

Gini Coefficients of Educational Attainment: Age Group Specific Trends in Educational (In)equality

Jesús Crespo-Cuaresma * Samir K.C. † Petra Sauer ‡

April 3, 2012

1 Introduction

The broad concept of human capital comprises aspects inherent in humans, which are - as in the case of congenital abilities, skills and talent - either given or - as in the case of education, experience and health - develop over time. In this context education obtained through the formal schooling system takes on an essential role in linking those two components of human capital. On the one hand, education is able to compensate for congenital differences as well as educational gaps arising in early childhood. Equal access to education therefore secures equality of opportunities. On the other hand, education constitutes the foundation of individuals' professional careers and affects, among other things, life-time income and health - thus well-being over the whole life-cycle. Individuals not only interact with each other and with other economic agents by engaging in production, supplying labor and consuming goods but also participate in political-decision making. The aggregate level of education hence crucially shapes economies' welfare. This is true for industrialized countries, where human capital is vital for technology driven, sustainable development and for developing countries,

*Institute for Fiscal and Monetary Policy, University of Economics and Business; Wittgenstein Centre

†World Population Program, International Institute for Applied Systems Analysis (IIASA); Wittgenstein Centre)

‡RI for Human Capital and Development/Wittgenstein Centre, Vienna University of Economics and Business

where education is an essential factor for hauling societies out of poverty. However, even if access to education is among the basic human rights, huge educational gaps between various groups exist within countries, whereby education is prevented from unfolding its welfare enhancing effects entirely.

Until now, the main focus has been on measuring the average level of education, while not much attention has been devoted to its distributional dimension. Yet, allowing for the second moment of education should greatly improve our understanding of population dynamics as well as the mechanics and channels linking education to economic outcomes in a broad sense, including not only economic growth but also inequality, poverty alleviation, democracy and political instability.

We calculate Gini Coefficients of Educational Attainment for 175 countries for 1960-2000. Using the recently constructed IIASA/VID dataset of populations by age, sex, and levels of education enables us to incorporate the demographic dimension into our analysis. Thereby we are able, not only to analyze age group-specific trends but also to distinguish an unequal distribution across age groups - as it arises from educational expansion - from an unequal distribution within age groups. Thus, we aim at providing a measure which gives a comprehensive picture of the degree of inequality in educational attainment within and across societies.

In section 2, we provide an overview on how the measurement of educational inequality has evolved and the prime implications which can be deduced therefrom. The methodology of calculating Education Gini Coefficients as well as their expected behavior is described in sections 3 and 4, respectively. After describing the education dataset in section 5 we present our results in section 6. Section 7 summarizes and concludes.

2 Measuring Educational Inequality

Two measures have been used primary in order to investigate the distributional dimension of education.¹ The standard deviation of schooling has been used to explore the impact of the distribution of education on income growth and poverty reduction (e.g. Birdsall and

¹Fan *et al.* (2002) also calculate Theil indices of educational attainment and Castelló & Doménech (2002) additionally report the distribution of education by quintiles.

Londoño 1997; López et al. 1998) as well as income inequality (e.g. Lam and Levison 1991; Inter-American Development Bank 1999). Furthermore, similar to the concept of income inequality, standard deviations have been applied for testing the existence of an Education Kuznets Curve, i.e. an inverted U-shape relation between the distribution and the average level of schooling. By relating the standard deviation to average years of schooling of 140 countries in 2000, Fan et al. (2002), confirm the findings of Londoño (1990) and Ram (1990) that educational inequality first increases as the average level of schooling rises, and, after reaching a peak, starts to decline.

