Accumulation of human capital

in developing regions.

Guido van Steen[1]

CPB Netherlands Bureau for
Economic Policy Analysis.

24-05-2000

[1]The author wants to thank Arie ten Cate, Nico van Leeuwen, Arjan Lejour and Hans Timmer for their help, although he alone wishes to be responsible for all the remaining errors.

DRAFT VERSION: NOT TO BE QUOTED!

Contents

DRAFT VERSION: NOT TO BE QUOTED!

1. Introduction

One of the main determinants of a nation's wealth is its amount of accumulated human capital. Frederick Harbison once said it like this: '[A] country which is unable to develop the skills and knowledge of its people and to utilize them effectively in the national economy will be unable to develop anything else.' This paper discusses schooling and human capital formation in developing regions. It proposes a method to construct a stock of human capital essentially based on enrollment and population data. We will present some of the results, which will be used in WorldScan, a global model intended for the analysis of trade patterns between different regions.[2] The core version of the model divides the world into 12 regions, some of them being relatively rich, others much poorer. In the current version of WorldScan (see CPB, 1999) schooling and human capital formation are modelled in a rather aggregated way. This makes it hard to analyse the effects of schooling policies. The human capital stocks constructed in this paper will enable a better assessment of potential policy effects.

2. Modelling human capital accumulation

Policy makers can influence the formation of human capital mainly through schooling. The latter concept cannot easily be grasped in numbers, due to its qualitative nature. Quality differences between schooling systems are often hard to measure. Different countries may have school days of unequal length. Class sizes may vary as well as the qualifications of teachers. Moreover, students are not taught the same curriculum in every region.
In many countries schooling extends not only to young people who attend school but also to workers in their professional careers. In the latter phase the schooling process is even harder to monitor. What, for instance, is the importance of 'on-the-job training' or 'learning-by-doing'. How do refresher courses contribute to the knowledge of professional workers? Questions like these are difficult to answer, if only because of a lack of empirical data. Moreover, different cultural backgrounds may cause varying attitudes towards certain aspects of learning. An example is the respect for the elderly, which has had important consequences for the educational systems of some cultures.

It is equally hard to monitor human capital in a such way that differences between countries can be measured consistently. Experienced farmers may be redundant in Singapore, whereas they could be highly productive in Canada. Even within countries it is difficult to track the value of human capital stocks. Depreciation of human capital depends strongly on a country's economic circumstances. If the structure of the economy changes slowly, the depreciation of human capital may be small. Contrarily, quick and substantial structural changes call for high depreciation rates. These measurement problems mentioned notwithstanding, we will define human capital to consist of accumulated years of schooling. It depreciates by a mechanism to be specified later.

WorldScan currently describes the skill level by dividing the workforce into a high-skilled and a low-skilled part. This split is based on data provided by Barro and Lee (1996) and Ahuja and

[2] WorldScan has been developed at CPB Netherlands Bureau for Economic Policy Analysis.

Filmer (1995). CPB (1999) describe a procedure that yields a share of high-skilled workers in the workforce for each of the twelve WorldScan regions. People are classified high-skilled if they have at least completed secondary education. The high-skilled part of the population in each of the various regions may vary by the year.
However, this procedure does not fully take account of an important characteristic of human capital. The largest part of it is formed while people are young, the main reason for this being the longer pay-back period that young people expect ahead of them. This makes their *investment* more cost-effective. Besides, it might be argued that people are more flexible learners when they are young. By consequence, the age-structure of the population plays an important role in the dynamics of human capital formation.
Non-OECD countries are often characterized by relatively large shares of younger cohorts. This is caused by high birth and death rates. These nations can therefore, in principle, upgrade their populations' skill levels swiftly. The dialectics of progress may be reinforced because older people, who may be more inclined to block technological progress, form a larger share of the population in the OECD. (*cf.* Canton *et al.*, 1999). On the other hand the potential of education in non-OECD regions is partially offset by the limited means available for schooling, both directly and indirectly. Governments and households in these countries often lack sufficient income for large investments in education.

