## Human capital in PWT 9.0

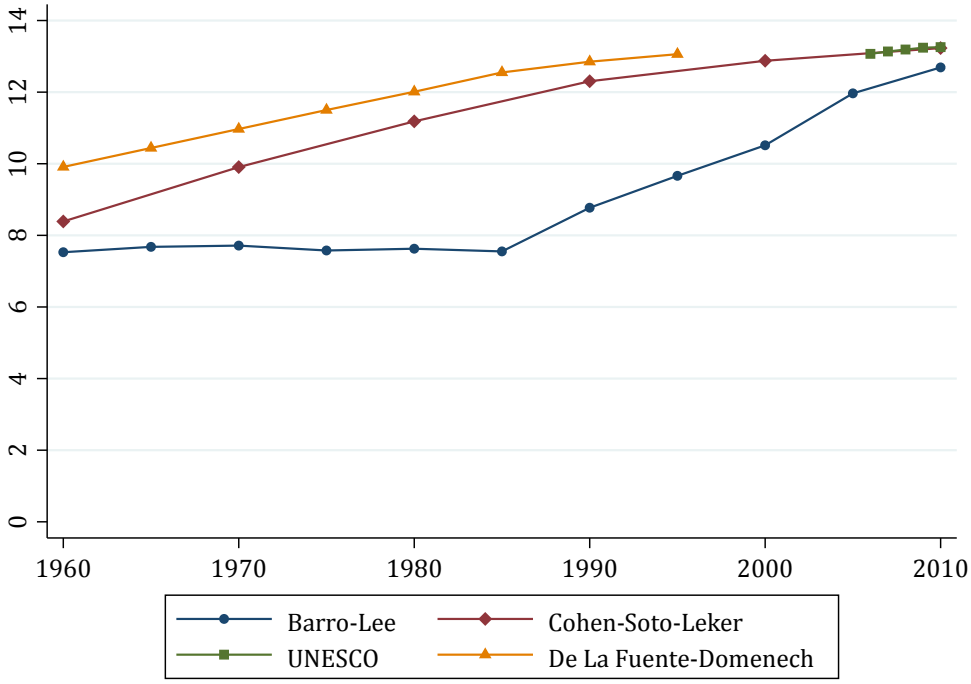
Following a common approach in the literature (e.g. Caselli, 2005), PWT version 8 introduced a human capital index based on the average years of schooling from Barro and Lee (BL, 2013) and an assumed rate of return to education, based on Mincer equation estimates around the world (Psacharopoulos, 1994). The construction of data on the average years of schooling has attracted considerable attention in the literature, with authors such as De La Fuente and Doménech (DD, 2006) and Cohen and Soto (2007) arguing that (earlier versions of) the BL data had undesirable features related to their inconsistent use of source data. The general challenge in constructing data on the average years of schooling in the population is to combine information from (decadal) population censuses with information on school enrolment, in the face of inconsistencies in classification systems between sources or censuses.

Since the release of PWT 8.0 in July 2013, Cohen and Leker (2014) have constructed an alternative dataset for average years of schooling, which, they claim, is superior to the BL data and which is constructed in a similar fashion as the Cohen and Soto (2007) data; it will be referred as the CSL data. Though we take no stand on which specific data construction approach is most desirable, the availability of the CSL data provides an opportunity for comparison to the BL data and, as we will argue, there is a clear case for combining data from both sources in compiling the human capital index in PWT 9.0.

A *prima facie* case for examining the different sources in more detail is given in Figure 1. The figure plots data for Germany from BL (available for 146 countries at five-year intervals from 1950 to 2010), CSL (95 countries, ten-year interval, 1960-2020), DD (21 countries, five-year interval, 1960-1995) and from Unesco (125 countries, annual but short time series concentrated in the 2000s). According to the BL data, the average years of schooling in Germany was only 7.6 years as late as 1985, while neighbouring countries like Austria (8.4 years), the Netherlands (9.9) and Poland (8.3) all had notably higher levels of schooling. After 1985, BL show rapidly increasing educational attainment until average years of schooling hit 12 years in 2005, exceeding the level in Denmark, the Netherlands and Poland. In contrast, the other three sources show higher levels of schooling throughout the period and more gradual increases over time. Moreover, both

in levels and trends, the CSL, DD and UNESCO data are fairly close. This argues against using the BL data in the case of Germany.