However, the standard deviation is only a measure of absolute dispersion and it does not provide a consistent picture of the distribution of education, especially for countries with very low and high levels of average schooling. As a measure of relative inequality, the Education Gini Coefficient is therefore seen as a more consistent and robust measure of the distribution of education. Some earlier studies used schooling enrollment figures (e.g. Maas and Criel 1982, Rosthal 1978 and Sheret 1988) or education finance data for calculating Education Gini Coefficients for small samples of mostly developing countries. These data bases do not accurately reflect the existent stock of human capital, though. Enrollment ratios are flow variables that add to future stock of human capital. Even if they constitute an indicator of access to education or equality of opportunity, they do not capture the degree of inequality in educational outcomes. Due to the availability of datasets, which, by reporting attainment figures for various education levels, provide a more appropriate picture of the actual distribution of education, more recent studies calculate the Education Gini based on educational attainment of the concerned population. López *et al.* (1998) were the first who derive Gini coefficients for 12 countries from attainment data. Fan *et al.* (2001) provide a detailed description of the underlying methodology, calculate Education Gini's for 85 industrialized and developing countries for the period from 1960 to 1990 and relate them to average educational attainment, educational gender-gaps and real GDP per capita. They further extend the sample to 140 countries from 1960 to 2000 in their subsequent work (see Fan et al. 2002). Thereafter, their approach has been utilized for deriving a consistent indicator of the distribution of education, that can be related to income distribution (e.g. Checchi 2000) and income growth. Non-conforming with earlier results, plotting Gini Coefficients against average educational attainment consistently does not support an Education

Kuznet's Curve but reveals a strong negative relation between the degree of inequality and the average level of educational attainment.

3 Lorenz Curves and the Education Gini Coefficient

The Gini coefficient, named after the Italian statistician and sociologist Corrado Gini, is a principal measure of statistical dispersion. Even if this summary measure is being used for describing the distribution of a wide range of socio-economic factors, e.g. wealth, consumption, land or health, its main application has turned out to be determining of the degree of inequality in the distribution of income. The computation of the Gini index is based on the Lorenz Curve as its geometrical representation. In the case of income, the Lorenz curve is a continuous function which results from plotting the cumulative percentage of the population against the cumulative percentage of income. The surface of the area between the forty five-degree line, along which everybody earns the same amount of income, and the Lorenz Curve determines the degree of inequality in the income distribution. The bigger the surface, the more unequal the distribution of income. The Gini Coefficient is then calculated as the ratio between this surface and the whole triangle below the "*egalitarian line*".

In general, education comprises formal and informal as well as qualitative aspects. Individuals will hence differ, among other things, according to the quantity and quality of their formal education, post-school learning and experience as well as the informal knowledge existing in their social environment. It is not possible to observe and measure all aspects of peoples educational achievement, though. Even with data from individual or household surveys, one is almost always restricted to information on formal schooling careers. That is we observe if a person did not experience any education, has attained some basic or higher schooling and we can estimate the years associated with the respective education level. From this it follows that formal schooling is a discrete rather than a continuous variable. It has lower boundary at zero and an upper boundary, given by the cumulative duration of tertiary education.

If we define the proportion of the population for which the education level i is the highest attained as $p_i = \frac{pop_i}{pop}$ and the corresponding cumulative duration of schooling in years as $y_i = \sum_{j=1}^i dur_j$, the cumulative probability of attaining i levels of schooling or the

population share of those attained years of schooling less than or equal to y_i is given by

$$F_i = \sum_{j=1}^i p_j \quad (1)$$

$\mu = \sum_{i=1}^n p_i * y_i$ and $\Gamma = \sum_{i=1}^n pop_i * y_i$ are mean and total years of schooling respectively.

Thus, the cumulative percentage of education up to level i is

$$S_i = \frac{1}{\Gamma} * \sum_{j=1}^i y_j * pop_j = \frac{1}{\mu} * \sum_{j=1}^i y_j * p_j \quad (2)$$

for $i = 1, \dots, n$. F_0 and S_0 are defined as zero while $F_n=1$ and $S_n=1$.

