutions were available, we devised source-specific strategies to estimate the missing data. We used the augmented dataset to construct a pair of composite variables that measure educational attainment for two age groups that constitute the standard groupings for country-level educational attainment data. **E_ATTAIN$_{15+}$** refers to educational attainment (in years) for the proportion of the population that is fifteen years of age and over; **E_ATTAIN$_{25+}$** refers to educational attainment for the proportion of the population that is twenty-five years of age and over. Our efforts to address the missing data problems in the available educational attainment data – which involved a three-wave procedure – are described, illustrated and analyzed in the following sections.

The first section discusses the source of the root data for educational attainment: UNESCO data for four educational attainment variables at five-year intervals – as modified by Barro and Lee (2001). This section also describes the modifications to the 2001 Barro and Lee dataset that were made to integrate it with the SID project, which involved adding some countries and deleting others. The second section outlines our "wave-one" efforts to address the missing data. The addition of more than forty countries to the Barro and Lee dataset, and its extension to 2005, resulted in missing data for about 36% of the country-years. Most of our efforts in the first wave focus on estimating

missing data for the 15 and older group, for reasons rooted in the structure of one of the most powerful of the estimation techniques (the perpetual inventory method).[1]

The third section reports on the procedures used to transform the four educational variables in each dataset (15 and older, 25 and older) to generate a preliminary version of the **E_ATTAIN$_{15+}$** and **E_ATTAIN$_{25+}$** variables. The fourth section reports our "wave two" efforts, which involved using the composite variables to predict missing values in one composite from available data in the other (e.g., **E_ATTAIN$_{15+}$** is used to predict **E_ATTAIN$_{25+}$** , where **E_ATTAIN$_{25+}$** is missing; and vice versa). These efforts reduced the missing data to 6.2 % of the country-years for the period from 1960 to 2005. Thus, for this period we have complete data on 162 countries, all but 13 of the 175 countries in the SID project.

The fifth section explains our "wave-three" efforts. In this phase we developed procedures to "backcast" the 1960-2005 dataset to 1950 for the ninety-two countries that were independent before 1960. This effort is made possible because of availability of fairly complete data for the composite variables in the 1960-2005 era. The sixth section summarizes the role of the different estimation strategies, illustrates the SID educational attainment data by displaying it across countries and over time; and assesses the credibility of the augmented dataset by comparing it with the Barro and Lee in an examination of economic growth.

The Barro and Lee Refinements to the UNESCO Educational Attainment Data

Description of the Barro-Lee Data Data

A review of the pertinent literature suggested that the work of Barro and Lee in refining and extending the original UNESCO data has been the most sophisticated and ambitious. Their work is reported in Barro and Lee (2001). Because of its standing in the field, we began our effort by obtaining the Barro-Lee dataset from the web site of the Center for International Development at Harvard University (BL2000).[2] The Barro and Lee archive is actually composed of two datasets. The first contains attainment data for the population 15 and older (15+); the second contains attainment data for the population 25 and older (25+). Each dataset gauges a country's educational attainment by listing the proportion of the population in each of four categories:

[1] Focusing on the 15 and older group generates the most complete set of data possible because deploying the perpetual inventory method requires less data to produce estimates, as noted below. This is an efficient strategy because, once the available data for the 15 and older group is transformed into a composite variable (**E_ATTAIN$_{15+}$**) it is possible to generate refined estimates for the 25 and older group (**E_ATTAIN$_{25+}$**).

[2] http://www.cid.harvard.edu/ciddata/ciddata.html.

- proportion with no education;
- proportion whose highest education level was primary education;
- proportion whose highest education level was secondary education; and
- proportion whose highest education level was tertiary education.

Each data set contains observations for each country at five year intervals from 1950 through 2000. The variable names and descriptions of the data sets are listed below.

Population 15 and Older

COUNTRY	Country Name
YEAR	Year
POP_15	Population 15 and older (in thousands)
NOEDUC_15	Percent of 15 and older population with no education
PRIM_15	Percent of 15 and older population attained primary
SEC_15	Percent of 15 and older population attained secondary
TERT_15	Percent of 15 and older population attained tertiary

Population 25 and Older

COUNTRY	Country Name
YEAR	Year
POP_25	Population 25 and older (in thousands)
NOEDUC_25	Percent of 25 and older population with no education
PRIM_25	Percent of 25 and older population attained primary
SEC_25	Percent of 25 and older population attained secondary
TERT_25	Percent of 25 and older population attained secondary

BL2000 data has observations for 141 countries for the 15+ group and the 25+ group. There are missing data for about 1% of the four attainment variables in both datasets.

Modifications to the Barro and Lee Dataset

A number of modifications were made to the BL2000 data to integrate it with the SID project. The first was to delete observations for countries with populations less than 500,000 (i.e., the micro-states). Ten microstates were eliminated from the 15+ group and eleven were deleted from the 25+ group; Appendix E-1 contains a list of the countries and years deleted because of their population size. The second modification pertains to the 25+ group. Thirty-seven countries contained observations of attainment data for either 1950 or 1955 and, in one case (Philippines), data for both 1950 and 1955. To simplify the generation of missing data for the 1960-2005 period, these observations were deleted. These observations were reintegrated into the dataset after estimates for the 1960-2005 period were made, as noted above. Appendix E-2 contains a list of the countries and years deleted because they were before 1960.

The third modification was to add the following variables to each data set: 1) a country identification variable containing the country's Correlates of War country code (**COWCODE**); 2) a dummy variable indicating whether or not the country was

independent (**INDEP**) in a given year; [3] and 3) a dummy variable (**BLEE**) indicating whether or not the country-year observation was in the original BL2000 dataset. The addition of these variables facilitates merging these data with other datasets, identifying the missing data that needed to be estimated (we only provide estimates for independent nations), and differentiating the original Barro and Lee data from the estimates provided here. Having the capacity to differentiate the original data from the augmented data provides the basis for assessing the impact of the missing data estimated here.[4]

A fourth modification was necessary because the BL2000 datasets are not balanced panels (i.e., some countries that existed in the 1960-2000 time-frame do not have observations on attainment data for every year). For example, while Libya was an independent country between 1960 and 2000, it has only three observations (1965, 1975, and 1985). Thus, the BL2000 datasets were converted into balanced panel datasets, with an entry for every country for every year. Next, the timeframe for the newly balanced educational attainment datasets was expanded to include an observation for 2005. Balancing these data and adding the 2005 observation had the effect of increasing the amount of missing data to about seventeen percent in both datasets.[5]

The final modification was to include the countries in the SID project that were missing from the BL2000 data; SID includes all countries with a population of at least 500,000 in 2004. This required the addition of 43 countries to the 15 and older dataset and 44 countries to the 25 and older dataset. Appendix E-3 contains a list of the countries added. After adding the addition countries and balancing the panel, the BL2000 data provides attainment data for 64% of the balanced panel.

Wave One Strategies for Missing Data Reduction

We used seven different strategies in our wave-one efforts. We began by scouring the UNESCO data archive for data and developing perpetual inventory model estimates where possible. We also used literacy rates as estimate of "no education" rates, interpolated data values where possible, and extrapolated to fill in missing values in initial years of a country's time-series – where needed and possible. We also developed strategies to replace missing values in African countries and post-Soviet states. Finally, we extended the series to 2005. The details for each of these strategies are reported below and in the accompanying appendices.

[3] The independence variable is based on the Gleditsch and Ward's release 3.2 dated 22 January 2007.

[4] We also include a dummy variable for each estimation technique developed. This allows users to be selective in which of the estimation techniques they are comfortable in using, if any.

[5] The percent of missing data was calculated by first computing the number of possible data points in each year. The number of data points in any given year is equal to the number of independent countries multiplied by four. The total number of possible data points is the sum of these yearly figures. The percent missing is the ratio between the existing data points and the total possible data points.

UNESCO Data, Country Archives, and Perpetual Inventory Model Estimates

We began our efforts to address the missing data problem by searching the UNESCO Statistical Yearbooks from 1963 to 1997 for attainment data not included in the BL2000 data set. At the same time we also collected enrollment ratios for primary, secondary, and tertiary schooling. Where available, the net enrollment ratios were collected.[6] If net enrollment ratios were unavailable, gross enrollment ratios were collected.[7] Enrollment ratios after 1999 were collected from the UNESCO Institute for Statistics Data Center.[8]

In countries with at least one available data point for the attainment variables and available enrollment ratios, we followed Barro and Lee by using perpetual inventory methods to estimate missing attainment data.[9] These methods are used to estimate capital stock for a particular point in time given information on: (1) capital stock at a previous or later point in time, (2) investment, and (3) depreciation of capital stock. Estimates of capital stock using perpetual inventory methods may be either forward-flow or backward-flow estimates (Barro and Lee, 1993). Estimates of capital stock for a given time based on earlier capital stock values are referred to as forward-flow estimates. Equation (1) contains an example of a forward-flow estimate where K_t is the value we wish to estimate (capital stock at time t), I_t is information on investment at time t, K_{t-1} is the value for capital stock at time $t-1$, and r is information pertaining to depreciation of capital stock.[10]

$$K_t = K_{t-1}(1-r) + I_t \qquad (1)$$

Equation (2) contains an example of a backward-flow estimate in which the value of capital stock at a previous point is based on later capital stock values, investment, and depreciation.

$$K_{t-1} = \frac{K_t - I_t}{(1-r)}(1-r) \qquad (2)$$

There are four attainment variables that can be used as measures for human capital stock for each population group: the number of persons with no education (H_0), the number of

[6] UNESCO defines net enrollment ratios (NER) as follows: The enrollment of the official age group for a given level of education expressed as a percentage of the corresponding population.

[7] UNESCO defines gross enrollment ratio (GER) as follows: The total enrollment in a specific level of education, regardless of age, expressed as a percentage of the eligible official school-age population corresponding to the same level of education in a given school year. When the net enrollment ratio is compared to the gross enrollment ratio, the difference between the two highlights the incidence of under-aged and over-aged enrollment in a given level of education.

[8] http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=143&IF_Language=eng

[9] Estimates for missing data in the 25 and older data set followed the methodology in Barro-Lee (1993); estimates for missing data in the 15 and older data set followed the methodology in Barro-Lee (2000).

[10] Notation is from Baffes and Shah (1998, p. 294).

persons who attained a primary education (H_1), the number of persons who attained a secondary education (H_2), and the number of persons who attained a tertiary education (H_3). We draw from Barro and Lee (2000) for an example of how we estimated missing values in the education attainment data for the population 15 and older. We know the number of people in the population without education at time *t-5* ($H_{0,t-5}$) and wish to estimate the number of people with no education at time *t* ($H_{0,t}$).[11] Substituting the variable representing the population with no education into equation (1) yields our forward-flow estimate of missing data.

$$H_{0,t} = H_{0,t-5}(1-r) + I_t \tag{3}$$

Investment in human capital (as measured by education attainment) does not immediately enter the labor force, as does investment in some physical capital. Rather, there is a delay between the actual investment and when the fruit of such investment enters the labor force and is reflected in our measure of human capital stock (Nehru, 1995). At time *t*, the cohort of persons between the ages of 15 and 19 becomes visible to our measure of human capital stock. We assume that anyone with no education at time *t* was not enrolled in primary schooling at time *t-5*. Thus, our measure of investment in human capital is the number of people in the population between age 15 and 19 at time *t* who *were not enrolled* in primary school at time *t-5* where $L15_t$ is the population age 15-19 at time *t* and PRI_{t-5} is the portion of primary school aged children who were enrolled in primary school at time *t-5*.

$$H_{0,t} = H_{0,t-5}(1-r) + \left[L15_t\left(1 - PRI_{t-5}\right)\right] \tag{4}$$

There is no universally known rate of depreciation in human capital stock; thus, a common measure of depreciation is some mortality rate (δ). We can then substitute the mortality rate in equation (4) as our rate of depreciation in human capital stock.

$$H_{0,t} = H_{0,t-5}(1-\delta) + \left[L15_t\left(1 - PRI_{t-5}\right)\right] \tag{5}$$

Barro and Lee (2000) specify the morality rate as the ratio of the population 15 and older at time *t-1* (L_{t-5}) and the sum of the population 15-19 at time *t* ($L15_t$) and the population 15 and older at *t-5* (L_{t-1}) minus the population 15 and older at time *t* (L_t).

$$\delta = \frac{\left(L15_t + L_{t-5} - L_t\right)}{L_{t-5}} \tag{6}$$

To make the education attainment measure comparable across countries, we want to convert the raw number of people 15 and older with no education (H_0) into the percent of the population 15 and older with no education (h_0). This is calculated by dividing the number of people 15 and older with no education at a given time by the number of people in the population 15 and older at the same point in time. More formally,

[11] It is important at this point to remember that these data are in country-year format at five year intervals.

$$h_{0,t} = \frac{H_{0,t}}{L_t} \quad \text{and} \quad h_{0,t-5} = \frac{H_{0,t-5}}{L_{t-5}} \tag{7}$$

Substituting equation (6) into equation (5) and converting the raw numbers of persons with no education into proportions of the population yields equation (8), our forward-flow estimate of the percent of the population 15 and older at time *t* with no education (see Appendix E-4 for the proof).

$$h_{0,t} = h_{0,t-5}\left(\frac{L15_t}{L_t}\right) + \left(\frac{L15_t}{L_t}\right)\left(1 - PRI_{t-5}\right) \tag{8}$$

This example uses the estimator from Barro and Lee (2000) for the percent of the population 15 and older with no education. Appendix E-5 contains both the forward and backward-flow estimators for the remaining measures of human capital for a population 15 and older as well as those for a population 25 and older. These methods account for 3% of the 15+ dataset and 4% of the 25+ dataset.[12] Appendix E-6 contains a list of the countries and years added.

Literacy Rates and Educational Attainment

Barro and Lee (1993, 2000) use a country's illiteracy rate as a proxy for the no-education attainment variable and we followed their lead in this regard. This technique was applied to both data sets and accounts for 0.7% of the 15+ data and 1.3% of the 25+ data. Appendix E-7 contains a list of the countries and years added using this technique.

Interpolation

There were 100 instances in which a country was missing data between temporal data points (e.g., a country had an observation for 1985 and 1995 but was missing an observation for 1990). In these cases we used a linear interpolation to estimate the missing value. This technique accounts for 0.2% of the 15+ data and 0.8% of the 25+ data. Appendix E-8 contains a list of the countries and years added.

Extrapolation: Estimating Missing Initial Years

In many countries missing data existed for the initial year in their time-series. The years affected varied because not every country was independent at the beginning of the UNESCO series (1960). Missing data problems on the end-points of a time-series are

[12] The reason for the greater reduction in missing values for the 15 and older data compared to the 25 and older data lies in the formula of the perpetual inventory models. For example, to estimate a value for 2005 in the 15 and older data requires the primary enrollment ratio for 2000 (t-5) and the secondary and tertiary enrollment ratios for 2005 (t). The same estimate for the 25 and older data requires the primary enrollment ratio for 1990 (t-15), the secondary enrollment ration for 1995 (t-10), and the tertiary enrollment ratio for 2000 (t-5). There are a number of countries that came into existence after 1990 with the breakup of the USSR and Yugoslavia. Enrollment ratios for these countries do not appear in the data until 1995 or later. This lack of data prevented a forward flow estimate for 2005.

more challenging to address than those within a series and estimates could not be generated for all countries. However, the values of the existing data and the distribution of missing data for seven countries allowed us to make reasonable estimates. The countries affected were Bahamas, Benin, China, Egypt, Libya, Mauritania, Mongolia, the Democratic Republic of Vietnam, and the Arab Republic of Yemen. Moreover, the estimates for the Democratic Republic of Vietnam and the Arab Republic of Yemen were used as estimates for two other countries that were later subsumed within them: the Republic of Vietnam (previously South Vietnam) and the People's Republic of Yemen.

A combination of factors made the estimation of missing values tractable for these countries: the number of variables with missing data, the values of the variables for the first year with good data, and the distribution of values in subsequent years. Consider, for example, Egypt, which had missing data for three of the four educational attainment variables for both age groups for 1960 and 1965. In this case, the values for these variables in 1970 made it feasible to extrapolate back to 1960 and 1965. Egypt's value for **NOEDUC_15** in 1970 is 89.2. Thus there was not much room for error since the value of **NOEDUC_15** is bounded by 100. Egypt's value for **PRIM_15** in 1970 is 7.7; **SEC_15** in 1970 is 2.4. Here again, there is not much room for error since these numbers cannot be less than 0 and the sum of the four educational variables must equal 100. The initial values for the Bahamas, Benin, Vietnam and Yemen were similar to Egypt's.

China presents a somewhat different problem. It had missing data for all of the educational variables but **TER_15** for 1960, 1965 and 1970. But its 1975 values for the variables that were missing during these years are more moderate. For example **NOEDUC_15** in 1975 is 40.2, **PRIM_15** is 27.5, and **SEC_15** in 1970 is 31.4. What made China's missing values problem tractable was that the average change in **PRIM_15** and **SEC_15** in the years after 1975 are relatively stable: 2.4 and 4.8, respectively. These rates of change were used to "backcast" the values for China in 1960-1970. The fact that we had available data for **TER_15** in these years, and that the four variables had to equal 100, made it possible to estimate **NOEDUC_15**. The distribution of Mongolia's values was somewhat similar to China's. Moreover, missing data existed in Mongolia only for 1960.

Backcasting the educational attainment data accounts for 1.2% of the 15+ data and 0.7% of the 25+ data. The countries and years affected are reported in Appendix E-9.

African Countries with Minimal Educational Attainment Data

Fourteen countries in Sub-Saharan Africa for which we have virtually no information on educational attainment are an important, and troubling, source of missing data. These countries are: Angola, Burkina Faso, Cape Verde, Chad, Comoros, Djibouti, Equatorial Guinea, Ethiopia, Eritrea, Gabon, Guinea, Madagascar, Nigeria and Somalia. While we were able to identify illiteracy rates for many years in these countries (see Appendix E-10) and estimate the data for the missing years,[13] the UNESCO archive and the Barro and

[13] The estimation of the missing data points for the illiteracy data was relatively straightforward because we had fairly complete data for the "book-ends" of the time-series for most of these countries. This, of course, bounded the generation of estimates and made possible a good deal of simple interpolation. For example, we had illiteracy values for the start-points in six of the fourteen countries (Cape Verde, Djibouti, Eritrea,

Lee datasets have no data on the standard attainment variables for them. This situation is troubling because these countries account for 28% of Sub-Saharan Africa nations that contain 38% of the 1990 population in Sub-Saharan Africa (4% of the 1990 world population).[14] Missing data for this many countries in a continent as economically, politically, and socially distinctive as Africa has the potential to skew findings concerning the developmental role of educational attainment. However, using the illiteracy rates as a proxy for **NOEDUC_15** makes it possible to address this situation in a methodologically sound manner, largely because of the availability of educational attainment data in thirty-six other African nations.

Two characteristics of the available African educational data make it possible to generate estimates for **PRIM_15**, **SEC_15**, and **TER_15** in these fourteen countries. The first is that the values of most of the education variables fall within relatively narrow bounds, particularly **TER_15**. The second is that a clear structure exists among the educational attainment variables within African countries with available data. These two factors make it possible to: (1) generate refined estimates for **TER_15**; and (2) allocate the remaining proportion of the population across **PRIM_15** and **SEC_15**. The next two sections address these points; the third section outlines the procedures used to estimate the missing data.

Bounded African Educational Attainment Data: TER_15

When compared to global averages, the thirty-six African nations with educational attainment data lie within a relatively narrow range at the lower end of the continuum, particularly **TER_15**. This observation has important implications for estimating educational attainment data for the fourteen African countries lacking data because it suggests that these estimates can be bounded within a fairly narrow range, for two

---

Gabon, Nigeria and Somalia) and we had end-points for all countries but Somalia. We were missing 1960 entries for three other countries (Chad, Ethiopia and Guinea). However, the rate of illiteracy in 1965 for these countries is so high (above 90% in each country) that reasonable estimates could be made for 1960. Burkina Faso was missing illiteracy data for the first three time-points in the series (1960, 1965, 1970), but it had a high rate of illiteracy in 1975 (91.3%). Moreover, Burkina Faso had available data for every year after 1975. Both factors made backcasts straightforward for Burkina Faso.

