

The Role of Age and Gender in Education Expansion

Sauer, Petra

DOI:

[10.57938/498b2179-e019-4f87-ae04-99f79d9b34b5](https://doi.org/10.57938/498b2179-e019-4f87-ae04-99f79d9b34b5)

Published: 01/01/2016

Document Version

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Sauer, P. (2016). *The Role of Age and Gender in Education Expansion*. INEQ Working Paper Series No. 4
<https://doi.org/10.57938/498b2179-e019-4f87-ae04-99f79d9b34b5>



Working Paper Series

#4

Petra SAUER

The Role of Age and Gender in Education Expansion

The Role of Age and Gender in Education Expansion

Within- and Between-group Specific Trends in Education (In)equality

Petra Sauer*

Abstract

Using the IIASA/VID dataset of populations by age, sex and level of education, I calculate education Gini coefficients and decompose the overall degree of educational inequality into age, sex and within-group components. I analyze the relative relevance of these components for inequality reduction and investigate the distributional outcomes of education expansion. I find that, on average, equalization between males and females, younger and older cohorts as well as within these subgroups of the population has significantly contributed to declining educational inequality over the observed sample period around the globe. But the relative role of these components fluctuates in the process of education expansion. First, as improvements are initiated by enhancing the educational opportunities of the youth, the gap between cohorts widens in transition phases but vanishes thereafter. Second, gaps between sexes have been reduced but are predicted to widen again if either males or females are the first to enter higher education levels. To a lesser extent, this is also true for gaps within population subgroups which can be due to the ethnic background or the social and economic status of people.

Keywords: Education, Inequality, Age Distribution

JEL codes: I24,I25

*Institute for Macroeconomics and INEQ Research Institute Economics of Inequality (WU Vienna University of Economics and Business), WIC Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID, WU)

1 Introduction

“... an overall trend of education expansion, accompanied by a reduction of inequality in the distribution of education has taken place ..., however, ... huge differences across countries still persist.” (Sauer & Zagler, 2014) Diverseness in the distributional outcomes of education expansion result from the extent to which policies are able to enlarge the group of people who participate in education. This is done by improving the educational opportunities of women as well as of people from disadvantaged socio-economic backgrounds. In both respects, providing for enhanced schooling prospects of the youth secures educational outcomes of future generations. Consequently, differential education expansion trajectories result from differential magnitudes of human capital accumulation and equalization among the youth as well as between men and women.

In order to examine comparative distributional implications of education expansion, I decompose overall education inequality into age, gender and within-group components. Using a matrix algebra approach, Silber (1989) shows how to decompose the Gini index into three components, a within- and a between-group component, and a residual term. I adapt his method so that it can be applied to aggregate education data instead of individual income data. Doing so enables me to separate inequalities which are due to inequality across age groups as well as between males and females from disparities within these groups for 171 countries over the time span from 1970 to 2010 in 5-year intervals. Moreover, I provide an intuitive interpretation of the residual term which relates to the degree of within-group inequality. Analyzing these new indicators provides novel insights into comparative education expansion trajectories around the globe.

The remainder of this papers is structured as follows. Section 2 surveys existing work on the distribution of educational attainment. Thereafter, I use matrix algebra to derive the education Gini coefficient as a measure of between-category inequality and describe its decomposition in Section 3. I present the data I use in Section 4 and discuss the results of my decomposition analysis in Section 5. In Section 6, I test for non-linearity in the relationship between average educational attainment and the obtained inequality components. Finally, Section 7 concludes.

2 Measuring Education Inequality - An Overview

Two measures have been used primarily in order to investigate the distributional dimension of education.¹ The standard deviation of schooling was used to explore the impact of the distribution of education on income growth and poverty reduction (e.g. Birdsall and 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 were applied to test for the existence of an Education Kuznuts 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), confirmed the findings of Londoño (1990) and

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

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 (e.g. Maas and Criel 1982, Rosthal 1978 and Sheret 1988) used schooling enrollment figures 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.

Like in its application to income inequality, the education Gini coefficient is a measure of mean standardized deviations between all possible pairs of persons and lies in a range between zero and one. A value of zero indicates a perfectly equally distributed education structure, with the opposite being true for a value of one. The former case corresponds to a situation in which the whole population attains the same education level, irrespective of which. In the latter case, one person completes for example tertiary education, while the rest of the population does not attain any formal schooling. López *et al.* (1998) were the first to 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 (e.g. Castelló & Doménech, 2002 and Sauer & Zagler, 2014). Non-conforming with earlier results, plotting Gini Coefficients against average educational attainment does not support an education Kuznets curve but reveals a strong negative relation between the degree of inequality and the average level of educational attainment.