**Figure 1, Average years of schooling Germany, 1950-2010.**




Note: for source of data series, see main text. Observations for earlier years (Barro-Lee) or later years (Cohen-Soto-Leker) are omitted for brevity.

**Table 1, Descriptive statistics on average years of schooling, BL and CSL data**

| | Mean | St. dev. | Min | Max |
|---|---|---|---|---|
| BL data | 5.4 | 3.3 | 0.1 | 13.4 |
| CSL data | 5.6 | 3.5 | 0.1 | 13.4 |

Notes: descriptive statistics cover 525 country/year observations covered in both datasets for countries with more than one observation.

More in general, this argues for a more systematic comparison of data from these sources. Since BL and CSL provide the broadest coverage in terms of years and countries, we focus on contrasting these two for the 525 overlapping country/year observations, including all countries with at least two observations. Table 1 shows that the two datasets are, overall, broadly comparable with only small differences in descriptive statistics, i.e. a mean of 5.4 versus 5.6 average years of schooling. The correlation across all countries and years is very high at 0.94 and even for Germany, the correlation is high at 0.80. This

suggests that for cross-country comparisons, the choice of data source will not be highly consequential.

For growth in the average years of schooling, and thus in human capital, the choice turns out to be much more consequential. To illustrate this, we compute a human capital index based on both series as:[1]

$$\phi(s)=\begin{cases} 0.134\cdot s & \text{if } s\le 4 \\ 0.134\cdot 4+0.101(s-4) & \text{if } 4<s\le 8 \\ 0.134\cdot 4+0.101\cdot 4+0.068(s-8) & \text{if } s>8 \end{cases},$$

where $s$ is the average years of schooling from either dataset. The correlation of decadal growth rates, $\phi(s_{it})/\phi(s_{it-10})-1$, is 0.37, much lower than for the level of schooling. Figure 2 illustrates this correlation in a scatter plot, indicating the wide variation in growth rates and the low degree of similarity between the two series. This will have its main impact for estimates of growth of total factor productivity (TFP) – reflected in PWT variable *RTFPNA*.

**Figure 2, 10-year growth rates of human capital index, BL versus CSL data**


[1] The assumed rates of return are from Psacharopoulos (1994) as implemented by (e.g.) Caselli (2005).

To that end we set out to determine the most appropriate series on average years of schooling for every country. We apply several heuristic principles in determining which series to use:

1. Only the BL and CSL data cover a sufficient period of time for use as the main data source, so we choose between these two sources whenever both are available. There are 56 countries covered by BL but not by CSL and 5 countries covered by CSL but not by BL, so for 61 countries we choose the only available series by default.
2. We choose the BL or CSL series if either is closer in terms of the level or trend over time to the DD and/or the UNESCO data.
3. If neither series is available as a cross-check, the trend in years of schooling is compared to enrolment data from UNESCO. If the average years of schooling starts increasing rapidly before enrolment rates start to increase rapidly, this counts against the BL or CSL series.
4. The BL series shows declines in the average years of schooling at higher (though still very low) rate than the CSL data. If a series for a country shows such a decline, that counts against (typically) the BL series.
5. If the other criteria provide no clear basis for preference of one series over the other, we choose the BL series as its data start in 1950 versus 1960 for the CSL data.

The result of applying these rules is that PWT 9.0 will include human capital data for 150 countries. For 95 countries, data are based primarily on BL, for the other 55 countries the data are based primarily on CSL. The appendix lists the choice for each country and provides the motivation for that choice.