In figure 1 Education Lorenz Curve are drawn for $n = 4$. The resulting function is a kinked line, with the kinked points corresponding to each of four education categories. Moreover, if a proportion of the population does not attain any education, i.e. $y_1 = 0$, the Lorenz Curve is truncated along the horizontal axis.

The ratio of the surface A between the Lorenz Curve and the forty five-degree line and the whole triangle $A + B = \frac{1}{2}$, with B depicting the area blow the Lorenz Curve, can be written as

$$\begin{aligned} R &= \frac{A}{A+B} \\ &= 2A \\ &= 1 - 2B \end{aligned} \quad (3)$$

As the area B is derived from a sum of triangles, the Education Gini Coefficient can be computed as follows

$$GINI_E = 1 - \sum_{i=1}^4 (F_i - F_{i-1})(S_i + S_{i-1}) \quad (4)$$

Gini (1912), and thereafter many others², showed the geometrical approach to be related to the relative mean difference being a statistical measure of dispersion. Accordingly, the Education Gini “...measures the ratio to the mean (average years of schooling) of half of the average schooling deviations between all possible pairs of people.” (Fan *et al.* 2001, 7)

$$GINI_E = \frac{1}{2\mu} \sum_{i=2}^4 \sum_{j=1}^{i-1} |y_i - y_j| p_i p_j \quad (5)$$

²For a survey on the respective literature see e.g. Xu (2004).

4 Expected Behavior of the Education Gini Index

From (4) and (5) it becomes clear that the Education Gini always lies in a range between zero and one, indicating perfect equality and perfect inequality respectively. By resorting to the simple case of two education categories, Thomas & Wang (2008) demonstrate, on the one hand, that the Education Gini Coefficient asymptotically moves towards one as average years of schooling move toward zero. On the other hand, as the average level of education approaches its upper boundary, the Education Gini asymptotically moves towards zero. “In other words, when nearly no one has the opportunity to obtain education, the society is the most unequal one. ..., when almost everyone has the opportunity of obtaining education, and has indeed done so, the society is a perfectly equal one in regard to the distribution of schooling.” (Thomas & Wang 2008, 48) This results in a linearly negative, theoretical relationship between the Education Gini and average years of schooling. However, if we expand the set of possible education outcomes, an increase in average schooling need not simultaneously imply a decrease in the degree of inequality but the distributional effect will depend on the existing level of educational attainment and the initial state of its distribution.

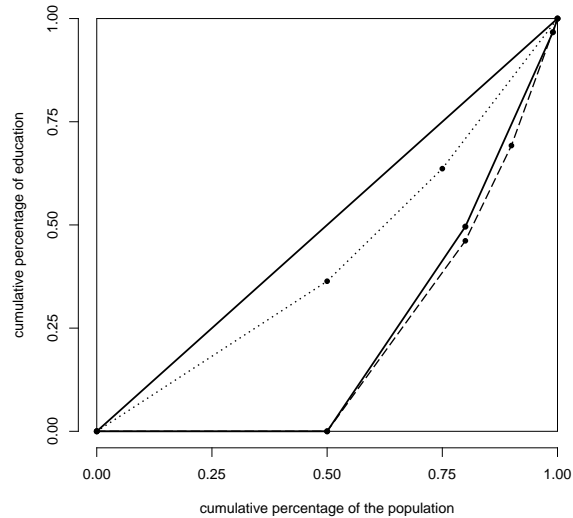


Figure 1: Education Lorenz Curve

As it is evident from figure 1, the concept of Education Lorenz Curves and Gini Coefficients has the nice feature that perfect equality will never be reached as long as there are people without any education. Allowing for universal primary education, represented by the Lorenz Curve shifting inwards, substantially decreases the degree of inequality in the education distribution and will simultaneously increase average educational attainment. A similar effect with respect to average schooling could be obtained by a policy of providing higher education to a few, thereby generating a more unequal distribution of years of schooling. Equal levels of average educational attainment therefore involve different compositions of the educational structure, reflected in varying degrees of inequality. The analysis of the historical, political, institutional and socio-economic causes and effects of such variations in the Education Gini Index, holding average schooling constant, are of prime interest for research. In our work, the demographic dimension of inequality is examined in detail.

