Policy Research

WORKING PAPERS

Socioeconomic Data

International Economics Department
The World Bank
April 1993
WPS 1124

A New Database on Human Capital Stock

Sources, Methodology, and Results

Vikram Nehru
Eric Swanson
and
Ashutosh Dubey

A brief review of the method and data sources used to prepare the International Economics Department's estimates of the stock of education. Analysis suggests that it is not unreasonable to use education stock as a proxy for human capital in production function analysis.

Policy Research

WORKING PAPERS

Socioeconomic Data

WPS 1124

This paper describes a new database on human capital stock in developing and industrial countries prepared by the International Economics Department (IEC) and undertaken as part of a larger IEC research project on total factor productivity growth. Copies of this paper are available free from the World Bank, 1818 H Street NW, Washington, DC 20433. Please contact Moira Coleridge-Taylor, room S8-049, extension 33704 (April 1993, 19 pages).

Nehru, Swanson, and Dubey describe the techniques and data adopted for the construction of a new series of estimates of the stock of education in 85 countries over 28 years (1960-87). It covers all the important developing regions except the republics of the former Soviet Union.

IEC continues a well-established trend in growth research of using educational stock (measured as mean school years of education of the labor force) as a proxy for human capital. The series are built from enrollment data using the perpetual inventory method, adjusted for mortality.

Estimates are corrected for grade repetition among school-goers and country-specific drop-out rates for primary and secondary students. Enrollment data series used start as far back as 1930 for most countries, and even earlier for others. This reduces the need for backward extrapolation of enrollments to provide the initial estimates of the investment inventory.

A New Database on Human Capital Stock: Sources, Methodology, and Results

Vikram Nehru
Eric Swanson
Ashutosh Dubey

A brief review of the methodology and data sources used in the preparation of IEC's education stock estimates, and an analysis of the results.

A New Database on Human Capital Stock in Developing and Industrial Countries: A Brief Commentary on the Methodology, Sources, and Results

Table of Contents

A New Database on Human Capital Stock in Developing and Industrial Countries: A Brief Commentary on the Methodology, Sources, and Results

Introduction

1. Increasing importance is being given to human capital in economic theory as reflected in the literature on the contribution of human capital to economic growth that has appeared in recent years. [1] This literature emphasizes the importance of human capital formation in the long term growth of output, especially in developing countries. In addition, it highlights the influence of national economic policies on long term economic growth, a feature absent in neoclassical models. According to this view, changes in the rate of human capital investment lead to changes in the long term rate of output growth rather than simply to changes in the level of output. The allocation of expenditures between human and physical investments, including government expenditures, therefore has critical ramifications for future growth and development.

2. The quest for a better understanding of the determinants of growth has stimulated fresh interest in improving estimates of human capital stock. Strictly speaking, the measurement of human capital should cover the range of investments that human beings make in themselves and in others, including formal and informal education, on-the-job-training, health, nutrition, and social services. So far, no researchers have put together such a composite measure, although efforts are underway. [2] Instead, proxies for human capital used in growth research include such variables as enrollment rates, adult literacy rates, and health indicators. Each of these faces several conceptual and empirical drawbacks. To overcome them, the trend has been to develop education stock estimates based on the mean school years of education per working person in an economy. [3] While this measure is also one-dimensional in nature and subject to other weaknesses, it has the advantages of being a stock measure and of requiring for its construction data sets that are relatively more complete and extensive

3. These advantages have prompted several research efforts aimed at estimating the education stock of countries using different data sets and techniques. [4] This paper presents the results of one such effort. It describes the techniques and the data adopted for the construction of a new series of education stock estimates covering 85 countries for 28 years. The research is part of a project to calculate total factor productivity growth for a large number of industrial and developing countries. The education stock estimates produced in this study will be used together with physical capital stock estimates (forthcoming) to estimate production functions that can be used to derive total factor productivity growth for these countries. It is intended to put in place systems that will update these estimates as and when additional data become available.

4. The basic approach to measuring human capital investment in this paper is similar to that of Lau, Jamison, and Louat (1991), Psacharopolous and Arriagada (1986, 1992), and others who take years of schooling as a proxy measure of human capital. The series are built from enrollment data using the perpetual inventory method

---

[1] See bibliography.

[2] For example, Lavy, Victor (1991) proposes an aggregate measure of human capital investments based on total national expenditures on education, health, nutrition, and family planning.

[3] Education may be the most important component of human capital partly because it also increases the ability of people to live healthier lives and learn more rapidly on-the-job once they enter the labor force.

[4] The parallel research efforts are by Kyriacou (1992), Barro and Lee (1992), Psacharopoulos and Arriagada (1992), and Lau, Jamison, and Louat (1991).

adjusted for mortality. Lau, Bhalla, and Louat (1991) made similar calculations as background for the WDR 1991, but we correct for estimated rates of grade repetition among school-goers and employ country-specific drop-out rates at the primary and secondary levels. Accounting for grade repetition is particularly important in developing countries where enrollments may otherwise be overstated by as much as 25 percent. In addition we have located new sources of historical school enrollments -- as far back as 1930 for many countries and even earlier for others. This reduces the need for backward extrapolation of enrollments to provide the initial estimates of the investment "inventory."

5. The following section describes the perpetual inventory method and its application to the estimation of education stock. A complete, computational "model" is presented in order to show clearly where primary data have been used and where, because of lack of data, estimates, averages, and simplifying assumptions have been employed.

6. The section after that describes the principal sources of data used to construct the education stock series. The results presented in the final section include estimates of the average years of schooling by region. The detailed estimates for primary, secondary, and tertiary stages, by country, covering the period 1960 through 1987 will be published in electronic format.

Measuring Education Stocks

8. We follow Lau, Jamison, and Louat (1990), Psacharopolous and Arriagada (1986), and others in associating human capital with the accumulated years of schooling present in the working age population. The stock of human capital is, therefore, built up from past "investments" in schooling. Unlike physical capital, educational investment is not placed immediately into service. It enters the capital stock when its bearer enters the labor force and is withdrawn when he or she retires. Because we have no evidence on the rate of obsolescence of human capital -- it is generally assumed to be very long-lived compared to physical capital -- we discount investment, prior to its planned retirement, only by the rate of mortality. [5]

9. Psacharopoulos and Arriagada provide estimates of the mean years of schooling in the labor force for 99 countries using census data. For a given country, the generic form of their measure of education stock is given as

$$L = \sum_{i=1}^{n} l_i S_i \qquad (1)$$

where, l_i is the share of persons in the labor force with the ith level of schooling; S_i is the average number of years of education received in the ith level of schooling; i designates the classifications of illiterates (or no education), primary incomplete, primary complete, secondary incomplete, secondary complete, and tertiary education. In this analysis, all levels of schooling are weighted equally.