The BL data are available only once every 5 years and the CSL data only once every 10 years. In addition, the CSL data are not available before 1960, while 2010 is the latest year in the BL data; the CSL data include projections up to 2020. Regardless of the chosen series, we interpolate linearly between observations. If CSL data are used, the trend in average years of schooling for the period 1950-1960 is used from the BL data (if available) to extrapolate. If BL data are used, the trend in average years of schooling for the period 2010-2020 is used from the CSL data to extrapolate. If CSL data are not available, the trend growth from 2005-2010 from the BL data is used to extrapolate until 2014, the last year covered in PWT 9.0. An analogous procedure is also followed for the

few countries in the CSL data without projections to 2020, but then using the 2000-2010 growth from the CSL data.

## References


Barro, Robert J. and Jong-Wha Lee (2013), "A new data set of educational attainment in the world, 1950-2010" *Journal of Development Economics* 104: 184–198.

Caselli, Francesco (2005), "Accounting for cross-country income differences" in Phillipe Aghion and Steven N. Durlauf (eds.) *Handbook of Economic Growth, Volume 1A*, Elsevier: 679–741.

Cohen, Daniel and Marcelo Soto (2007), "Growth and human capital: good data, good results" *Journal of Economic Growth* 12(1): 51–76.

Cohen, Daniel and Laura Leker (2014), "Health and Education: Another Look with the Proper Data", *mimeo* Paris School of Economics.

De la Fuente, Angel, and Rafael Doménech (2006), "Human capital in growth regression: How much difference does quality data make?" *Journal of the European Economic Association*, 4(1): 1–36.

Psacharopoulos, George (1994), "Returns to investment in education: A global update" *World Development* 22(9): 1325–1343.

## Appendix: choice of source by country

| ISO code | Country | Source | Argument |
|---|---|---|---|
| ALB | Albania | BL | only source |
| DZA | Algeria | CSL | enrolment takes off in the 1970s, suggesting BL trends are too optimistic |
| AGO | Angola | CSL | only source |
| ARG | Argentina | BL | closer to UIS in level |
| ARM | Armenia | BL | only source |
| AUS | Australia | CSL | DD matches CSL trend |
| AUT | Austria | CSL | DD matches CSL trend and closer in levels |
| BHR | Bahrain | BL | only source |
| BGD | Bangladesh | BL | UIS supports BL level over CSL |
| BRB | Barbados | BL | only source |
| BEL | Belgium | BL | similar series in CSL, but BL is longer |
| BLZ | Belize | BL | only source |
| BEN | Benin | BL | hard to decide between BL and CSL based on enrolment data |
| BOL | Bolivia | CSL | levels closer to UIS, but similar trend as BL |
| BWA | Botswana | BL | only source |
| BRA | Brazil | BL | decline in 2010 in CSL and BL and UIS show (short) similar trend |
| BRN | Brunei Darussalam | BL | only source |
| BGR | Bulgaria | CSL | trends is more steady than BL, but fairly similar |
| BFA | Burkina Faso | CSL | only source |
| BDI | Burundi | CSL | enrolment takes off in the 1980s, suggesting BL trends are too optimistic |
| KHM | Cambodia | BL | only source |
| CMR | Cameroon | CSL | UIS closer to CSL |
| CAN | Canada | CSL | BL shows several declines; CSL closer in level to DD |
| CAF | Central African Republic | CSL | enrolment takes off in the 1980s, suggesting BL trends are too optimistic |
| CHL | Chile | CSL | slightly smoother trend than BL, but very similar |
| CHN | China | BL | similar series in CSL, but BL is longer; BL closer to UIS for single observation |
| HKG | China: Hong Kong SAR | BL | level similar to UIS |
| MAC | China: Macao SAR | BL | only source |
| COL | Colombia | CSL | UIS closer to CSL |
| COG | Congo | BL | only source |
| CRI | Costa Rica | BL | UIS closer to BL |
| CIV | Côte d'Ivoire | CSL | enrolment takes off in the 1990s, suggesting BL trends are too optimistic |
| HRV | Croatia | BL | only source |