5 Data

Using the demographic method of multistate back and forward projection, a group of researchers at the International Institute for Applied Systems Analysis (IIASA) and the Vienna Institute of Demography (VID) has recently constructed population data for 175 countries by age, sex and level of educational attainment from 1960 to 2010. (Lutz & KC 2011) This dataset gives the full educational attainment distributions for four educational categories (see table 1) by five year age groups of men and women. The demographic approach allows for considering educational mortality, fertility and migration differentials. Moreover, the definitions of categories are based on UNESCO's ISCED (International Standard Classification of Education) categories and are strictly consistent over time.

In order to compute the Education Gini Coefficient according to (5), we combine the time series of education data with country-specific information on the time it takes to reach each educational level from UIS (Unesco Institute for Statistics). However, the four broad categories of educational attainment consist not only of people who have completed but also of people who did not complete the respective level. Using the total duration for completion would overestimate the years an individual spent in school on average and, thus, therefrom arising differences between all pairs of people. Hence, some corrections have to be made to

Table 1: Education Categories and Adjusted Durations of Formal Schooling Cycles

IIASA/VID education categories	mean	min	max
E1 no schooling	0	0	0
E2 primary education	3.921	1.750	8.935
E3 secondary education	12.02	8.75	16.48
E4 tertiary education	15.71	12.00	18.00

Values are computed from the whole sample (i.e. all 175 countries, time intervals, sexes and age groups).

arrive a consistent estimate of the dispersion of schooling. We hence calculate the years of schooling in each education category by age and sex as follows.³

$$y_1 = 0$$

$$y_2 = 0 + 0.5 * dur_2 + 0.25dur_2 * \left(1 - \frac{E1}{E1 + E3}\right)$$

$$y_3 = 0 + dur_2 + 0.5 * dur_3 + 0.25 * dur_3 * \left(1 - \frac{E2}{E2 + E4}\right)$$

$$y_4 = 0 + dur_2 + dur_3 + dur_4$$

Firstly, we assume a minimum and a maximum to exist at the lower and upper quintile of the duration of each education category (dur_i), respectively. The actual years further depend on the weights given by surrounding education levels. That is, the higher attainment in the preceding and the lower attainment in the subsequent category, the lower the level-specific years of schooling. Vice versa, the lower attainment in the preceding and the higher attainment in the subsequent category, the more people will have completed the respective schooling cycle and ,thus, the higher years of schooling. Finally, as category four comprises only people who have completed higher education, we compute the total years it takes to attain tertiary education.

³This approach has also been used in order to derive an estimate for mean years of schooling from the educational attainment distributions provided by the IIASA/VID dataset. (see KC *et al.* (2010, 403)

6 Results

We use formula (5) derived in section 3 for calculating Education Gini Coefficients for 175 countries by year, agegroup and gender. Preliminary analysis of our data reveals three general trends. First, huge educational gaps exist within as well as across countries. Moreover, substantial education expansion, accompanied by a reduction of inequality in the distribution of education amongst individuals as well as between men and women, has taken place. In order to depict the vast information contained in our education dataset, we firstly present results for two country examples - India and South Korea - which are of interest in their own right. We proceed at the regional level in order to analyze differences across regions and to study the development over time.

6.1 Country Examples: India and South Korea

Multistage age pyramids, as presented in figure 2, are a very illustrative way for studying the history of country-specific educational trends. Females' educational attainment is shown on the right side of the pyramid and males' on the left in five-year age groups above age 15. The colors in each age group show the numbers of men and women without any formal education as well as with some primary, at least completed junior secondary and completed tertiary education.