In addition to having relatively complete data on the book-ends, we also had a significant number of data points for most countries. Within the time frame being estimated (1960 to 2000) there is a maximum of 10 time-points; Angola does not enter the series until 1970; Djibouti in 1980; and Eritrea in 1995. We had data for at least half of these time-points in ten of the countries. The distribution of these time-points can be seen in Appendix E-10, which reports the matrix of illiteracy data used as proxies for **NOEDUC_15** in these countries. The bolded entries represent the actual data; the other entries are estimates. As Appendix E-10 makes clear, there are only three problematic countries. The first is Angola, which is missing the first two observations and has only three data points altogether. The second is Somalia, which only has two observations altogether and none after 1980. The third is Comoros, for the pre-1980 period. Despite the paucity of data for these countries we were able to use data on the inter-period changes in conjunction with the available data to generate serviceable estimates.

[14] These population figures are somewhat misleading as most of these fourteen countries are relatively small. Nigeria and Ethiopia are the only truly populous countries. If they are excluded from the calculations, remaining twelve countries account for only 8% of Sub-Saharan African population and less than 1% of the world population.

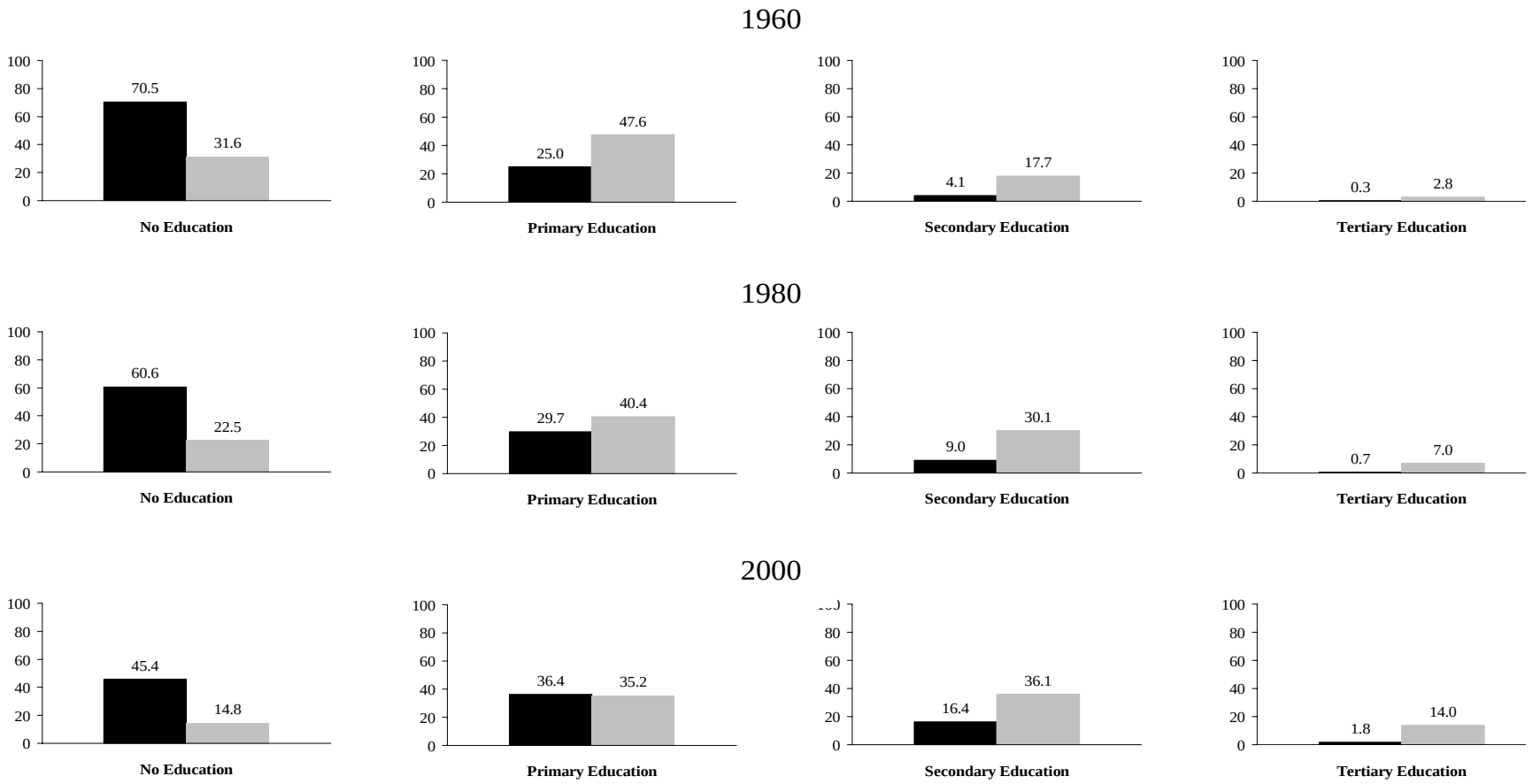
reasons. First, it is unlikely that the countries with no data will have systematically higher levels of educational attainment than those with data. This is particularly true since several of these nations (Angola, Eritrea, Ethiopia, Nigeria, Somalia, etc.) have experienced a good deal of civil strife since independence. This means that the available means can be used to provide reasonably good information on the *upper bounds* of the estimates. The *lower bounds* are set by natural limits (e.g., you cannot have less than 0% of the population without a college degree; you cannot have less than 0% of the population without a secondary degree, etc.).

Graph 1 compares the African means for **NOEDUC_15**, **PRIM_15**, **SEC_15**, and **TER_15** with the global means (excluding African countries) for three years (1960, 1980, and 2000). It illustrates, with the exception of **PRIM_15**, the enduring distinctiveness of African countries on these variables throughout the time frame. For example, the average for **NOEDUC_15** in the African states is more than twice as large as that for the rest of the world in 1960; it is about three times as large in 1980 and 2000. While the average value of **PRIM_15** in Africa is roughly half that for the rest of the world in 1960, by 2000 the African states match global norms. This is largely due to sustained drops in **PRIM_15** in other parts of world, which is a reflection of the fact that larger proportions of citizens are completing secondary education. The average value of **SEC_15** in Africa is less than a quarter of that for the rest of the world in 1960, about a third as large in 1980, and half as large in 2000. African nations fare even worse in terms of higher education attainment: their averages for **TER_15** are one-ninth, one-tenth, and one-seventh of the world averages for 1960, 1980 and 2000, respectively. Just as important as the distinctiveness of the African averages for **TER_15** are their exceptionally small values. The **TER_15** averages increase from only .3% of the population in 1960 to .7% in 1980 to 1.8% in 2000. These small averages, when combined with the natural limit of '0,' define a fairly narrow range that tightly structures the generation of estimates for **TER_15**.

Structure of the African Data

To generate more refined estimates for **PRIM_15, SEC_15** and **TER_15** we sought to identify some structure within the available data that would help to narrow the parameters of the estimates. We conducted two separate analyses. The first focused on **TER_15**; it examined the distribution of **TER_15** *across the African nations with available data*. Thus, it was concerned with differences in higher educational attainment within African countries. A number of structural factors were considered (e.g., wealth, colonial heritage, existence of enduring civil strife, religious profile, political regime, etc.). After evaluating these alternatives, as well as the quality of the data available to measure them, we decided to use per capita GDP. It is difficult to argue that any other factor would have a closer association with educational attainment; moreover, the *quality* of the data for per capita GDP greatly exceeded that for its competitors. Moreover, in analyzing the distribution of **TER_15**, it became clear that it was highly structured across levels of African economic development. While various economic groupings were examined (e.g., quintiles, quartiles, etc.), the small number of countries within each year (which ranged from18 to 36) limited the options significantly.

Graph 1
Sub-Saharan African Countries and World Education Attainment Rates for 15 and Older Age Group



The clearest and most consistent differences across **TER_15** emerge with a crude dichotomy, which divides the African "haves" (those in the top one-third in terms of per capita GDP in a given year), from the "have-nots" (those in the bottom two-thirds in terms of per capita GDP in a given year). The means and standard deviations of **TER_15** for these groupings are reported in Table 1. The overall mean of **TER_15** for the least developed group is 10% that of the more developed group. Moreover, the gap does not change much over time. Given the size of the means (which suggest that there is little room for noise) and the clear differences across the economic groupings, the data in Table 1 provide the basis for sound estimates of **TER_15**. Thus, they were used, in conjunction with information on the year and the country's economic status, as estimates of **TER_15** for the fourteen African countries with missing educational attainment data.

Table 1
Distribution of African Educational Attainment Variables, by Year and Economic Grouping

a. Economically Most Developed One-third

YEAR	TER_15	SD	N
1960	2.1	0.4	6
1965	2.2	0.2	7
1970	3.5	1.0	8
1975	5.4	1.1	8
1980	7.5	1.4	9
1985	8.1	1.0	10
1990	9.0	1.1	11
1995	10.4	2.1	11
2000	12.3	2.3	11

b. Economically Least Developed Two-thirds

YEAR	TER_15	SD	N
1960	0.3	0.4	12
1965	0.3	0.3	12
1970	0.3	0.3	10
1975	0.5	0.6	22
1980	0.5	0.4	22
1985	0.6	0.4	21
1990	0.8	0.5	23
1995	0.9	0.6	21
2000	1.2	0.8	25

With sound estimates available for **NOED_15** and **TER_15**, the remaining issue for completing the estimation of missing data is to apportion the proportion of the remaining population above 15 across **PRIM_15** and **SEC_15.** More concretely, the issue is – if the value of a country on **NOED_15** is 75.2 and its value on **TER_15** is 1.8 – how can we apportion the remaining 23% of the population across **PRIM_15** and **SEC_15**? To allocate this residual population in a methodologically responsible manner we took advantage of the relationship between **PRIM_15** and **SEC_15** across different levels of **NOED_15**. These variables are related because, in African countries with very low levels of educational attainment (i.e., high levels of **NOED_15**), the ratio of **PRIM_15** to **SEC_15** is much higher than in countries with high levels of educational attainment. This structure exists because educational attainment necessarily progresses incrementally. Age cohorts must initially attain a primary education; as development progresses more proceed to complete a secondary education.

The relationship between **PRIM_15** and **SEC_15** across different levels of **NOED_15** for the thirty-six African countries with available data can be seen in Table 2. If 30% or less of a country's population has not attained a primary education, the average ratio of **PRIM_15** to **SEC_15** is 1.6; if **NOED_15** is between 30% and 40% the ratio nearly doubles, to 3. However, if **NOED_15** is between 40% and 60%, the ratio increases to 4.4; if **NOED_15** is above 60% the ratio increases to 5.1. The differences across these ratios provide the capacity to generate empirically grounded estimates for **PRIM_15** and **SEC_15**.

Table 2
Distribution of the Ratio of PRIM_15 to SEC_15,
by Groupings of NOED_15

NOED_15	PRIM_15/SEC_15	SD	N
30% or Less	1.6	0.7	31
30%-39%	3	1.9	36
40% to 59%	4.4	3.1	72
60% of More	5.1	3.4	126

African Estimation Procedure

As noted above, the averages for **TER_15** reported in Table 1 were used as country-year specific estimates for **TER_15**. This procedure was straightforward for ten of the fourteen countries: they are consistently in either the most developed one-third of the nations or the least developed two-thirds. Djibouti and Gabon are always in the top one-third; Angola, Burkina Faso, Chad, Eritrea, Ethiopia, Madagascar, Nigeria, Somalia are always in the lower two-thirds. The other four countries (Cape Verde, Comoros, Equatorial Guinea, and Guinea) switched categories over time. Cape Verde vacillated between the categories before 1985 but remained in the top category beginning in 1985. Comoros slipped into the bottom tier in 1970 and never re-emerged. Equatorial Guinea remained in the bottom tier for most of the period (with the exception of 1975) but slipped into the top category in 2000. Finally, Guinea was in the top tier in every year but 1985 and 1990. To

deal with this variation in the categorization of these countries a weighted average was used. For example, Cape Verde was in the top economic grouping for six of the nine years and in the bottom grouping for three years. Thus, a weighted average of the year-specific averages was calculated, with the average for the top category given twice the weight of the average for the bottom category.

The estimates for **TER_15** were used in conjunction with the estimates of **NOEDUC_15** (i.e., the illiteracy rates) to define the educational attainment rates for two of the four variables. We used the estimates of **NOED_15** and **TER_15** to calculate a residual figure that is, by definition, the sum of **PRIM_15** and **SEC_15**. To allocate that residual across **PRIM_15** and **SEC_15** we used the ratios reported in Table 2. For example, the following algebraic formula was used to estimate **PRIM_15**:

if (**NOED_15** le .3) then **PRIM_15** = .63*(1-(**NOED_15** + **TER_15**));
if (**NOED_15** gt .3 and **NOED_15** lt .4) then **PRIM_15** = .75*(1-(**NOED_15** + **TER_15**));
if (**NOED_15** gt .4 and **NOED_15** lt .6) then **PRIM_15** = .77*(1-(**NOED_15** + **TER_15**));
if (**NOED_15** ge .6) then **PRIM_15** = .80*(1-(**NOED_15** + **TER_15**));

An analogous formula was used to calculate estimates for **SEC_15**.

Using the data reported in Table 2, in conjunction with the estimates for **NOED_15** and **TER_15**, allowed us to generate a complete set of educational attainment for the fourteen African nations with minimal data. This estimation procedure integrates a great of information on African educational attainment and takes advantage of both the structure of the data and natural limits to the values of the four variables. The estimation procedures account for 7.8% of the 15+ data.

Post-Soviet States with Minimal Data

Because they did not achieve independence until 1991, there is very little attainment data on the former soviet socialist republics in the Barro and Lee data. These countries include the Baltic States (Estonia, Latvia and Lithuania), several Eastern European countries (Belarus, Moldova, and Ukraine), a handful of South Asian republics (Armenia, Azerbaijan and Georgia), and the "Stans" (Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan, and Turkmenistan). Six of these countries had one data point in the Barro and Lee 2000 data (Estonia, Kazakhstan, Latvia, Lithuania, Moldova, and Tajikistan).

The paucity of educational attainment data for these countries is troubling because, like the African nations discussed earlier, they are geographically concentrated and culturally distinctive. However, we did not have to rely upon elaborate estimation procedures for these countries, both because of the short time span involved and the general availability of educational attainment data from country or region-specific data archives. These included such sources as the Commonwealth of Independent States Statistical Committee, the Estonian statistical database, and the Ukraine population census website. Native language speakers were retained to identify data sources and provide translations. In most cases educational attainment data were available before independence, which facilitated the interpolation of values where post-independence data were available. In

some cases, however, the category labels differed from the UNESCO terms and had to be reconciled with the four educational variables used here.15 The estimates for the post-Soviet states account for 2.8% of the 15+ data.

Miscellaneous Countries

The abovementioned efforts notwithstanding, base educational attainment data were still missing for a number of countries. Fortunately, data for a number of these countries – Cambodia, Laos, Macedonia, Maldives, and Saudi Arabia – are available from a project recently completed by Lutz et. al (2010).16 Their project provides educational attainment data for the 15 and older population in five year increments for the years 1970-2000. Moreover, their dataset is broken down by gender and age group. Appendix E-11 provides a breakdown of the data obtained from the Lutz et al. (2010).

Educational Attainment for 2005

The UNESCO Institute for Statistics published a Global Education Digest in 2008. The digest contained new attainment data for the 25+ group. This seemed to be an ideal source of data for 2005, one that would make possible the interpolation of the data between 2000 and 2005. However, the attainment categories in the 2008 report differed

---

[15] Where the category labels differed from UNESCO labels, the educational attainment variables used here were constructed as indicated below.

Baltic States

NOED_15 = No primary school (literate) + No primary school (illiterate) + Vocational education + Education unknown
PRIM_15 = Primary + General basic education + Vocational basic
SEC_15 = General secondary education + Vocational secondary education + Vocational education after secondary education + Professional secondary/ technical education + Professional secondary/technical education after secondary education
TER_15 = Higher education + Master's degree + Candidate of sciences/doctor's degree

Azerbaijan and Kyrgyzstan

NOED_15 = Total 15 and Older Population – Total Educated Population
PRIM_15 = Incompleted Secondary
SEC_15 = Incompleted Higher + Vocational + Secondary
TER_15 = Higher Education

Ukraine

NOED_15 = Letterless
PRIM_15 = Secondary total – Secondary complete + No Primary
SEC_15 = Higher Ed total – Higher Ed complete + Secondary complete
TER_15 = Higher Ed complete

Uzbekistan, Georgia, Turkmenistan

NOED_15 = Illiterate Population
PRIM_15 = Total Population – Tertiary – Secondary – Illiterate Population
SEC_15 = Secondary
TER_15 = Higher

[16] For information in regards to how the Lutz et al. (2010) dataset was constructed refer to this website: <http://www.iiasa.ac.at/Research/POP/edu07/index.html>

from the four categories employed in earlier reports. Moreover, comparing the attainment data from 2000 to 2005 revealed some unusually large changes in the attainment variables for countries where large changes would not be expected given historical patterns. For example, **SEC_25** increased by almost twenty points in Germany, a thirty-six point increase was registered in Italy, a twenty-six point decrease in Singapore, a forty point increase in Hungary, and a twenty-three point increase in Venezuela. These are remarkably large increases for a five-year period that led us to question the comparability of the current data with the historical data on attainment. Consider that the average absolute change between 2000 and 2005 is 6.5 points in **NOEDUC_25**, 10.3 in **PRIM_25**, 12.6 in **SEC_25**, and 4.1 in **TERT_25**. For the 1960-2000 period the average absolute change is 3.4 in **NOEDUC_25**, 3.5 in **PRIM_25**, 3.0 in **SEC_25**, and 1.5 in **TERT_25**. Thus, on average, the average changes for the 2000-2005 period are 2-3 times as large as the average five-year changes in these variables for the 1960-2000 era.

Pending a clarification of these unusual increases from UNESCO, we opted to generate 2005 estimates for both the 15+ and 25+ groups using the perpetual inventory method. These models account for approximately 80% of the 2005 observations in the 15+ data and 74% of the 2005 observations in the 25+ data. The countries affected by these estimations are reported in Appendix E-12.

Scale Construction: Creating Preliminary Versions of $\mathbf{E\_ATTAIN_{15+}}$ and $\mathbf{E\_ATTAIN_{25+}}$

To reduce the two sets of educational attainment variables into a composite measure that can be useful as a gauge of human capital we used a modification of Psacharopoulos and Ariagada's (1986) formula for average years of schooling:[17]

$$Average\ Years\ of\ \ Schooling = \sum_{j} \left(YR_j\right)\left(HS_j\right) \qquad (9)$$

where j is the level of schooling (primary, secondary, tertiary), and YR_j is the number of years of schooling represented by level j, and HS_j is the portion of the population that attained j as their highest level of education.

The length of primary and secondary education varies throughout the world and over the course of the post WWII era. But information is available in the UNESCO statistical yearbooks prior to 1998 and on-line at the UNESCO Institute for Statistics from 1999 to present day.[18] The number of years to complete tertiary education is assumed to be 16 across all countries and times. This assumption was based on descriptions of tertiary

[17] The modification was to use levels of educational attainment for primary, secondary, and tertiary. The original forms included categories for incomplete primary and broke secondary attainment into lower and upper categories.

