3. Basic data and measurement issues: standard proxy estimates of human capital

1. INTRODUCTION

It is important, before turning to the historical and economic analyses of education and human capital, to start with brief overview of the available data and problems with their collection, interpretation, and estimation for India, Indonesia, and Japan. The data as treated in this chapter are mainly the basic, non-transformed data. These can essentially be divided in the number of children enrolled in education and the expenditure on education. We will start in the following section with an overview of the enrolment figures. In addition, we will also present the gross enrolment ratio (the number of children enrolled at a certain level of education, divided by the relevant age-group). In section 3 we continue with two related variables, namely attainment (the percentage of the population of 15 years and older with a certain education level) and the average per capita years of education in the population. In section 4 we turn to the government and private expenditure on education. The interpretation of the figures presented here and the more demanding estimation of an alternative stock of human capital that conforms to the definition presented in the previous chapter will be treated in chapters 4 and 5 respectively.

2. ENROLMENT DATA

2.1 Definitions and sources

Educational enrolment figures are at the start of many analyses of education systems. They are also the point of departure for almost all more demanding estimates of indicators of human capital. They indicate the number of persons enrolled at a certain level of education in a certain year. As these figures are of interest for national and international governments from a policy making and a budgetary point of view, these data were among the first ones to be collected when national statistical bureaus started their work. When international organizations such as the United Nations were founded, they also soon started to collect these statistics and make them comparable. Within the United Nations this is specifically done by the UNESCO (United Nations Economic and Social Committee).

These data are thus relatively abundantly available in both national and (since the 1950s) international statistical publications (see table 3.1). However, some

Table 3.1: General sources on educational enrolments used in this study with statistics of Japan, India and Indonesia, 1880-2000*

Organisation	Source	Country	Time
Centraal Bureau voor de Satistiek (CBS)	Jaarcijfers voor het Koninkrijk der Nederlanden: koloniën/Statistisch Jaaroverzicht Nederlandsch-Indië	Indonesia	1893-1930
Centraal Kantoor voor de Statistiek (CKS)	Algemeen Verslag van het Europeesch en Inlandsch Onderwijs	Indonesia	1880-1914; 1915-1938
-	Colonial Report	Indonesia	1880-1929
Centraal Kantoor voor de Statistiek (CKS)	Indisch Verslag	Indonesia	1931-1939
Hollandsch-Inlandsche Onderwijs Commissie (HIS)	Report of the Dutch Indies Education Commission, no.2, 3, and 10.	Indonesia	1930
Badan/Biro Pusat Statistik (BPS)	Statistical Yearbook of Indonesia	Indonesia	1976-2000
Department of Commerical Intelligence and Statistics	Statistical Abstract for British India	India	1880-1945
Central Statistical Organisation (CSO)	Statistical pocketbook India	India	1969-1990
Central Statistical Organisation (CSO)	Statistical Abstract India	India	1950-2000
Bureau of Statistics	Japan Statistical Yearbook	Japan	1980-2000
Statistical Bureau, Japan Statistical Association	Historical Statistics of Japan (vol. 5)	Japan	Circa 1870-1985
UNESCO	UNESCO Statistical Yearbook	International	1964-1999
International Historical Statistics: Africa and Asia	Mitchell	International	Circa 1850-1988

* This excludes stray statistics from individual publications. Those can be found, however, in the references to this chapter.

difficulties remain with their collection. First, during colonial rule, often no statistics on the indigenous education system were collected. Second, private schools which were not eligible for government subsidies were often ignored in the statistics. Of course, in many cases the previous two categories concerned the same schools. Third, there are problems in making the enrolment figures comparable over time and among countries. For example, practical (or vocational) education, if given at all, was often taught at the primary level at the start of the twentieth century while after 1950 it could mainly be found at the secondary level.

To counter the problem of comparability of these data, we took into account for all countries the number of pupils enrolled in the public and the subsidized and/or recognized private schools. This is especially important in India and Indonesia because in those countries there was a difference in education system between the (often subsidized) European private education and the often non-subsidized private indigenous education. In Japan, due to its more homogenous population, this problem was less pressing and all school types were included in the statistics. However, the so-called ‘wild’ or ‘unrecognized private’ schools (the non-subsidized non-recognized

schools) in India and Indonesia have been left aside. This has two reasons.[55] First, it is almost impossible to get enough data on their numbers. Some very crude estimates are, however, possible. In the next chapter (section 4.1) we argued that in India a literacy rate of 10-20% for men may be acceptable in the 1830s-1840s. Given that in section 3 of this chapter we will estimate attainment in India at 4.3% in 1890 (an average of males and females), given that attainment for males is substantially higher than that of females, and given that European and subsidized education was much lower in 1840 than it was in 1890, the share of indigenous education was probably close to 5-15 percentage points of total literacy around 1840, hence about three times as large as European and sbsidized education. Towards the end of the nineteenth century indigenous education probably declined in line with the rise of European education. In table 3.4 in this chapter we can see that literacy in 1891 according to the census was around 5.3% compared to an attainment of 4.4%. Given that there might have been a relapse into illiteracy after following formal education, this suggests that around 2% of the literates (or around 40% of all schooling) still took place in non-subsidized private schools. This figure, however, dropped fast in the following decennias. The same pattern can we see in Indonesia. Reid (1988, 218) argues that in the Dutch 1930 Census for Indonesia it was especially in those regions such as the Lampung districts of Southern Sumatra where the 'modern' European education was not widespread that the highest literacy was recorded (45% for males and 34% for women). Given that we estimate total attainment in 1930 at 16%, this suggests a share of non-recognised non subsidized education of around 50%. However, in other parts of Indonesia non-recognised education was much smaller. Hence, the total share of non-recognised non-subsidized schools in Indonesia was probably less than 25% and strongly declining towards independence.

Second, the non-recognised, non-subsidized schools were almost all either *madrasahs, pesantrens* or Hindu schools and had an almost exclusive religious curriculum. This did not enhance the participation of their students on the labour market, especially on the European one. This was recognised in Indonesia both in the 1950s when the education system of the Republic of Indonesia got shape and also during the 1936 conference at Padang-panjang where Muslim organisations discussed the future structure of religious education. However, the secular changes that were

[55] A third reason may be that it is common in the literature to ignore these schools. See for example Meyer, Ramirez, and Nuhoglu Soysal (1992, 132-133).

enacted in religious education following these conferences took mainly place after independence. In that period, these schools were accounted for in the national statistics anyway so no special modification has to be made. We also left out kindergartens as they are in general either not aimed at the acquiring of human skills or the skills acquired are so basic that they are not regarded as skills anymore in the labour market.

Besides the question of which schools to include, we have to make a division into levels of education. This division into primary, secondary (general and vocational), and higher education is largely made according to the standards of the UNESCO. The UNESCO in the 1970s developed the first International Standard Classification of Education (ISCED), which was revised in 1997. However, these standards were more a reflection of already existing systems. They leave room for national and cultural deviations of educational structures.

The main criterion for primary education as indicated in the ISCED is the 'beginning of systematic apprenticeship of reading, writing and mathematics.' Subsidiary criteria are the 'start of compulsory education', and 'entry in nationally designated primary institutions.' In general it is considered that primary education does not start before age 5 or after age 7. When it forms a part of basic education only the first part (or the first six years) is considered primary education. A final criterion for primary education is that it is program and not subject based. In other words, it is aimed at giving children a comprehensive schooling and laying a basis for possible further subject-oriented education. For India the data on the primary education level correspond simply to primary education as given in its statistical abstracts. This exists of compulsory primary education between 6 and 14 years, which can be split into primary and upper primary schools. Only the first part thus forms the primary level. Although primary education lasts eight years in twenty States and Union Territories and seven years in twelve (see International Bureau of Education (IBE) 2001) this is not a problem because of the flexibility of ISCED. In Indonesia this definition means that, for the colonial period, the European Primary School, Dutch-Chinese School, Dutch-Indonesian School, the Advanced Elementary Education, the Link School, the Standard School, the Continuation School, and the Village School can all be considered primary education. They are programme oriented and in general have a duration of between 3 and seven years. The advanced elementary education is also counted with primary education because it is also programme oriented and because it

was considered an end to formal schooling and not a step to a 'lifelong learning', which is considered characteristic for secondary education. For the period after independence the *sekolah rakyat,* later renamed *sekolah dasar* (SD), is considered primary education with its entrance age of 6 or 7 years, and its six-year duration. It also has a clear program orientation with the exception that local languages are allowed during the first three years. In Japan, the *terakoya* (primary education for commoners) was replaced after the Meiji Restoration in 1868. In a following transition phase ending in 1882 a primary level was created which consisted of a 4-year Ordinary Elementary School after which children could continue in a 2-year Higher Elementary School. It was only in 1908 that a single 6-year Ordinary Elementary School was established which would continue afterwards.