---

[5] This assumption bears a striking resemblance to the "one-hoss shay" assumption often assumed in the calculation of the gross physical capital stock. But such an assumption appears to ignore two important factors affecting human capital formation and decay. First, "learning-by-doing" can be an important source of human capital acquisition. And second, the quality of education tends to improve with time, thereby leading to a productivity differential between young and old workers. It is assumed in this analysis that these two factors are offsetting.

10. The advantage of their method is that in 66 countries they are able to obtain information directly on the educational characteristics of the current labor force. In the remaining 33 they use information on the educational attainment of the population by age and sex to estimate the schooling profile of the labor force. Furthermore, because they are measuring educational attainment and labor force participation contemporaneously, no correction for expected mortality is required.

11. One problem with census-based measures is that the true value of S_i is not known for those who completed only part of each schooling stage; consequently, analysts are forced to make arbitrary estimates (Psacharopoulos and Arriagada (1986, 1992), Barro and Lee, (1993)). [6] Because repeater rates and drop-out rates tend to vary considerably across countries, education stock estimates based on census survey data are subject to measurement error. Another problem with census-based measures is that they are available only at discrete intervals. In only 34 countries do Psacharapoulos and Arriagada obtain more than one year of data. Barro and Lee (1993) have estimated time series for 129 countries using similar census data, but in their case 77 countries have three or more observations. Perhaps the most serious concern with the Barro and Lee estimates is that they refer to the population aged 25 and over. This can lead to a serious downward bias in the estimates of the education stock because in most developing countries the segment of the population between the ages of 15 and 25 is usually large and growing over time.

12. It was noted earlier (para. 3) that the ultimate objective of estimating education stocks in developing countries was to derive total factor productivity growth estimates for these countries. And measuring the productivity of investment in education over long periods of time requires an unbroken time series of estimates of the education stock. Given sufficiently long series on enrollments, the perpetual inventory method can be used to accumulate a continuous series of estimates of the stock of education. In this paper, the stock of education is defined as the sum of person-school years. Let S_{gt} be the addition to our education stock as a result of 1 year of education in grade g in year t, then the cumulative investment in education that takes place in grades $G = [g_1, g_2]$ between the years $T = [t_1, t_2]$ is

$$H_{GT} = \sum_{G}\sum_{T} S_{gt} \qquad (2)$$

where the summation operators act over the range of index sets G and T.

13. Note that S_{gt} is not necessarily a count of enrollments but could instead measure the "quality" of education or human capital investment. For example, one might specify that

$$S_{gt} = q_{gt} \cdot E_{gt} \qquad (3)$$

where E_{gt} are the enrollments in grade g in year t and q_{gt} is a measure of the "quality" of the additional year of education received in grade g in year t. Lacking plausible, *a priori* measures of quality (either between grades and years in a given country, or between countries), we measure S_{gt} as the total of net enrollments:

$$S_{gt} = E^{*}_{gt} \qquad (4)$$

---

[6] Barro and Lee (1993) have recently produced quinquennial estimates of average years of schooling based on census data supplemented by the perpetual inventory method. Their estimates, however, like those of Psacharopoulos and Arriagada (1986), depend upon an arbitrary assumption concerning S_i -- that census respondents who say they have attended a particular stage of school have completed it. In some censuses, respondents answer that they have partially completed a particular stage of schooling. In such cases, Barro and Lee assume that half the number of years of schooling have been completed in that stage. Since S_i tends not only to fluctuate over time but also differs considerably between countries, these assumptions can lead to an overestimation or underestimation of the level of the education stock.

14. The difference between "gross" enrollments and "net" enrollments is the number of repeaters and dropouts in each grade [7]:

$$E^{*}_{gt} = E_{gt} - R_{gt} - D_{gt} \qquad (5)$$

where R_{gt} and D_{gt} are measures of the number of repeaters and dropouts respectively by grade and year. If we have repetition and drop-out rates, equation (5) can be rewritten as

$$E^{*}_{gt} = E_{gt}(1 - r_{gt} - d_{gt}) \qquad (6)$$

where r_{gt} is the ratio of repeaters to total (gross) enrollments in grade g in year t and d_{gt} is the drop-out rate from grade g in year t.

15. We assume that repeaters acquire the equivalent of one full year of schooling no matter how long they spend in a grade. For simplicity in accounting, we attribute all of the "credit" for completing the grade to the first year of enrollment and deduct all subsequent re-enrollments in the same grade.

16. The proper accounting of dropouts is somewhat problematic. A student who attends part of a year should presumably be credited with a partial enrollment, in which case d_{gt} might measure the average portion of a school year attended by dropouts. But it is open to question whether a dropout has acquired any useful education during the year in which he or she leaves school (Hartley and Swanson (1988)), in which case dropouts should be treated like repeaters and be fully netted out of enrollments in the year in which they dropout. In many school systems, however, dropouts are not reported until the beginning of the following year, at which point they are no longer counted as enrolled. Because practices are so uneven and reporting so imprecise, our procedure is to treat dropouts as if they had completed the year in which they were last enrolled in cases where enrollment data are available by grade. In such instances, no adjustment for dropouts is required in equations (5) or (6) to obtain net enrollments -- dropouts are simply non-enrollments.[8] This may result in a slight overestimate of total years of schooling, especially in school systems characterized by very high rates of dropping out. [9]

17. To determine when human capital is put into service and therefore relevant to determining aggregate output, we need to know when an individual enters or becomes eligible to enter the labor force. In developed countries there are well established statutory limits on the age of entry into the labor force and, in many, the retirement age is also defined either by statute or custom. Practices differ widely in developing countries (Psacharapoulos and Arriagada, 1991). In all countries, there are systematic differences with respect to the sex, education, and social class of the individual. To simplify matters, we count all persons between the ages of 15 and 64 inclusive as constituting the labor force. For this purpose, we ignore cyclical contractions and expansions of the "economically active" population as well as differences in the participation rates of different sub populations.

---

[7] In many estimates, only repeaters are netted out of gross enrollments. However, given our treatment of dropouts in the following analysis, it is convenient to exclude them also from net enrollments.

[8] Given data on enrollments and repetition by grade and assuming that transfers into and out of the system are negligible, the number of dropouts in each year can be estimated using the "grade transition" or "reconstructed cohort" method.