In India, on average 55.2 % of people aged 20-24 did not attain any formal education in 1970. The gender gap was huge, with the respective share being 71.3 % for females and 40.1 % for males. Beyond that, only a negligible share of males reached the tertiary level. These people moved up the age pyramid and represent the cohort of 50-54 years olds in 2000.⁴ Then, educational attainment of young age groups is comparatively high. More people reached the primary or secondary level and the share of tertiary educated increased for males as well as for females. Yet, in 2000 41.5 % of females and 20.2 % of males have still been illiterate. In contrast, the population pyramids for South Korea impressingly reveal the country's educational expansion over the past decades. Among the young age groups, almost everybody attains secondary or tertiary education in 2000. On the other hand, still persisting illiteracy rates and gender gaps amongst the elderly reflect overall

⁴This cohort is smaller in size in 2000 because some people have died over the course of the years.

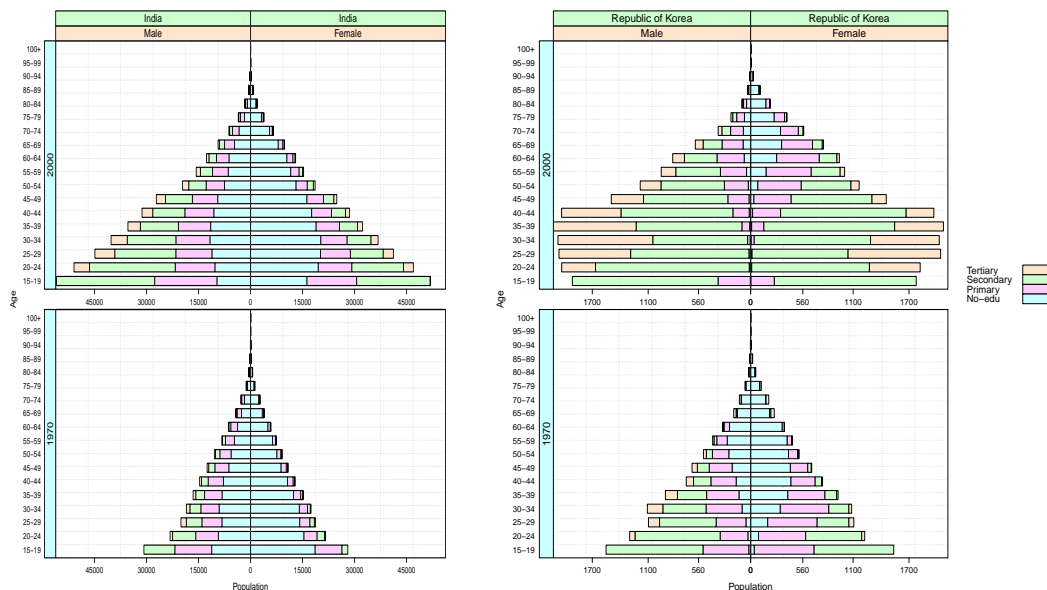


Figure 2: Population Pyramids

lower educational attainment in preceding periods.

How these cohort- and gender-specific structures in the educational distribution translate into our measure of inequality in educational attainment is depicted in figure 5. In general, the degree of inequality is much lower among the youth than among the elderly. Moreover, education is more equally distributed among men than among women. This gender gap is quite pronounced in India. While the Education Gini for males ranges from 0.3 in the lowest age group to 0.65 in the cohort of people aged 65 and above, it lies between 0.43 and 0.88 for females. On the contrary, the gap disappears as education amongst South Korea's youth became almost perfectly equally distributed, with the Education Gini leveling off at 0.03. The steep slope of the curve reveals that education expansion in South Korea was accompanied by a substantial decline in the degree inequality in the distribution of education.