Following the ISCED, the secondary level consists of two parts which largely share the same characteristics. First, there is lower secondary (or the end of compulsory) education. This generally has a duration of 3 years after primary education. Second, there is higher secondary education which either finishes the educational program by preparing for the labour market or prepares for higher education. This level starts generally at age 15 or 16 after finishing lower secondary education. These phases show an increase from still largely programme orientation at lower secondary, to more subject orientation at higher secondary education. Both general and pre-vocational/technical eduction focus on a broader range of subjects either to prepare the pupil for the labour market or for further education. Vocational/technical education trains for a specific occupation. These levels are clearly recognisable in India, Indonesia, and Japan, especially after World War II. In India the secondary level consists of the Upper Primary or Middle school and the Secondary school. This is either four or five years and follows directly on the primary level. In Indonesia after independence there was a standard lower and higher secondary education. However, prior to that there was a HBS (a former Dutch high school), a Lyceum, and general secondary schools all with a duration of between 3 and 5 years. The Japanese system, however, consisted of middle and vocational schools before the War and of lower and higher secondary schools afterwards, introduced during the occupation period. Yet, in all three countries the secondary level was often subject oriented. The stronger programme orientedness in higher secondary schools can, for example, be seen from the fact that, especially after the War, in India, just as in Indonesia, vocational education was largely given in Senior

Secondary Schools. However, in the colonial period there were also primary vocational courses. Here the enrolment was, however, not large and it therefore does not significantly alter the figures. This is contrary to Japan where most vocational education was given at the secondary level. The already relatively high enrolment ratios at the end of the nineteenth century caused primary education to be focussed on general skills while further specialisation had to await the secondary level. Until the Sino-Japanese War (1894-1895) vocational education had remained almost entirely private. However, combined with the developing Japanese technical industry, the war made clear that the demand for technically trained people was increasing. Therefore a Vocational Education Law was drafted and passed in 1894. Additional steps were taken in 1899 when fishery, forestry, and agricultural vocational schools were established at lower secondary level (Passin 1965: 97). At the start of the twentieth century also technical education at the upper secondary level expanded rapidly. This development continued during the first half of the twentieth century. As a consequence, already at the start of the twentieth century in Japan both a vocational and a general secondary education system had emerged which in India and Indonesia had to wait until after the Second World War.

The tertiary level in the ISCED is basically described as the remaining education. For simplicity, we include in this category also the post-secondary non-tertiary education because they share the same characteristics. Generally, the entrance requirement for tertiary education is completed secondary education. In addition, this level has a strong subject orientation leading to either an occupation or a research qualification. In the three countries of this study, higher education was scattered over different institutions of different ethnicity, religion and public/private denomination. In Japan this changed when many private institutions were recognised around 1919. However, as in India and Indonesia secondary education was relatively underdeveloped, this could only be a limited canal for following higher education. In Indonesia higher education was completely absent in the period before the 1920s while afterwards it was limited to technical, law, and medicine colleges with a duration of between 4 and 7 years. The number of universities in India was much larger, already enacted in the mid nineteenth century because the colonial government focused more on higher than primary education. However, both India and Indonesia lagged considerably behind Japan.

2.2 Estimates of levels of enrolment

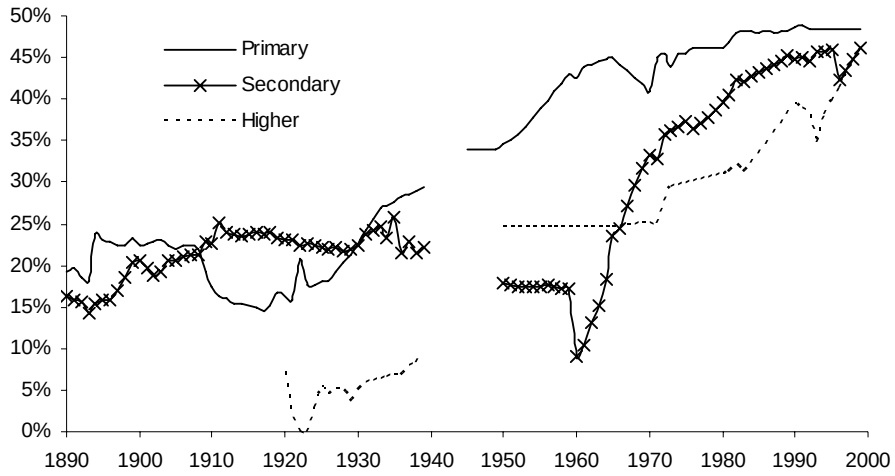
Now we have made a distinction between three levels of education, it becomes possible to collect historical data on enrolment. The term 'enrolment' refers to the number of persons enrolled in (i.e. following) a certain education level in a certain year. Although enrolment rates are generally straightforward in that they are explicitly given in the statistical sources, three points have to be stressed. The first point concerns school attendance. If we look at the enrolment levels in the tables A.6.1-A.6.3 in appendix A.6, we notice that (given the size of the population) Japan has by far the highest enrolment at the start of the twentieth century. However, even in Japan actual school attendance was much lower. In India and Indonesia, the drop-out rates were also very large. In addition, these two countries had their own educational systems prior to the colonial period. However, these indigenous systems had deteriorated strongly in the eighteenth and nineteenth centuries. This was one of the main reasons why the colonial governments in both countries during the nineteenth century started to set up an alternative education system which would later be continued by the newly elected governments after independence. As a consequence, most of the data, except for some occasional statistics, were collected only on the new, colonial, education systems. This led to an underestimation of enrolment figures, because enrolments in the indigenous education system were largely excluded, although this problem was already relatively small at the start of the twentieth century.

The second problem with which we are confronted when trying to estimate enrolment figures is that in Indonesia there was a strong tendency to underenumeration in the population censuses at the start of the twentieth century. Contrary to the population data from the surveys and censuses (see for example Van der Eng 1996: 271), however, we opted not to correct the education data for underenumeration. As the educational reports in contrast to the population surveys were not based on estimates or corrections from the village head, the margin of error will be smaller. Furthermore, the data between 1914 and 1940 and after 1970 are relatively complete although only after 1970 for the first time the total number of schools and pupils was collected instead of being inferred from a sample. Before 1914, especially for the village schools, the differentiation of the data in sexes was not always given. This was partly due to the situation that the government experiment with these schools only started on a small scale around 1906. As a consequence, only

limited statistics were reported for these schools in the following years. Moreover, the continuous changes in Indonesian primary education with, for example, the introduction of the Dutch-Indonesian School made data collection difficult. For the colonial period about 90% of the figures could be directly obtained from the sources. Where data were missing of, for example, a differentiation between boys and girls

Figure 3.1

Percentage girls per level of education in Indonesia, 1890-2000



Source: Appendix A.6, table A.6.1.