The share of the high-skilled in the workforce is currently exogenous in WorldScan. It is thus independent of the age structure of the population. We want to be able to incorporate some of the ideas mentioned above in WorldScan in a more explicit way.[3] Detailed information is necessary for this purpose. We will use a method that in fact follows the enrollment records of each one-year cohort at every one-year interval. We will for example approximate the share of the Angolans that will be 56 in 2008 who were still enrolled at the age of twelve. As another example we will try and assess how many of Indonesians who were 38 twelve years ago, were still enrolled when they were 15.

Such data can be used to construct indicators of human capital in two ways. First, they could be used to construct a mean number of years that a representative worker has spent at school. Second, one could compute a high-skilled ratio, like the one currently used in WorldScan. In this approach everyone who attained a specific level of schooling would be classified high-skilled.

The first method yields continuous variables. Let h_i be the mean number of years that a cohort born in year i spent at school. If N_i is the number of people in cohort i, then H_i, the total number of years that cohort i has spent at school, can be written as:

$$H_i = h_i * N_i \qquad (1)$$

where h_i is the age-specific number of years spent at school. The mean number of year spent at school for the workforce is then:

[3] CPB (1999) does take account of the age structure in an implicit way, because they use the Ahuja and Filmer data set.

$$h_T = \Sigma H_i / \Sigma N_i \qquad (2)$$

Information necessary for equation (1) may be unavailable for some generations. If information on a specific year or on a specific cohort is missing, one might determine the schooling level as:

$$h_{i,t} = h_{i-1, t-1} + u_{i,t} - d_{i,t} + s_{i,t} \qquad (3)$$

Recursive application of this formula leads to expressions like $h_{i,t} = h_{i-2, t-2} + (u_{i,t} + u_{i-1,t-1}) - (d_{i,t} + d_{i-1,t-1}) + (s_{i,t} + s_{i-1,t-1})$. In (3) $u_{i,t}$ denotes knowledge formation and upgrading at time t. $d_{i,t}$ is the depreciation of human capital and $s_{i,t}$ is the effect of demographic factors on the mean years spent at school. As we noted before, $u_{i,t}$ is high for young generations, due to a higher expected present value of schooling. Besides, people belonging to older generations are often less flexible. They may for example have a family and therefore may not be able to afford the expenses necessary for schooling. Therefore, for most older cohorts the value of $u_{i,t}$ will be small compared to $h_{i-1, t-1}$. Moreover, $u_{i,t}$ is compensated by the term $d_{i,t}$, which increases as generations get older. As we saw before, this effect becomes stronger when the structure of the economy changes more rapidly. However, relatively high depreciation rates may endogenously be compensated by a workforce that takes more refresher courses. The factor $s_{i,t}$ reflects the effect of the composition of the workforce. It will have a positive effect on the average education level, because higher levels of education will later lead to higher incomes, which enable educated people to live longer (Najman, 1993). This effect will lead to an increase of a cohort's mean years of education as the cohort gets older.
Thus $s_{i,t}$ and $u_{i,t}$ tend to raise the skill level, whereas $d_{i,t}$ causes a tendency the other way around. The net effect is therefore hard to predict. For the sake of simplicity one might assume that for grown-up cohorts that $u_{i,t} - d_{i,t} + s_{i,t}$ is equal to 0, so that equation (3) reduces to:

$$h_{i,t} = h_{i-1, t-1} + e_{i,t} \qquad (4)$$

Here $e_{i,t}$ is the part of $u_{i,t}$ that is caused by formal education only. This makes it possible to concentrate on the enrollment levels of young cohorts in the population, which after all remain the most important force behind future human capital formation. Another advantage of this approach is that hard information about $u_{i,t}$, $d_{i,t}$ and $s_{i,t}$ is very difficult to obtain, whereas information about school enrollment $e_{i,t}$ has been recorded relatively well.