[18] Primary and secondary education data were extracted from the UNESCO statistical yearbooks for a given year with the following exceptions: data for 1975 was extracted from the 1977 yearbook, data for 1965 was extracted from the 1968 yearbook, and data for 1960 was extracted from the 1963 yearbook. Duration data for 2000 and 2005 were extracted from the UIS on-line data base. The schooling duration data for 1955 and 1950 are assumed have the same value as the data for 1960.

education contained in UNESCO's International Standard Classification of Education (1977), dated May 2006.

While there are non-trivial differences, at the margins, in what countries define as a primary or secondary education, there is a surprising level of consistency across countries. For example, while the length of what is considered to be a primary education ranges from three to ten, in 80% of the country-years the range is between five and seven. The data on secondary education are similar. In 85% of the country-years the length of a secondary education is between 11 and 13 years. The duration figures for primary and secondary education change across time in many countries. To address this issue we use the methodology cited in Barro and Lee (2002). Their approach assumes that changes in the length of a primary education affect the 15 and over group in 10 years and the 25 and over group in 20 years. Changes in the length of a secondary education affect the 15 and older population in 5 years, while they affect the 25 and over group in 15 years.

The resulting scale has a theoretical range from 0 to 16. Countries in which large portions of the population are uneducated will score close to zero, while countries where large percentages of the population attained a tertiary education will score closer to 16. Table 3 present summary statistics on the two composite variables.

Table 3
Summary Statistics for the Composite Educational Attainment Variables

	E_ATTAIN_{15+}	E_ATTAIN_{25+}
Mean	6.02	6.00
Variance	3.27	3.23
Median	5.99	6.05
25^{th} percentile	3.30	3.29
75^{th} percentile	8.74	8.64
Minimum	0.02	0.05
Maximum	13.5	13.6

Wave Two Estimations

Cross-composite Estimation

There are a number of countries with complete education attainment data in one population group (the 15+ group, or the 25+ group), but are missing data in the other. Correlations of the education scales for countries with data in both population groups vary year to year between 0.97 and 0.99. As a result of this strong relationship we can use the values of the education scale in one group to predict values in the other and replace any missing data with the predicted values.

One concern with using one scale to predict values for another is the possibility that a country with a low score on one of the scales would receive a predicted value that was less than zero. Negative predicted values were generated in a handful of cases. To address this problem a logarithmic transformation of the two education composites was done prior to estimating the models. The specific steps used to perform these cross-variable estimations are as follows:

- Perform a logarithmic transformation on **E_ATTAIN$_{15+}$** and **E_ATTAIN$_{25+}$**;
- Regress **E_ATTAIN$_{25+}$** on **E_ATTAIN$_{15+}$** at every time point from 1960-2005
    - Generate predicted values for **E_ATTAIN$_{25+}$**
    - Replace missing values in **E_ATTAIN$_{25+}$** with the predicted values
- Repeat the above steps to replace the missing values for **E_ATTAIN$_{15+}$**
- Transform the augmented **E_ATTAIN$_{15+}$** and **E_ATTAIN$_{25+}$** measures data back to their original scales, which has a theoretical range of 0-16.

The cross-composite estimation technique provided the data accounts for 4.3% of the 15+ data educational attainment and 14.1% of the 25+ data.

Country-year Expansion

Barro and Lee's data, as well as the augmentations described above, are in five-year increments. To expand these data into a country-year format we filled in the gaps largely by using a linear interpolation (e.g., if a country had a data point in 1960 and a data point in 1965, the gap in the data between 1960 and 1965 was filled by linear interpolation). In a special subset of cases, however, a linear extrapolation had to be used: when the year of a country's independence fell in between the five year data points. To illustrate, if a country attained independence in 1993, the values for 1993 and 1994 were filled by extrapolation of the 1995 data. In three cases this generated a negative value on the education scale (People's Republic of Yemen in 1967, 1968 and Guinea-Bissau in 1974) in both the 15+ and 25+ data. In all three cases the negative values were replaced with the last positive value.

Wave Three: Estimating Pre- 1960 Educational Attainment

ARIMA Backcasts

The SID project is concerned with societal development in the post WWII era and the time-series for many of its data collections extend until at least 1950. Thus, it was important to investigate the possibility of extending the educational attainment data to 1950 for countries that were independent prior to 1960. After considering and testing a number of alternative approaches, we decided to employ dynamic univariate backcasting models to generate estimates for the period between 1950 and 1960.[19] Our efforts here

[19] Backcasting was used between 1950-1970 for some of the Lutz et. al. (2010) education attainment data. In particular, Cambodia, Laos, and Saudi Arabia used a backcast for this period. Backcasts were also done for Maldives between the years 1965 and 1969.

were aided by the fact that the Barro and Lee 2000 dataset includes educational attainment data for the 25+ population group in 1955 for a limited number of countries (24) – though none for the 15+ population. The available data were used to construct values for **E_ATTAIN$_{25+}$** for 1955 for these 24 countries.[20] An exploratory effort was then undertaken to determine the utility of making backcasts. The point of this effort was to compare the actual values of **E_ATTAIN$_{25+}$** for these countries with the values produced by our estimation procedure (i.e., dynamic univariate backcasting).

The results were encouraging as the correspondence between the backcasts and the actual values were quite good for all but a few years in a handful of countries, most of which had very low scores on **E_ATTAIN$_{25+}$**. The residual between the forecast and actual values of **E_ATTAIN$_{25+}$**, expressed as a proportion of **E_ATTAIN$_{25+}$**, has a mean of .15 and a median of .04. To illustrate, consider a nation with a value of 6 on **E_ATTAIN$_{25+}$**. Using the median, most of the forecasts would be between 5.76 and 6.24; using the mean, most of the forecasts would be between 5.1 and 6.9. The reason for the discrepancy was the existence of a handful of outliers, most of which had low scores on **E_ATTAIN$_{25+}$**. Thus, while 30 forecasts had a value of 0 or .01 on this residual measure, 24 had a value greater than .25. Twelve of these 24 had values on **E_ATTAIN$_{25+}$** that were less than 1. Thus, a very small residual will produce a very high ratio. Graph 2 presents three overlays of the backcasts and **E_ATTAIN$_{25+}$** for a country with a very good backcast (U.S.) an average fit (Romania), and a relatively poor fit (Japan).

Because of the encouraging results of the exploratory analysis, we generated backcasts for the countries that were independent prior to 1960 but were missing educational data. To produce the backcasts, the Box-Jenkins (1976) Autoregressive Integrated Moving Average (ARIMA) procedure was used to select an appropriate model for each country. The underlying notion behind this approach is parsimony. From a Box-Jenkins perspective, parsimonious models produce better models than over-parameterized models; a parsimonious model will be more efficient because it will fit the data well without adding any additional and unnecessary coefficients.

Before estimating the ARIMA models, a visual inspection of country-specific graphs for **E_ATTAIN$_{25+}$** revealed two data issues that needed to be addressed before generating the backcasts. The first is rooted in the fact that the country-year data used in the ARIMA procedure are based on educational attainment data generated every five years. The interpolations that filled in the gaps between these five-year intervals produced distributions that often had peaks corresponding to these five-year intervals. These peaks reflect the manner in which the data were collected rather than the underlying pattern of educational attainment in a nation. Moreover, they led the ARIMA procedure to identify an artificial seasonal component in the time-series. Thus, we decided to smooth the peaks. We experimented with a number of moving average procedures to do this and settled on a seven-year moving average. It generated a reasonably smooth distribution of **E_ATTAIN$_{25+}$** for most countries for most timeframes.

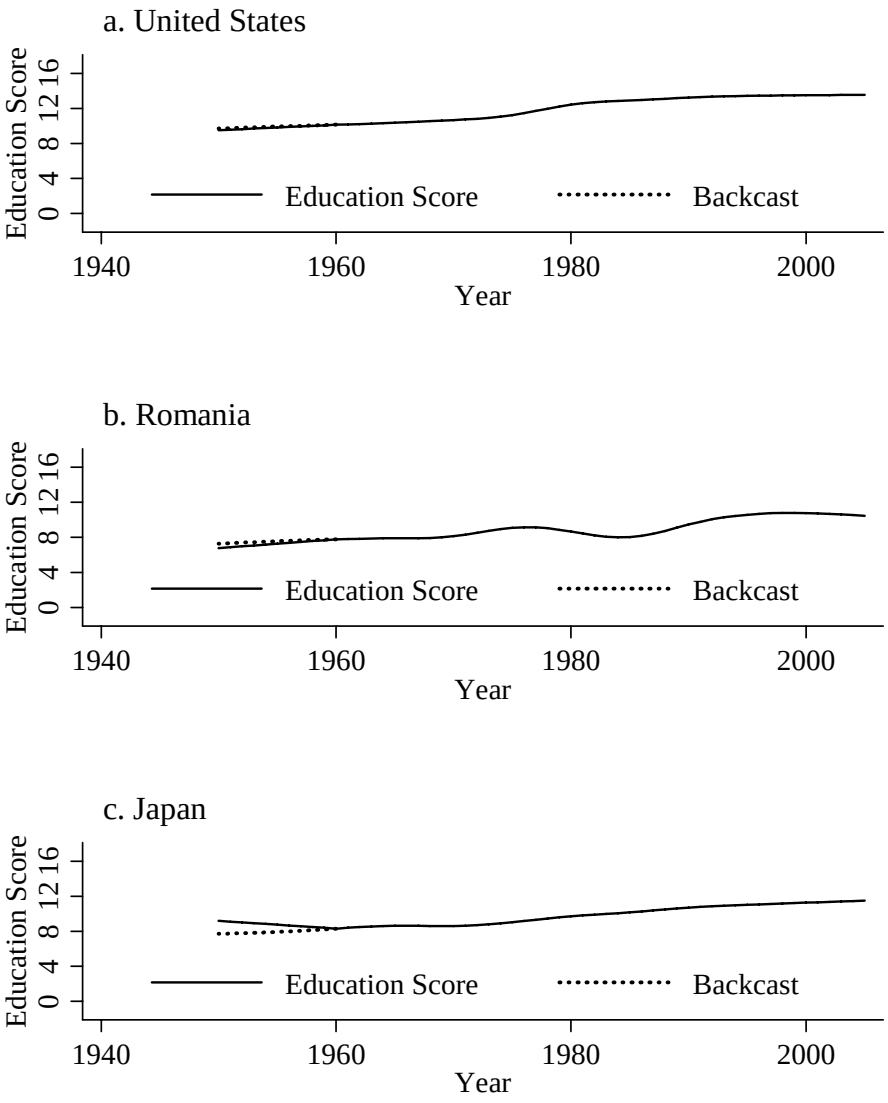
[20] To compute the education scale for pre-1960 data, we assume the education duration figures for 1960 apply to 1955 and 1950.

The second issue revealed by the visual examination of the country-specific graphs for **E_ATTAIN$_{25+}$** is that the trend line for most countries is increasing over most of the post-1960 era, as would be expected. However, these trends cause two problems for the backcasting procedure. First, for countries with a low score and an increasing trend, ARIMA backcasting procedure can produce negative values in early years. Thus, we performed a logarithmic transformation on **E_ATTAIN$_{25+}$** prior to estimating the ARIMA models. This addressed the problem because re-transforming the data with its anti-log generates only positive numbers. The second problem is that the ARIMA models require the data to be "stationary." That is, the data must have a constant mean, variance and autocorrelation structure over time. If there is a trend in the data used to generate the ARIMA model, the data must be transformed to make it stationary. Thus, the education data were differenced until they became stationary.

After these data transformations were completed the next step was to identify the appropriate ARIMA model to use in generating the backcasts, which had to be done on a country-by-country basis. We routinely examined three to four potential models based on the ACF and PACF results. While the model with the lowest Akaike Information Criterion score was normally chosen to model the backcast, several other pieces of information were also used in the model selection process. First, a test for the existence of residual correlation was done using a Ljung-Box test. Second, the normality of the model was assessed by using a quantile-quantile (Q-Q) plot. Finally, the residuals of the models were plotted using ACF and the PACF to determine if the model residuals were white noise. If the model failed any of these three tests, a model with the next lowest AIC was examined and evaluated until a satisfactory model was identified.

Once an ARIMA model was specified for a country, it was used to generate backcasts. To generate these backcasts, dynamic methods were used; that is, the backcasts were generated one year at a time. Thus, we began with a one-step backcast for 1959. Using the 1959 backcast, a backcast was produced for 1958; using the 1958 backcast, a backcast was produced for 1957; and so forth. This method was used to estimate **E_ATTAIN$_{25+}$** for the years between 1950 through 1959. If a country became independent after 1950, the backcast would stop on the year the country became independent. Altogether, 529 additional education scores for 57 countries were obtained using this method; these are reported in Appendix E-13.

Graph 2
Education Scores and Time Series Backcasts for Select Countries



Final Cross-composite Estimation

The last step in the estimation process was to estimate **E_ATTAIN$_{15+}$** for the pre-1960 period using the backcasted values of **E_ATTAIN$_{25+}$**. Thus, the final SID educational attainment data set contains the four education attainment variables for both age groups for the period from 1960 to 2005 (**NOEDUC_15**, **PRIM_15**, **SEC_15**, **TERT_15** and **NOEDUC_25**, **PRIM_25**, **SEC_25**, **TERT_25**) and **E_ATTAIN$_{15+}$** and **E_ATTAIN$_{25+}$** for the period between 1950 (or year of independence) and 2005. **E_ATTAIN$_{15+}$** and **E_ATTAIN$_{25+}$** are smoothed versions of the raw scores that were generated by a seven-year moving average procedure; the unsmoothed data for **E_ATTAIN$_{15+}$** and **E_ATTAIN$_{25+}$** are also available for the 1960-2005 period.

These variables are available for 162 countries. Table 4 lists the countries and the years for which we are lacking data.

Table 4
SID Countries without Educational Attainment Data

Country	Years Missing	Country	Years Missing
Albania	1950-1959, 1961-2005	North Korea	1950-2005
Bhutan	1950-2005	Oman	1950-2005
Bosnia-Herzegovina	1992-2005	Solomon Islands	1978-2005
East Timor	2002-2005	Suriname	1975-2005

The SID Educational Attainment Archive: An Overview and Assessment

This section summarizes the techniques used to augment the Barro and Lee data, examines the contours of the data across time and space, and assesses the credibility of the estimated data.

Summary of Estimation Procedures and Coverage Comparisons

Chart 1 and 2 summarize the contributions of the various data sources to the SID educational attainment archive for the 15+ and 25+ age groups, 1950-2005. The original Barro and Lee data provided 64.6% of the data for the 15+ group (see Chart 1) while the African estimates and the extrapolations to 2005 contributed 16.9%. The next most important contributor was the cross-composite estimates (4.4%) followed by the UNESCO data and perpetual inventory estimates (2.9%) and the procedures used to address the missing data for the post-Soviet states (2.8%). The use of literacy rates, interpolation and extrapolation account for less than 2% each. Just 6.3% of the data are missing for E_ATTAIN15+.

For the 25+ group (Chart 2) the Barro and Lee data provide 64.1 % of the total, while the cross composite estimations provide 14.1%. The next largest contributor is the extrapolations to 2005 (8.5%), followed by the UNESCO data and perpetual inventory

estimates (4.2%).The use of literacy rates, interpolation and extrapolation account for less than 2% each. Just 6.2% of the data are missing for E_ATTAIN25+.

Chart 1
Source of Education Data – 15 and Older

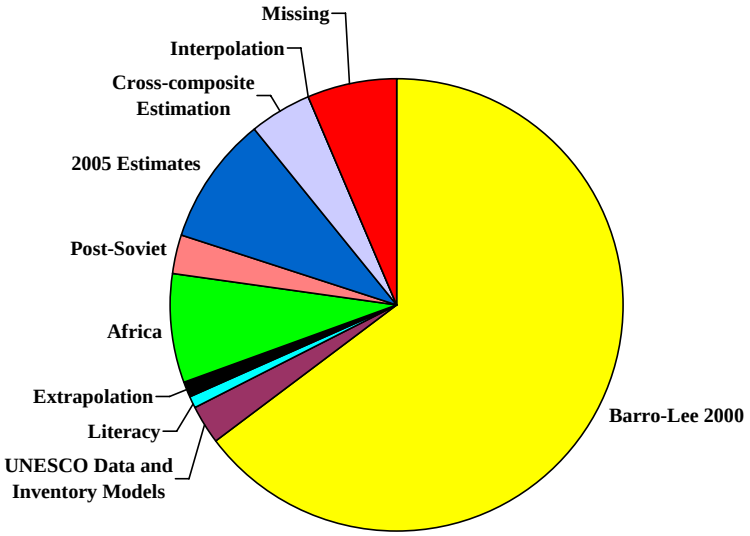
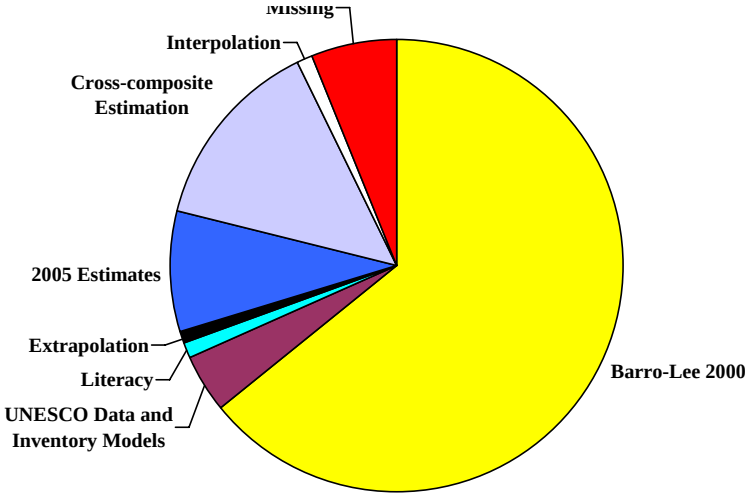


Chart 2
Source of Education Data – 25 and Older



Appendix E-14 provides a year by year tally of the contributions of each of these sources to the total for each series.

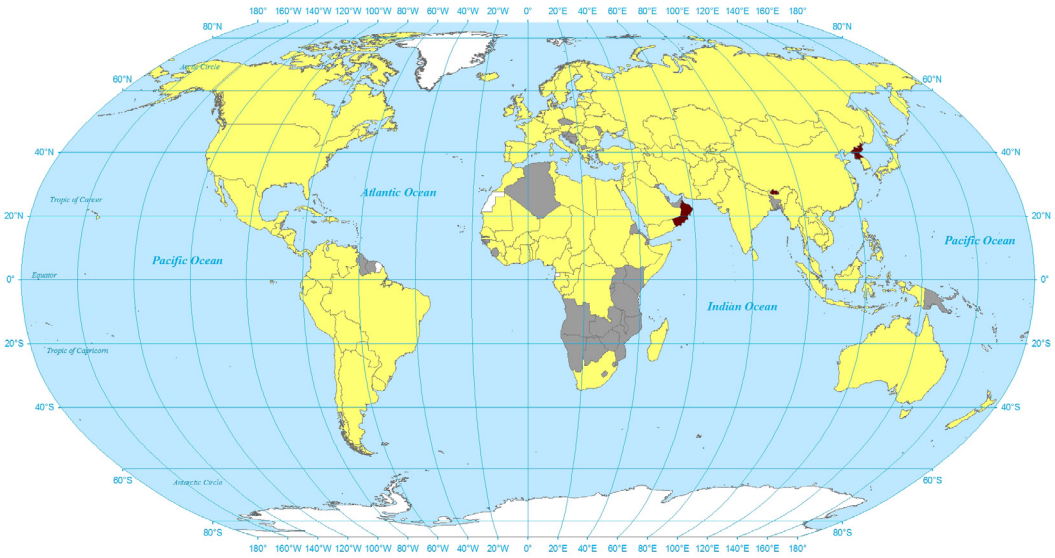
Maps 1-3 provide a succinct, geo-political perspective on the augmented educational data by comparing the coverage of the SID data with the Barro and Lee data for 1960, 1980 and 2000. In 1960 the SID archive provides more coverage for Africa, Southeast Asia and Mongolia; a similar picture exists in 1980 (Map 2), with the exception that there are more independent African nations for which SID data exist. Map 3 underscores the contributions of SID efforts to complete the educational data for the post-Soviet states.