[9] However, for several countries and over different periods, enrollment data by grade are not available. In such cases, we use the drop-out rate in the calculation of net enrollments (see equation 10).

18. To calculate the total stock of education created in a particular stage of schooling and embodied in the labor force in a given year T we must begin with the first year in which the oldest cohort enrolled in the first grade and continue summing E^*_{gt} by cohort through to the last year in which the youngest cohort entered the final grade of the stage. For the moment, let us consider only the stage of primary education. In most school systems, the primary stage includes the first six grades and typically children enter school at age six. In year T, the oldest cohort in the labor force began school in the year T-64+6, while the youngest cohort began in T-15+6. The total net enrollments of the 50 cohorts who entered the primary stage between T-58 and T-9 are calculated by

$$S_{PT} = \sum_{T-58}^{T-9} \sum_{g=1}^{6} E^*_{g,T-g-1} \qquad (7)$$

19. Equation (7) gives the total number of years of schooling (after correcting for drop-outs and repeaters) that were acquired by the population who lived and enrolled in school between the years T-58 and T-9. But we need to "depreciate" the stock of education by the expected losses in each year due to mortality. Assuming that we know the age-specific mortality rates for population in each of the years T-58 through T-9, we can calculate the probability of each enrollee surviving until the year T. Because age in school is closely related to grade, we can associate a probability of survival to the year T with each enrollee in grade g in the year t. Let this probability be $\theta_{g,T}$, then the expected number of surviving enrollments embodied in the work force in the year T is given by

$$\hat{E}_{PT} = \sum_{T-58}^{T-9} \sum_{g=1}^{6} \theta_{g,T-g-1} E^*_{g,T-g-1} \qquad (8)$$

which provides the measure of primary education stock in year T.

20. Expanding net enrollments E^*, which is a function of the gross enrollment level, the retention rate, and the drop-out rate (see equation 6), equation (8) can be now rewritten as

$$\hat{E}_{PT} = \sum_{T-58}^{T-9} \sum_{g=1}^{6} \theta_{g,T-g-1}(1-r_{g,T-g-1}) E_{g,T-g-1}(1-d_{g,T-g-1}) \qquad (9)$$

21. Assuming $r_t=r$ for all t and $d_{g,t}=d$ for all g,t, equation (9) can be simplified to

$$\hat{E}_{PT} = \sum_{T-58}^{T-9} \sum_{g=1}^{6} \theta_{g,T-g-1}(1-r) E_{g,T-g-1}(1-d) \qquad (10)$$

This is the equation used to estimate primary education stock. The same approach is used to calculate secondary and tertiary education stock. These results were subsequently normalized by the working age population to obtain the mean school years of education

Data Sources

22. The education stock estimates discussed in this paper are based on enrollment data acquired from UNESCO sources. Education systems vary from country to country, but UNESCO has drawn up a standard classification -- the International Standard Classification of Education (ISCED) -- and recommendations concerning statistical presentation to ensure that international statistics are as comparable as possible. Primary (or first level) education -- ISCED level 1 -- is defined as having its main function as providing the basic elements of education. Secondary education -- ISCED levels 2 and 3 -- is based upon primary education of at least four years, and can be general or specialized. Therefore, in addition to middle and high schools, secondary education can also

cover vocational and technical courses and teacher training of non-university level. Tertiary education -- ISCED levels 5, 6, and 7 -- is defined as requiring a minimum condition of admission, successful completion of secondary education, or proof of equivalent qualifications (for example, from a university, teachers' college, or higher professional school).

23. If education stock estimates based on the perpetual inventory method are to start in 1960, and it is assumed that the labor force comprises all those between the ages of 15 and 64, then enrollment data series need to begin in 1902. The only previous systematic effort to develop education stock estimates for developing countries based on the perpetual inventory method was by Lau, Jamison, and Louat (1991). They used enrollment data from UNESCO sources, kept 1960 as their first year for most countries, and created the series between 1902 and 1960 using statistical methods.

24. One of the key departures of this study was to try and use actual data on enrollment for the years before 1960, to the extent that this was feasible. Fortunately, data sources were found for primary and secondary education that allowed the construction of a gross enrollment series from 1902 onward for 50 countries and from 1930 onward for 26 more countries. [10] Interpolation techniques were used to fill gaps in the data, but the use of this approach was kept to a minimum. Data from 1950 onward were used for 6 countries, and from 1960 onward for 4 countries -- most of these being in Sub-Saharan Africa.

25. In putting together the longer series, due care was taken to account for national boundary changes. For example, data on pre-independent India had to be divided into components for Bangladesh and Pakistan, and pre-1971 data for Pakistan had to be split into West and East Pakistan, with the latter being added to the Bangladesh series. Data on East Africa, where boundary changes occurred following the Second World War, already incorporated these adjustments in the original data sources, and therefore no further adjustments were made to these series.

26. Where gaps existed between 1902 and the starting year of the series, country-specific growth rates were used to extrapolate the series. The reference periods for the calculation of these growth rates usually covered more than a decade, and were chosen carefully to avoid the inclusion of unusual conditions (such as wars, sudden policy changes, etc.). In no case did the series reach zero when extrapolated backward. [11]

27. The tertiary enrollment series were more difficult to put together and required considerably greater use of interpolated and extrapolated estimates as well as spliced data series from different data sources. The anchoring series was obtained from UNESCO data sources housed in the Bank's Economic and Social Database (BESD). These were supplemented by various UNESCO yearbooks, Mitchell (1982), UNESCO (1958), UNESCO (1961),

---

[10] For primary and secondary enrollment data, the principal data sources were:

- UNESCO Educational Statistics available in the World Bank Economic and Social Database (BESD);
- UNESCO. 1958. World Survey of Education, Volume III (New York: United Nations);
- UNESCO. 1961. World Survey of Education, Volume IV (New York: United Nations);
- Mitchell, B.R. 1980. European Historical Statistics, 1750-1970, Second Edition (New York: Facts on File);
- Mitchell, B.R. 1982. International Historical Statistics, Africa and Asia, Second Edition (New York: New York University Press);
- Mitchell, B.R. 1983. International Historical Statistics, The Americas and Australasia, (Detroit: Gale Research Cmpany);

[11] In the paper by Lau, Jamison, and Louat (1991), the enrollment series reached zero in a number of cases.

and a variety of national sources. [12] In cases where the national sources and UNESCO series did not match (usually for definitional reasons), the two were spliced by applying the annual growth rate implicit in the national series. Again, due account was taken of changes to national boundaries.