Note: The straight lines of the ratio of females in secondary and higher education after independence are caused by the assumption that the ratio remained constant in that period. We could have assumed that there was a linear in- or decrease in the ratios from before World War II, however, it is unlikely that the male-female composition just after the War changed considerably.

enrolled, the sex ratios of children enrolled at the begin and at the end of the period with missing data were taken. These ratios were interpolated and by multiplying this result with the total enrolment the number of girls enrolled was obtained. For the period after independence about 70% of the yearly data on enrolments were directly available. Of course, most data were missing in the 1940s during the period of WWII and the following struggle for Independence. The gaps in the data for these periods were either filled by using ratios with the available data or by linear interpolation. An especially difficult subject was higher education before 1977. The largest part of the total enrolment was available but the sex ratio was given only in a few years. For

these missing data on boys and girls enrolled, again the interpolation of the available ratios was used. Consequently, the ratio of female students to total students remains about the same for the first years after the War, as we can see in figure 3.1. One could linearly interpolate the ratio from girls to total enrolments in higher education between 1940 and 1970. However, it is unlikely that this ratio changed strongly in the first years after the War and, consequently, we assumed that most of this increase in the ratio took place already before World War II. Furthermore, figure 3.1 shows that the enrolment in higher education started only in 1920. Before that time, those wanting to pursue higher education generally had to go to the Netherlands.

Compared to Indonesia, the Indian data are relatively complete both before and after independence, but we are confronted with significant changes in territory over the twentieth century. First, in 1937 Burma (present day Myanmar) was split of from the statistics of British India. After independence in 1947, British India was split in India and Pakistan.[56] To correct the enrolment figures for this split in territory, it was necessary to either add the Bangladeshi and Pakistani figures to the Indian totals after independence in order to obtain the totals for Undivided India or to filter the Indian totals from the figures before independence. We opted for the latter possibility. The reason was mainly that it would be difficult to obtain enough comparable data for these three countries. One problem was that the statistical methods and definitions differed substantially, making it very difficult to obtain comparable figures in order to create an aggregate figure. In addition, it would be almost impossible to integrate the educational systems of India, Pakistan, Bangladesh, and possibly Burma, into one system. In other words, these systems are so pluriform that it would be difficult to make any generalizing comments on them.[57] Furthermore, focussing on India would better reflect the actual situation without reverting to a 'theoretical construct of British India' instead of the present-day situation of three (or four) heterogenous countries. Finally, within British India, India was by far the largest part, so the error of removing the other countries from the data will probably not be very large.

The removal of Burma, Pakistan, and Bangladesh from the enrolment data is done in a very general way by calculating their ratio with total enrolments of British India around 1948 and assuming that this ratio remained constant in the period

[56] In 1971 Pakistan was split into Pakistan and Bangladesh.
[57] At least when treating it as one geographical aggregagate unit. Of course, if one wants to make a comparative analys, these three countries (India, Bangladesh, and Pakistan) could suffice.

between 1880 and 1948.[58] Although not always perfectly accurate, the outcome is probably quite reliable, mainly because India was so much larger than the other territories. Therefore it is necessary to come to a percentage of the different school

Table 3.2: Percentage of total number of students per level of education, sex, and country around 1948

	Pakistan	Bangladesh	India
Total Population*	39,448,000	45,646,000	359,000,000
Distribution Population*	8.88%	10.28%	80.84%
Primary Education			
Boy	4.24%	16.96%	78.80%
Girl	2.67%	6.14%	91.19%
Total	3.86%	14.34%	81.80%
Secondary Education			
Boy	14.67%	15.40%	69.94%
Girl	10.59%	9.53%	79.89%
Total	12.53%	12.97%	74.50%
Vocational Education			
Boy *	1.07%	0.76%	98.17%
Girl *	3.37%	0.63%	96.00%
Total *	1.37%	0.75%	97.89%
Higher Education			
Boy **	8.33%	5.12%	86.54%
Girl **	7.52%	3.80%	88.69%
Total **	8.23%	4.96%	86.82%

*Figures Pakistan and Bangladesh 1950, India 1948. Population figures 1950.

**Figures Pakistan, Bangladesh, and India from 1953.

types in India and Pakistan/Bangladesh. The 1948 division is given in table 3.2. These figures are used to correct enrolment for the colonial period back to 1880. We could have opted for a more refined way by estimating the enrolment levels for smaller areas and than deducting it from the total when they were outside the territory of contemporary India. However, this is unlikely to give significantly better results due to lack of enough detailed data.

2.3 Estimates of the Gross Enrolment Ratio

After having obtained the historical enrolment per education level it is now necessary to go one step further by estimating the Gross Enrolment Ratio (GER). The gross enrolment ratio is the number of persons enrolled at a certain education level, divided by the relevant age group. In other words, if 10 children are enrolled in primary education, which lasts from age 6 to age 12, and the total number of persons in the

[58] Until 1937 Burma was included in the general statistics for British India. Yet, as they were also given separately, these data could be deducted from the total level directly.

population between age 6 and age 12 is 20, then the gross enrolment ratio is 50%. In this way the gross enrolment ratio is given in tables A.6.1-A.6.3 (appendix A.6) as the enrolment as '% of the relevant age group'. It is important to note that the gross enrolment ratio calculates all persons enrolled in a certain level of education, not only the children which belong to that age class. As a consequence, the gross enrolment ratio may exceed 100%. If we would only include all children enrolled in a certain education level who belong to the relevant age class, we would get the net enrolment ratio. However, this data is generally not available for the period prior to 1960.

The reasons for the estimation of the gross enrolment ratio are clear when one considers the difficulty in comparing the enrolment data between countries or over time. If one compares for example India with the Netherlands Indies, it is obvious that the enrolment levels are far higher in India simply because India is a far larger country. In the same way, it makes a comparison of the enrolment data within one country over time possible because it corrects for the growth in population and the (associated) increase in enrolment. A final point to note is that the gross enrolment ratio also corrects for changes in the educational system. If, for example, primary education is extended from 4 to 6 years, than the enrolment level may increase dramatically. However, as the relevant age class is also broadened, this is not necessarily the case for the gross enrolment ratio. This is especially visible for Indonesia after the War where we reduced the age class for primary education from a weighted average of 5-12 and 5-10 years to 6-11 years.

Thus, after obtaining the enrolment numbers for Japan, India and Indonesia, it is desirable to also calculate the gross enrolment ratio for these countries. As the enrolment data were already calculated, we now need population figures. The data construction for India and Japan was relatively straightforward as the data were readily available. For India, we simply took the census data for 1891, 1901, 1921, 1931, 1941, 1951, 1961, 1971, 1981, 1991-2000 and the total population by Sivasubramonian (2000) and Bina Roy (1996). For each census year we calculated the relevant age groups per level of education. Using the ratios of the relevant age group to the total population of British India, we used the total population figures from Sivasubramonian (2000) and Roy (1996) for the Indian Union, to calculate the relevant age group for the Indian Union solely for each census year. Then the next step was to interpolate the ratios between the age groups and the total population. Using these interpolated data, we could use the total population figures for India to

calculate the age groups in the years between the censuses. Then, by dividing the number enrolled by the relevant age group we arrived at the gross enrolment ratio. The same was done for Japan, where we used the five-yearly population data per age group from the Historical Statistics of Japan, in addition with the Statistical Yearbook of Japan. The total population was obtained from Pilat (2002).

However, the estimates for Indonesia were somewhat more complex because the enrolment figures before 1941 are also divided between ethnic lines. On the one hand this caused problems because of the different school duration in the indigenous and European schools. On the other hand, it also creates the possibility of calculating the gross enrolment ratio per ethnicity before independence. The latter we did in appendix A.5. The results are reported in table A.5.1 in appendix A.5. This table shows the large difference in enrolment among the different ethnic groups in Indonesia. Whereas in 1890 almost 1.5 times as many Europeans followed education as there were children in the relevant age class (a GER of 150)[59], only 2 out of 100 Chinese and 1.4 out of 100 Indonesian children followed primary education. Around 1940 these figures had to some extent converged, but a large gap remained.