Illustrating the Distribution of the SID Educational Attainment Data

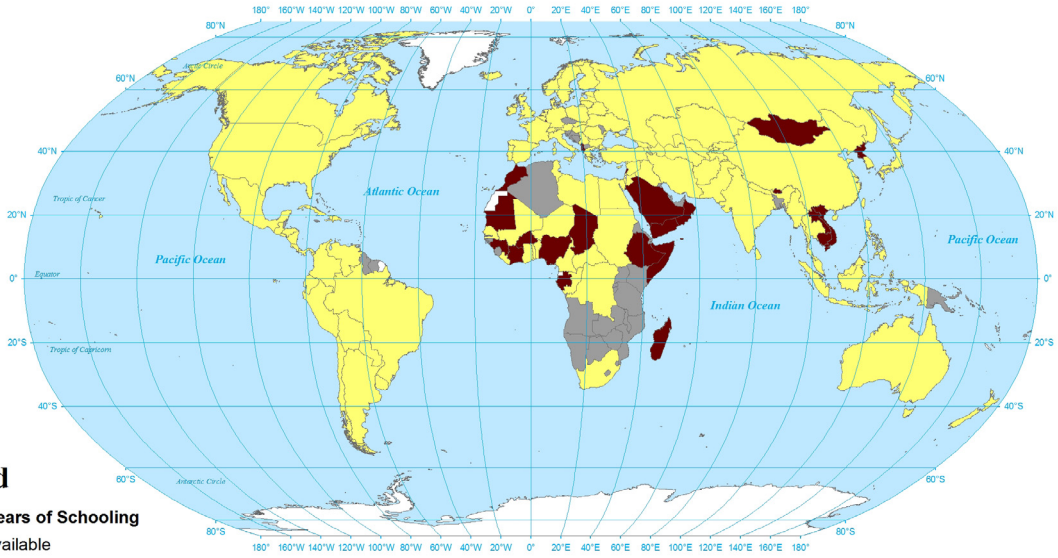
Graph 3 displays an overlay of $\mathbf{E\_ATTAIN_{25+}}$ for six countries selected for the diversity of their patterns in educational attainment. Unsurprisingly, the highest levels of attainment are for two Northern countries, Canada and Norway. Canada begins the series with an average attainment rate of 10 years and it took more than two decades (1972) to reach 11 years. By 1980, however, it reached 12 years and it surpassed 13 years just before the end of the century (1998). Norway begins with a lower rate than Canada (7.2) and it did not move to 9 years until almost 1975. But Norway experienced a sustained increase during the 1980s and by 2005 its average educational attainment was nearly as high as Canada's (12.5 years). Bolivia begins the series with a modest educational attainment rate of 6.4

Map 1
Comparison of Cline Center and Barro-Lee
Available Average Years of Schooling Data, 1960

a. Cline Center Countries



b. Barro-Lee Countries



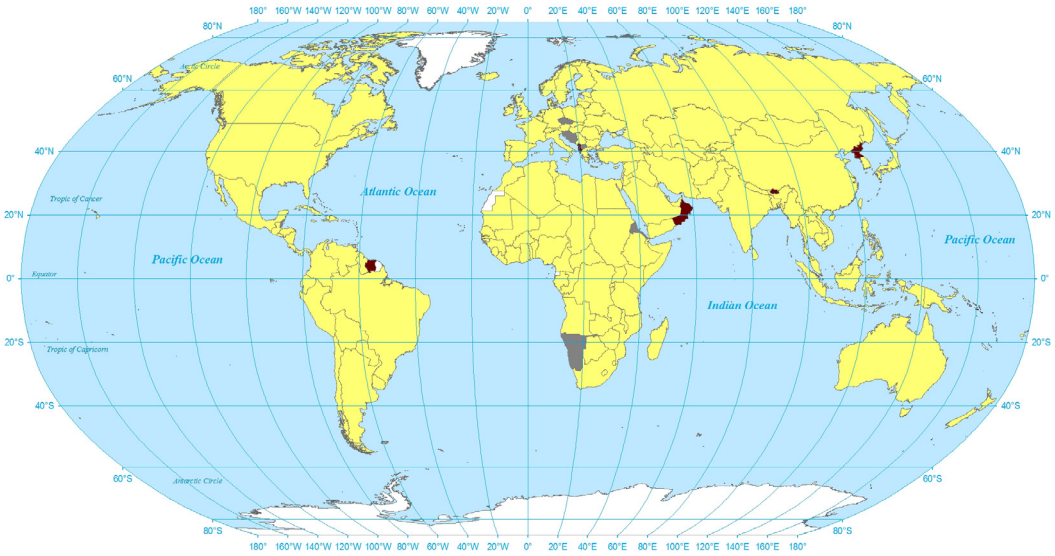
Legend

Average Years of Schooling

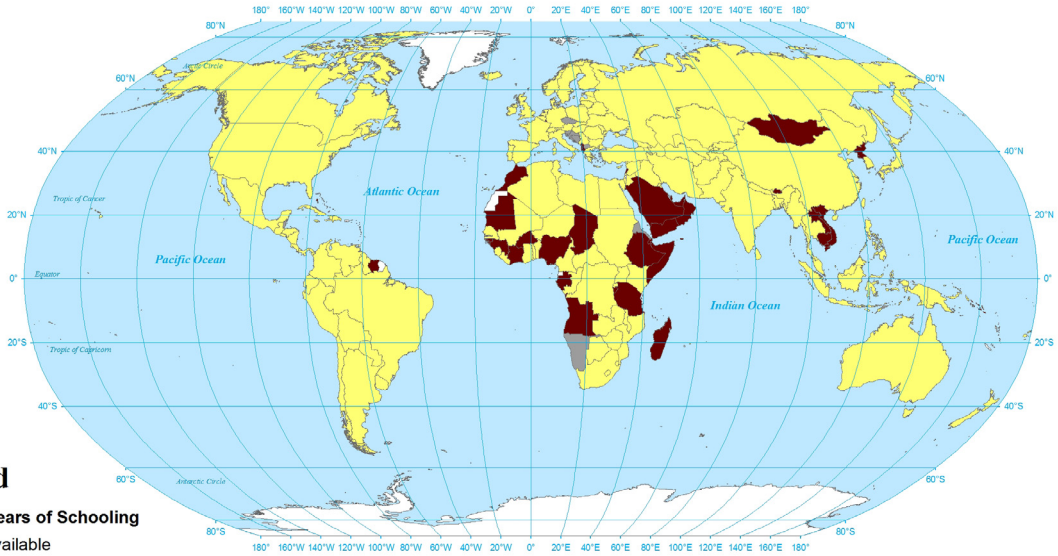
- Unavailable
- Available
- Not Independent

Map 2
Comparison of Cline Center and Barro-Lee
Available Average Years of Schooling Data, 1980

a. Cline Center Countries



b. Barro-Lee Countries



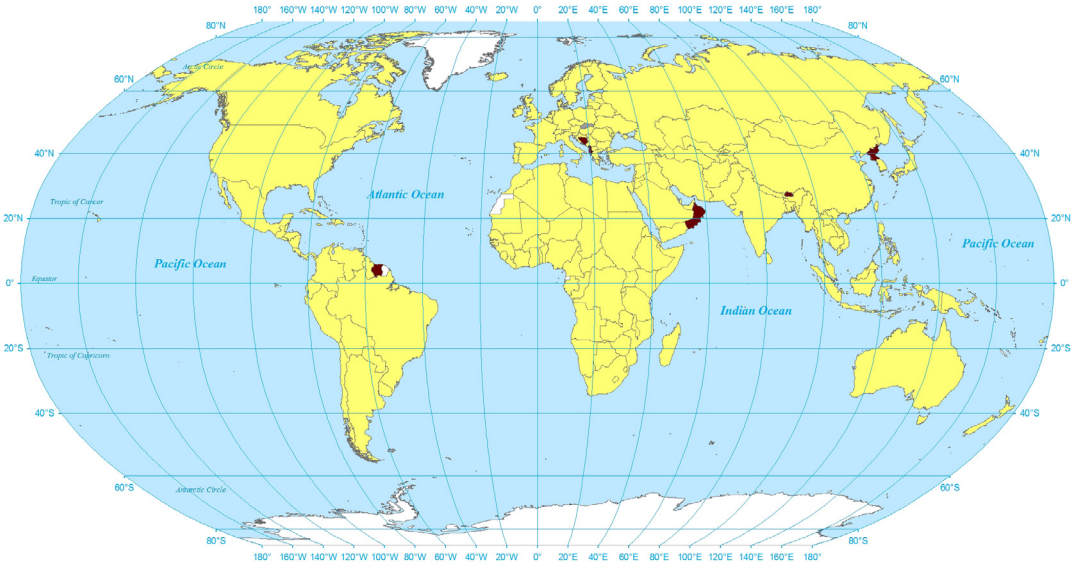
Legend

Average Years of Schooling

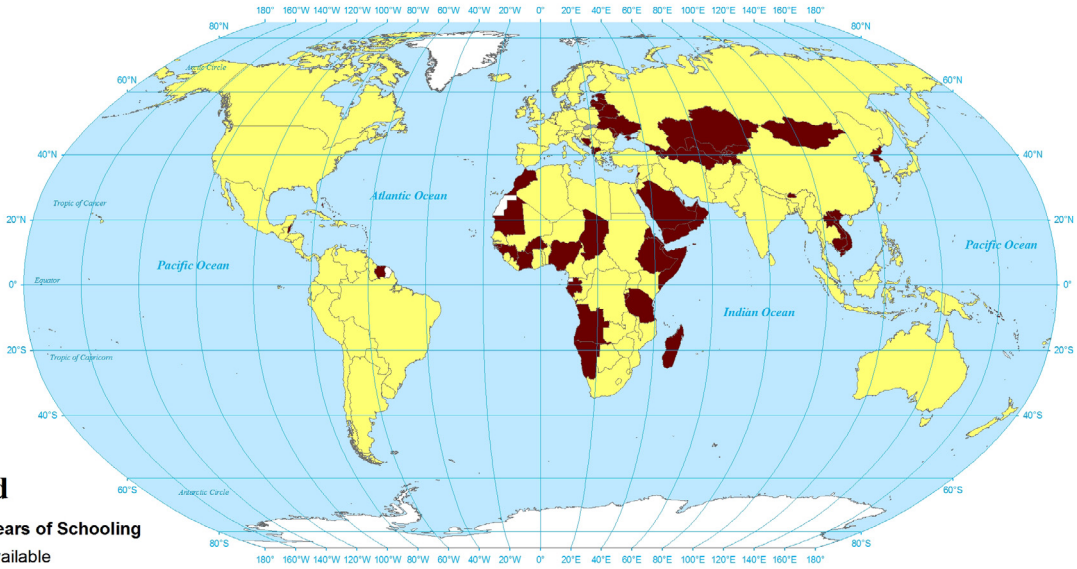
- Unavailable
- Available
- Not Independent

Map 3
Comparison of Cline Center and Barro-Lee
Available Average Years of Schooling Data, 2000

a. Cline Center Countries



b. Barro-Lee Countries



Legend

Average Years of Schooling

- Unavailable
- Available
- Not Independent

years and it vacillates between 6.4 and 5.8 years until 1980. It experiences some sustained growth after 1980 but it stabilizes at about 7.4 year toward the end of the series. The Congo begins its series in 1960 with a relatively low attainment rate of 2.2 years. While it experiences steady increases it never exceeds 6.5 years. Indonesia has an average educational attainment rate in 1950 of just 1.8 years but by 1960 it is very close to the Congo. Indonesia grows more rapidly than the Congo from the late 1960s through the early 1980s and achieves rates somewhat higher than the Congo until about 1988. At that point it falls behind the Congo though it finishes the series at near parity. In contrast, Afghanistan, which begins the 1950s with a value comparable to Indonesia's (2.2 years), evidences a steadily decline for more than twenty years. By 1970 its educational attainment rate is under 1.4 years. It begins a slight incline around 1975 but it ends the series with an educational attainment rate that is less than its value in 1950, 2.1 years.

Graph 4 illustrates the relationship between the two educational attainment variables in the SID archive, **E_ATTAIN$_{15+}$** and **E_ATTAIN$_{25+}$**. As the data for the four countries depicted in Graph 4 (Algeria, China, Mexico and South Korea) illustrate, there is a close correspondence between the two. This demonstrates why it was possible to use the cross-composite estimations to fill in missing gaps in one or the other. As would be expected, the values for **E_ATTAIN$_{25+}$** are always somewhat higher than for **E_ATTAIN$_{25+}$**, as educational attainment increases somewhat in the decade between 15 and 25 years of age.

The last set of illustrative materials is presented in Table 5, Table 6, and Maps 4-6. Table 5 lists the average educational attainment (**E_ATTAIN$_{25+}$**), weighted by population, for each region of the world and the world as a whole. These materials illustrate the steady increases in educational attainment over the post-WWII era – as well as marked

Table 5
Educational Attainment (E_ATTAIN$_{25+}$), by Year and Region

Year	North America	South America	Europe	Africa	Asia	Middle East	Oceania	World
1950	7.6	4.2	7.4	0.3	2.3	0.9	8.0	3.8
1955	7.8	4.2	7.6	0.3	2.4	1.3	8.3	3.8
1960	8.5	4.3	7.7	0.8	2.5	1.6	8.4	4.0
1965	8.9	4.2	7.5	1.3	2.8	1.8	8.4	4.1
1970	8.9	4.4	8.3	1.5	3.2	2.1	8.9	4.5
1975	9.3	4.8	8.8	1.9	3.5	2.4	9.2	4.7
1980	10.1	5.1	8.8	2.3	3.9	2.9	9.5	5.1
1985	10.5	5.6	8.8	2.8	4.3	3.4	9.3	5.4
1990	11.0	6.6	9.6	3.3	4.7	4.0	9.3	5.8
1995	11.2	7.4	9.5	3.9	5.2	4.6	9.3	6.2
2000	11.4	7.7	9.7	4.3	5.6	5.1	9.5	6.5
2005	11.5	8.0	10.0	4.6	6.0	5.7	9.7	6.8

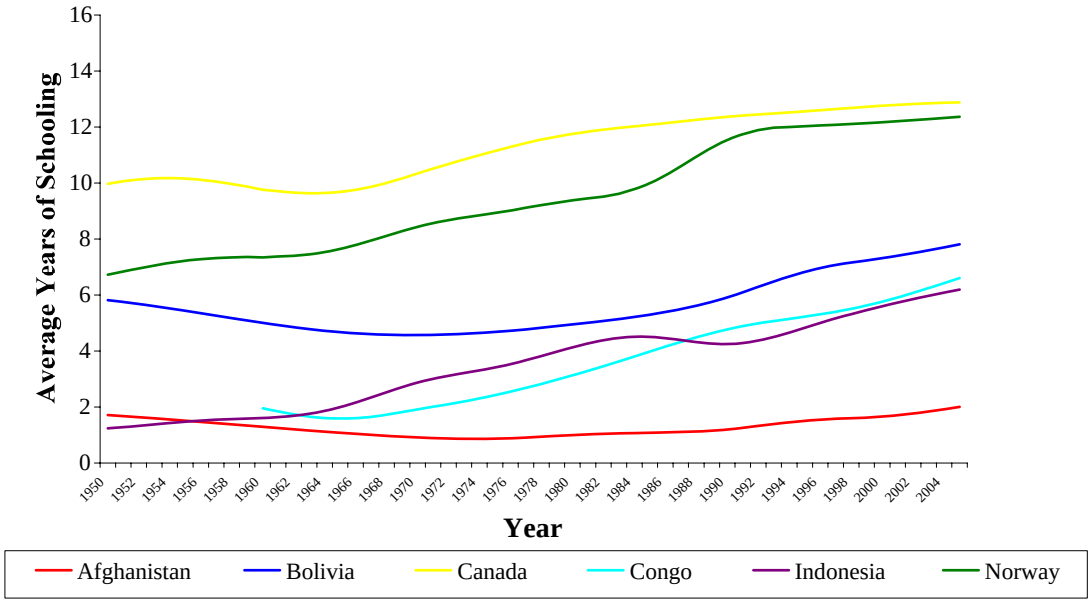
disparities and deficits across countries. The weighted global average for **E_ATTAIN$_{25+}$** increases from 3.8 years in 1950 to 4.7 years in 1975 to 6.5 years in 2000; the weighted average is 6.8 years in 2005. At the same time, there are markedly different patterns across both continents and countries. As reported in Table 5, Oceania begins the Post WWII era with the highest level of educational attainment, but it grows relatively slowly. By the end of the period (2005) it ranks third, behind North America and Europe. At the lower end of the continuum are Africa and the Middle East, which begin at .3 and .9. Despite steady progress during the latter part of this period, both maintain their relative ranking in 2005. In the middle of this grouping in 1950 is South America and Asia, with weighted averages of 4.2 and 2.3 years, respectively. While Asia grows somewhat more rapidly than South America, they end the period in roughly comparable relative positions.

Maps 4-6 display the distribution of educational attainment rates by country for 1960, 1980 and 2000 and they illustrate the within-region differences in educational attainment. Table 6 summarizes some of these differences; the country-specific data are reported in Appendix E-15. As expected, the greatest changes in educational attainment between 1950 and 2005 came in developing regions while areas with more established education systems, North America, South America, and Europe, have lower rates of change in educational attainment.

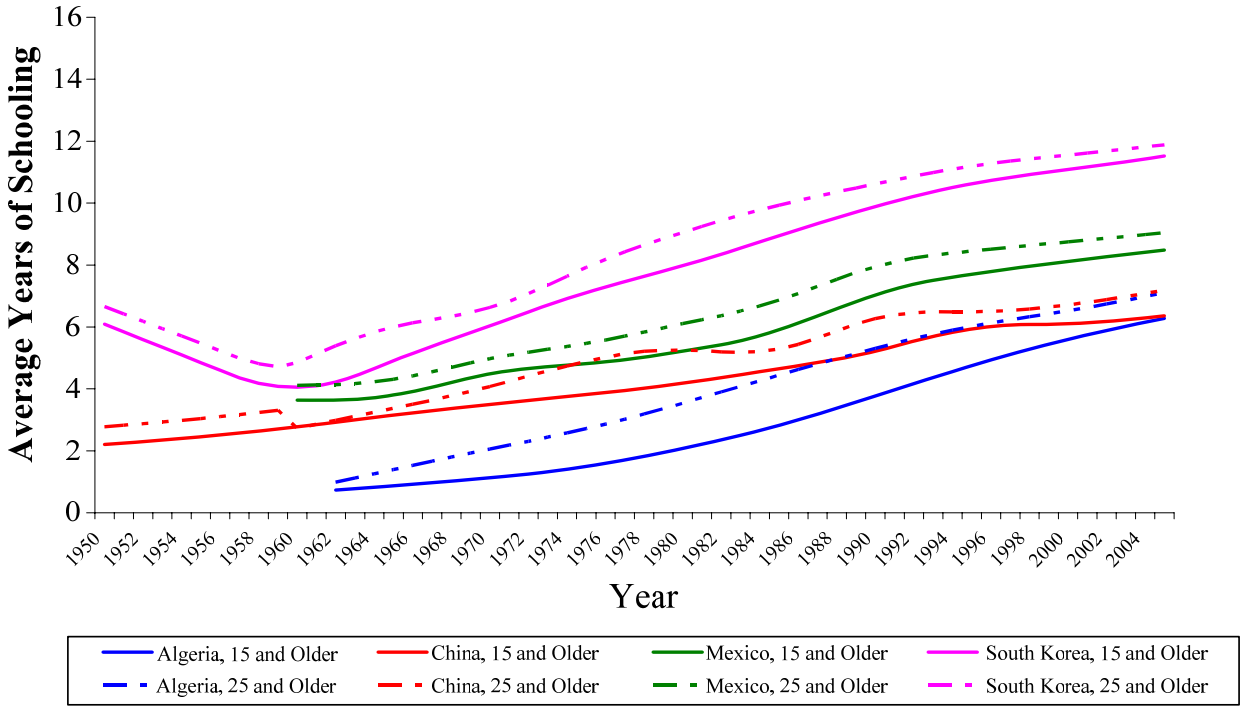
Table 6
Illustrative Changes by Continent over Time

Region	1950		1980		2000		2005		Average Change	Rank
	E_ATTAIN$_{25+}$	Rank	E_ATTAIN$_{25+}$	Rank	E_ATTAIN$_{25+}$	Rank	E_ATTAIN$_{25+}$	Rank		
Africa	0.3	7	2.3	7	4.3	7	4.6	7	0.35	3
Asia	2.3	5	3.9	5	5.6	5	6	5	0.35	4
Europe	7.4	3	8.8	3	9.7	2	10	2	0.24	6
Middle East	0.9	6	2.9	6	5.1	6	5.7	6	0.39	2
North America	7.6	2	10.1	1	11.4	1	11.5	1	0.34	5
Oceania	8	1	9.5	2	9.5	3	9.7	3	0.44	1
South America	4.2	4	5.1	4	7.7	4	8	4	0.15	7
World	3.8	.	5.1	.	6.5	.	6.8	.	0.27	.