The individual values of $h_{i,t}$ can also be used to compute high-skilled ratios, which is in line with the second approach to compiling human capital indicators. One may define for example those workers that at least attended school until their 15th birthday as high-skilled. If one sums these workers over all the age cohorts in the workforce and divides this number by the total workforce again, one obtains a high-skilled ratio. Although this criterion is not exactly equal to the secondary-education-completed criterion used in WorldScan, it does give an indication of the share of the population that is relatively high-skilled.

This second approach is the approach allows an easier comparison to the methodology currently followed in WorldScan. We will pursue it in the rest of this paper. Applications of the first methodology, which is based on mean years of schooling, will be postponed to later occasions.

3. Data

Many researchers have investigated stocks of human capital. A few important studies like Ahuja and Filmer (1995) and Barro and Lee (1996) are based on the publication UNESCO (1993), which contains trends of enrollment data.[4] These enrollment data are reported as shares of age groups. They were compiled for 108 countries in developing regions. The trends in the report consist of both realisations and projections. They include observations from the sixties until the beginning of the nineties. For most countries the trends are extrapolated for the period between 2000 and 2015. The data are not complete for all years. For many countries data points are reported for the years 1960, 1980, 1990, 2000, 2005 and 2015 only. In more favourable cases the year 1991 or 1992 is also included. For a few countries, like Liberia and Cambodia, no projections are given at all. They consist of merely 2 or 3 data points.
Enrollment rates are mostly available for three aggregated age groups: the first contains the cohorts between 6 and 11, the second those between 12 and 17, and the third those between 18 and 23 old. These respective age groups are assumed to be enrolled in primary, secondary and in tertiary education. Each of the age groups in UNESCO (1993) thus spans 6 years.[5]

As we said in the last section we would like to follow these cohorts in more detail. One may for example require the expected enrollment rate in primary education in Peru in 1999. If we have no other prior information our best guess (or interpolation) would be 2000's expectation, perhaps slightly biassed towards the observations made in the beginning of the nineties. In order to impute all missing points we will apply the so-called cubic-splines technique. This method, which is often encountered in the literature, draws cubic polynomials between all pairs of neighbouring data points.[6] [7]

Cubic splines have some disadvantages. A major drawback is that one cannot guarantee that all interpolations fall within a range considered valid. In our case we regard the numbers between 0 and 100 as the valid range for enrollment rates. Without precautions cubic splines may not respect such boundaries. For instance, in the case of three consecutive observations 91%, 99%, 99%, a cubic spline may easily assign a percentage above 100% between the second and the third

[4]UNESCO (2000), an update of UNESCO (1993), will be available soon.

[5]In some cases the actual age borders reported in UNESCO (1993) are slightly different from the description here. We have not taken these exceptions into account.

[6]A pair of neighbouring data points is a pair of data point without any other data points between them.

[7]Technically, each non-boundary data point x is intersected by two functions f and g, in such a way that $f'(x) = g'(x)$ and $f''(x) = g''(x)$. Together with the obvious condition '$f(x) = g(x)$' this yields enough information to determine f and g as cubic polynomials. For the boundary points one may use the DeBoor's so-called not-a-knot condition (DeBoor, 1978), which requires that the first two line segments are described by the same cubic polynomial. A similar requirement applies to the two last line segments. In this paper we will use DeBoor's method.

observation. To avoid such complications one can apply a suitable monotonic (*e.g.* logit) transformation from the allowed range into the real line. After splining the transformed series, one can reconstruct the original scale by applying an inverse transformation. This guarantees that interpolated values falls within the feasible set.

UNESCO (1993) contains no information on enrollment rates after 2015. DeBoor (1978) warns against the use of splining methods when it comes to the extrapolation of trends. In this paper we will make all projections that to be extrapolated equal to the last available UNESCO (1993) data point. Past observations before the interpolation period will be made equal to the first available UNESCO (1993) data point.