To arrive at a gross enrolment ratio for the entire population of Indonesia, we have to add these gross enrolment figures for Chinese, Indonesians, and Europeans, weighted for their population shares, for the period prior to 1940 (see appendix A.5). For the years hereafter, we use the same method as for India and Japan. We used the population figures from the census data, where we corrected the 1961 census for the omission of Irian. As the duration of each level of education changed over time, we also used different population cohorts for eah level of education. For 1941-1969 we took the cohorts aged 6-11, 12-17, and 18-22, as did the *UNESCO Statistical Yearbook* for the period 1960-1970.[60] After 1970 the cohorts 7-12, 13-18, and 19-23 were used. The number of children in these cohorts were estimated from the census data and interpolated with the total population figures from Van der Eng (1996; 2002). Next, we divided the total enrolment per level of of education (see section 2.2 above) by the relevant age class to arrive at the gross enrolment ratio of that level.[61]

[59] This is possible when also older and younger children enter education.
[60] For the period 1950-1960 the UNESCO used 5-14 and 15-19, but these are implausible cohorts.
[61] We also distinguished between boys and girls. This had to be done also for the period prior to 1941 as we did not distinguish by sex. To do this, we estimated the age classes of boys and girls from the censuses (1890, 1895, 1900, 1905, 1920, 1927, 1930, 1961, 1971, 1980, and 1990-2000) and took the ratio with the total population figures. The ratios of the in-between years were interpolated and then multiplied with the total population.

Furthermore, we used some assumptions from the literature to arrive at the total enrolment rates during the War[62] and the division into male and female enrolments.[63]

3. ATTAINMENT AND AVERAGE YEARS OF EDUCATION

Although, as we have seen in chapter 2, some earlier analytical studies on the relation between human capital and economic growth used gross enrolment ratios or enrolment figures, present-day work prefers variables that are a better indication of the stock of human capital. Two related measures have become very popular. The first one is attainment, which has become especially popular since the work of Barro and Lee (1993; 2001). With attainment in a certain level of education we mean, following Barro and Lee, the percentage of the population of 15 years and older who have been enrolled[64] in that specific level of education and no more than that. So if, for example, primary attainment is 10%, this means that 10% of the population of 15 years and older has once attended primary education. Please note, however, that these are only those persons who did not pursue any further education. This means that, if primary attainment is 10%, secondary attainment is 15% and higher education is 5%, in total 30% of the population of 15 years and older has attended primary education as persons must first have completed primary education before attending secondary or higher education. The second, related, measure of the human capital stock is average years of education per capita. This variable is strongly linked to attainment because in fact it is calculated as attainment per level of education (including 'no education'), multiplied by the years of education per level of education, and finally divided by 100.

[62] We now only miss the gross enrolment ratios for the period 1941-1944. This can be solved by calculating the missing enrolment data. They are estimated in the following way. First, total primary enrolment in 1943 was estimated by using the 1944 enrolment minus the total Europeans in primary education in 1940. The main idea is that it is unlikely that there was a strong increase in education of the indigenous population between 1943 and 1944. In addition, the Europeans were put in camps, so no school attandence is likely in that period. The 1942 enrolment figure for primary education was estimated by taking the 1944 figure - Europeans - (0.5*private education in 1940) as private education was strongly restricted in these years. The 1941 figure was a linear extrapolation of 1939 and 1940.

[63] The number of boys and girls in 1940 and 1941 were then calculated by the 1939 ratio as nothing really changed in those years. For 1942 the number of girls was calculated as the number of girls in 1943 - (% European girls*Total number of girls in 1940)-(0.5*girls in private education) as the number of girls in private education was as a % larger than in public education. For 1943 and 1944 the ratio of 1945 was used. These data were of course divided by the population in the relevant age. The gross enrolment ratio for boys in higher education was assumed constant for 1940 and rise to 0.01 for 1941. For 1942 there was an almost total drop of student numbers, which led to an enrolment ratio in 1943 of almost 0, and in 1944 a small rise.

[64] But not necessary completed.

There are several ways to calculate attainment. However, as we already noticed in the previous chapter, the methold of Barro and Lee outperforms the alternatives. For example Krueger and Lindahl (2001, 1117), in their overview article on the micro and macro growth literature, estimated that the reliability of the Barro &Lee data as 0.577 compared to 0.195 for the Kyriacou (1991) data. We will therefore start with a brief overview of the methodology of Barro and Lee (1993; 2001).

Barro & Lee estimated attainment at five-year intervals since 1950. They used census figures as benchmarks and, as most censuses are held once every 10 years, they used a formula to fill in every missing fifth year. This use of benchmarks is contrary to, for example, Nehru (1995) who relies solely on mortality and enrolment figures to calculate average years of education. Although the use of census data improves the quality of the attainment figures, it remains questionable how reliable the Barro & Lee estimates are.

Unlike their earlier estimates, in their more recent work Barro and Lee (2001) use net enrolment ratios, keep track of repeaters, and adjust them for later entries into the specified education levels. However it is quite likely that in this way they will underestimate the attainment as the percentage enrolled at a higher age may be large. This is especially true for developing countries such as India and Indonesia in the period after independence. Therefore we use the gross enrolment ratio, adjusted for the duration of official education. Although this includes repeaters, the importance hereof diminishes in secondary and higher education. Furthermore, other data are not available prior to 1950 in India and Indonesia. In addition early as well as late entry is also important. This means that there was no clear entry age, especially in indigenous education, in the colonial period. As a consequence, using the net enrolment ratio excludes both the children that enter before and after the specified age class. Ignoring this is likely to understate enrolment and, as a consequence, attainment. Furthermore, Barro and Lee use the available census data after 1950 as benchmarks. In many countries, especially in India and Indonesia, these censuses may have a strong bias, taking things as the political situation and literacy campaigns into account. In addition, it is noteworthy that in general, the Barro and Lee interpolation tends to underestimate the attainment figures for the years between censuses (Portela, Alessi, and Teulings 2004, 5). Therefore we used for all three countries as a starting point the census around 1960 and the 1965 data of Barro and Lee. Furthermore there is not

much fluctuation in attainment in this period in all countries so it can be interpolated. In this way it is possible to obtain yearly figures.

From this starting point, our first step is to use the perpetual inventory method of Barro and Lee (1993; 2001) with the gross enrolment ratio as input and the years 1960-1965 as the most important benchmark years. The formula used to extrapolate the data back to circa 1890 is based on the formula used by Barro and Lee for the population 15 and over:

$$h_{1t} \equiv \left[1-\left(L15_t / L_t\right)\right]* h_{1t-5} + \left[\left(L15_t / L_t\right)*\left(PRI_{t-5} - SEC_t\right)\right] \qquad (3.1)$$

Here *h* is the attainment of *1* (primary education) in year *t*. *L15* is the population 15-19 and *L* the total population of 15 years and over. *PRI* is the gross enrollment ratio of primary and *SEC* of secondary education. In the same way secondary and higher attainment are calculated as:

$$h_{2t} \equiv \left[1-\left(L15_t / L_t\right)\right]* h_{2t-5} + \left(L15_t / L_t\right)* SEC_t - \left(L20_t / L_t\right)* HIGH_t \qquad (3.2)$$

$$h_{3t} \equiv \left[1-\left(L15_t / L_t\right)\right]* h_{3t-5} + \left[\left(L20_t / L_t\right)* HIGH_t\right] \qquad (3.3)$$

Here, *2* and *3* are secondary and higher education, *HIGH* is the gross enrolment ratio of higer education, and *L20* is the population aged 20-24.

Equations (3.1)-(3.3) indicate that for each level of education a duration of five years is assumed. However, this can be easily adapted to the different age cohorts in the different periods for the gross enrolment ratio. But even if we adapt the equations, for example by using longer time lags, mortality can remain the same. Mortality is calculated by Barro and Lee as the number of persons surviving from age 15 to age 19. As we assume that the death rate is inversely correlated with the length of education, and as the duration of education increases, there is a relative decline in mortality (because mortality remains to be estimated over a five year period). Thus there is no pressing need to alter the mortality assumptions from Barro and Lee. Finally, these figures were smoothed using a five-year moving average. As Barro and Lee estimated five yearly figures and because we adapted the five year period for changing school duration, the fluctuations became in our yearly data rather hectic. Therefore, we applied a five year moving average.