28. The greater use of "statistically manufactured" gross enrollment data in creating the tertiary education stock series gives these data lower information content and makes them less reliable in regression estimates. Although the data appear to behave in line with *a priori* reasoning, as subsequent sections of this paper will show, they should be used with some caution.

29. Data on repeater rates by grade were available in five-yearly intervals between 1960 and 1985 for most countries and were restricted to primary and secondary education. Using these data as benchmarks, annual data series were constructed by simple interpolation for the years between 1965 and 1985. It was assumed that for the years before 1965, the repeater rate remained at the 1965 rate, and for the years after 1985 at the 1985 rate. Finally, owing to the dearth of data on enrollments by grade, it was necessary to construct a weighted average repeater rate (the weights being the enrollments by grade in the few years when such data were available). It should be noted that education stock estimates are particularly sensitive to repeater rate levels -- if repeater rates were to have doubled throughout the period from 1902 to 1985, the education stock estimate would have halved, giving an elasticity of -0.5. Obtaining a longer and more accurate series of observations on repeater rates is, therefore, of some importance in future research work on human capital stock estimation.[13]

30. Data on the other two variables used in the construction of the net enrollment series -- age-specific mortality rates and drop-out rates -- were even more difficult to acquire. Data on age-specific mortality rates are sparse, so series were developed for a representative country in each region and then applied to all the countries of that region. [14] When compared to errors in the repeater rate estimates, errors in the mortality rate estimates are of less consequence to the final education stock estimates -- a doubling of the mortality rate, for instance, tends to reduce the education stock by between 2 and 3 percent, an elasticity of -0.02 to -0.03. In the absence of data on drop-out levels, drop-out rates were calculated using available information on gross enrollments, mortality rates, and repeater rates.

31. Apart from shortcomings in the methodology and the dearth of data on enrollments, mortality rates, repeater rates, and drop-out rates (discussed above), the data are subject to three important, but well known, weaknesses that need to be kept in mind. First, the data do not measure the quality of education, and this makes inter temporal as well as cross-country comparisons difficult to interpret. Unfortunately, no good indicator of the quality of education is available easily for developing countries. One popular measure often used for this purpose, the teacher-student ratio, does not appear to be strongly related to the value added of the schooling system (see Barro, 1991; Behrman and Rosenzweig, 1992). [15] Second, enrollment data suffer from the same problem as other developing country data -- their reporting tends to get more accurate with development, making inter temporal

---

[12] Details can be provided upon request.

[13] It should be noted that the paper by Jamison, Lau, and Louat (1991) did not use repeater rate rate data to derive net enrollment levels from the gross enrollment series. Given the large differences in repeater rates across countries, this obviously leads to some differences between their data series and ours.

[14] The representative countries are the same as the ones used by Lau, Louat, and Bhalla (1991) -- Egypt for the Middle East and North Africa, the Philippines for East Asia, Sri Lanka for South Asia, and Brazil for most Latin American economies. The data source for mortality rates was various issues of the United Nations Demographic Yearbook between 1945 and 1987.

[15] We would like to explore further the use of drop-out rates as a proxy for education quality.

comparisons subject to error. In the case of the education stock series, with its database on enrollments stretching back to 1902, this problem could be potentially serious. And third, years of schooling as a proxy for education is subject to error in cross-country analysis because the number of days and hours of schooling per year can vary substantially across countries. Although the ISCED system developed by UNESCO attempts to take into account such differences in its data compilation, some differences in national definitions inevitably remain.

32. This study is not alone in having to deal with data and methodological problems, however. Other studies using alternative databases and methodological approaches to estimating education stock also confront major, perhaps prohibitive, shortcomings. For example, studies that estimate education stock on the basis of UNESCO data on the distribution of the population 25 years of age and older by levels of educational attainment (see, for example, Barro and Lee (1992); and Psacharopoulos and Arriagada (1986)) have several weaknesses. These problems are documented in Behrman and Rosenzweig (1992), who note that such studies are based on a limited number of national surveys and censuses conducted over a wide variance of years and using a wide variety of definitions. These problems are catalogued in detailed notes to the data presented by UNESCO. For example, in 14 countries the age ranges differ from the standard definitions, in 5 countries the answer "not stated" is combined with "no schooling", and in 9 countries illiteracy is interpreted to be "no schooling". A number of smaller definitional differences also render comparisons difficult. Moreover, this data set suffers from the standard problems of ignoring quality differences across countries, variations in length of school days and school years, and the importance of non-schooling education. And finally, census surveys do not report the number of years of schooling of individuals, merely whether they attended primary/secondary/tertiary educational institutions. The high incidence of repeaters and dropouts apparent in enrollment data but not captured in census surveys suggests that education stock figures based on census survey data alone would tend to be overestimated.

33. Since education stock estimates tend to be based on sparse data of uneven quality, testing the final results would provide some indication of their quality and reliability. In the section that follows, the database on education stock presented in this paper is analyzed briefly and compared to measures of education stock prepared by other researchers.

A Brief Description of the Results

34. The education stock series presented in this paper are available for 85 countries for the years 1960-87. The average education stock measures the mean school years of education of the working age population (defined as the population between the ages 15 and 64), and is the sum of primary, secondary, and post-secondary average education stock. All the important countries in each developing region are covered except for the republics of the former Soviet Union. The data are particularly weak for Sub-Saharan Africa, especially in the case of post-secondary education, but also for primary and secondary education. For the most part, data for the other developing regions are based on longer time series on enrollment rates and tend to have stronger background documentation.

35. A comparison of the average education stock between industrial and developing countries, and between different developing regions, provides some interesting insights (Table 1). As one would expect, the mean school years of education in developing countries is less than half that of industrial countries. But the overall growth of average education stock in industrial countries appears to have slowed, owing largely to the fact that the primary education stock has declined marginally. The most rapidly expanding component of the education capital stock in industrial countries has been in post-secondary education. But, despite such growth, the mean school years of education in post-secondary educational institutions stands at less than 1.0. Similarly, the mean school years of education in secondary schools is below 3.0 (compared to a potential maximum of 6.0), indicating significant room for further expansion.