## Appendix: choice of source by country

| ISO code | Country | Source | Argument |
|---|---|---|---|
| CYP | Cyprus | CSL | BL shows declines (though CSL may underestimate level given UIS) |
| CZE | Czech Republic | BL | only source |
| COD | D.R. of the Congo | BL | only source |
| DNK | Denmark | CSL | BL shows declines and a low level compared to DD and UIS |
| DOM | Dominican Republic | BL | levels closer to UIS, broadly similar trend |
| ECU | Ecuador | CSL | slightly smoother trend than BL, but fairly similar |
| EGY | Egypt | CSL | slightly smoother trend than BL, but fairly similar |
| SLV | El Salvador | CSL | closer to UIS in level |
| EST | Estonia | BL | only source |
| ETH | Ethiopia | CSL | only source |
| FJI | Fiji | CSL | decline in BL data |
| FIN | Finland | CSL | smoother trend and closer to DD (in trend and level) than BL |
| FRA | France | CSL | smoother trend and closer to DD (in trend and level) than BL |
| GAB | Gabon | BL | hard to decide between BL and CSL; longer series for BL |
| GMB | Gambia | BL | only source |
| DEU | Germany | CSL | CSL close to DD and UIS, BL is too low and distorted trend |
| GHA | Ghana | BL | trend is similar, though higher in BL; longer time series |
| GRC | Greece | CSL | BL shows declines in early years, levels and trends similar from 1970s onwards |
| GTM | Guatemala | BL | CSL short time series, few UIS observations; both similar to BL |
| HTI | Haiti | CSL | enrolment takes off in the 1980s, suggesting BL trends are too optimistic |
| HND | Honduras | CSL | level similar to UIS |
| HUN | Hungary | CSL | BL shows declines and a high level compared to UIS |
| ISL | Iceland | BL | only source |
| IND | India | BL | similar series in CSL, but BL is longer |
| IDN | Indonesia | CSL | level similar to UIS |
| IRN | Iran (Islamic Republic of) | CSL | enrolment takes off in the 1980s, suggesting BL trends are too optimistic |
| IRQ | Iraq | BL | similar trend in CSL, but BL is longer |
| IRL | Ireland | CSL | closer to DD in level |
| ISR | Israel | BL | only source |
| ITA | Italy | CSL | very similar series, but CSL closer to DD |

## Appendix: choice of source by country

| ISO code | Country | Source | Argument |
|---|---|---|---|
| JAM | Jamaica | CSL | enrolment more consistent with flattening profile around 2000 rather than BL's acceleration |
| JPN | Japan | CSL | decline in BL data, CSL closer to DD |
| JOR | Jordan | BL | hard to decide between BL and CSL; longer series for BL |
| KAZ | Kazakhstan | BL | only source |
| KEN | Kenya | CSL | level similar to UIS, but similar trends in CSL and BL |
| KWT | Kuwait | BL | only source |
| KGZ | Kyrgyzstan | BL | only source |
| LAO | Lao People's DR | BL | only source |
| LVA | Latvia | BL | only source |
| LSO | Lesotho | BL | only source |
| LBR | Liberia | BL | only source |
| LTU | Lithuania | BL | only source |
| LUX | Luxembourg | BL | only source |
| MDG | Madagascar | CSL | only source |
| MWI | Malawi | CSL | level similar to UIS |
| MYS | Malaysia | BL | similar series in CSL, but BL is longer |
| MDV | Maldives | BL | only source |
| MLI | Mali | BL | hard to decide between BL and CSL; longer series for BL |
| MLT | Malta | BL | only source |
| MRT | Mauritania | BL | only source |
| MUS | Mauritius | CSL | decline in BL data |
| MEX | Mexico | CSL | level similar to UIS, but similar trends in CSL and BL |
| MNG | Mongolia | BL | only source |
| MAR | Morocco | BL | hard to decide between BL and CSL; longer series for BL |
| MOZ | Mozambique | BL | short CSL time series |
| MMR | Myanmar | BL | similar series in CSL, but BL is longer |
| NAM | Namibia | BL | only source |
| NPL | Nepal | BL | hard to decide between BL and CSL; longer series for BL |
| NLD | Netherlands | CSL | CSL close to DD and UIS, BL is too low and distorted trend |
| NZL | New Zealand | BL | closer to DD in level, despite some declines in BL |
| NIC | Nicaragua | BL | similar series in CSL, but BL is longer |
| NER | Niger | CSL | enrolment rates seem more consistent with CSL levels |
| NGA | Nigeria | CSL | only source |