The second step in estimating attainment concerns the extrapolation backwards over 70 years (from 1960 back to 1890) using the method of Barro and Lee. This makes it possible that there is an increasing error over time even though the data were corrected for school duration and population growth. The reason for this

divergence may be, as we already indicated in chapter 2, that the Barro and Lee method may underestimate actual attainment due to the fact that in their method mortality is independent of education. In phases with strong educational growth, the survival chance of younger (more educated) persons is underestimated as is, as a consequence, attainment (Portela, Alessie, and Teulings 2004, 5). However, Barro and Lee use their method in forward extrapolation. As we go backward in time, using this method may overestimate actual attainment.

We therefore, want to correct the attainment estimates for the bias in the mortality rates from Barro and Lee for all three countries. We used the adapted method of Portela *et al.* (2004) who assume that there is a bias when the data of Barro and Lee are extrapolated backwards, forwards, or when they are interpolated. However, we are only concerned with backward extrapolation. We can thus modify the formula of Portela to include only backward extrapolated data. We get:

$$Edu_{it} = \alpha + \beta_1 Before_{it} + \eta_i + \varepsilon_{it} \qquad (3.4)$$

Here, *Edu* is attainment in country *i* at time *t* for primary, secondary, or higher education respectively, *Before* is the number of years from the 1960 observation of Barro and Lee to the first census observation, η is the between group (country) effect, and ε_{it} is the whte noise error term. Following Portela *et al.* (2004) we estimate a fixed effect model.

Now using a sample of 112 countries from the Barro and Lee dataset, we collected the attainment figures for primary, secondary, and higher attainment for 1960 and following years up and until the first census. The results of these (fixed effect) estimations are presented below.

prim= -0.199**Before* - 9.656**Dasia* - 0.205*(*Before*D1960)* - 0.0211*(*Before*D1965)*
(SE) (0.449) (1.15) (0.56) (0.376)

+ 0.09031*(*Before*D1970)* + 0.164*(*Before*D1975)* + 9.8 + ε_{it}
(0.19) (0.159) (1.27e-014)

No. obs. 229
R^2 0.93

sec= - 0.320**Before* - 12.39**Dasia* + 0.657*(*Before*D1960)* + 0.119*(*Before*D1965)*
(SE) (0.222) (16) (0.531) (0.097)

+ 1.7 + ε_{it}
(2.71e-014)

No. obs. 228
R² 0.87

high= + 0.054**Before* - 3.599**Dasia* - 0.102*(*Before*D1965)* -0.113*(*Before*D1970)*
(SE) (0.056) (5.53) (0.0503) (0.047)

+ 0.4 + ε_{it}
(6.66e-015)

No. obs. 228
R² 0.87

Here $Before * D_t$ is the variable *Before* multiplied with a time dummy. Now, if we followed the methodology of Portela *et al.*, we would use the equation $PEdu_{it} = Edu_{it} - \beta_1 Before_{it}$, where β_1 is the beta coefficient from equation (3.4) and $PEdu_{it}$ is the corrected attainment. However, we have to be careful that the β_1 says that on average over a five year period the estimation of attainment ought to be an x-percentage lower, i.e. was biased upwards. For Portela *et al.* this does not matter because they only correct interpolation at a five-year interval, but we want to correct over the entire 40-year period.

Therefore, we used the following formula taking primary attainment as an example. Primary attainment has a decrease in percentage attainment (so an overestimation of attainment by the Barro and Lee method) of -0.199-0.2045 (on average)= -0.404% percentage points yearly decline over a five year period on average. However, we need the yearly percentage decline if we want to create a yearly correction factor. The reason is that the basis is the present period. So, simply using longer time lags would overestimate actual decline. We thus deduct -0.404%/38.8% (the average primary attainment around 1960 in 112 countries) = -0.01041. Now, we can estimate for every year how much has to be deducted by calculating $(1+\alpha)^t$, where *t* is the number of years until the basis, and a the yearly percentage. For a fourty year period, we thus have to subtract from primary

attainment $(1-0.01041)^{40} = 65.8\%$, 100%-65.80% = 34.2% of the estimated attainment using the Barro and Lee method 40 years before 1960. The correction

Table 3.3: Correction factors for the backward extrapolation with the Barro and Lee (2001) method

years	years backward	Primary education	Secondary education	Higher education
1959	1	0.990	0.975	0.982
1955	5	0.949	0.881	0.914
1950	10	0.901	0.776	0.835
1945	15	0.855	0.684	0.763
1940	20	0.811	0.602	0.697
1935	25	0.770	0.531	0.637
1930	30	0.731	0.468	0.582
1925	35	0.693	0.412	0.532
1920	40	0.658	0.363	0.486
1915	45	0.624	0.320	0.444
1910	50	0.593	0.282	0.406
1905	55	0.562	0.248	0.371
1900	60	0.534	0.219	0.339
1895	65	0.507	0.193	0.310
1890	70	0.481	0.170	0.283

Estimation method: see text.

factors for each fifth year are given in table 3.3. We calculate back from 1960, so the first year with a correction factor is 1959. The fifth extrapolated year (see column 2) is 1955, etc. Column 3-5 give the factor with which the attainment figure for each level of education as estimated with the Barro and Lee method has to be multiplied in order to correct for the bias in the estimation method.

The corrected results for primary, secondary, and higher attainment are presented in table A.7.1 in appendix 7. They seem to conform rather well to the expected values. For example, table 3.4 compares the literacy rates with total attainment (primary plus secondary plus higher attainment). Unfortunately, no literacy rates for Japan are available.[65] However, as we will argue in the next chapter, it is likely that it rises from 30-40% in 1890 to at most 100% in 1960, a figure which we also found in our estimates. For India, table 3.4 shows that the literacy and attainment figures follow the same pattern. However, both for India and Indonesia, attainment figures are somewhat higher than literacy figures. This might be because, as Mayhew (1926, 228) argued, "… school enrolment figures under the present system in India

[65] Figure 1.3 in chapter 1 reports the total attainment figures as an indication of literacy in Japan.

Table 3.4: A comparison between literacy and total attainment in India and Indonesia, 1891-1951

	India		Indonesia	
	Literacy	Attainment	Literacy	Attainment
1891	5.3%	4.4%		4.4%
1901	5.4%	5.8%		5.9%
1911	5.9%	7.9%		8.4%
1921	7.2%	10.4%	5.4%	11.8%
1931	9.5%	14.0%	9.0%	15.9%
1941	16.1%	18.3%	12.7%*	21.0%
1951	18.3%	24.8%		27.1%

* Calculated here. The 1930 illiteracy figure was used. From this, 1/40 multiplied with the gross enrolment ratio , PRI(t-40), was subtracted and 1/40 multiplied with he gross enrolment ratio, PRI(t-3), was added (the latter because we estimate illiteracy for persons aged 15 and over). This is done for each year after 1930 and so we arrive at 100-87.3%=12.7% literay in 1941. This method is also almost ideally suited to get the 28.9% 'no school attainment' in 1961 according to the Unesco Statistical yearbook 1974 (we arrived at 26.7% literacy).
Source: India: Statistical Yearbooks; Indonesia: Indisch Verslag 1931 and 1935. Unesco Statistical Yearbook 1974.

mean very little. The education given in very many of our primary schools ends, as an official reporter once remarked, with the cradle and allows a relapse of 39 per cent of its beneficiaries into illiteracy within five years." This means that it is likely that attainment figures are higher than literacy figures in the first half of the twentieth century. The same can be argued for Indonesia.