36. In the developing regions, the bulk of the average education stock derives from primary school education; the mean school years of education in secondary and tertiary schools, when taken together, are less than 1.0. At the same time, the most rapidly expanding components of the average education stock are the tertiary and secondary components. Developing Europe, Latin America, and East Asia have the highest average education stock among the developing regions, Sub-Saharan Africa and South Asia the lowest. In general, the regions where the average education stock is high, such as Latin America and Developing Europe, the growth of the stock has been low; and in regions where the average stock is low, such as in Sub-Saharan Africa, growth has been rapid. Asia, however, presents an exception. In South Asia, growth in the average education stock has been relatively low despite its level also being low; in East Asia, both the level and the growth rate are high.

Table 1: **Level and growth of average education stock in industrial and developing countries** (school years of education per person between the ages of 15 and 64)

	Stock in 1987				Growth 1960-87 (% per year) [a]			
	Primary	Secondary	Tertiary	Total	Primary	Secondary	Tertiary	Total
Industrial	6.53	2.60	0.88	10.0	-0.5	2.2	4.9	0.3
Developing	3.70	0.72	0.06	4.48	3.2	6.0	5.3	4.0
East Asia	4.38	0.72	0.03	5.13	3.9	9.2	3.4	4.2
South Asia	2.39	0.88	0.12	3.39	2.9	4.3	6.4	3.3
Latin America	4.65	0.56	0.31	5.52	1.5	5.3	6.7	2.0
Sub-Saharan Africa	2.33	0.19	0.02	2.54	3.9	9.7	12.6	4.2
Dev. Europe	4.39	0.88	0.23	5.50	1.6	4.0	6.0	2.0
MENA	3.24	1.13	0.41	4.79	2.2	1.9	6.3	2.3
World	4.38	1.17	0.29	5.85	1.0	2.9	4.4	1.4

[a] OLS growth rates.
Source: IEC

Table 2: **Correlation coefficients of average education stock using alternative estimation procedures**

	IEC	PA	B-L	K
IEC	1.0	..	..	..
PA	0.84	1.0	..	..
B-L	0.81	0.92	1.0	..
K	0.89	0.86	0.89	1.0

Source: IEC: International Economics Department, The World Bank
PA: Psacharapoulos and Arriagada (1991)
BL: Barro and Lee (1993)
K: Kyriacou (1991)

37. The IEC data were compared to other databases on education stocks that have been prepared using different

techniques. The correlation coefficients between IEC data and the databases of Psacharapoulos and Arriagada (1991), Barro and Lee (1992), and Kyriacou (1992) all exceed 0.8 (Table 2). [16] The correlation coefficient between the data prepared by Barro and Lee and that prepared by Psacharopoulos and Arriagada have a correlation coefficient of 0.92, reflecting the use of similar data (census surveys) and methodologies. Both the IEC and the Barro-Lee data are correlated to a similar degree with the Kyriacou database; this is significant, since the three use altogether different techniques for estimating education stock.

38. It was noted earlier that the education stock data presented in this paper were based on an annual enrollment series that went as far back as 1930 and in several cases to 1902. The series that stopped in 1930 had to be extrapolated backward to 1902, and this could have introduced a measurement bias. Prima facie, the countries with incomplete series tended to be low and middle income countries and the countries with complete series tended to be high income. The measurement bias, if any, could therefore be expected to be correlated with per capita GNP.

39. The IEC data were, therefore, compared to the data series generated by Psacharopoulos and Arriagada, and the difference was regressed against per capita GNP as well as against time; this was done for the entire sample of countries and for the low and middle income group only. In no case did the coefficient of the GNP per capita variable exceed 0.0003, indicating virtually no association between the differences between the alternative data sets and per capita income (Table 3). A similar exercise was conducted by comparing the IEC and the Barro-Lee data set, and the findings were identical.

Table 3: **Regressions estimates relating differences between average education stock data from alternative sources and per capita income**

Independent variable / Dependent variable	Constant	GNP per capita	Time	Adj. R^2	Prob>F
Difference between IEC data and:					
Barro-Lee data					
All countries	0.327**	-0.0001**		0.033	0.000
All countries	-11.516	-0.0001**	0.0060	0.036	0.002
Only LMICs	0.051	0.0002**		0.019	0.017
Only LMICs	30.230	0.0003**	-0.1530	0.022	0.024
Psacharopoulos data					
All countries	-0.083	-0.0001**		0.047	0.055
All countries	113.371	-0.0001	-0.0574	0.073	0.007
Only LMICs	-0.411	0.0002		0.004	0.038
Only LMICs	109.635	0.0003*	-0.0557	0.038	0.085

** Significant at the 95 percent confidence level.
* Significant at the 90 percent confidence level.
Source: Psacharopoulos and Arriagada (1986); Psacharopoulos and Arriagada (1992); Barro and Lee (1992); authors' estimates.

[16] The correlation between the Kyriaou and Psacharopoulos databases could not be calculated because of an insufficient number of overlapping years.

40. Simple tests were also conducted on the IEC education stock data to check if there was any association with a wide range of social, especially health, indicators . The results appear encouraging. The correlation coefficient between average education stock and a variety of social indicators of development, including such indicators as the fertility rate, birth rate, adult literacy rate, and infant mortality rate seem to have the right signs and orders of magnitude that one would expect (see Table 4 and Appendix Table 1). [17] For example, the overall correlations across countries are high -- all above 0.8 -- but there is considerable differentiation once countries are divided into different income groups. [18] For low income countries, the correlation coefficients are often not as high as in the case of middle income countries, suggesting one of three possibilities: education standards are not as high in low income countries; other factors associated with low income levels tend to prevent education from raising social indicators; or the data on average education stock (as well as the social indicators with which they are being compared) are inferior in quality in low income countries. In the case of high income countries, the correlation coefficients are small and, in some cases, of the wrong sign. This is not altogether surprising; a clear and strong relationship between education and health would not be expected to prevail in economies at the higher end of the per capita income spectrum.