The geometrical representation of the Education Gini Coefficient in the form of Lorenz Curves support our hitherto existing findings. The younger the age cohort, the smaller the area between the Lorenz Curve and the "equality line". Beyond that, figure 4 reveals an interesting fact about the demographic dimension of educational inequality. The black Lorenz

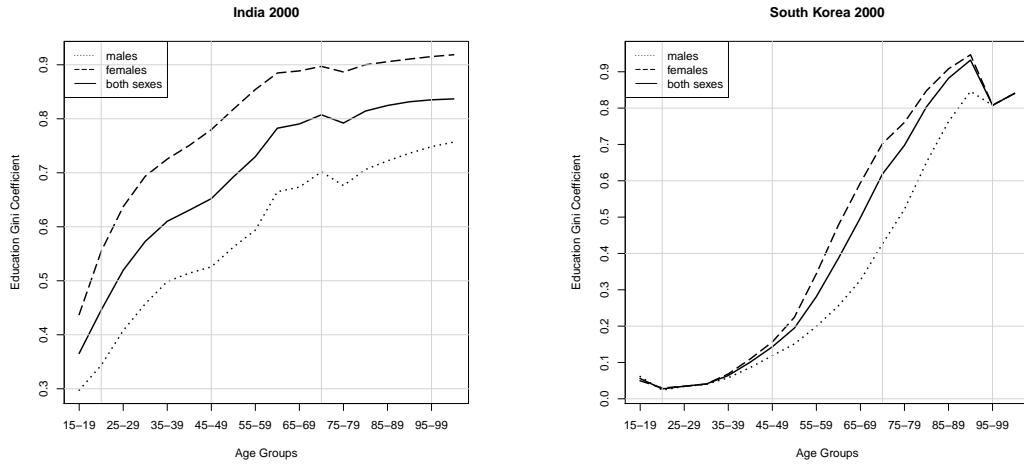


Figure 3: Education Gini Coefficients Across Age Groups

Curve depicts the distribution of education among the population aged 15 and above. Even if it should measure the average degree of inequality across the whole population, it is slanted towards the educational distribution of the oldest age cohort. The Gini coefficient which is derived therefrom thus overestimates the overall degree of inequality in the distribution of education. This is especially true for countries which, as South Korea, have experienced a history of enormous educational improvement. A policy of educational expansion is conducted by investing in the education of the youth in order to increase their attainment level. In earlier phases, educated young age cohorts hence opposite older cohorts with lower educational attainment. The resulting unequal distribution between the youth and the elderly produces an upward bias in the estimate of the degree of educational inequality in the total population.

6.2 Regional Trends

It is evident from figure 5 that our finding of age cohort-specific degrees of educational inequality carries over to the regional level. Beyond that, studying the behavior of the Education Gini across regions and over time sheds light into the dynamics of educational improvement. In general, the distribution of education has become more equal over the past decades. However, not only the range of the Education Gini but also cohort-specific trends