For the post World War II period we can compare our estimates with some alternatives. Yet, as these are mostly given as 'average years of education in the population', we will first convert our attainment figures to obtain average years of education in the population of 15 years and older. Indeed, the indicator 'average years of education' is closely related to attainment. We used the attainment figures to calculate the average years of education as $(h_{1t} * Years_{1t} + h_{2t} * Years_{2t} + h_{3t} * Years_{3t})/100$, where h is attainment (%) of *1* (primary), *2* (secondary), or *3* (higher) education. It is important to divide by 100 to include also persons with no education (primary + secondary + higher attainment does not necessarily sum to 100). The results of these exercises on attainment, and average years of education are presented in table A.7.1.

For all three countries, our estimates of 'average years of education' are above those of Barro & Lee (see figures 3.2-3.4). This is a pattern which can be found in many countries. For example, figure 3.4 shows that also in Japan the Barro and Lee figures are seriously lower than the three alternative measures. Second, we notice,

together with Portela *et al.* (2004), that the data by Barro & Lee for the periods between the surveys seems to be somewhat underestimated. This is extremely well visible in the case of Indonesia. At least for 1965 and 1975, it is clear that the

Figure 3.2

Average years of education in Indonesia, 1950-2000

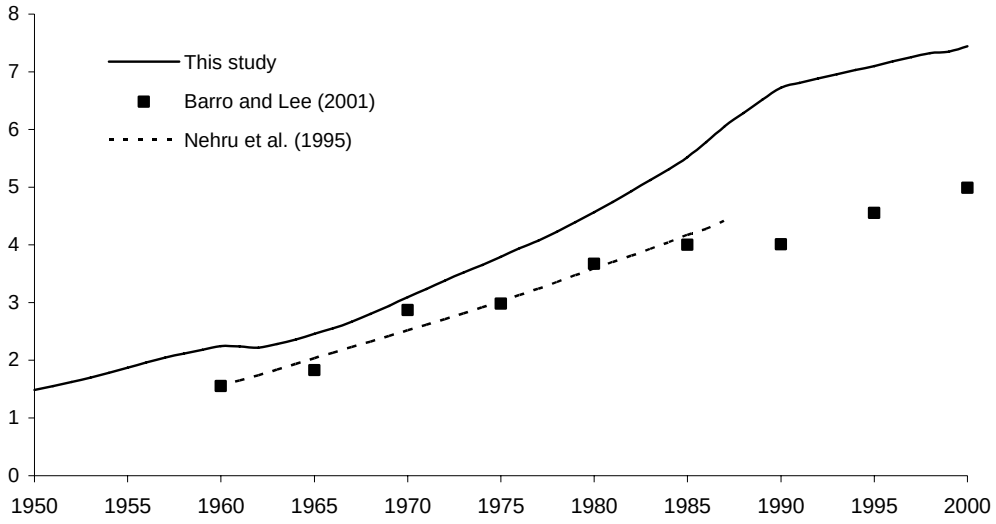


Figure 3.3

Average years of education in India, 1950-2000

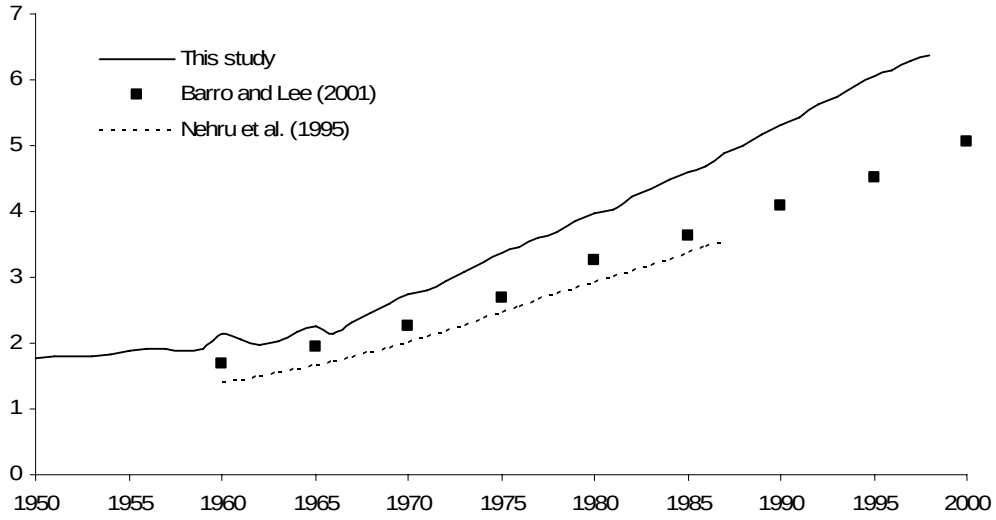
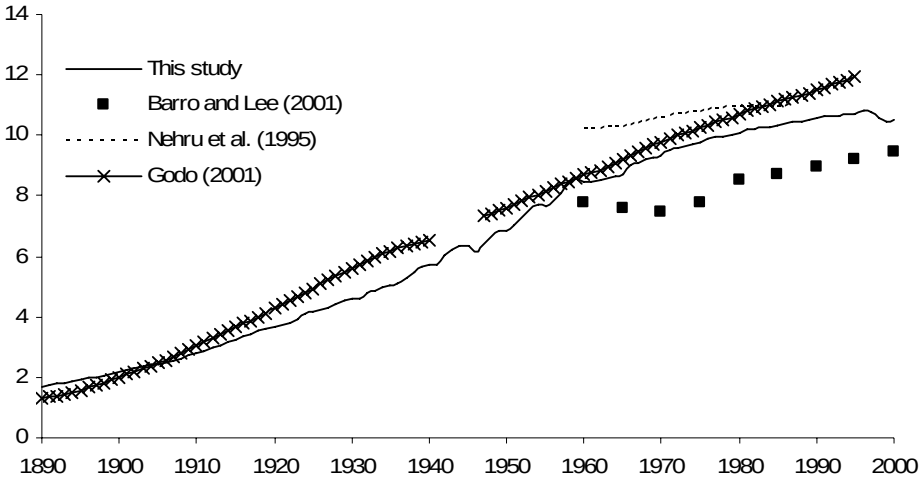


Figure 3.4

Average years of education in Japan, 1890-2000



Barro and Lee estimates are unexplainable low. Third, our estimates, although somewhat lower in the case of India and Indonesia and somewhat higher in the case of Japan, move in the same direction as the Nehru *et al.* (1995) estimates. Finally, for Japan, also some historical estimates are available from Godo (2001, table C1). The levels of these series corresponds quite well to ours. The main difference is that it shows somewhat higher growth rates in the late nineteenth and at the end of the twentieth century. Yet, we think this might be an overestimate. Especially at the end of the century, with already high literacy rates, one might expect that the growth of average years of education declines.

4. PRIVATE AND GOVERNMENT EXPENDITURE ON EDUCATION

We end this chapter with an overview of the collection of data on government and private expenditure on education. Here, government expenditure entails all government expenditure related to education, i.e. expenditure on students, school buildings, teacher salaries, and textbook materials. Private expenditure includes, as far as possible, all expenditure done by households on education, i.e. both school fees and expenditure on stationary (writing materials and textbooks). These data are on the one hand easier to collect than the enrolment data because government expenditure is mostly well documented. On the other hand, the difficulties may be far larger. One

reason is that private expenditure is often badly documented, especially in India and Indonesia prior to independence. Therefore we opted not to distinguish private expenditure by level of education. This would create data which were far too unreliable.

Concerning government expenditure on education, a first problem is that not all financing was done at the national level. Often sums were also spent at regional levels. This was especially true in Japan at the start of the twentieth century. In addition, we are again confronted with the situation that the territory of India changed in the twentieth century making it necessary to correct the obtained data to arrive at the figures for the Indian Union. Finally, we have reported both the government and the private expenditure on education in tables A.8.1-A.8.3 (appendix A.8) in current prices. However, they could be deflated by using the consumer price indices of Indonesia, India, and Japan respectively as reported in appendix A.1. Although one could argue that it would be preferable to use the wholesale price index for government expenditure, we decided against this because a large part concerns for example teacher's salaries which have an obvious relation with the consumer prices.