Table 4: Correlation coefficients of average education stock and selected social indicators of development

	Fertility rate	Birth rate	Adult literacy rate	Infant mortality rate
Low income	-0.48	-0.51	0.61	-0.68
Middle income	-0.63	-0.65	0.81	-0.65
High income	-0.07	-0.15	0.17	-0.16
All	-0.80	-0.82	0.84	-0.82

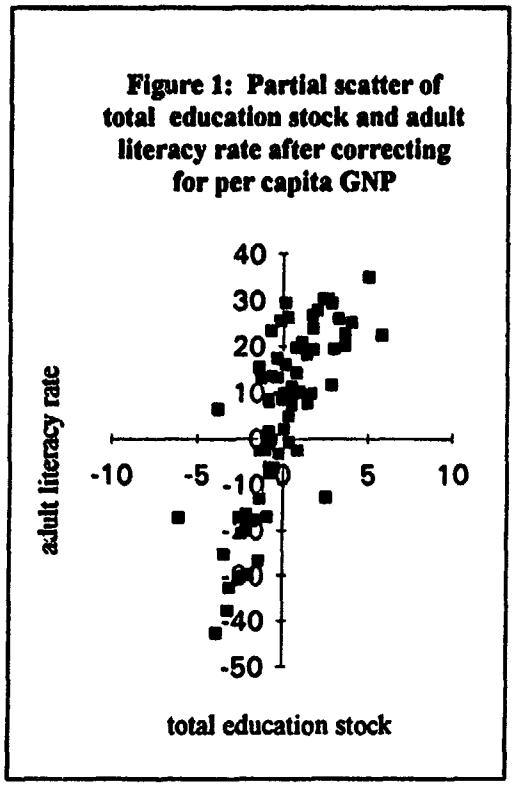
Source: IEC and BESD.

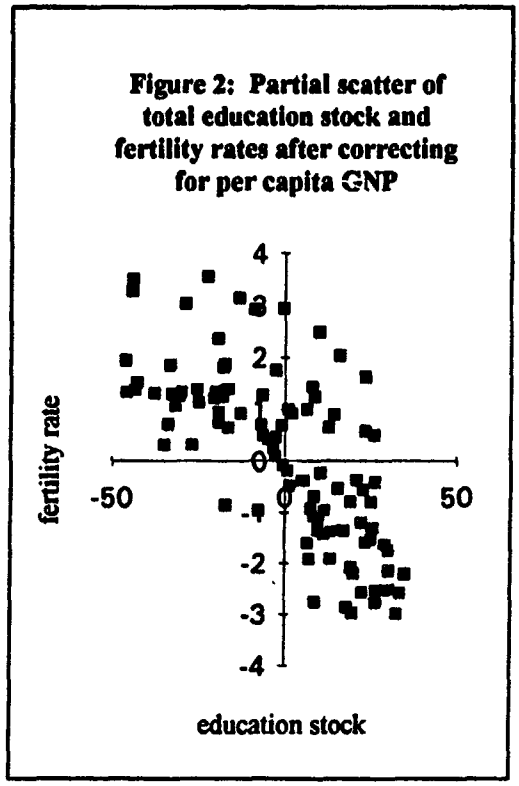
Table 5: Correlation coefficients of fertility rates and average education stock by level of education

Country group	Primary	Secondary	Tertiary	Total
Low income	-0.31	-0.73	-0.33	-0.48
Middle income	-0.63	-0.34	-0.33	-0.63
High income	0.10	-0.28	0.06	-0.07
All	-0.72	-0.72	-0.65	-0.80

Source: IEC and BESD

[17] The correlation coefficients were calculated using cross section data for one particular year. For education stock, this year was 1987; for the other social indicators, the latest available data was used from the World Bank's BESD database.





41. When average education stock is broken down into primary, secondary, and tertiary (i.e. post-secondary), the correlation coefficients reinforce these initial impressions. For example, the association between average education stock and fertility rates appears to be relatively strong at the secondary level stage for low income countries; but in the case of middle income countries, the association appears strongest at the primary education stage (see Table 5). This tends to support the view that owing to the poor quality of education in low income countries, an additional number of years of education are required to make a significant dent in fertility rates. Among high income countries, as noted earlier, the relationship between average education stock and fertility rates appears weak.

42. Since both average education stock and other social indicators would be expected to be associated strongly with income levels (at least among low and middle income economies), the relatively high correlation coefficients described above should come as little surprise. To eliminate the effect of per capita income, both average education stock as well as adult literacy rates and fertility rates were regressed against per capita income. [19] The orthogonal components from these regressions were then plotted against each other (see Figures 1 and 2). The data shows that the relationship between average education stock and the adult literacy rate, and between average education stock and fertility rates, is extremely strong even after correcting for the influence of per capita income. Similar results were achieved when other social indicators were used.

---

[18] The income groups used here -- low, middle, and high -- use standard Bank definitions as they appear in World bank (1992b) and World Bank (1992a).

[19] The per capita income measure used was the latest per capita GNP calculated according to the World Bank's Atlas method.

43. In assessing the comparisons described above, it is important to recognize that the measures of social development that were used to test the average education stock estimates are themselves subject to considerable error. For example, the infant mortality rate estimates are generally of poor quality (especially for low income countries): they are often based on observations from old censuses; the sample used in the surveys are often not representative of the entire population; and so-called *indirect estimation techniques* often make assumptions about past trends that may not be accurate. [20] Given concerns about the accuracy of the education stock data as well as the social indicators data, the relatively strong association between them found in this paper provides some source of comfort. Of course, any full fledged analysis of the interrelationships between education and other social indicators of development would require more detailed econometric work which is outside the scope of this paper.

Conclusion

44. This paper presents a description of the data and methodology used in deriving annual education stocks for 85 industrial and developing economies over the period 1965-87. The data compare favorably with other data series measuring the same variable prepared by other authors using different methods and data sources. The paper also finds a striking association between average education stock and other indicators of social development, especially once the effects of per capita income are corrected for. The weak points of the data lie in the estimates of repeater and drop-out rates, as well as mortality rates, which can be improved considerably if more original data were available. Nevertheless, the inclusion of such variables in estimating net enrollment levels was itself an advance over previous efforts at calculating education stocks. In addition, the estimates of education stock are particularly uncertain for ten countries for which the enrollment series begin after the Second World War. Here, again, the availability of additional data would significantly improve these estimates.

45. The education stock data presented in this paper were prepared as part of a larger research effort to estimate total factor productivity growth for a wide range of countries. It was intended that they (the education stock data) would be used as proxies for human capital stock. The high correlation between average education stock and other indicators of human capital -- especially a variety of health indicators -- suggest that the use of such a proxy would be a reasonable step. A similar effort has been launched to estimate stocks of physical capital in industrial and developing countries. The results of this effort will be described in a forthcoming paper.

---

[20] See Bos, E., M. Vu, and P. Stephens (1992).