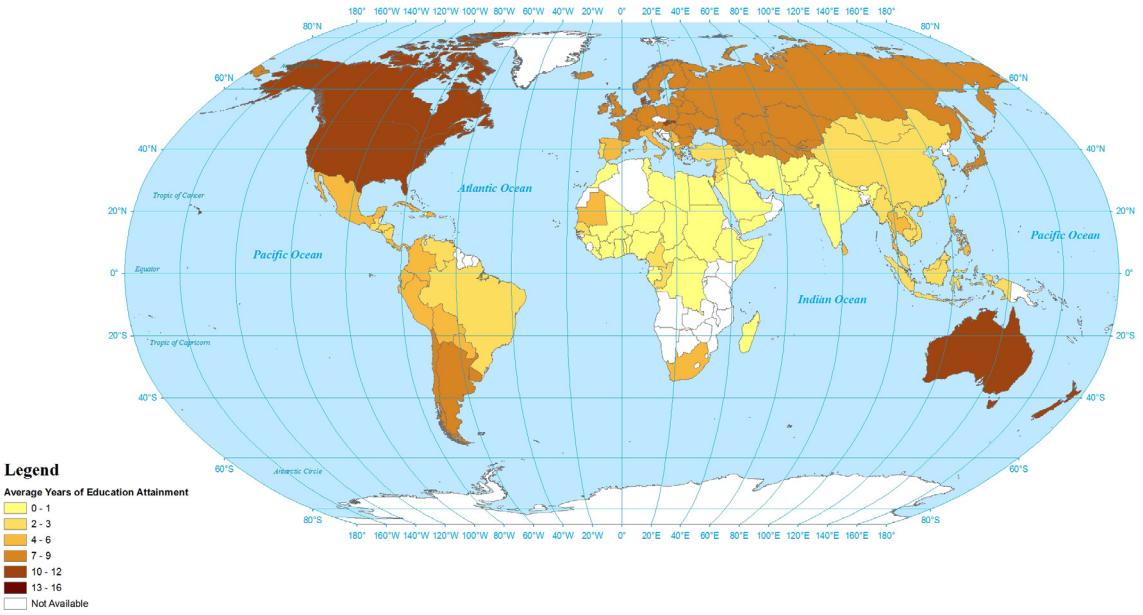
Graph 3
E_Attain25+ for Select Countries, 1950-2005



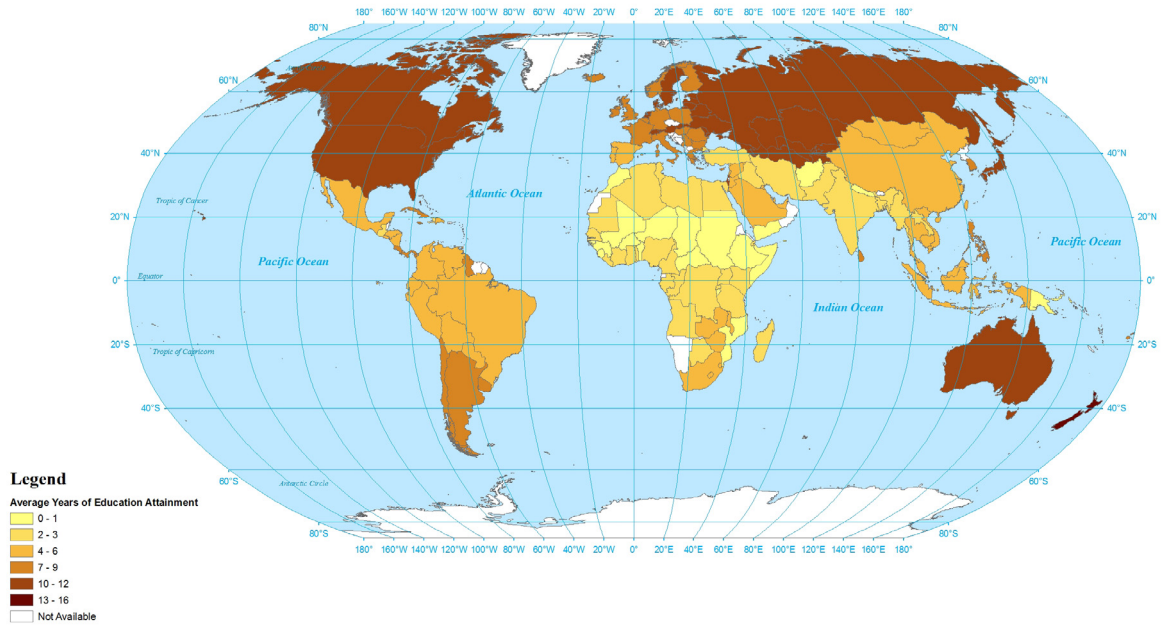
Graph 4
Overlay of E_Attain15+ and E_ATTAIN25+, Select Countries



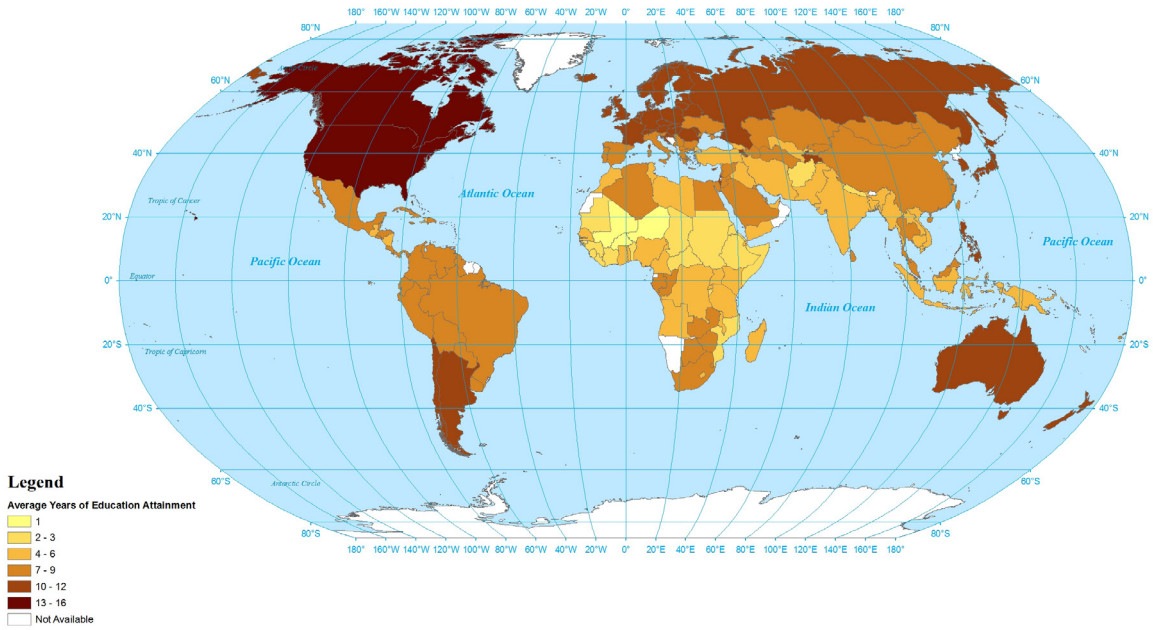
Map 4
Average Years of Schooling for 25 and Older, 1960



Map 5
Average Years of Schooling for 25 and Older, 1980



Map 6
Average Years of Schooling for 25 and Older, 2000



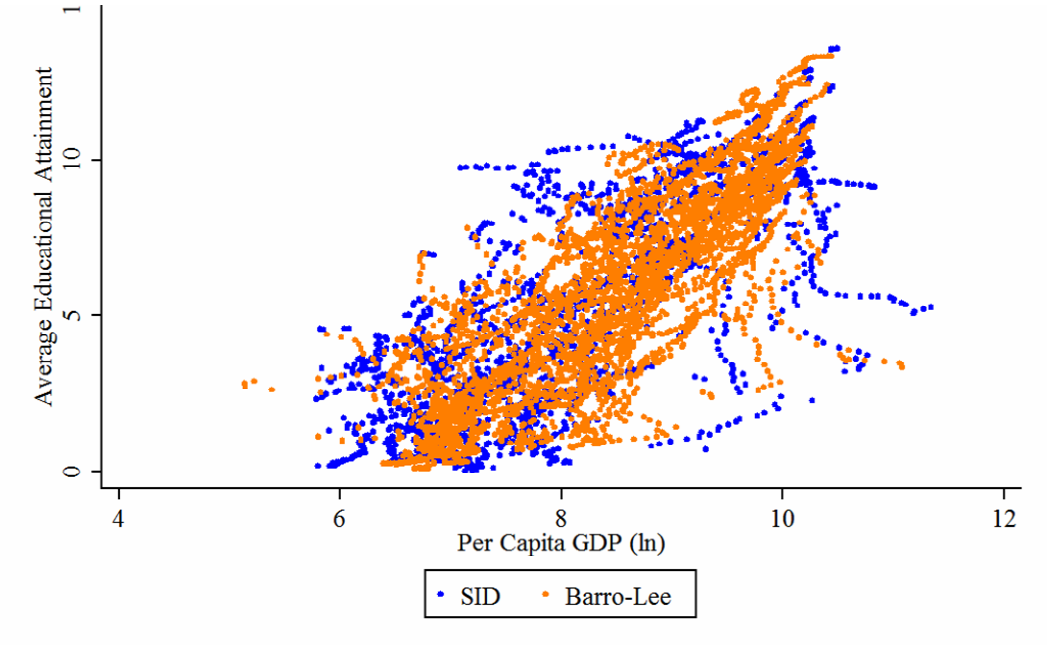
Assessing the Credibility of the Augmented Educational Attainment Data: A Comparative Examination of Education and Economic Growth

To assess the credibility of the augmented SID educational attainment data we conducted a comparative analysis of the relationship between education and economic growth. The analysis focused on the form of the relationship and compared the Barro and Lee data with the augmented data described above. To facilitate this analysis the different sources of data are color-coded: the Barro and Lee data are depicted in orange while the SID estimates are depicted in blue.[21] As noted earlier, the major sources of the augmented educational attainment data come from three sources: the African estimates, the backcast estimates for the 1950s, and the inventory model estimates for the post-2000 cases.

The first step in this analysis was to produce a scatter plot of per capita GDP and **E_ATTAIN$_{25+}$**. This scatterplot is depicted in Graph 5 and it clearly indicates a strong positive relationship between a country's human capital stock and its economic well-being. Indeed, the correlation between educational attainment and GDP is .81 for the Barro-Lee data and .75 for the SID data. However, there are also a number of "tentacles" stretching out from the main cloud in the scatterplot. These tentacles are of some concern because if they are the result of the estimation procedures employed here they would call into question the utility of the augmented data. However, closer scrutiny reveals that

[21] Because the original Barro and Lee data were in five year intervals from 1960 through 2000 we expanded it to a country-year format using linear interpolation. Once reformatted, the Barro and Lee data contain a little over 4200 cases. The SID data contains a little over 3000 cases.

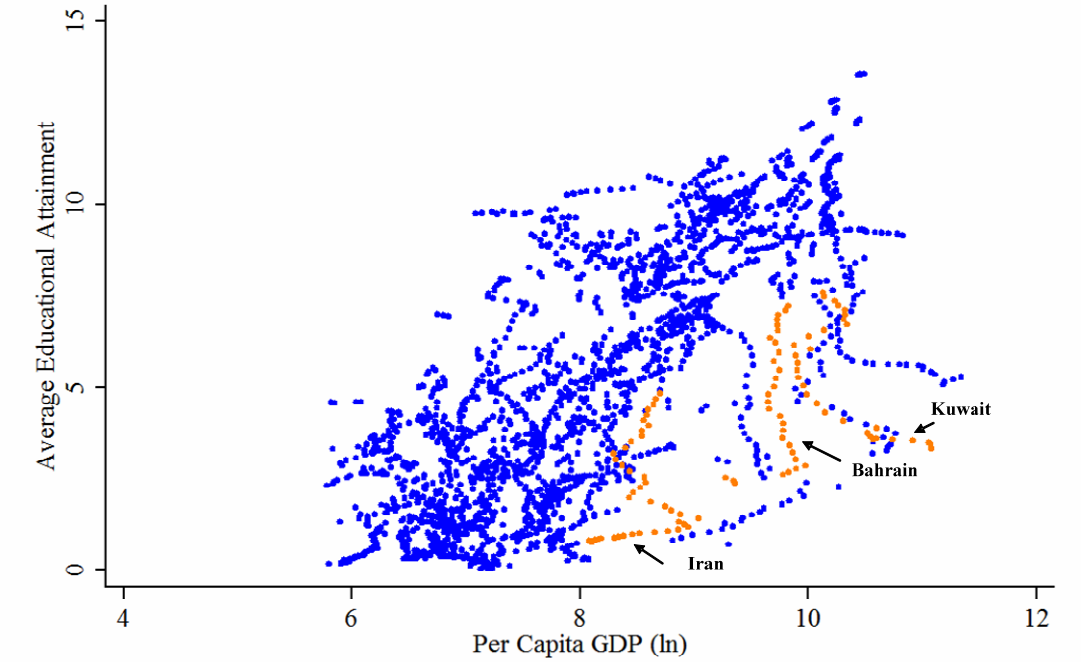
Graph 5
Barro-Lee and SID Educational Attainment



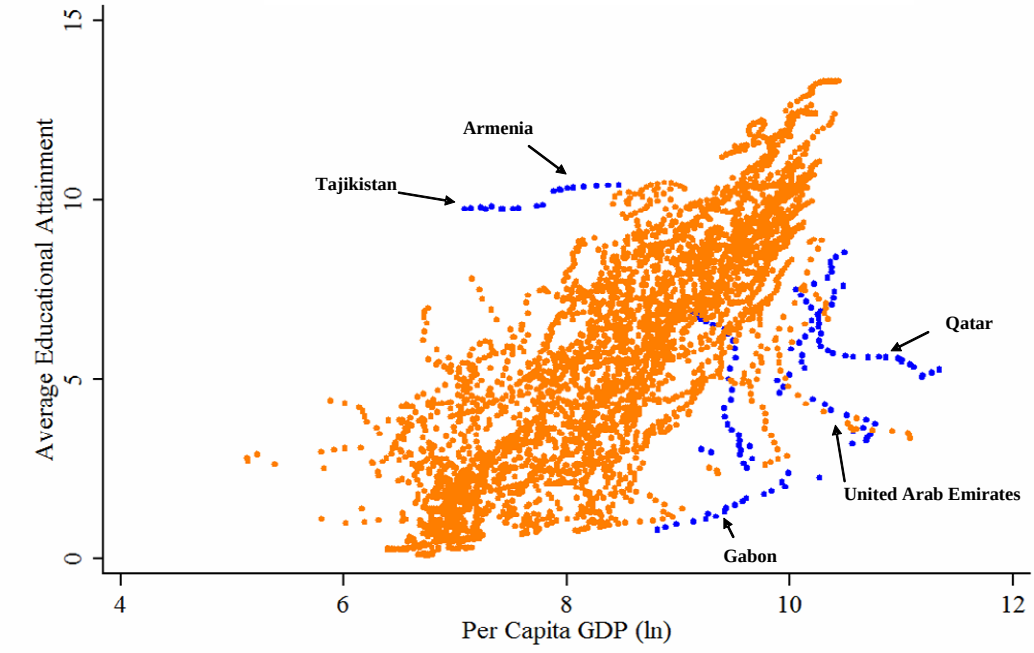
there are tentacles in both the Barro and Lee data and the SID data. Graphs 6 and 7 contain the "tentacles" attributable to the Barro-Lee data and the SID data respectively. The Barro and Lee tentacles belong to three Middle Eastern countries: Bahrain, Iran and Kuwait; the SID tentacles belong to two Middle Eastern countries (Qatar and United Arab Emirates), two central Asian countries (Armenia and Tajikistan), and one African country (Gabon). It should be noted that the source of the SID data for four of the five countries come from archival sources or perpetual inventory model estimation techniques. For example, the "hard" data points for Armenia and Tajikistan were obtained from the Commonwealth of Independent States Statistical Committee while "hard" data points for Qatar and the United Arab Emirates were obtained from UNESCO.

Thus, based on the review of the SID tentacles it is clear that, rather than producing outliers, SID efforts at data augmentation incorporated a number of countries with distinctive relationships between education and economic growth. Thus, including them in cross-national analyses will provide more comprehensive and meaningful results. Indeed, the inclusion of an additional five countries with unconventional relationships between education and economic growth are, in part, responsible for the lower correlation using the SID data. But they do not account for it entirely and it is important to examine the impact of the main estimation techniques on the observed relationship. As mentioned

Graph 6
Barro-Lee Tentacles

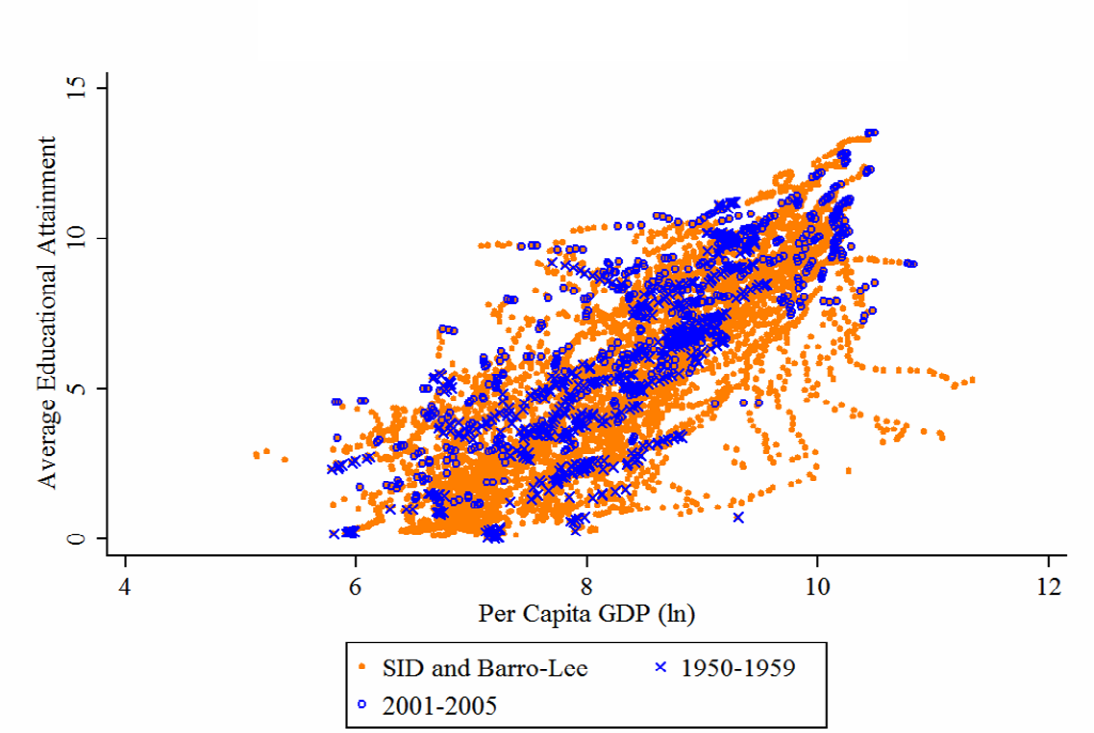


Graph 7
SID Tentacles



above, the largest sources of estimated data are the pre-1960 estimates, the post 2000 estimates, and the African estimates. Graph 8 contains a scatterplot highlighting the pre-1960 and post-2000 cases. Neither group of estimations can be fairly depicted as outliers as the preponderance of the data points lie within the main cloud. Indeed, the correlation between the pre-1960 educational attainment and per capita GDP is .78 while the correlation between the post-2000 educational attainment and GDP is .83. These correlations are quite similar to that using the original Barro and Lee data (.81). Thus, we turn to an examination of the African estimates.