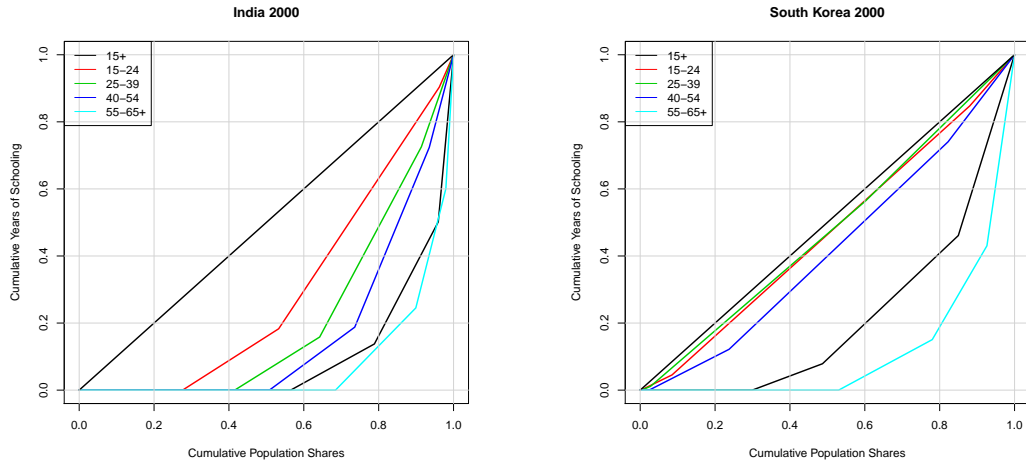


Figure 4: Education Lorenz Curves

differs greatly across regions. On the one hand, the Education Gini remains at a high level in Sub-Saharan Africa and in South Asia. The time series are relatively flat, indicating little improvement in the degree of inequality in the distribution of education. Only after 1980 did the distribution of education among the youngest age group become slightly more equal in South Asia. This trend will carry over to older age groups in subsequent periods when this cohort is moving up the age pyramid (see figure 2). The Middle East & North Africa as well as Eastern Asia & the Pacific, on the other hand, have been experiencing a time of relatively high dynamics in their educational structure. The steep slopes of the functions corresponding to young age groups indicate great improvement in the distribution of education within the youth which will be passed on to older cohorts in the future. As the degree of inequality decreases further, like in Latin America & the Caribbean, in South America as well as in Central Asia & Europa, the time series of young age groups becomes flatter. Simultaneously, the function of older cohorts is steeper, reflecting past movements in the educational composition. Finally, in Advanced Economies the gap between different age groups has narrowed substantially as the series are eventually horizontal and the Education Gini levels off at about 0.1.

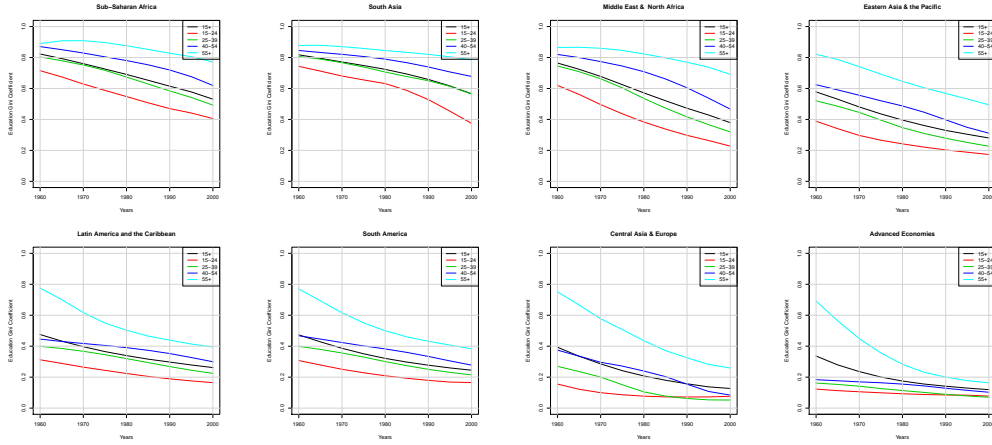


Figure 5: Education Gini Coefficients by Region & Time

7 Summary and Conclusion

The focus of our work has been the distributional dimension of education. Consequently, we used the IIASA/VID dataset for calculating Gini Coefficients of Educational Attainment for 175 countries by year, sex and five-year age groups. In doing so, we find the educational composition to vary greatly with age groups as well as with gender. Young people turn out to be not only more educated than the elderly but also is formal education more evenly distributed among younger age cohorts. Beyond that, computing Education Gini Coefficients for the total population aged 15 and above might overestimate degree of educational inequality. As has been shown, this is especially true for countries experiencing a phase of rapid educational expansion which temporary generates an unequal distribution of education between the youth and the elderly. Finally, comparing the age group-specific trends in educational inequality across regions revealed quite interesting features of inequality dynamics. That is, a process towards an equal distribution of education has been initialized by improving the educational composition of the youth which carries over to older age cohorts in subsequent time periods. In addition, age group-gaps are narrowing as the degree of inequality levels off at a low level.