By far the easiest collection of the data is again for Japan. The data are readily available in the *Historical Statistics of Japan* and the *Japan Statistical Yearbook* (various issues). The only important point to note concerns private expenditure. These data were also available in the *Historical Statistics of Japan* and in the *Estimates of Long-Term Economic Statistics of Japan since 1868* (Ohkawa, Shinohara, and Umemura 1967, Vol. 6, table 95). In general we took school fees and stationary. However, no data were available for 1941-1946. Therefore we assumed that private expenditure (in constant 1990 prices) remained constant from 1941 until 1943. This is not unlikely as most War-restrictions in education took place from that year on. We filled in the years 1944-1946 by linear interpolation.

Both the collection of government and private expenditure on education was far more complex in Indonesia. Fortunately, for government expenditure, in the colonial period the largest share of expenditure came from the Education Ministry as education was largely centralized (with exception, until the end of the 1920s, of the village schools). An important source concerning the educational finances in the Netherlands Indies were the reports of the Dutch-Indies Education Commission, especially Report No. 3 on government expenditures on education. This report presented data on government expenditure on education for the period 1911-1929 and

in some cases even went back to 1900. The remaining years could be obtained from the Colonial Reports, Educational Reports, Budgets of the Volksraad, and, for the period prior to 1900, from the Budget of the Netherlands Indies.

Based on above observations and sources, we may conclude that prior to independence the government expenditure figures are, although somewhat more centralized than after independence, still very fragmented. It was only in 1930 when the Report no. 3 of the Dutch Indies Education Commission *De Overheidsuitgaven voor Onderwijsdoeleinden in Nederlandsch-Indïe* was published that these data were more or less systematically collected. The Report divided the finance data in data of the central government and that of the provinces. The central government was again divided by department, the most important one being the Education and Religion Department (Departement van Onderwijs en Eeredienst). Before 1911 there are no data available on the actual expenditure. Therefore the Commission only gave data on the Education budget. Because the expansion of education took place only after 1911 this should not cause much trouble. Therefore, from 1900 till 1911 only *begrotingscijfers* (budget figures) have been used. From 1911 the actual expenditure is available. Moreover, other departments than Education supervised some branches of vocational education. By far the largest was medical education. This has been added to the total current education figures.

For the period up until 1971 (and even 1998) the data are still not totally coherent. The data given for this period largely reflect the budget of the Education Ministry, which is the largest source of funding. In essence there are now two levels at which the education in Indonesia is financed by the government. The first level is the national level. Here three Ministries are involved: Education, Religion, and Home Affairs. The second level involves the provinces. There are three sources from which data can be obtained. The first one is the central government budget (recurrent, development, and, for public universities, self-generated funds). The second source is the Ministry of Home Affairs. This ministry pays teachers' salaries in public and 70% of teachers' salaries in the private primary schools. Here data are, however, not readily available although there are some years for which they are estimated. The third source is the Ministry of Religion. Here the data are only partly available. The main costs are the teachers' salaries in public and private religious schools. These data are, however, only available on a regional level from the offices of the Ministry of Religion. In sum, in 1995-1996 the Ministry of Education paid about 51%, the

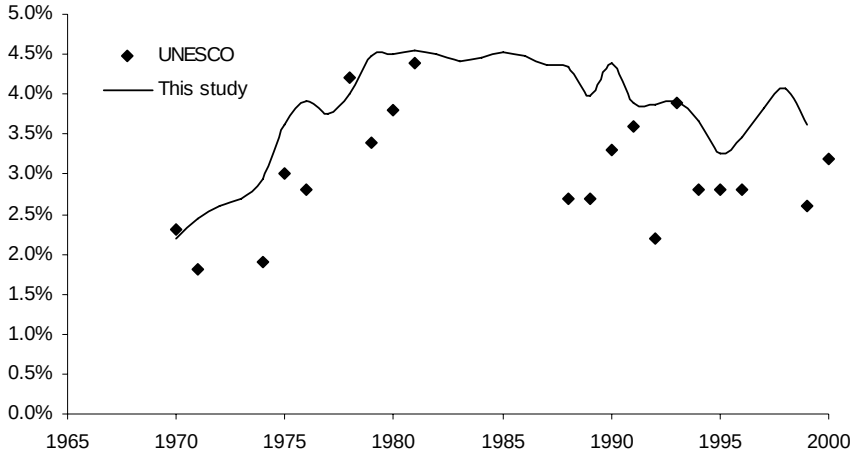
Ministry of Home Affairs about 38%, and the Ministry of Religion about 4% of total centralised government expenditure on education (Clark *et al.* 1998, 37-38). For the expenditure by lower government levels one has to turn to the Provincial accounts. Especially after independence the data on government expenditure in Indonesia are therefore not even partly available. Nevertheless two studies have been done into the education finances of Indonesia. Both studies were supported by (or took place within) a large project to obtain insight into the government finances. Therefore, they had access to data that, for other years, was not available. The first study, of Ruth Daroesman, appeared in two parts in the Bulletin of Indonesian Economic Studies in 1971/72 and in a separate draft report (Daroesman 1971; 1971; 1972). This study is largely based on a survey of the author herself as the government data were incomplete and unreliable. The second work on Indonesian education finance was performed by Clark *et al.* (1998). This study shows that between the study of Daroesman in the 1970s and the 1990s not much has changed in educational finances. The study of Clark *et al.* claims to be the first really comprehensive study of Indonesian educational finance in 1995/96. But even in this study, the almost 40,000 private *pesantrens* and kindergartens were not taken into account. Equally, many short-cut estimates had to be made just as Ruth Daroesman had done for 1970. One example constitutes the finances for teachers' salaries paid by the Ministry of Home Affairs.

Therefore it is not feasible to calculate for each year the actual expenditure on education for the post-colonial period. As mentioned, however, there are two works available in which the actual amount of government expenditure on education is calculated. Furthermore the data on the expenditure at the end of the colonial period were available. These were also divided into school level and to source of funding. With these benchmarks some available data on development and current budgets and on current GDP were used to obtain, with an interpolation of the ratios between total government expenditure in 1970 and 1995 and the data on government expenditure and total government expenditure, the data on government expenditure on education for the entire period. These data seem to be fairly accurate. If we compare them to the available expenditure figures obtained from the IMF Government Finance Statistical Yearbook (which only has a few data for the last years of the twentieth century) or with the UNESCO data, there are some differences but the pattern is the same (see figure 3.5).

Private expenditure is even more problematical in Indonesia than is government expenditure. The best data available are for the period after

Figure 3.5

Government expenditure on education as a percentage of GDP in Indonesia, 1967-2000



Note: The expenditure data from the UNESCO were as much as possible corrected for changes in definition.

Source: UNESCO Statistical Yearbook (various issues); Appendix A.8, table A.8.2.

independence. Two important sources are available. First, the input-output tables provide data on private expenditure on education for 1971, 1975, 1980, 1985, 1990, 1995, and 2000. These data are relatively reliable as they provide information on all streams of income, production and expenditure in the Indonesian society. As such, we will use these figures to provide benchmarks. Further, we have data on private expenditure on education from the Indonesian household surveys (SUSENAS) which are available since the 1960s. However Bina Roy (2003, 9) has remarked for India that expenditure surveys are generally much less reliable than commodity flow methods. This is partly because (often richer) households, underreport actual consumption. As the expenditure on education is often not underreported, education as a percentage of total consumption is overestimated. Consequently, we made the SUSENAS results comparable to the input-output tables and than estimated ratios for the given years between private consumption on education to total private consumption obtained from the Indonesian national accounts (see also appendix

A.4).[66] These ratios were interpolated. Multiplying these results with the total private consumption, gave the private expenditure on education between 1960 and 2000.