Crespo-Cuaresma *et al.* (2013) integrated the demographic dimension into the analysis of education inequality. They have constructed a dataset of education Gini measures by age group and gender for 175 countries from 1960 to 2010 in 5-year intervals based on the first version of the IIASA/VID² global dataset of populations by age and sex as well as four levels of education.³ Investigating differential trends of education inequality within population subgroups, they show that education is more equally distributed among

²International Institute for Applied Systems Analysis/Vienna Institute of Demography

³Benaabdelaali *et al.* (2012) also computed education Gini coefficients by age group based on the Barro-Lee 2010 education dataset.

the youth than among the elderly and, among men than among women. In this work, I build on and add to the analysis of Crespo-Cuaresma *et al.* (2013). First, I use the most recent version of the IIASA/VID education dataset which provides a breakdown of populations into six education categories, incorporating uncompleted primary as well as lower secondary schooling (see Section 4). Moreover, I decompose the overall education Gini coefficient in order to examine the relative relevance of inequality within and between age groups and gender.

3 Matrix Algebra and the Education Gini

In general, education has formal and informal dimensions and comprises aspects of quality. 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, one observes if a person did not experience any education, has attained some basic or higher schooling and one can estimate the years associated with the respective education level. From this it follows that formal schooling is a categorial rather than a continuous variable. It has a lower boundary at zero, an upper boundary given by the duration to complete tertiary education and categories which correspond to formal education levels.

In an early work Silber (1989) presents a matrix approach to the computation of the Gini index of income inequality. In the following, I demonstrate how this method is adapted to the categorial structure of aggregate education data. In particular, I derive the Gini coefficient of educational attainment in matrix notation as a measure of between-category inequality. Using matrix algebra also enables to decompose the Gini index by population subgroups into three components: a within-group component equal to the weighted sum of within-group inequality, a between-group component equal to the weighted sum of between-groups inequality, and a residual term which can be interpreted in relation to the ranking of individuals within subgroups.

3.1 The Education Gini as a Measure of Between-category Inequality

According to Silber (1989), for individual data, the Gini index of inequality can be written in matrix notation as

$$I_G^E = e'Gs \tag{1}$$

where e' is a row vector with n elements equal to $1/n$. n is the number of observed individuals. If educational attainment is measured by years of schooling, one element, s_i , of the column vector s is the share of individual i 's attainment in total years of schooling ($Y_T = \sum_{i=1}^n Y_i$) in/of the concerned population. The n elements of s are sorted

in descending order according to individual ranks in the education distribution, so that

$$s_1 \geq s_2 \geq \dots \geq s_i \geq \dots \geq s_n \quad (2)$$

The linear operator, introduced by Silber (1989), is the G-matrix; which is an $n \times n$ matrix with upper-diagonal elements g_{ij} when $j > i$ equal to -1, lower-diagonal elements when $i > j$ equal to 1 and diagonal elements when $i = j$ equal to 0.

If the available information is limited to the formal duration it takes to complete an education level, individual variation within these categories vanishes. In order to demonstrate that the education Gini is thus reduced to a measure of between-category inequality, I partition the relevant vectors as well as the G-matrix. The quantity of subvectors of e and s is given by the number of categories, c . The amount of subvector elements, in turn, depends on the number of individuals, n_h , for which h is the highest education level attained. The partitioned matrix G consists of c^2 submatrices and thus has the following form

$$\begin{pmatrix} G(n_1, n_1) & \dots & G(n_1, n_q) & \dots & G(n_1, n_c) \\ \vdots & \ddots & & & \vdots \\ G(n_p, n_1) & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ G(n_c, n_1) & \dots & \dots & \dots & G(n_c, n_c) \end{pmatrix} \quad (3)$$

The main-diagonal submatrices of dimension $n_h \times n_h$ capture within-category inequality, with zeros in their main diagonal, -1 in their upper right and 1 in the lower left triangle. Submatrices, $G(n_p, n_q)$, for which $q > p$, consist of identical elements equal to -1. If $p > q$, the elements are equal to 1. Summing over partitioned elements, the education Gini can thus be written as consisting of a within- and between-category component,

$$I_G^E = \sum_{h=1}^c e'(n_h)G(n_h, n_h)s(n_h) + \sum_{p=1}^c \left[\sum_{q \neq p}^c e'(n_p)G(n_p, n_q)s(n_q) \right] = I_W^E + I_B^E \quad (4)$$

If no information about within-category variation is available, the within component is redundant, the overall Gini index reduces to its between-category component, i.e. $I_G^E = I_B^E$, and the degree of inequality is generally underestimated. Further inspection of I_B^E enables to show how the between-category measure of education inequality can be computed using population shares and category averages of years of schooling.