Graph 8
SID Data 1950-1959 and 2001-2005



In contrast to non-African countries, African countries have a weaker relationship between educational attainment and per capita GDP in both the Barro-Lee and SID data. The correlation between per capita GDP and Barro-Lee's educational attainment data for Africa (0.54) is slightly higher than that between per capita GDP and the SID educational attainment estimates for Africa (0.47); for non-African countries the correlation is 0.77 using the reformatted Barro and Lee data and 0.65 using the SID data. Graph 9 contains a scatter plot of GDP and joins the Barro and Lee African data with the SID African data. A closer inspection of these data reveals that there are a number of countries where there is virtually no correlation between per capita GDP and educational attainment in both data sets (see Graph 10). Table 7 reports the countries with minimal correlations between educational attainment and growth. The SID data contains more countries where the relationship between educational attainment and GDP is exceptionally weak than are found in the Barro-Lee data, which accounts for the slightly lower correlation between

human capital stock and economic well-being in the SID data for Africa. If these countries are removed, the correlation is 0.51 using the Barro and Lee data and 0.60 using the SID data.

Graph 9
Barro-Lee and SID African Data

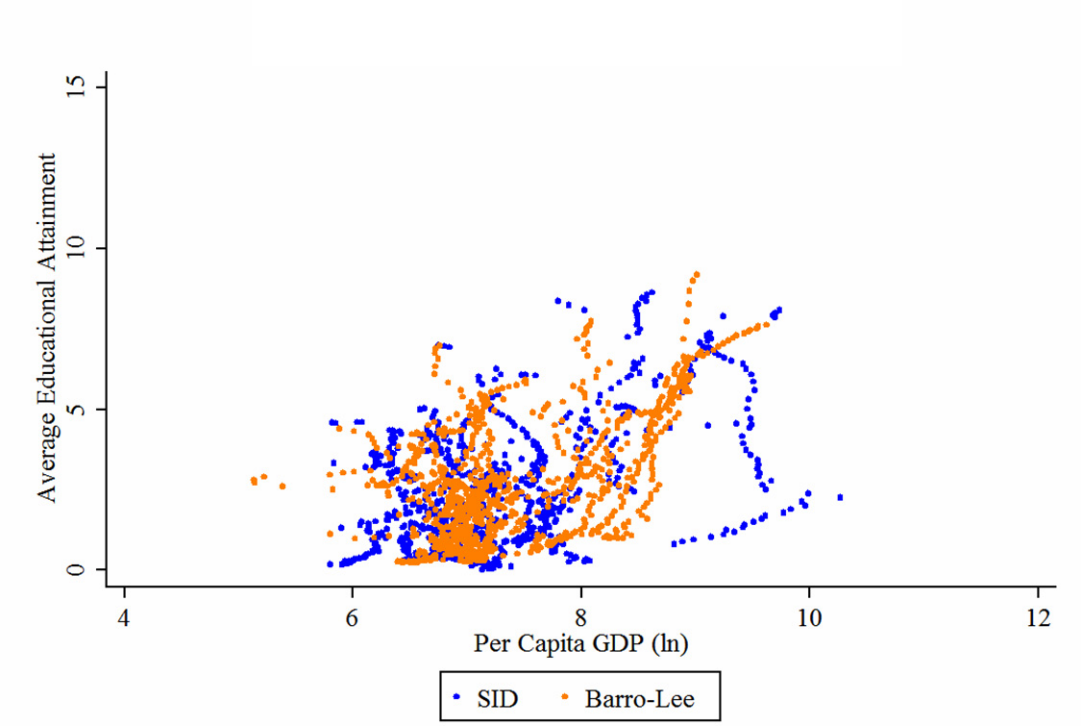
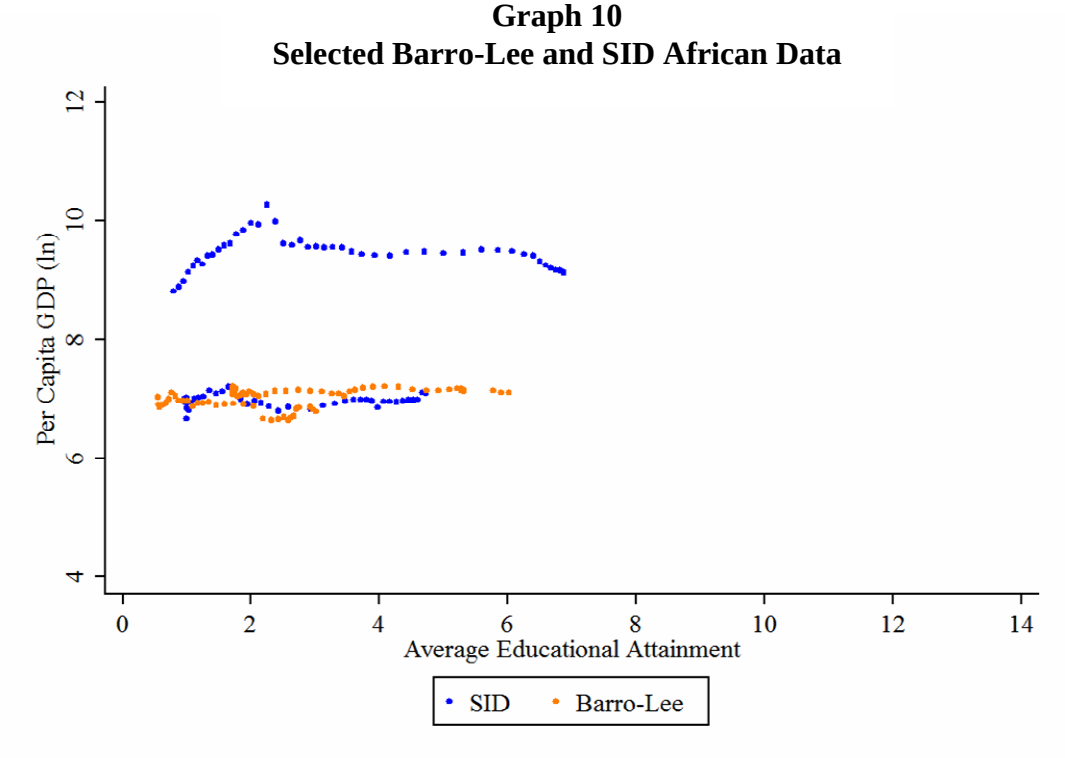


Table 7
Countries with No Correlation Between GDP and Educational Attainment

Barro-Lee	SID
Benin	Eritrea
Congo	Gabon
Mozambique	Guinea
Sudan	Liberia
Rwanda	Nigeria
	Sudan
	Zimbabwe

Graph 10
Selected Barro-Lee and SID African Data



Given the overall relationship of educational attainment and per capita GDP we would expect the educational attainment measure to be a reasonably good predictor of GDP. For example, we would posit that:

$$GDP_{PC} = a + \beta_1 * E\_ATTAIN_{25+} + \mu_i + \upsilon_{it} \quad (10)$$

To test this relationship we model per capita GDP using a random effects model where GDP is a function of educational attainment (β_1), an unobserved country-specific disturbance (μ_i), and case specific error term (υ_{it}) (see Baltagi 2001, Chapter 2).

Table 8
Model Coefficients

	Barro-Lee	SID	Merged Barro-Lee and SID	Barro-Lee Africa	SID Africa	Merged Barro-Lee and SID Africa
E_ATTAIN_{25+}	0.17	0.21	0.19	0.06	0.08	0.07
Intercept	7.41	6.98	7.21	7.23	7.12	7.16
N	3825	2259	6084	960	977	1937
R^2 (within)	0.34	0.42	0.37	0.04	0.10	0.08
R^2 (overall)	0.65	0.56	0.61	0.28	0.20	0.24

The Barro-Lee measure of educational attainment has a slightly lower coefficient than the SID measure of educational attainment (see Table 8). In both sets of data, educational attainment coefficient is much smaller for African countries and the explanatory power of the model is quite weak. This finding demonstrates the importance of estimating attainment data for the large number of African countries missing from the Barro and Lee data. It provides a more meaningful basis for estimating the developmental role of educational attainment, as well as the opportunity to study countries that evidence relationships between education and wealth that are distinctive from global norms. Understanding these distinctive relationships can provide insights into the factors that affect the role of education in development.

Conclusion

This paper develops, summarizes and documents an extensive effort to extend, in a methodologically defensible manner, cross-national data on one of the variables most crucial to our understanding of societal development: educational attainment. It demonstrates the feasibility of using an integrated set of strategies to devise source-specific solutions to eliminating missing data. In some cases we were able to find new sources of data; in other cases we were able to piece data together in ways that allowed us to estimate gaps in national time-series. We also used established data estimation procedures (perpetual inventory flow methods) where possible. To address other issues we developed new procedures that allowed us to capitalize on existing data to fill gaps (cross-variable estimation) and extend the time series (dynamic univariate backcasting).

In addition to being transparent in each of the methods used, where and when they were used, and what contribution they make to the overall estimation effort (see Chart 1 and 2; Appendix 14), we are meticulous in documenting the country-years affected by each procedure (see Appendices 6-12). This makes it possible for others to eliminate estimated data when they have doubts about specific estimation procedures. The result of our efforts is a pair of composite variables that measure educational attainment for two age groups that are standard in the literature (15+; 25+). These composites compare favorably with the Barro-Lee data when their distributions are examined and they perform comparably to the Barro-Lee data in explaining per capita GDP. Their value-added lies in capturing more countries for a longer period of time. As documented above, this is valuable because it captures more marginal countries that have distinctive relationships between educational attainment and development. The hope is that contributions such as these can advance our efforts at understanding developmental processes, particularly for those nations most in need of development.

References

Baffes, John and Anwar Shah. 1998. "Productivity of Public Spending, Sectoral Allocation Choices, and Economic Growth," *Economic Development and Cultural Change*, 46(2): 291-303.

Baltagi, Badi H. 2001. *Economic Analysis of Panel Data*. Chichester, UK: John Wiley and Sons Ltd.

Barro, Robert J. and Jong-Wha Lee. 1993. "International Comparisons of Educational Attainment," *Journal of Monetary Economics*, 32(3): 363-394.

Barro, Robert J. and Jong-Wha Lee. 2001. "International Data on Educational Attainment: Updates and Implications," *Oxford Economic Papers,* 3:541-563.

Lau, Lawrence J., Dean T. Jamison, and Frederic F. Louat. 1991. "Education and Productivity in Developing Countries: An Aggregate Production Function Approach," Policy Research Working Paper Series 612, The World Bank.

Nehru, Vikram, Eric Swanson, and Ashutosh Dubey. 1995. "A New Database on Human Capital Stock in Developing and Industrial Countries: Sources, Methodology, and Results," *Journal of Economic Development*, 46(2): 379-401.

Psacharopoulos, George and Ana Marie Arriagada. 1986. "The Educational Composition of the Labor Force: An International Comparison," *International Labor Review*, 125(5): 561-574.

Psacharopoulos, George and Ana Marie Arriagada. 1992. "The Educational Composition of the Labor Force: An International Update," PHREE Background Paper Series No. PHREE/92/49 (Education and Employment Division, Population and Human Resources Department, The World Bank, Washington, DC).

Appendices

Appendix E-1
Countries with Populations Less Than 500,000
Deleted From Barro-Lee 2000 Data Set

15 and Older

Antigua and Barbuda	1960
Dominica	1960, 1970, 1980
Hong Kong	1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000
Reunion	1965
Seychelles	1960, 1970, 1985
St. Kitts and Nevis	1960, 1980
St. Lucia	1960, 1970, 1980
St. Vincent and Grenadines	1960, 1980
Vanuatu	1980
Western Samoa	1965, 1970, 1975, 1980

25 and Older

Antigua and Barbuda	1960
Dominica	1970, 1980
Hong Kong	1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000
Puerto Rico	1990
Reunion	1955, 1965
Seychelles	1960, 1970, 1985
St. Kitts and Nevis	1960, 1980
St. Lucia	1970, 1980
St. Vincent and Grenadines	1980
Vanuatu	1980
Western Samoa	1965, 1970, 1975, 1980

Appendix E-2
Countries with Observations Prior to 1960 Deleted From Barro-Lee 2000 Data Set

25 and Older

Algeria	1955	Malta	1950
Argentina	1950	Mauritius	1950
Bulgaria	1955	Mexico	1950
Canada	1950	Myanmar (Burma)	1955
Chile	1950	Nicaragua	1950
Colombia	1950	Norway	1950
Costa Rica	1950	Panama	1950
Croatia	1955	Paraguay	1950
Ecuador	1950	Philippines	1950, 1955
El Salvador	1950	Romanía	1955
Finland	1950	Sudan	1955
France	1955	Turkey	1950
Greece	1950	United Kingdom	1950
Guatemala	1950	United States	1950
Haiti	1950	Venezuela	1950
Iran	1955	Yugoslavia (Serbia)	1955
Israel	1955	Zaire	1955
Italy	1950		
Japan	1950		
Korea, South (Rep.)	1955		

Appendix E-3
SID Countries Not Included in Barro-Lee 2000 Data Set

15 and Older

Albania	East Timor	Malta
Angola	Equatorial Guinea	Mongolia
Armenia	Eritrea	Morocco
Azerbaijan	Gabon	Nigeria
Bahamas	Georgia	Oman
Belarus	Guinea	Qatar
Bhutan	Korea, North	Saudi Arabia
Bosnia-Herzegovina	Kyrgyzstan	Somalia
Burkina Faso	Laos	Suriname
Cambodia	Lebanon	Turkmenistan
Cape Verde	Luxembourg	Ukraine
Chad	Macedonia	Uzbekistan
Comoros	Madagascar	Vietnam, Republic of
Cote d'Ivoire	Maldives	Yemen PDR (South, Aden)
Djibouti		

25 and Older

Albania	East Timor	Mongolia
Angola	Equatorial Guinea	Morocco
Armenia	Eritrea	Nigeria
Azerbaijan	Gabon	Oman
Bahamas	Georgia	Qatar
Belarus	Guinea	Saudi Arabia
Bhutan	Guinea-Bissau	Somalia
Bosnia-Herzegovina	Korea, North	Suriname
Burkina Faso	Kyrgyzstan	Tanzania
Cambodia	Laos	Turkmenistan
Cape Verde	Lebanon	Ukraine
Chad	Luxembourg	Uzbekistan
Comoros	Macedonia	Vietnam, Republic of
Cote d'Ivoire	Madagascar	Yemen
Djibouti	Maldives	Yemen, PDR

Appendix E-4
Sample Proof of "No Education" Estimator

Given:

$$H_{0,t} = H_{0,t-5}(1-\delta) + \left[L15_t(1-PRI_{t-5})\right]$$

$$\delta = \frac{(L15_t + L_{t-5} - L_t)}{L_{t-5}}$$

$$h_{0,t} = \frac{H_{0,t}}{L_t} \quad \text{and} \quad h_{0,t-5} = \frac{H_{0,t-5}}{L_{t-5}}$$

Then:

$$H_{0,t} = H_{0,t-5}\left[1-\left(\frac{(L15_t + L_{t-5} - L_t)}{L_{t-5}}\right)\right] + \left[L15_t(1-PRI_{t-5})\right]$$

$$H_{0,t} = H_{0,t-5}\left[1-\left(\frac{L15_t}{L_{t-5}} + \frac{L_{t-5}}{L_{t-5}} - \frac{L_t}{L_{t-5}}\right)\right] + \left[L15_t(1-PRI_{t-5})\right]$$

$$H_{0,t} = H_{0,t-5}\left[1-\left(\frac{L15_t}{L_{t-5}} + 1 - \frac{L_t}{L_{t-5}}\right)\right] + \left[L15_t(1-PRI_{t-5})\right]$$

$$H_{0,t} = H_{0,t-5}\left(\frac{L_t}{L_{t-5}} - \frac{L15_t}{L_{t-5}}\right) + \left[L15_t(1-PRI_{t-5})\right]$$

$$H_{0,t} = H_{0,t-5}\left(\frac{1}{L_{t-5}}\right)(L_t - L15_t) + \left[L15_t(1-PRI_{t-5})\right]$$

$$H_{0,t} = h_{0,t-5}(L_t - L15_t) + \left[L15_t(1-PRI_{t-5})\right]$$

$$H_{0,t}\left(\frac{1}{L_t}\right) = h_{0,t-5}\left(\frac{1}{L_t}\right)(L_t - L15_t) + \left(\frac{1}{L_t}\right)\left[L15_t(1-PRI_{t-5})\right]$$

$$h_{0,t} = h_{0,t-5}\left(1-\frac{L15_t}{L_t}\right) = \left(\frac{L15_t}{L_t}\right)(1-PRI_{t-5})$$

Appendix E-5
Forward and Backward-flow Estimators, by Age Category

Population age 15 and older

Where:
h_0 = percent of population 15 and older with no education
h_1 = percent of population 15 and older attained primary education
h_2 = percent of population 15 and older attained secondary education
h_3 = percent of population 15 and older attained tertiary education
L = population 15 and older
L15 = population 15-19
L20 = population 20-24
PRI = net primary enrollment ratio
SEC = net secondary enrollment ratio
HIGH = net tertiary enrollment ratio

Forward-flow

$$h_{0t} = h_{0t-5}\left(1-\frac{L15_t}{L_t}\right)+\left[\left(\frac{L15_t}{L_t}\right)\left(1-PRI_{t-5}\right)\right]$$

$$h_{1t} = h_{1t-5}\left(1-\frac{L15_t}{L_t}\right)+\left[\left(\frac{L15_t}{L_t}\right)\left(PRI_{t-5}-SEC_t\right)\right]$$

$$h_{2t} = h_{2t-5}\left(1-\frac{L15_t}{L_t}\right)+\left[\left(\frac{L15_t}{L_t}\right)\left(SEC_t\right)\right]-\left[\left(\frac{L20_t}{L_t}\right)\left(HIGH_t\right)\right]$$

$$h_{3t} = h_{3t-5}\left(1-\frac{L15_t}{L_t}\right)+\left[\left(\frac{L20_t}{L_t}\right)\left(HIGH_t\right)\right]$$