For the period prior to 1960 we followed the same method as outlined in appendix A.4. We used household expenditure data from several surveys in the period

Table 3.5: Private expenditure on education per household category based on the household surveys, 1924-1959 (mln current guilders)

(non-)-agriculture	household category	1924	1932	1937	1939	1941	1942	1953	1959
Agriculture	Agricultural employee household				0.4	1.1			45.6
	Operator, land owner 0,0-0,5 ha agriculture household	8.4					0.4		58.5
	Operator, land owner 0,5-1 ha agriculture household		0.9				1.7		120.4
	Operator, land owner >1 ha agriculture household	10.5					10.6		
non-agricultural	Non agricultural lower level rural household	1.7	1.4		0.4				
	Non labour force rural household								
	Non agricultural higher level rural household				0.8				
	Non agricultural lower level urban household			0.4				26.0	107.8
	Non labour force urban household			2.3				15.4	110.4
	Non agricultural higher level urban household								30.6
Total*		**18.5**	**17.5**	**13.3**	**16.6**	**14.2**	**27.6**	**81.8**	**1,087.3**

* Estimation method (see text)
Source: Household surveys (see table A.4.1 in Appendix A.4); total educational expenditure, see Appendix A.8.

1880-1960. An overview of these surveys is given in table A.4.1 in appendix A.4. From these surveys we obtained the educational expenditure of several household categories for several years (see table 3.5) (for a description of the household classes see Appendix A.4.). Please be aware that the figures in table 3.5 include a considerable margin of error.

Because for the period prior to 1960 only data on a few household classes and years were present, we used these data to calculate the ratios with the educational expenditure of other classes. This gave total educational expenditure for some years. The missing years were imputed where government expenditure on education, skilled and unskilled wages, and and educational enrolments were used as independent variables. However, imputation leaves the original data points of the dependent variable intact. For private expenditure these may show strong fluctuation as not all

[66] The total private consumption expenditure statistics were obtained from Badan Pusat Statistik, *Statistik Indonesia (Statistical Yearbook of Indonesia)*, Jakarta: BPS 1976-2003. Further they were obtained from Pusat Penelitian dan Perkembangan Statistik, Biro Pusat Statistik, *Pendapatan Nasional Indonesia 1960-1968 (National Income of Indonesia 1960-1968)*, Djakarta: Pusat Penelitian dan Perkembangan Statistik, Biro Pusat Statistik 1970. Finally, the input-output tables were used.

differences between household surveys could be removed. Therefore, just as in appendix A.4, we regressed the imputed values of educational expenditure on the other variables such as government expenditure on education, skilled and unskilled wages, and educational enrolment. We used the resulting coefficients to predict the private education expenditure variable. Yet, because of lack of suitable household surveys, we miss data on private educational expenditure for the period prior to 1928. Therefore we used the ratio in 1928 from private educational expenditure with total private expenditure to estimate the expenditure on education between 1880 and 1927.[67]

For India the data are easily obtainable from the statistical yearbooks. The correction for the separation of Pakistan and Bangladesh can be done in the same way as we did for the enrolment figures. The expenditure on education in India in 1950 was 89.06% of total expenditure of India, East and West Pakistan and Bangladesh (undivided India). This figure was used for all expenditure back to 1890, of course after substracting the figures for Burma. For the period after independence, it is important to note that, just as in Indonesia, the central government, state governments, local authorities and a variety of private sources financed education. In addition, the budget for the state and provincial governments is divided in the development and the maintenance budget (development and current budget in Indonesia) (see Bordia 1995, 436). Private expenditure on education was arrived at by calculating the expenditure on education by individuals and private funds. This same method was used by Bina Roy (2003) when she estimated total private expenditure and private expenditure on education in India between 1900 and 1950. We extended these series to include 1880-2000.[68]

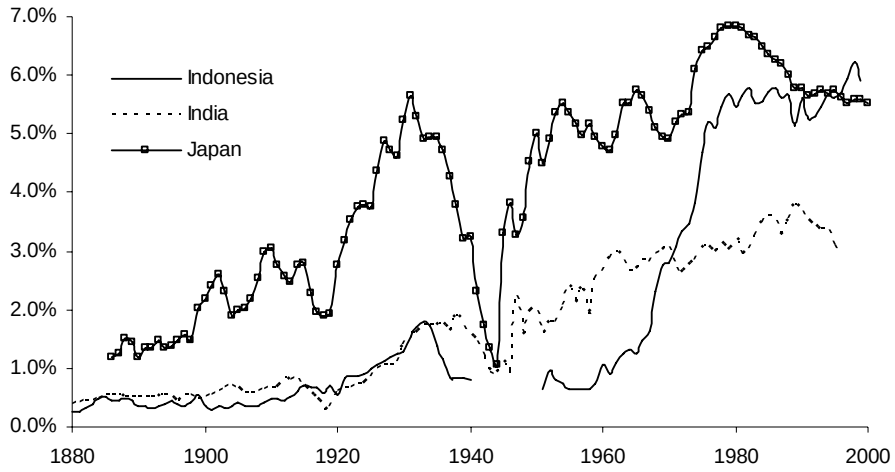
The results of both the government and private expenditure on education for Indonesia, India, and Japan are presented in tables A.8.1-A.8.3 in appendix A.8. The total educational expenditure, that is private plus government expenditure, as a percentage of the GDP is reported in figure 3.6. One point to note is that already in 1890, the share of educational expenditure in Japan was far higher than that in India and Indonesia. We may also notice that in the 1970s a strong rise of the share of

[67] This means that the percentage private education expenditure in total private expenditure is assumed constant between 1880 and 1927.
[68] However, these do only partly include expenditure on stationary. Therefore, we think they might be somewhat underestimated. However, the difference is likely to be marginal.

educational expenditure took place in Indonesia. That this is largely caused by an increase in government expenditure can be seen from a comparison with figure 3.5. Combined with an equally strong rise in GDP due to the oil boom in those years, this means that the growth of educational expenditure was really astonishing. India, however, experienced a gradual rise of the share of education in GDP from the 1920s onwards. This is probably caused largely by on the one hand the political focus on

Figure 3.6

Government plus private expenditure on education as a percentage of GDP in India, Indonesia, and Japan, 1880-2000 (in current prices)



industrialisation and less so on education and on the other hand the lack of political will to invest more in education. Indeed, the 1950s and 1960s in India seem to be largely dominated by plans for educational reform that never materialised. A more extensive description of the educational developments in India, Indonesia, and Japan is presented in chapter 4.

5. CONCLUSION

The present chapter was intended to give a brief overview of the basic variables that are the starting point of any study on human capital. Although data for the construction of these variables such as attainment, average years of education, and (as an indicator of human capital formation) enrolment are in principle available, there

are several obstacles when one wants wants to collect them and make them comparable. We saw that there were three main problems. One is lack of data, especially in Indonesia during the Second World War and the following period of decolonization. This requires the use of alternative assumptions and estimation methods to get an idea of the movement of the educationan variables during that period. The second difficulty was the split of British India in India and Pakistan (the latter was subsequently split in Pakistan and Bangladesh). Although it would have been possible to collect data for small administrative regions and use it to correct the national figures for the border changes, we decided to calculate the ratio between on the one hand India and on the other Pakistan and Bangladesh and to perpetuate this ratio backwards. Finally, we were confronted with an overestimation of educational attainment when using the Barro and Lee-method to bring the attainment estimates back in time. Using a slightly adapted method of Portela *et al.* (2004) we constructed a set of correction factors to correct for the overestimation of attainment at each level of education.

The results from our estimates indicate that Japan was clearly more developed in the field of education at the start of the twentieth century than were India and Indonesia. Japan had a far greater share of educational expenditure in GDP and had far higher levels of attainment and enrolment at all levels of education (see next chapter for a description). But also between India and Indonesia there were differences. We briefly touched upon some of these such as the steady increase in educational expenditure as a percentage of GDP in India during the period 1940-1960 while in Indonesia there was a decline followed by a boom in educational expenditure. Yet, a more extensive description of the educational development of these three countries is given in the next chapter where we will try to connect this development to their institutional structure.