Appendix Table 1: Correlation Coefficients Between Average Education Stock Estimates and Social Development Indicators

	Fertility rate 1988	Birth rate 1988	Death rate 1988	Low birth weight babies %, 1988	Health expend. as percent of GDP	Daily calorie supply % of requirements	Urban population (%, 1988)	Adult literacy rate (%, 1985)	Scientists and technician per '000 of population	Infant mortality ra per '000 live birth
All countries										
Primary	-0.72	-0.73	-0.59	-0.56	0.59	0.60	0.68	0.83	0.50	-0.77
Secondary	-0.72	-0.77	-0.30	-0.53	0.74	0.71	0.61	0.64	0.73	-0.69
Post-secondary	-0.65	-0.67	-0.42	-0.50	0.59	0.63	0.68	0.66	0.55	-0.65
Total	-0.80	-0.82	-0.54	-0.61	0.71	0.71	0.73	0.84	0.66	-0.82
Low income countries										
Primary	-0.31	-0.35	-0.56	-0.19	0.33	0.24	0.22	0.65	0.16	-0.59
Secondary	-0.73	-0.75	-0.68	0.19	-0.15	0.49	0.16	0.19	0.35	-0.60
Post-secondary	-0.33	-0.34	-0.43	-0.12	0.00	0.68	0.40	-0.06	0.86	-0.28
Total	-0.48	-0.51	-0.68	-0.12	0.25	0.37	0.26	0.61	0.33	-0.68
Middle income countries										
Primary	-0.63	-0.62	-0.50	-0.10	0.33	0.11	0.46	0.79	0.47	-0.61
Secondary	-0.34	-0.43	-0.30	-0.40	0.24	0.48	0.24	0.41	0.52	-0.49
Post-secondary	-0.33	-0.35	-0.31	0.03	0.15	0.03	0.49	0.61	0.41	-0.34
Total	-0.63	-0.65	-0.51	-0.17	0.34	0.21	0.48	0.81	0.59	-0.65
High income countries										
Primary	0.10	0.07	0.13	-0.13	0.19	0.49	-0.27	0.01	-0.03	0.00
Secondary	-0.28	-0.42	0.33	0.09	0.23	0.08	-0.28	0.31	0.32	-0.36
Post-secondary	0.06	0.12	-0.15	0.34	-0.03	0.30	-0.15	0.06	-0.08	0.16
Total	-0.07	-0.15	0.23	0.01	0.24	0.40	-0.34	0.17	0.13	-0.16

Source: IEC; BESD

Appendix Table 2: Distribution of Countries by Their Rankings

Ranking	No. of Countries	Number of years for which enrollment data extrapolated
1	50	None
2	26	28 (1902-1929)
3	6	48 (1902-1949)
4	4	58 (1902-1960)

The following is the list of countries by region for which the educational stock series has been constructed. The numbers indicate the ranking of the countries.

AFRICA	RANK
Angola	2
Cameroon	2
Cote d'Ivoire	1
Ethiopia	3
Ghana	1
Kenya	2
Liberia	4
Madagascar	3
Malawi	2
Mali	4
Mauritius	1
Mozambique	2
Nigeria	1
Rwanda	4
Senegal	4
Sierra Leone	2
Sudan	2
Tanzania	2
Uganda	3
Zaire	2
Zambia	2
Zimbabwe	1

LATIN AMERICA	RANK
Argentina	1
Bolivia	1
Brazil	1
Chile	1
Colombia	1
Costa Rica	1
Ecuador	1
El Salvador	1
Guatemala	1
Honduras	1
Haiti	1
Jamaica	2
Mexico	1
Panama	2
Peru	1
Paraguay	1
Uruguay	1
Venezuela	1

EAST ASIA	
China	
Hong Kong	3
Indonesia	1
Japan	1
Korea, Republic of	2
Malaysia	1
Philippines	1
Singapore	1

SOUTH ASIA	
Bangladesh	1
India	1
Myanmar	1
Pakistan	1
Sri Lanka	1

Taiwan	1
Thailand	2

EUROPE, MIDDLE EAST, NORTH AFRICA	RANK
Algeria	1
Egypt	1
Greece	1
Iran	2
Iraq	2
Israel	2
Jordan	3
Morocco	2
Portugal	1
Spain	2
Syrian, Arab Republic	2
Tunisia	1

HIGH INCOME COUNTRIES	RANK
Australia	1
Austria	2
Belgium	1
Canada	1
Cyprus	1
Denmark	1
Finland	1
France	1
Germany	1
Ireland	2
Italy	1
Netherlands	1
New Zealand	1
Norway	1
Sweden	2
Switzerland	2
United Kingdom	2
United States	1
Turkey	2

Appendix Table 3 : Reference Period Used for Estimation of Growth Rates of Enrollments

COUNTRY	PRIMARY	SECONDARY
Angola	1929-1950	1930-1950
Austria	1917-1950	1918-1950
Cameroon	1910-1950	1910-1950
China	1930-1950	1949-1960
Ethiopia	1930-1950	1948-1970
Hong Kong	1930-1950	1950-1960
Iran	1920-1950	1920-1950
Iraq	1927-1950	1927-1950
Ireland	1924-1950	1924-1950
Israel	1920-1950	1920-1950
Ivory coast	1936-1950	Not needed
Jamaica	1932-1950	1950-1960
Jordan	1930-1950	1952-1970
Kenya	1926-1950	1926-1950
Korea	1910-1950	1912-1950
Liberia	1950-1970	1950-1970
Madagascar	1930-1950	1945-1950
Malawi	Not needed	19451960
Mali	1948-1970	1948-1970
Morocco	1913-1950	1914-1950
Mozambique	1926-1950	1920-1950
Panama	1930-1950	1934-1950
Rwanda	1950-1970	1950-1970
Senegal	1948-1970	1950-1970
Sierra Leone	1936-1950	Not needed
Spain	1920-1950	1920-1950
Sudan	1930-1950	1930-1950
Sweden	1920-1950	1920-1950
Switzerland	1920-1950	1920-1950
Syria	1927-1950	1927-1950
Tanzania	1921-1950	1920-1950
Thailand	1913-1950	1913-1950
Turkey	1923-1950	1923-1950
Uganda	1950-1970	1930-1950
United Kingdom	1920-1950	1930-1950
Zaire	1930-1950	1930-1950
Zambia	1927-1950	1926-1950

Bibliography

1. Barro. R. 1991. "Economic Growth in a Cross-Section of Countries", in The Quarterly Journal of Economics, May 1991, pp. 407-449.

2. Barro, R. and J. Lee. 1993. International Comparisons of Educational Attainment, paper presented at the conference on "How Do National Policies Affect Long-Run Growth", World Bank, Washington D.C.

3. Becker, Gary. 1964. Human Capital (New York: Columbia University Press).

4. Becker, G., K. Murphy, and R. Tamura. ___. "Human Capital, Fertility, and Economic Growth" in Journal of Political Economy, Vol. 98, No.5, pt.2, S12-S37.