The procedure just described provides enrollment rates in every year, for primary, secondary and tertiary education, in 108 developing nations. A next step is to find cohort-size statistics that can be matched with these enrollment rates. Population statistics from UN (1994) provide such data, organised in age cohorts that mostly span 5 years.[8] They are available for each year between 1950 and 2050.[9] They can, however, not be matched with UNESCO (1993), since the latter source reports age groups that span 6 years. Explicit information about the underlying one-year cohorts in UN (1994) is not available. We will therefore again construct the missing information through the use of splining methods.

A complication arises if we analyse cohort data, because interpolated age cohorts do not necessarily add up to the original values. Therefore, in the case of cohort data it is advisable to transform the original series into a cumulative counterpart. This ‘integration’ process ensures that the interpolations, once they are ‘differentiated’ again, will satisfy the cumulation constraint. In this way the UN population statistics, which originally span 5 years, can be transformed in age cohorts that span one year.[10]

These data can in turn be aggregated into cohorts spanning 6 years of age, which can be multiplied by our splined enrollment rates. In this way we construct cohorts enrolled in primary, secondary and secondary education. In practice people rarely reverse the decision to stop going to school. Therefore one may expect that enrollment rates are inversely related to the age of the cohort. This is confirmed in the data on the aggregated age groups of UNESCO (1993). We assume that before the age of 6 nobody is enrolled. On top of that we know the numbers of people enrolled between 6 and 12 and between 12 an 18. Therefore, we have three observations of the cumulative number of children at school in the age group between 6 and 18.[11] We can

[8]The last age cohort represents those older than 80, and thus has no upper border. Instead of infinity we chose 88 as a technical upper limit.

[9]Recently a new data set for the year 1998 became available. We will update our calculations as soon as UNESCO (2000) is published.

[10]There are data on single-year age groups between 5 and 24 in UN (1994). These data do not differ in significant way from the single-year age group data that we construct.

[11]The number of people older than 18 who are still enrolled is negligible in less developed

spline them into 13 consecutive cumulative observations, which amounts to drawing a parabola through the three available points. Once these series is 'differentiated', it yields 12 age groups each spanning one year. A restriction here is that the size of the enrolled cohort must not exceed the total cohort size. One may verify this on a case-by-case basis and apply corrections if necessary. Dividing these enrolled numbers by the corresponding full cohort sizes generates 12 separate enrollment rates.

One may be particularly interested in *productive* human capital. For this one's attention could be restricted to the economically active. The size of this part of the population may vary according to the definition used. In Western eyes an age range between 16 and 65 may sound reasonable. In many developing countries, however, both younger and older age cohorts often contribute their labour services to the economy as well. Cohort-specific participation rates can be obtained from ILO (1996). They can be interpolated to yield single-year cohorts as well. In this paper we will do so, and define the labour force as those in between 10 and 70 years old. A problem that we do not take into account is the interaction between age-specific participation and skill levels. We will postpone such a modelling exercise until a future research project.

4. Results

WorldScan divides the world is into 12 regions. Eight of them can in some way be called developing regions. These regions are the Eastern Europe (EAE), the Former Soviet Union (FSU), the Middle East and North Africa (MNA), Sub-Saharan Africa (AFR), Latin America (LAT), China (CHI), South East Asia (NIC), and South Asia (IND). Neither Ahuja and Filmer (1995) nor UNESCO (1993) give projections for the former socialist countries in the EAE and FSU regions. Therefore, these regions are excluded from our computations. In our base calculations we define a person to be high-skilled if he has at least attended school until his 14th birth day. For the 6 WorldScan regions this yields the following results:

Table 1. Percentage of the population that has more than 8 years of schooling

	MNA	AFR	LAT	CHI	NIC	IND
2000	45.6%	30.1%	55.3%	50.9%	46.2%	28.1%
2010	53.3%	31.4%	59.2%	49.9%	51.6%	32.4%
2020	59.1%	32.2%	62.8%	49.1%	56.3%	35.9%

MNA = Middle East and North Africa, AFR = Sub-Saharan Africa LAT = Latin America, CHI = China, NIC = South East Asia, IND = South Asia.