## Appendix: choice of source by country

| ISO code | Country | Source | Argument |
|---|---|---|---|
| NOR | Norway | CSL | CSL close to DD and UIS, BL is too low and shows decline |
| PAK | Pakistan | BL | only source |
| PAN | Panama | CSL | closer to UIS in level |
| PRY | Paraguay | BL | closer to UIS in level |
| PER | Peru | BL | similar series in CSL, but BL is longer |
| PHL | Philippines | BL | similar series in CSL, but BL is longer |
| POL | Poland | BL | only source |
| PRT | Portugal | BL | closer to DD in level |
| QAT | Qatar | BL | short UIS time series |
| KOR | Republic of Korea | CSL | decline in BL data |
| MDA | Republic of Moldova | BL | only source |
| ROU | Romania | BL | closer to UIS in level |
| RUS | Russian Federation | BL | only source |
| RWA | Rwanda | BL | only source |
| SAU | Saudi Arabia | BL | only source |
| SEN | Senegal | CSL | closer to UIS in level |
| SRB | Serbia | BL | only source |
| SLE | Sierra Leone | BL | similar series in CSL, but BL is longer |
| SGP | Singapore | BL | similar series in CSL, but BL is longer |
| SVK | Slovakia | BL | only source |
| SVN | Slovenia | BL | short UIS time series |
| ZAF | South Africa | CSL | decline in BL data |
| ESP | Spain | CSL | levels closer to UIS (late sample) and similar level and trend as DD |
| LKA | Sri Lanka | BL | only source |
| SDN | Sudan | BL | similar series in CSL, but BL is longer |
| SWZ | Swaziland | BL | only source |
| SWE | Sweden | CSL | closer to DD and UIS in level |
| CHE | Switzerland | CSL | BL much lower than CSL and DD, plus decline in BL series |
| SYR | Syrian Arab Republic | CSL | enrolment suggests continued increases in attainment, rather than flattening |
| TWN | Taiwan | BL | only source |
| TJK | Tajikistan | BL | only source |
| TZA | Tanzania | CSL | enrolment more consistent with flat profile before 1990 |
| THA | Thailand | CSL | decline in BL data |
| TGO | Togo | BL | only source |
| TTO | Trinidad and Tobago | BL | similar series in CSL, but BL is longer |
| TUN | Tunisia | BL | hard to decide between BL and CSL; longer series for BL |
| TUR | Turkey | BL | similar series in CSL, but BL is longer |

## Appendix: choice of source by country

| ISO code | Country | Source | Argument |
|---|---|---|---|
| UGA | Uganda | BL | hard to decide between BL and CSL; longer series for BL |
| UKR | Ukraine | BL | only source |
| ARE | United Arab Emirates | BL | only source |
| GBR | United Kingdom | CSL | levels closer to UIS (late sample) and similar level and trend as DD |
| USA | United States | BL | closer to DD and UIS in level |
| URY | Uruguay | BL | closer to UIS in level |
| VEN | Venezuela | BL | very similar series, but BL is longer |
| VNM | Vietnam | BL | only source |
| YEM | Yemen | BL | only source |
| ZMB | Zambia | BL | very similar series, but BL is longer |
| ZWE | Zimbabwe | BL | very similar series, but BL is longer |

Notes: BL: Barro-Lee (2013) data; CSL: Cohen-Soto (2007)/Cohen-Leker (2014) data; UIS: UNESCO data; DD: De la Fuente and Doménech (2006) data; enrolment data through World Bank Education statistics. Listed are the 150 countries covered in PWT 9.0 and either the BL or CSL data.