Backward-flow

$$h_{0t-5} = \frac{h_{0t} - \left[\left(\frac{L15_t}{L_t}\right)\left(1 - PRI_{t-5}\right)\right]}{\left(1 - \frac{L15_t}{L_t}\right)}$$

$$h_{1t-5} = \frac{h_{1t} - \left[\left(\frac{L15_t}{L_t}\right)\left(PRI_{t-5} - SEC_t\right)\right]}{\left(1 - \frac{L15_t}{L_t}\right)}$$

$$h_{2t-5} = \frac{h_{2t} - \left[\left(\frac{L15_t}{L_t}\right)\left(SEC_t\right)\right] + \left[\left(\frac{L20_t}{L_t}\right)\left(HIGH_t\right)\right]}{\left(1 - \frac{L15_t}{L_t}\right)}$$

$$h_{3t-5} = \frac{h_{3t} - \left[\left(\frac{L20_t}{L_t}\right)\left(HIGH_t\right)\right]}{\left(1 - \frac{L15_t}{L_t}\right)}$$

Population Age 25 and Older

Where:
h_0 = percent of population 25 and older with no education
h_1 = percent of population 25 and older attained primary education
h_2 = percent of population 25 and older attained secondary education
h_3 = percent of population 25 and older attained tertiary education
L = population 25 and older
L25 = population 25-29
PRI = net primary enrollment ratio
SEC = net secondary enrollment ratio
HIGH = gross tertiary enrollment ratio

Forward Flow Equations

$$h_{0t} = \left[h_{0t-5}\left(1 - \frac{L25_t}{L_t}\right)\right] + \left[\left(\frac{L25_t}{L_t}\right)\left(1 - PRI_{t-15}\right)\right]$$

$$h_{1t} = \left[h_{1t-5}\left(1 - \frac{L25_t}{L_t}\right)\right] + \left[\left(\frac{L25_t}{L_t}\right)\left(PRI_{t-15} - SEC_{t-10}\right)\right]$$

$$h_{2t} = \left[h_{2t-5}\left(1 - \frac{L25_t}{L_t}\right)\right] + \left[\left(\frac{L25_t}{L_t}\right)\left(SEC_{t-10} - HIGH_{t-5}\right)\right]$$

$$h_{3t} = \left[h_{3t-5}\left(1 - \frac{L25_t}{L_t}\right)\right] + \left[\left(\frac{L25_t}{L_t}\right)\left(HIGH_{t-5}\right)\right]$$

Backward Flow Equations

$$h_{0t-5} = \frac{h_{0t} - \left[\left(\frac{L25_t}{L_t}\right)\left(1 - PRI_{t-15}\right)\right]}{1 - \left(\frac{L25_t}{L_t}\right)}$$

$$h_{1t-5} = \frac{h_{1t} - \left[\left(\frac{L25_t}{L_t}\right)\left(PRI_{t-15} - SEC_{t-10}\right)\right]}{1 - \left(\frac{L25_t}{L_t}\right)}$$

$$h_{2t-5} = \frac{h_{2t} - \left[\left(\frac{L25_t}{L_t}\right)\left(SEC_{t-10} - HIGH_{t-5}\right)\right]}{1 - \left(\frac{L25_t}{L_t}\right)}$$

$$h_{3t-5} = \frac{h_{3t} - \left[\left(\frac{L25_t}{L_t}\right)\left(HIGH_{t-5}\right)\right]}{1 - \left(\frac{L25_t}{L_t}\right)}$$

Appendix E-6
Data Added by Perpetual Inventory Estimates and UNESCO Yearbook Search

15 and Older (108 data points)

Afghanistan
Algeria
Argentina
Australia
Austria
Bahrain
Bangladesh
Barbados
Belgium
Benin
Bolivia
Botswana
Brazil
Bulgaria
Cameroon
Canada
Central African Republic
Chile
China
Colombia
Congo
Costa Rica
Croatia
Cuba
Cyprus
Czech Republic
Denmark
Dominican Republic
Ecuador
Egypt
El Salvador
Fiji
Finland
France
Gambia
German Federal Republic
Ghana
Greece
Guatemala
Guinea-Bissau
Guyana
Honduras
Hungary
Iceland
India
Indonesia
Iran
Iraq
Ireland
Israel
Italy
Jamaica
Japan
Jordan
Kenya
Korea, South (Rep.)
Kuwait
Lesotho
Liberia
Malawi
Malaysia
Mali
Mauritius
Mexico
Mozambique
Nepal
Netherlands
New Zealand
Nicaragua
Niger
Norway
Pakistan
Panama
Papua New Guinea
Paraguay
Peru
Philippines
Poland
Portugal
Romania
Russia (USSR)
Rwanda
Senegal
Sierra Leone
Slovakia
Slovenia
South Africa
Spain
Sri Lanka
Sudan
Swaziland
Sweden
Switzerland
Syria
Tanzania
Thailand
Togo
Trinidad and Tobago
Tunisia
Turkey
Uganda
United Kingdom
United States
Uruguay
Venezuela
Zaire
Zambia
Zimbabwe

Appendix E-6
Data Added by Perpetual Inventory Estimates and UNESCO Yearbook Search (Continued)

25 and Older (324 data points)

Albania	1960 (all)
Bahamas	1985 (**SEC_25**, **TERT_25**), 1990 (all), 1995-2005 (**NOEDUC_25**, **PRIM_25**)
Belize	1995-2005 (**NOEDUC_25**, **PRIM_25**)
Brunei	1965, 1975, 1985 (**NOEDUC_25**, **PRIM_25**)
Cote d'Ivoire	1960-1985, 1995-2005 (**NOEDUC_25**, **PRIM_25, SEC_25**)
Guinea-Bissau	1975, 1985, 1990, 1995 (**NOEDUC_25**, **PRIM_25**), 2000 (**NOEDUC_25**)
Lebanon	1960-2005 (all)
Libya	1960 (**NOEDUC_25**, **PRIM_25**), 1970, 1980, 1990-2000 (all)
Luxembourg	1960-1990 (**PRIM_25, SEC_25**, **TERT_25**), 1995 (**PRIM_25**)
Mongolia	1965-2005 (all)
Morocco	1960-1965 (**NOEDUC_25**, **PRIM_25, SEC_25**), 1970-2005 (all)
Qatar	1975-2005 (all)
Solomon Islands	1980 (**NOEDUC_25**, **PRIM_25**), 1985 (**NOEDUC_25**), 2000 (all), 2005 (**NOEDUC_25, PRIM_25**)
Tanzania	1965 (**SEC_25**, **TERT_25**), 1970-1975 (all), 1985-2005 (**SEC_25, TERT_25**)
UAE	1970 (**NOEDUC_25, PRIM_25**), 1980-2005 (all)
Yemen	1965, 1970, 1980, 1985, 1990 (**NOEDUC_25**, **PRIM_25**)
Yugoslavia (Serbia)	1995, 2000 (all), 2005 (**NOEDUC_25, PRIM_25**)

Appendix E-7
Illiteracy Rate as Proxy for No Education

15 and Older (140 data points)

Albania	2000, 2005
Angola	1985, 2000, 2005
Armenia	1990, 2005
Belarus	1990, 2005
Bhutan	2005
Morocco	1980, 1995, 2005
Bosnia-Herzegovina	2000
Brunei	1990, 2000, 2005
Burkina-Faso	1975, 1980, 1985, 1995, 2000, 2005
Cambodia	2000, 2005
Cape Verde	1960, 1970, 1980, 1990, 1995, 2000, 2005
Chad	1960, 1975, 1985, 1995
Comoros	1960, 1965, 1970, 1980, 1990, 1995, 2000, 2005
Djibouti	1970, 1980, 1990, 1995, 2000, 2005
Equatorial Guinea	1970, 1980, 1990, 1995, 2000, 2005
Eritrea	1970, 1980, 1990, 1995, 2000, 2005
Ethiopia	1965, 1990, 1995, 2000, 2005
Gabon	1960, 1990, 1995, 2000, 2005
Ghana	2005
Guinea	1965, 1970, 1980, 1985, 1990, 1995, 2005
Kyrgyzstan	2000, 2005
Laos	1995, 2000, 2005
Lebanon	2005
Madagascar	1970, 1980, 2000, 2005
Macedonia	1995, 2000
Maldives	1975, 1985-2005
Malta	1985, 1995, 2005
Mongolia	2000, 2005
Namibia	2005
Nigeria	1960, 1970, 1980, 1990, 1995, 2000, 2005
Oman	2000, 2005
Qatar	1985, 1995, 2005
Saudi Arabia	1990, 2000, 2005
Sierra Leone	2005
Singapore	2005
Somalia	1960, 1980
Surinam	2005
Thailand	2005
Turkmenistan	1995, 2005
Ukraine	2000, 2005
Uzbekistan	2005
Vietnam	1980

Appendix E-7
Illiteracy Rate as Proxy for No Education (Continued)

25 and Older (110 data points)

Albania	2000, 2005	Malaysia	2005
Armenia	1990, 2005	Maldives	1975, 1985-2005
Belarus	1990, 2005	Mali	2005
Bhutan	2005	Malta	2005
Bosnia-Herzegovina	2000	Mauritania	2000, 2005
Botswana	2005	Mauritius	2005
Brunei	1990-2005	Moldova	2005
Burkina Faso	1975, 1990-2005	Mozambique	2005
Cambodia	2000, 2005	Namibia	1990, 2000-2005
Cape Verde	2005	Nepal	2005
Chad	2000, 2005	Nicaragua	2005
Chile	2005	Niger	2005
China	2005	Oman	2000, 2005
Comoros	2005	Panama	2005
Cyprus	2005	Papua New Guinea	2005
Dominican Republic	2005	Philippines	2005
Ecuador	2005	Poland	2005
Egypt	2005	Portugal	2005
Eritrea	2000, 2005	Russia	2005
Estonia	2005	Saudi Arabia	1990, 2000-2005
Gabon	1995, 2005	Senegal	2005
Ghana	2005	Sierra Leone	2005
Guinea	2005	Singapore	2005
Honduras	2005	Slovenia	2005
India	2005	Sri Lanka	2005
Indonesia	2005	Surinam	2005
Iran	2005	Swaziland	2005
Jamaica	2005	Tajikistan	2000, 2005
Jordan	2005	Thailand	2005
Kazakhstan	2000, 2005	Trinidad and Tobago	2005
Kyrgyzstan	2000, 2005	Tunisia	2005
Laos	1995-2005	Turkmenistan	1995, 2005
Latvia	2000, 2005	Ukraine	2000, 2005
Liberia	2005	Uzbekistan	2000
Lithuania	2000, 2005	Vietnam	1980, 2000
Macedonia	1995, 2000`	Yemen	1995, 2005
Malawi	2005	Zambia	2005

Appendix E-8
Data Added by Interpolation

15 and Older (27 data points)

NOEDUC_15

Armenia	1995, 2000	Malta	1990, 2000
Belarus	1995, 2000	Morocco	1985, 1990, 2000
Brunei	1965, 1975	Qatar	1990, 2000
Burundi	1985	Saudi Arabia	1995
Maldives	1980	Turkmenistan	2000

PRIM_15

Brunei	1965, 1975		

SEC_15

Belize	1985	Yugoslavia (Serbia)	2000
Brunei	1965, 1975		

TERT_15

Belize	1985	Yugoslavia (Serbia)	2000
Brunei	1965, 1975		

Appendix E-8
Data Added by Interpolation (Continued)

25 and Older (73 data points)

NOEDUC_25

Albania	1965-2000	Maldives	1980
Armenia	1995-2000	Mauritania	1995
Belarus	1995-2000	Moldova	1995-2000
Belize	1975, 1985	Morocco	1970
Brunei	1995	Namibia	1995
Burkina Faso	1980-1985	Nigeria	1995-2000
Estonia	1995-2000	Saudi Arabia	1995
Ethiopia	2000	Solomon Islands	1990-1995
Gabon	2000	Tajikistan	1995
Guinea-Bissau	1980	Turkmenistan	2000
Kazakhstan	1995	Vietnam	1985, 1995
Latvia	1995	Yemen	2000
Lithuania	1995	Yemen, PDR	1975

PRIM_25

Belize	1975, 1985	Solomon Islands	1985-1995
Guinea-Bissau	1980	Yemen, PDR	1975
Morocco	1970		

SEC_25

Belize	1975, 1985	Solomon Islands	1980-1995
Brunei	1965, 1975	Tanzania	1980
Morocco	1970		

TERT_25

Belize	1975, 1985	Solomon Islands	1980-1995
Brunei	1965, 1975	Tanzania	1980
Namibia	1965-1985		

Appendix E-9
Data Added by Extrapolating to Missing Early Years

15 and Older (71 data points)

NOEDUC_15

China	1960, 1965, 1970
Egypt	1960, 1965
Democratic Republic of Vietnam	1960, 1965, 1970, 1975
Republic of Vietnam (South Vietnam)	1960, 1965, 1970, 1975
Yemen (Arab Republic of Yemen)	1960, 1965, 1970
Yemen (People's Republic of Yemen)	1970, 1975, 1980, 1985

PRIM_15

China	1960, 1965, 1970
Egypt	1960, 1965
Democratic Republic of Vietnam	1960, 1965, 1970, 1975, 1980
Republic of Vietnam (South Vietnam)	1960, 1965, 1970, 1975
Yemen (Arab Republic of Yemen)	1960, 1965, 1970
Yemen (People's Republic of Yemen)	1970, 1975, 1980, 1985

SEC_15

China	1960, 1965
Egypt	1960, 1965
Democratic Republic of Vietnam	1960, 1965, 1970, 1975, 1980
Republic of Vietnam (South Vietnam)	1960, 1965, 1970, 1975
Yemen (People's Republic of Yemen)	1970, 1975, 1980, 1985

TERT_15

Democratic Republic of Vietnam	1960, 1965, 1970, 1975, 1980
Republic of Vietnam (South Vietnam)	1960, 1965, 1970, 1975
Yemen (People's Republic of Yemen)	1970, 1975, 1980, 1985

Appendix E-9
Data Added by Extrapolating to Missing Early Years (Continued)

25 and Older (42 data points)

NOEDUC_25	
Bahamas	1975, 1980, 1985
Benin	1965, 1970
China	1960, 1965, 1970
Egypt	1960, 1965, 1970
Mongolia	1960
PRIM_25	
Bahamas	1975, 1980, 1985
Benin	1965, 1970
China	1960, 1965, 1970
Egypt	1960, 1965, 1970
Mongolia	1960
SEC_25	
Bahamas	1975, 1980, 1995, 2000, 2005
China	1960, 1965, 1970
Egypt	1960, 1965, 1970
Mongolia	1960
TERT_25	
Bahamas	1975, 1980, 1995, 2000, 2005
Mongolia	1960

Appendix E-10
Matrix of Illiteracy Rates for African Countries

Country	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005
Angola	*	*	*	62.2	56.3	**50.4**	44.5	38.6	**35.6**	**32.6**
Burkina Faso	95.0	94.0	93.0	**92.5**	**91.3**	**90.3**	**89.3**	**87.2**	**81.8**	**76.4**
Cape Verde	**73.6**	68.4	**63.1**	56.3	**49.5**	43.4	**37.2**	**29.2**	**22.9**	**22.0**
Chad	98.1	**94.4**	**90.7**	87.1	**83.3**	76.8	**70.2**	87.8	**74.3**	**71.6**
Comoros	72.3	66.7	61.8	56.3	**52.1**	46.7	**41.3**	**37.3**	**31.2**	**27.6**
Djibouti	*	*	*	*	**59.5**	53.3	**47.0**	**53.8**	**35.4**	**29.7**
Equatorial Guinea	82.9	78.6	**73.4**	68.2	**63.0**	56.4	**49.8**	**42.4**	**35.8**	**28.6**
Eritrea	*	*	*	*	*	*	*	**48.9**	**47.5**	**38.5**
Ethiopia	96.2	**94.0**	**87.1**	83.6	**80.1**	**75.7**	**74.4**	**73.0**	**71.3**	**64.1**
Gabon	**87.6**	79.8	71.7	63.6	55.5	47.4	**39.3**	**27.8**	**22.6**	**20.0**
Guinea	94.6	**91.4**	88.4	85.3	82.2	79.1	**76.0**	**72.9**	71.9	**70.5**
Madagascar	85.3	80.7	**76.1**	64.3	**52.5**	47.6	41.4	34.9	**29.3**	**25.4**
Nigeria	**84.6**	82.3	**79.9**	72.9	**66.0**	55.3	**44.6**	**40.7**	**35.3**	**31.0**
Somalia	**98.5**	98.0	97.3	95.2	**93.9**	88.5	83.4	81.4	79.9	77.4

* indicates that the country was not an independent state

Appendix E-11
Data Added by Lutz et al. (2010) for both 15 and 25 and Older Age Groups

Country	Years
Cambodia	1970, 1975, 1980, 1985, 1990, 1995, 2000
Laos	1970, 1975, 1980, 1985, 1990, 1995, 2000
Macedonia	1995, 2000
Maldives	1975, 1980, 1985, 1990, 1995, 2000
Saudi Arabia	1970, 1975, 1980, 1985, 1990, 1995, 2000

Appendix E-12
Data Added by Extrapolation to 2005

15 and Older (73 data points)

NOEDUC_15

Angola
Djibouti
Equatorial Guinea
Haiti
Iraq
Macedonia
Madagascar
Myanmar
Somalia
Taiwan

PRIM_15

Angola
Brunei
Cape Verde
Comoros
Djibouti
Equatorial Guinea
Gabon
Guinea
Haiti
Iraq
Madagascar
Myanmar
Namibia
Nigeria
Papua New Guinea
Sierra Leone
Singapore
Somalia
Taiwan
Thailand

SEC_15

Angola
Cape Verde
Comoros
Djibouti
Equatorial Guinea
Gabon
German Federal Republic
Guinea
Haiti
Iraq
Madagascar
Malawi
Myanmar
Namibia
Nigeria
Papua New Guinea
Singapore
Somalia
Sri Lanka
Syria
Taiwan
Zambia

Appendix E-12
Data Added by Extrapolation to 2005 (Continued)

TERT_15

Angola
Cape Verde
Comoros
Djibouti
Equatorial Guinea
Gabon
German Federal Republic
Guinea
Haiti
Iraq
Madagascar
Malawi
Myanmar
Namibia
Nigeria
Singapore
Somalia
Sri Lanka
Syria
Taiwan
Zambia

Appendix E-13
Countries With a Backcast

Country	Number of Backcasts	Years for Backcasts
Afghanistan	10	1950-1959
Australia	10	1950-1959
Austria	10	1950-1959
Belgium	10	1950-1959
Bolivia	10	1950-1959
Brazil	10	1950-1959
Cambodia	20	1950-1969
Canada	10	1950-1959
China	10	1950-1959
Cuba	10	1950-1959
Czechoslovakia	10	1950-1959
Denmark	10	1950-1959
Dominican Republic	10	1950-1959
Egypt	10	1950-1959
Ethiopia	10	1950-1959
Finland	10	1950-1959
France	10	1950-1959
German Democratic Republic	10	1950-1959
German Federal Republic	10	1950-1959
Ghana	3	1957-1959
Guinea	2	1958-1959
Honduras	10	1950-1959
Hungary	10	1950-1959
Iceland	10	1950-1959
India	10	1950-1959
Indonesia	10	1950-1959
Iraq	10	1950-1959
Ireland	10	1950-1959
Jordan	10	1950-1959
Laos	20	1950-1969
Lebanon	10	1950-1959
Liberia	10	1950-1959
Libya	9	1951-1959
Luxembourg	10	1950-1959
Malaysia	6	1954-1959
Maldives	5	1965-1969
Mongolia	10	1950-1959
Morocco	4	1956-1959
Nepal	10	1950-1959

Appendix E-13
Countries With a Backcast (Continued)