5. Behrman, J.R. and M.R.Rozensweig. 1992. The Quality of Aggregate Inter-Country Time-Series Data on Educational Investments and Stocks, Economically Active Populations, and Employment; paper prepared for the Conference on Data Base of Development Analysis, Yale University, May 15-16, 1992.

6. Behrman, J. and A. Deolalikar. 1987. "Will Developing Country Nutrition Improve with Income? A Case Study for Rural South India", in Journal of Political Economy, Vol. 95, No.3, pp. 492-507.

7. Bos, E., M. Vu, and P. Stephens. 1992. "Sources of World Bank Estimates of Current Mortality Estimates", Policy and Research Working Paper Series No. 851, Population and Human Resources Department, World Bank, Washington D.C.

8. Cochrane, S., D. O'Hara, and J. Leslie. 1980. "Effects of Education on Health", World Bank Staff Working, No. 405, World Bank, Washington D.C.

9. Hartley, M.J. and E. V. Swanson. 1989. Retention of Basic Skills Among Dropouts from Egyptian Primary Schools, in Robert S. Mariano (ed.), Advances in Statistical Analysis and Statistical Computing, (London: JAI Press).

10. Kyriacou, G.A. 1991. Level and Growth Effects of Human Capital: A Cross-Country Study of the Convergence Hypothesis, mimeo., Department of Economics, New York University

11. Lau, L.J., S. Bhalla, and L.J.Louat. 1991. "Human and Physical Capital Stock in Developing Countries - Construction of Data and Trends", draft mimeo., World Development Report, The World Bank.

12. Lau, L.J., D.T. Jamison, and F. Louat. 1991. "Education and Productivity in Developing Countries: An Aggregate Production Function Approach", PRE Working Paper Series No.612, World Bank, Washington D.C.

13. Lavy, Victor. 1991. "Human Capital Accumulation and Economic Growth: An Empirical Evaluation", mimeo., Population and Human Resources Division, The World Bank.

14. Psacharopoulos, G. and A.M. Arriagada. 1986. "The Educational Composition of the Labor Force: An International Comparison" in International Labor Review, Vol. 125, No.5, pp.561-574.

15. Psacharopoulos, G. and A.M. Arriagada. 1992. "The Educational Composition of the Labor Force: An International Update", PHREE Background Paper Series No. PHREE/92/49, Education and Employment Division, Population and Human Resources Department, The World Bank, Washington D.C.

16. World Bank. 1992a. Global Economic Prospects and the Developing Countries, 1992 (Washington D.C: World Bank).

17. World Bank. 1992. World Development Report (New York: Oxford University Press).

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1102	Did the Debt Crisis or Declining Oil Prices Cause Mexico's Investment Collapse?	Andrew M. Warner	February 1993	M. Divino 33739
WPS1103	Capital Mobility in Developing Countries: Some Measurement Issues and Empirical Estimates	Peter J. Montiel	February 1993	R. Vo 31047
WPS1104	Trade Policy Reform in Latin America and the Caribbean in the 1980s	Asad Alam Sarath Rajapatirana	February 1993	J. Troncoso 37826
WPS1105	Estimating Quasi-Fiscal Deficits in a Consistency Framework: The Case of Madagascar	Philippe Le Houerou Hector Sierra	February 1993	N. Velasco 34346
WPS1106	Improving Women's Access to Higher Education: A Review of World Bank Project Experience	Halil Dundar Jennifer Haworth	February 1993	S. David 33752
WPS1107	Financial Reform Lessons and Strategies	Gerard Caprio, Jr. Izak Atiyas James Hanson	February 1993	W. Pitayatonakarn 37664
WPS1108	Public Output and Private Decisions: Conceptual Issues in the Evaluation of Government Activities and Their Implications for Fiscal Policy	Thanos Catsambas	February 1993	A. Correa 38549
WPS1109	Risk Management and Stable Financial Structures	Andrew Sheng Yoon Je Cho	March 1993	M. Raggambi 37664
WPS1110	What Would Happen If All Developing Countries Expanded Their Manufactured Exports?	Will Martin	March 1993	D. Gustafson 33714
WPS1111	Foreign Investment Law in Central and Eastern Europe	Cheryl W. Gray William Jarosz	March 1993	M. Berg 31450
WPS1112	Privatization, Concentration, and Pressure for Protection: A Steel Sector Study	Ying Qian Ronald C. Duncan	March 1993	S. Lipscomb 33718
WPS1113	The Lucky Few Amidst Economic Decline: Distributional Change in Côte d'Ivoire As Seen Through Panel Data Sets, 1985-88	Christiaan Grootaert Ravi Kanbur	March 1993	E. Vitanov 38400
WPS1114	Does Price Uncertainty Really Reduce Private Investment? A Small Model Applied to Chile	Anita George Jacques Morisset	March 1993	D. Bievenour 37899

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1115	Looking at the Facts: What We Know about Policy and Growth from Cross-Country Analysis	Ross Levine Sara Zervos	March 1993	D. Evans 38526
WPS1116	Implications of Agricultural Trade Liberalization for the Developing Countries	Antonio Salazar Brandão Will Martin	March 1993	D. Gustafson 33714
WPS1117	Portfolio Investment Flows to Emerging Markets	Sudarshan Gooptu	March 1993	R. Vo 31047
WPS1118	Trends in Retirement Systems and Lessons for Reform	Olivia S. Mitchell	March 1993	ESP 33680
WPS1119	The North American Free Trade Agreement: Its Effect on South Asia	Raed Safadi Alexander Yeats	March 1993	J. Jacobson 33710
WPS1120	Policies for Coping with Price Uncertainty for Mexican Maize	Donald F. Larson	March 1993	D. Gustafson 33714
WPS1121	Measuring Capital Flight: A Case Study of Mexico	Harald Eggerstedt Rebecca Brideau Hall Sweder van Wijnbergen	March 1993	H. Abbey 80512
WPS1122	Fiscal Decentralization in Transitional Economies: Toward a Systemic Analysis	Richard Bird Christine Wallich	March 1993	B. Pacheco 37033
WPS1123	Social Development is Economic Development	Nancy Birdsall	April 1993	S. Rothschild 37460
WPS1124	A New Database on Human Capital Stock: Sources, Methodology, and Results	Vikram Nehru Eric Swanson Ashutosh Dubey	April 1993	M. Coleridge-Taylor 33704