Defining the share of category h in total years of schooling, $s_{.h} = \sum_{i=1}^{n_h} Y_{ih}/Y_T$, as well as the mean individual share of years in category h , $\bar{s}_h = s_{.h}/n_h$, the between-category contribution of one pq -element for which $q > p$ can be written as

$$e'(n_p)G(n_p, n_q)s(n_q) = -\left(\frac{n_p}{n}\right)n_q\bar{s}_q \quad (5)$$

In turn, the contribution of an element with $p > q$ looks like

$$e'(n_q)G(n_q, n_p)s(n_p) = \left(\frac{n_q}{n}\right)n_p\bar{s}_p \quad (6)$$

So that overall between-category inequality can be written as the weighted average of its individual pq -contributions, I_{pq}^E , with weights equal to the two concerning categories' population and years-of-schooling shares respectively:⁴

$$\begin{aligned} I_B^E &= \sum_{p=1}^c \sum_{q>p}^c \frac{n_p + n_q}{n} (n_p \bar{s}_p + n_q \bar{s}_q) I_{pq}^E \\ &= \sum_{p=1}^c \sum_{q>p}^c \left(\frac{n_q}{n} \right) n_p \bar{s}_p - \left(\frac{n_p}{n} \right) n_q \bar{s}_q \end{aligned} \quad (7)$$

Both variants of Equation (7) use the mean of schooling years in each category in conjunction with population shares. They can thus be easily calculated based on aggregate data of educational attainment. Rearranging the second part of Equation (7) enables to obtain the familiar version of the education Gini index as a weighted sum of differences in category-specific years of schooling (see Section 2):

$$I_G^E = I_B^E = \frac{1}{\bar{Y}} \sum_{p=1}^c \sum_{q>p}^c (Y_q - Y_p) p_p p_q \quad (8)$$

where $\bar{Y} = Y_T/n$ is the mean of years of schooling in the concerned population, $Y_h = (\sum_{i=1}^{n_h} Y_i/n_h)/Y_T$ is the average duration it takes to complete education category h and $p_h = n_h/n$ is the corresponding population share.

3.2 Population Subgroups

In contrast to education categories, individuals from different sub-groups of the population cannot be ordered definitely, resulting in overlapping partitions of the education distribution. Nevertheless, the matrix approach provides an intuitive and straight forward method to decompose the education Gini index into population-subgroup contributions.

Following Silber (1989), I define an additional partitioned vector, v , which is ordered first, by subgroup averages of education attainment shares and, second, by individual attainment shares within subgroups. Thus,

$$\bar{v}_1 \geq \dots \geq \bar{v}_j \geq \dots \geq \bar{v}_g \quad (9)$$

and

$$v_{1,j} \geq v_{2,j} \geq \dots \geq v_{n_j,j} \quad \forall j \quad (10)$$

where $v_{i,j}$ is the education attainment share of individual i in group j and \bar{v}_j is the groups' mean attainment share. The number of subvectors depends on the number of groups g , and the quantity of elements, n_j , varies according to the group's population size. If also e and G are partitioned by population subgroup and defined just as in Section 3.1, a modified inequality index, $e'Gv$, can be decomposed into a within-group and a between-group component,

$$\sum_{j=1}^g e'(n_j)G(n_j, n_j)v(n_j) + \sum_{a=1}^g \left[\sum_{b \neq a}^g e'(n_a)G(n_a, n_b)v(n_b) \right] = I_W^E + I_B^E \quad (11)$$