The current skill level of the workforce in Sub-Saharan Africa and India is low. Based on our selection of countries the South-Asian region has the worst performance of the 6 regions listed.[12] Current Latin American and Chinese skill levels are relatively advanced. Latin America's record

countries.

[12]There is a large contrast between the skill levels of the average Indian worker and those of the Indians who migrate to the USA.

is probably due to its past wealth and European immigrants. After World War II China put much emphasis on the education of its population. However, its performance could have been much better if it had not experienced the so-called Cultural Revolution, which oppressed any kind of intellectualism during the seventies. The MNA region, which includes some important oil exporting countries, came from a relatively low position. The importance of oil for the world economy has increased enormously. A similar story applies to the economic relevance of the South East Asian region. This has enabled both regions to invest relatively large amounts in the education of younger people.

A priori one may expect high-skilled ratios to rise in most regions, because through the years relatively well-educated young cohorts will replace relatively low-skilled old cohorts. The differences between 2000 and 2010 and those between 2000 and 2020 are listed in Table 2:

Table 2. Growth of the percentage of the population that has more than 8 years of schooling

	MNA	AFR	LAT	CHI	NIC	IND
2000-2010	7.7	1.3	3.8	-1.0	5.4	4.3
2000-2020	13.5	2.0	7.5	-1.8	10.1	7.8

MNA = Middle East and North Africa, AFR = Sub-Saharan Africa LAT = Latin America, CHI = China, NIC = South East Asia, IND = South Asia.

Table 2 confirms this expectation of rising skill levels in 5 cases. However, it also shows rather large variations between the regions. China is an exception to the rule: its high-skilled ratio is expected to decrease. The main force behind this is the Cultural Revolution, which caused Chinese enrollment rates to plummet in the seventies. Moreover, UNESCO (19993) does not expect enrollments in China to increase significantly until 2015. Africa's prospects are not bright either. Due to civil wars and other disasters enrollment rates have not increased much since the eighties. Nevertheless a modest increase can be expected in this region. This is mainly due to the growth of the continent's population. Regions with moderately increasing high-skilled ratios are IND and LAT. Latin America remains the most skilled region listed. South Asia will surpass Africa, but this does not mean that this region can be satisfied with the future educational attainments of its workforce. The eventual skill level of the workforce remains extremely low by international comparison. Both IND and LAT will see their high-skilled ratios increase by approximately 4 points every ten years. The past investments in education in the NIC region is expected to result in a 5 points increase during the next 10 years. Within the MNA region the high-skilled ratio even increases by 7 points.

Another criterion would have lead to different high-skilled ratios. Nevertheless, the over-all picture would not have been very different. For example, if we had defined everybody who leaves school before his 16th birthday to be low-skilled, then Table 2 would look like this:

Table 3. Growth of the percentage of the population that has more than 10 years of schooling

	MNA	AFR	LAT	CHI	NIC	IND
2010	8.4	1.8	3.3	-2.5	5.6	4.2
2020	13.5	2.4	5.9	-7.3	10.5	7.6

MNA = Middle East and North Africa, AFR = Sub-Saharan Africa LAT = Latin America, CHI = China, NIC = South East Asia, IND = South Asia.

The growth in the high-skilled ratios in Table 2 and 3 is more or less the same. The only exception again is China, where the drop in skill levels seems even more serious now. This can be explained because the Cultural Revolution had more impact on secondary than on primary education.