Country	Number of Backcasts	Years for Backcasts
Netherlands	10	1950-1959
New Zealand	10	1950-1959
Norway	10	1950-1959
Pakistan	10	1950-1959
Peru	10	1950-1959
Poland	10	1950-1959
Portugal	10	1950-1959
Russia (Soviet Union)	10	1950-1959
Saudi Arabia	20	1950-1969
South Africa	10	1950-1959
Spain	10	1950-1959
Sri Lanka (Ceylon)	10	1950-1959
Sweden	10	1950-1959
Switzerland	10	1950-1959
Syria	10	1950-1959
Taiwan	10	1950-1959
Thailand	10	1950-1959
Tunisia	4	1956-1959
Uruguay	10	1950-1959
Vietnam, Democratic Republic of	6	1954-1959
Vietnam, Republic of	5	1955-1959
Yemen (Arab Republic of Yemen)	10	1950-1959
Total Number of Backcasts	529	

Appendix E-14
Data Source Summary

15 and Older Data

Year	Barro-Lee 2000	UNESCO Data and Inventory Models	Literacy	Interpolation	Extrapolation	Africa	Post-Soviet	2005 Estimates	Cross-Composite Estimation	Missing	Total
1960	0.75	0.02	0.00	0.00	0.04	0.09	0.00	0.00	0.04	0.07	1.00
1965	0.76	0.03	0.00	0.00	0.03	0.08	0.00	0.00	0.05	0.06	1.00
1970	0.76	0.04	0.00	0.00	0.03	0.07	0.00	0.00	0.04	0.06	1.00
1975	0.76	0.02	0.00	0.00	0.02	0.09	0.00	0.00	0.05	0.06	1.00
1980	0.75	0.03	0.00	0.00	0.01	0.09	0.00	0.00	0.05	0.06	1.00
1985	0.75	0.03	0.00	0.01	0.01	0.09	0.00	0.00	0.04	0.06	1.00
1990	0.76	0.02	0.00	0.00	0.00	0.09	0.00	0.00	0.05	0.07	1.00
1995	0.66	0.05	0.01	0.00	0.00	0.09	0.08	0.00	0.04	0.07	1.00
2000	0.66	0.05	0.01	0.00	0.00	0.09	0.08	0.00	0.04	0.06	1.00
2005	0.00	0.00	0.02	0.00	0.00	0.00	0.08	0.80	0.04	0.06	1.00

25 and Older Data

Year	Barro-Lee 2000	UNESCO Data and Inventory Models	Literacy	Interpolation	Extrapolation	Africa	Post-Soviet	2005 Estimates	Cross-Composite Estimation	Missing	Total
1960	0.74	0.02	0.00	0.00	0.03	0.00	0.00	0.00	0.14	0.06	1.00
1965	0.75	0.04	0.00	0.00	0.02	0.00	0.00	0.00	0.13	0.06	1.00
1970	0.76	0.04	0.00	0.01	0.01	0.00	0.00	0.00	0.12	0.06	1.00
1975	0.76	0.04	0.00	0.01	0.01	0.00	0.00	0.00	0.13	0.06	1.00
1980	0.74	0.06	0.01	0.01	0.01	0.00	0.00	0.00	0.12	0.06	1.00
1985	0.74	0.06	0.00	0.01	0.00	0.00	0.00	0.00	0.13	0.06	1.00
1990	0.75	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.11	0.06	1.00
1995	0.66	0.05	0.01	0.03	0.00	0.00	0.00	0.00	0.19	0.07	1.00
2000	0.65	0.05	0.04	0.02	0.00	0.00	0.00	0.00	0.18	0.06	1.00
2005	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.74	0.15	0.06	1.00

Appendix E-15
Illustrative Changes by Country and Continent over Time

Country	Continent	Year	E_ATTAIN$_{25+}$, 1950 or First Year of Independence	Rank Within Continent	Rank Within World	E_ATTAIN$_{25+}$, 2005	Rank Within Continent	Rank Within World	Average Change per Year	Rank Within Continent	Rank Within World
Afghanistan	Asia	1950	1.715	19	113	2.007	24	152	0.005	24	155
Algeria	Africa	1962	0.728	31	134	6.275	13	104	0.129	6	16
Angola	Africa	1975	1.495	21	116	3.512	32	136	0.067	28	92
Argentina	South America	1950	6.454	2	48	10.116	2	33	0.067	7	94
Armenia	Europe	1991	10.136	4	6	10.452	16	25	0.023	37	148
Australia	Oceania	1950	9.479	2	13	11.294	2	12	0.033	3	142
Austria	Europe	1950	8.143	18	29	10.433	17	26	0.042	32	133
Azerbaijan	Europe	1991	8.983	11	19	9.022	33	58	0.003	40	157
Bahamas	North America	1973	9.704	3	10	11.300	3	11	0.050	14	122
Bahrain	Middle East	1971	2.146	6	106	8.330	5	71	0.182	1	2
Bangladesh	Asia	1971	0.453	23	143	3.739	21	132	0.097	10	46
Barbados	North America	1966	11.751	1	1	9.962	4	39	-0.046	17	161
Belarus	Europe	1991	9.190	8	16	10.101	23	35	0.065	23	96
Belgium	Europe	1950	7.744	22	33	9.696	27	42	0.035	35	139
Belize	North America	1981	8.449	5	26	8.210	10	73	-0.010	16	159
Benin	Africa	1960	0.378	41	146	3.072	39	145	0.060	30	107
Bolivia	South America	1950	5.815	4	54	7.808	7	83	0.036	11	138
Botswana	Africa	1966	2.073	15	107	6.981	6	91	0.126	7	18
Brazil	South America	1950	3.486	7	79	7.517	9	85	0.073	6	83
Brunei	Asia	1984	7.523	5	35	10.123	4	32	0.124	7	20
Bulgaria	Europe	1950	6.832	27	41	10.900	9	16	0.074	18	77
Burkina Faso (Upper Volta)	Africa	1960	0.321	44	150	1.509	48	155	0.026	45	146
Burundi	Africa	1962	2.516	11	98	2.691	42	148	0.004	50	156
Cameroon	Africa	1960	1.916	17	110	5.188	19	117	0.073	27	84
Canada	North America	1950	9.969	2	7	12.879	2	2	0.053	13	118
Cape Verde	Africa	1975	1.715	18	112	6.826	11	98	0.170	1	4
Central African Republic	Africa	1960	0.510	35	139	3.111	38	144	0.058	31	109
Chad	Africa	1960	0.130	49	156	2.380	44	150	0.050	36	121
Chile	South America	1950	5.982	3	52	10.336	1	28	0.079	5	69
China	Asia	1950	2.206	17	104	6.359	17	103	0.076	13	73
Colombia	South America	1950	3.379	8	82	6.700	11	99	0.060	9	105
Comoros	Africa	1975	2.633	9	95	5.198	18	116	0.086	19	59

Appendix E-15
Illustrative Changes by Country and Continent over Time (Continued)

Country	Continent	Year	E_ATTAIN$_{25+}$, 1950 or First Year of Independence	Rank Within Continent	Rank Within World	E_ATTAIN$_{25+}$, 2005	Rank Within Continent	Rank Within World	Average Change per Year	Rank Within Continent	Rank Within World
Congo	Africa	1960	1.954	16	109	6.609	12	101	0.103	13	35
Costa Rica	North America	1950	5.218	7	57	7.912	11	79	0.049	15	125
Cote d'Ivoire	Africa	1960	0.765	30	132	4.141	30	130	0.075	25	74
Croatia	Europe	1991	10.526	2	4	8.124	36	75	-0.172	41	162
Cuba	North America	1950	4.429	10	69	9.398	5	45	0.090	3	50
Cyprus	Europe	1960	5.322	35	56	10.114	22	34	0.106	5	33
Czech Republic	Europe	1993	9.025	10	18	10.793	11	19	0.147	2	9
Czechoslovakia	Europe	1950	8.836	13	21	.	.	.	0.080	15	68
Democratic Republic of Congo (Zaire)	Africa	1960	0.918	28	129	4.598	24	124	0.082	23	66
Democratic Republic of Vietnam	Asia	1954	3.224	15	87	6.671	16	100	0.068	18	90
Denmark	Europe	1950	10.202	3	5	12.649	1	3	0.044	30	129
Djibouti	Africa	1977	2.275	14	103	4.712	23	122	0.087	18	57
Dominican Republic	North America	1950	3.244	12	86	6.835	13	97	0.065	8	95
Ecuador	South America	1950	3.354	9	83	8.223	6	72	0.089	2	55
Egypt	Africa	1950	0.021	51	160	6.857	9	95	0.124	8	19
El Salvador	North America	1950	2.166	15	105	7.082	12	90	0.089	4	54
Equatorial Guinea	Africa	1960	1.076	23	123	4.551	25	125	0.077	24	72
Eritrea	Africa	1993	2.805	7	93	3.134	37	143	0.027	44	145
Estonia	Europe	1991	9.556	6	11	10.589	14	22	0.074	20	79
Ethiopia	Africa	1950	0.214	47	154	2.174	45	151	0.037	40	137
Fiji	Oceania	1970	6.804	3	43	8.785	3	63	0.057	2	113
Finland	Europe	1950	6.038	32	51	11.473	4	9	0.099	10	40
France	Europe	1950	6.823	28	42	9.969	26	38	0.057	27	111
Gabon	Africa	1960	0.798	29	131	6.932	8	94	0.136	5	12
Gambia	Africa	1965	0.469	37	141	2.884	41	147	0.060	29	104
Georgia	Europe	1991	4.495	38	67	6.224	39	108	0.123	3	21
German Democratic Republic	Europe	1950	10.931	1	3	.	.	.	0.012	38	153
German Federal Republic	Europe	1950	8.505	17	25	10.789	12	20	0.042	33	134
Ghana	Africa	1957	0.975	24	125	5.686	17	114	0.098	14	42
Greece	Europe	1950	5.054	36	60	9.500	29	44	0.081	14	67
Guatemala	North America	1950	1.827	16	111	4.793	16	120	0.054	11	116
Guinea	Africa	1958	0.234	46	153	2.572	43	149	0.050	37	124
Guinea-Bissau	Africa	1974	0.471	36	140	1.974	47	154	0.048	38	126

Appendix E-15
Illustrative Changes by Country and Continent over Time (Continued)

Country	Continent	Year	E_ATTAIN$_{25+}$, 1950 or First Year of Independence	Rank Within Continent	Rank Within World	E_ATTAIN$_{25+}$, 2005	Rank Within Continent	Rank Within World	Average Change per Year	Rank Within Continent	Rank Within World
Guyana	South America	1966	5.193	4	58	8.451	5	70	0.084	3	62
Iceland	Europe	1950	6.397	31	49	10.662	13	21	0.078	16	71
India	Asia	1950	1.488	20	117	5.237	19	115	0.068	17	89
Haiti	North America	1950	0.740	17	133	3.703	17	133	0.054	12	117
Honduras	North America	1950	2.847	13	91	6.053	15	110	0.058	10	108
Hungary	Europe	1950	7.958	20	31	10.365	18	27	0.044	31	130
Indonesia	Asia	1950	1.237	21	120	6.198	18	109	0.090	11	51
Iran (Persia)	Middle East	1950	0.288	10	152	6.393	9	102	0.111	6	27
Iraq	Middle East	1950	0.054	11	158	5.920	10	112	0.107	8	32
Ireland	Europe	1950	7.455	24	36	10.295	19	29	0.052	28	119
Israel	Middle East	1950	7.364	1	37	10.891	1	17	0.064	12	98
Italy/Sardinia	Europe	1950	5.011	37	62	8.966	34	59	0.072	21	85
Jamaica	North America	1962	4.948	8	63	9.160	8	55	0.098	2	43
Japan	Asia	1950	9.194	2	15	11.489	2	8	0.042	23	132
Jordan	Middle East	1950	1.170	8	122	8.473	4	69	0.133	3	14
Kazakhstan	Asia	1991	8.322	3	27	9.289	7	50	0.069	15	87
Kenya	Africa	1963	1.617	20	115	6.258	14	105	0.110	11	28
Kuwait	Middle East	1961	2.620	5	97	7.880	7	80	0.120	4	23
Kyrgyz Republic	Asia	1991	8.171	4	28	9.263	9	52	0.078	12	70
Latvia	Europe	1991	8.857	12	20	10.083	24	36	0.088	13	56
Lebanon	Middle East	1950	3.394	3	81	8.774	3	64	0.098	9	44
Lesotho	Africa	1966	3.761	5	75	6.009	16	111	0.058	32	110
Liberia	Africa	1950	0.439	39	144	3.474	33	137	0.055	35	115
Libya	Africa	1951	0.332	42	148	9.248	1	53	0.165	2	5
Lithuania	Europe	1991	9.865	5	9	10.903	8	15	0.074	17	76
Luxembourg	Europe	1950	0.689	41	135	9.118	32	57	0.153	1	7
Madagascar (Malagasy)	Africa	1960	0.946	26	127	5.004	20	118	0.090	16	52
Malawi	Africa	1964	2.631	10	96	4.261	28	128	0.040	39	136
Malaysia	Asia	1957	2.845	16	92	9.159	10	56	0.132	5	15
Mali	Africa	1960	0.306	45	151	1.215	51	158	0.020	49	152
Malta	Europe	1964	7.536	23	34	8.736	35	65	0.029	36	143
Mauritania	Africa	1960	3.866	4	73	3.229	35	140	-0.014	51	160
Mauritius	Africa	1968	4.586	3	66	8.146	4	74	0.096	15	47

Appendix E-15
Illustrative Changes by Country and Continent over Time (Continued)

Country	Continent	Year	E_ATTAIN_{25+}, 1950 or First Year of Independence	Rank Within Continent	Rank Within World	E_ATTAIN_{25+}, 2005	Rank Within Continent	Rank Within World	Average Change per Year	Rank Within Continent	Rank Within World
Mexico	North America	1960	3.638	11	77	8.482	9	68	0.108	1	31
Moldova	Europe	1991	9.049	9	17	9.617	28	43	0.041	34	135
Mongolia	Asia	1950	0.361	24	147	7.853	12	81	0.136	4	13
Morocco	Africa	1956	0.200	48	155	4.209	29	129	0.082	22	65
Mozambique	Africa	1975	0.926	27	128	1.984	46	153	0.035	41	140
Myanmar (Burma)	Asia	1950	2.015	18	108	4.627	20	123	0.047	21	127
Namibia	Africa	1990	7.248	1	39	8.801	2	61	0.104	12	34
Nepal	Asia	1950	0.032	25	159	3.437	22	138	0.062	20	102
Netherlands	Europe	1950	5.901	33	53	11.041	6	13	0.093	12	49
New Zealand	Oceania	1950	11.027	1	2	12.256	1	5	0.022	4	151
Nicaragua	North America	1950	2.312	14	101	6.251	14	106	0.072	7	86
Niger	Africa	1960	0.321	43	149	1.328	49	156	0.022	47	149
Nigeria	Africa	1960	0.969	25	126	4.789	22	121	0.085	20	60
Norway	Europe	1950	6.728	30	45	12.361	2	4	0.102	6	36
Pakistan	Asia	1950	0.817	22	130	3.199	23	141	0.043	22	131
Panama	North America	1950	4.694	9	65	9.355	7	47	0.085	5	61
Papua New Guinea	Oceania	1975	1.296	4	119	3.543	4	135	0.075	1	75
Paraguay	South America	1950	4.052	5	71	7.484	10	86	0.062	8	100
People's Republic of Yemen	Middle East	1967	-0.013	13	162	.	.		0.047	13	128
Peru	South America	1950	3.605	6	78	8.926	4	60	0.097	1	45
Philippines	Asia	1950	3.252	14	85	9.267	8	51	0.109	8	30
Poland	Europe	1950	6.946	26	40	11.014	7	14	0.074	19	78
Portugal	Europe	1950	2.386	40	100	7.919	38	78	0.101	9	39
Qatar	Middle East	1971	5.056	2	59	8.793	2	62	0.110	7	29
Republic of Korea	Asia	1950	6.091	6	50	11.523	1	7	0.099	9	41
Rumania	Europe	1950	6.770	29	44	10.470	15	23	0.067	22	91
Russia (Soviet Union)	Europe	1950	9.458	7	14	10.066	25	37	0.011	39	154
Rwanda	Africa	1962	0.543	33	137	4.085	31	131	0.082	21	63
Senegal	Africa	1960	2.428	12	99	3.435	34	139	0.022	48	150
Sierra Leone	Africa	1961	0.551	32	136	3.055	40	146	0.057	33	112
Singapore	Asia	1965	4.064	9	70	9.846	5	40	0.145	3	11
Slovakia	Europe	1993	8.690	15	23	10.166	20	30	0.123	4	22
Slovenia	Europe	1991	8.708	14	22	10.126	21	31	0.101	8	38

Appendix E-15
Illustrative Changes by Country and Continent over Time (Continued)

Country	Continent	Year	E_ATTAIN$_{25+}$, 1950 or First Year of Independence	Rank Within Continent	Rank Within World	E_ATTAIN$_{25+}$, 2005	Rank Within Continent	Rank Within World	Average Change per Year	Rank Within Continent	Rank Within World
Somalia	Africa	1960	0.090	50	157	1.251	50	157	0.026	46	147
South Africa	Africa	1950	5.037	2	61	6.970	7	93	0.035	42	141
Spain	Europe	1950	3.754	39	76	9.341	30	48	0.102	7	37
Sri Lanka (Ceylon)	Asia	1950	4.946	7	64	8.643	11	66	0.067	19	93
Sudan	Africa	1956	0.408	40	145	3.151	36	142	0.056	34	114
Swaziland	Africa	1968	1.424	22	118	7.453	5	87	0.163	3	6
Sweden	Europe	1950	8.559	16	24	11.864	3	6	0.060	26	106
Switzerland	Europe	1950	8.002	19	30	11.373	5	10	0.061	25	103
Syria	Middle East	1950	1.014	9	124	7.355	8	88	0.115	5	25
Taiwan	Asia	1950	3.414	12	80	10.467	3	24	0.128	6	17
Tajikistan	Asia	1991	9.903	1	8	9.814	6	41	-0.006	25	158
Tanzania (Tanganyika)	Africa	1961	3.192	6	88	4.452	27	127	0.029	43	144
Thailand	Asia	1950	3.305	13	84	7.095	14	89	0.069	16	88
Togo	Africa	1960	0.456	38	142	4.511	26	126	0.090	17	53
Trinidad and Tobago	North America	1962	6.705	6	46	9.377	6	46	0.062	9	101
Tunisia	Africa	1956	0.522	34	138	6.239	15	107	0.117	9	24
Turkey/Ottoman Empire	Middle East	1950	1.211	7	121	5.916	11	113	0.086	10	58
Turkmenistan	Asia	1991	3.802	11	74	7.850	13	82	0.289	1	1
Uganda	Africa	1962	1.647	19	114	4.807	21	119	0.073	26	81
Ukraine	Europe	1991	7.874	21	32	9.194	31	54	0.094	11	48
United Arab Emirates	Middle East	1971	2.952	4	90	7.963	6	77	0.147	2	8
United Kingdom	Europe	1950	7.330	25	38	10.805	10	18	0.063	24	99
United States of America	North America	1950	9.502	4	12	13.556	1	1	0.074	6	80
Uruguay	South America	1950	6.581	1	47	9.339	3	49	0.050	10	120
Uzbekistan	Asia	1991	4.473	8	68	6.970	15	92	0.178	2	3
Venezuela	South America	1950	3.101	10	89	7.614	8	84	0.082	4	64
Vietnam, Republic of	Asia	1955	3.994	10	72	.	.	.	0.073	14	82
Yemen (Arab Republic of Yemen)	Middle East	1950	0.015	12	161	3.571	12	134	0.065	11	97
Yugoslavia (Serbia)	Europe	1950	5.352	34	55	8.089	37	76	0.050	29	123
Zambia	Africa	1964	2.292	13	102	6.853	10	96	0.111	10	26
Zimbabwe (Rhodesia)	Africa	1965	2.706	8	94	8.554	3	67	0.146	4	10