Thus, allowing for the demographic dimension turns out to be crucial in the study of education and its distribution. By providing a measure which gives a comprehensive picture

of the degree of inequality in educational attainment within and across societies, we aim at contributing to a better understanding of the causes and effects of socio-economic inequality.

References

- BANK, INTER-AMERICAN DEVELOPMENT. 1999. *Facing up to Inequality in Latin America: Economic and Social Progress in Latin America, 1998-99 Report*. Johns Hopkins University Press.
- BIRDSALL, NANCY, & LONDOÑO, JUAN LUIS. 1997. Asset Inequality Matters: An Assessment of the World Bank's Approach to Poverty Reduction. *American Economic Review*, **87**(2), 32–37.
- CASTELLÓ, AMPARO, & DOMÉNECH, RAFAEL. 2002. Human Capital Inequality and Economic Growth: Some New Evidence. *The Economic Journal*, **1128**, C187–C200.
- CHECCHI, DANIELE. 2000. *Does Educational Achievement Help to Explain Income Inequality?* Working Paper n.11.2000, Università degli Studi di Milano.
- FAN, XIBO, THOMAS, VINOD, & WANG, YAN. 2001. *Measuring Education Inequality: Gini Coefficients of Education*. World Bank Working Paper 2525.
- FAN, XIBO, THOMAS, VINOD, & WANG, YAN. 2002. *A New Dataset on Inequality in Education: Gini and Theil Indices of Schooling for 140 Countries, 1960-2000*. Mimeo, The World Bank.
- GINI, CORRADO. 1912. *Variabilità e Mutabilità*. Tipografia di Paolo Cuppini.
- KC, SAMIR, BARAKAT, BILAL, GOUJON, ANNE, SKIRBEKK, VEGARD, SANDERSON, WARREN, & LUTZ, WOLFGANG. 2010. Projection of populations by level of educational attainment, age, and sex for 120 countries for 2005-2050. *Demographic Research*, **22**, 383–472.
- LAM, DAVID, & LEVISON, DEBORAH. 1991. Declining Inequality in Schooling in Brazil and its Effects on Inequality in Earnings. *Journal of Development Economics*, **37**(1-2), 199–225.

- LONDOÑO, JUAN LUIS. 1990. *Kuznetsian Tales with Attention to Human Capital*.
- LÓPEZ, RAMÓN, THOMAS, VINOD, & WANG, YAN. 1998. *Addressing the Education Puzzle: The Distribution of Education and Economic Reform*. World Bank Working Paper 2031.
- LUTZ, WOLFGANG, & KC, SAMIR. 2011. Global Human Capital: Integrating Education and Population. *Science*, **333**, 587–592.
- RAM, RATI. 1990. Educational Expansion and Schooling Inequality: International Evidence and Some Implications. *The Review of Economics and Statistics*, **72**(2), 266–274.
- ROSTHAL, RICHARD A. 1978. *Measures of Disparity: A Note*.
- SHERET, MICHAEL. 1988. Equality Trends and Comparisons for the Education System of Papua New Guinea. *Studies in Educational Evaluation*, **14**(1), 91–112.
- THOMAS, VINOD, & WANG, YAN. 2008. Distribution of Opportunities Key to Development. In: HOLSINGER, DONALD B., & JACOB, W. JAMES (eds), *Inequality in Education*. Springer.
- V.L. MAAS, JACOB, & CRIEL, CEERT. 1982. *Distribution of Primary School Enrollments in Eastern Africa*. World Bank Staff Working Paper 511.
- XU, KUAN. 2004. *How Has the Literature on Gini's Index Evolved in the Past 80 Years?* Department of Economics, Dalhousie University, Halifax.