The trends discussed so far were constructed under the assumption that enrollment levels develop in a way according to statistics reported in UNESCO (1993). An interesting question is how the eventual skill levels of the workforce would change if all regions made an extra effort resulting in an extra increases of their enrollment rates by 5%. The effects of successfully carrying out such a policy should be a 5% increase in the long run. If the size of each of the cohorts were the same, such a long run should consist of 60 years, since in our computations the workforce consists of 60 one-year cohorts. Following this line of reasoning we would expect a 1.25 increase after 15 years and 2.1 after 25 years. However, even when the population size is stable, young cohorts will normally be then older ones. High birth and death rates will accentuate make this difference. In this case large relatively well-educated, cohorts of young people will replace smaller cohorts of retirees. This results in a faster upgrading of the skill level of the workforce. We simulated the following values:

Table 4. Growth of the percentage of the population that has more than 10 years of schooling if all enrollment ratios become 5% higher after 1995.

	MNA	AFR	LAT	CHI	NIC	IND
2010	2.0	2.3	1.8	1.5	1.7	1.9
2020	2.8	3.2	2.5	2.1	2.5	2.7

MNA = Middle East and North Africa, AFR = Sub-Saharan Africa, LAT = Latin America, CHI = China, NIC = South East Asia, IND = South Asia.

As predicted after 15 years the effect is above 1.25 points in all regions. In Africa population growth is expected to be higher than anywhere else. This explains why the effect of increasing enrollments is the highest in Africa. The relatively low value of 1.5 for China can be attributed to its one-child policy. Most results for 2020 are above 2.1. China is an exception again. This extreme result indicates that cohort size is not a decreasing function of age. This is also a consequence of the Chinese one-child policy. The relative distance between the ‘normal’ values 1.25 and 2.1 on the on hand and the simulation results on the other hand decreases. This reflects the expectation that population growth will slow down between 2010 and 2020.

5. Conclusions

The most important result in this paper is that the skill level of the workforce within most developing regions is expected to rise. This is caused by young, relatively well-educated, large cohorts that replace smaller old cohorts with less education. This effect is reinforced by past and current population growth. China is an exceptional case. The Cultural Revolution has depressed the education level attained by many of the younger cohorts. Moreover, its one-child policy has stabilized the population size.
Another important finding in this paper is that a policy directed towards the expansion of the formal education has modest effects in the short run. Less than 50% of the workforce will have benefited from such a policy, even after conducting it for 15 years. However, this does not make access to formal education less important for developing regions. Especially in countries where those below 18 form a large share of the population the accumulation of human capital due to rising enrollments will be relatively fast. Moreover, the increase in human capital currently expected is mainly due to investments in education that have been made since the fifties. If countries want their human capital stock to rise until 2050, they should make investments now.

6. References

Ahuja, V., and D. Filmer, "Educational attainment in developing countries: new estimates and projections disaggregated by gender", *background paper for the World Development Report*, 1995, The World Bank Washington D.C.

Barro, R.J., and J.W. Lee, "International measures of schooling years and schooling quality", *American Economic Review,* 1996. p. 363-394.

Canton, E., H. de Groot and R Nahuis, "Vested interests and resistance to technology adoption", *CentER Discussion Paper*, No. 99106, 1999, Tilburg University.

CPB Netherlands bureau for economic policy analysis, "WorldScan, the core model", 1999, The Hague.

DeBoor, C., "A Practical Guide to Splines", 1978 New York.

ILO International Labour Organisation, "The Economically active population, 1950-2010", 1996, Geneva.

Najman JA. "Health and puberty: past, present and prospects for the future" Social Science and Medicine, 1993, p. 157-66.

United Nations, "World Population Prospects, the 1994 Revision", 1994, New York.

UNESCO, "Trends and Projections of enrollments by level of education, by age and by sex, 1960-2025", 1993, Paris.

UNESCO Institute for Statistics, "Enrollment projections as assessed in 1998-1999", 2000, Paris.