⁴See Appendix A

Rewriting the within-group elements in terms of group attainment and population shares enables to obtain the within-group component as a weighted average of within-group inequality, with weights equal to the groups' population and attainment shares (v_j) respectively. Thus,

$$I_W^E = \sum_{j=1}^g \frac{n_j}{n} v_j I_{G_j}^E \quad (12)$$

where within-group inequality, $I_{G_j}^E$, is the between-category Gini index of subgroup j . Beyond that, in a similar manner as in Section 3.1, the between-group component of the education Gini can be shown to be equal to a weighted average of pairwise contributions, I_{ab}^E , with weights equal to the two groups' joint population and attainment shares:

$$\begin{aligned} I_B^E &= \sum_{a=1}^g \sum_{b>a}^g \frac{n_a + n_b}{n} (n_a \bar{v}_a + n_b \bar{v}_b) I_{ab}^E \\ &= \sum_{a=1}^g \sum_{b>a}^g \left(\frac{n_b}{n}\right) n_a \bar{v}_a - \left(\frac{n_a}{n}\right) n_b \bar{v}_b \end{aligned} \quad (13)$$

Defining mean years of schooling in subgroup j as $\bar{Y}_j = \sum_{i=1}^{n_j} Y_{ij} / n_j$, the second part of Equation (13) can be rearranged in order to obtain the between-group contribution as a weighted sum of differences in subgroups' mean years of schooling:

$$I_B^E = \frac{1}{\bar{Y}} \sum_{a=1}^g \sum_{b>a}^g (\bar{Y}_a - \bar{Y}_b) p_a p_b \quad (14)$$

Finally, the difference between the inequality measures obtained from using the definitely ordered versus the reordered attainment share vectors, $e'Gd = e'(s - v) = e'Gs - e'Gv$, builds the third component of the Gini index decomposition. This factor can be interpreted as the intensity of modifications necessary to rank individuals according to their groups' educational attainment, or, as the degree to which groups are overlapping.

4 Data

Applying the demographic method of multistate back and forward projection, researchers at the International Institute for Applied Systems Analysis (IIASA) and the Vienna Institute of Demography (VID) have constructed population data by *four* levels of educational attainment, 21 five-year age groups and sex for 120 countries (KC *et al.*, 2010 and Lutz & KC, 2011). More recently, the dataset has been updated and extended to cover 171 countries. Most importantly, the new version enables to assess the full educational attainment distribution over *six* educational categories which are based on UNESCO's International Standard Classification of Education (ISCED 1997) levels and are thus consistent over time and across countries. For the base year 2010, Bauer *et al.* (2012) use census data if available, survey data otherwise. They apply a set of standardization procedures in order to harmonize national education levels. KC *et al.* (2013) adopt demographic techniques

using education specific estimates of fertility, mortality and migration in order to provide back and forward projections of populations by educational attainment. Based thereupon, I construct age-group- and gender-specific population shares of six education levels which are depicted in Table 1.

Table 1: Categories of educational attainment

Category ^a	ISCED 1997 level
1 No education	No level of ISCED0, Grade 1 of ISCED 1 not completed
2 Incomplete Primary	Incomplete ISCED1
3 Primary	Completed ISCED1, Incomplete ISCED2
4 Lower Secondary	Completed ISCED2, Incomplete ISCED3
5 Upper Secondary	Completed ISCED3, Incomplete ISCED4 or 5B
6 Post-secondary	ISCED 4 & 5B ^b , ISCED 5A & ^c

^aSee Bauer *et al.*, 2012 and KC *et al.*, 2013.

^bfirst diploma, shorter post-secondary courses

^clonger post-secondary courses, post-graduate level

To compute education Gini coefficients and decompose them by population sub-group, I also require information on the standard duration of different levels of schooling. Potancoková *et al.* (2014) have assembled duration data which they apply for calculating an estimate of mean years of schooling based on IIASA/VID population shares. Potancoková *et al.* (2014) are the first to thoroughly adjust data on completed formal schooling cycles from the UNESCO Institute for Statistics (UIS) to allow for incomplete levels using region-specific correction factors based on survey estimates. In addition, they account for country-specific and time-varying education systems. I use their adjusted duration data for calculating level- specific and overall total and mean years of schooling.

Applying Equation (8) to each subgroup of the population, I obtain a dataset of education Gini coefficients for each of 18 five-year age groups above 15, males and females. I cover 171 countries over the time period from 1970 to 2010 in five-year intervals. Figure 1 presents the resulting structure of the data for the example of Iran in 2010. Plotting education Gini coefficients against age gives rise to an upward sloping curve. Thus, educational improvements start among the youth. If these not only affect the average level but also the spread, education is more equally distributed among the youth than among the elderly. Even if this relation holds globally, countries differ with respect to the slope, i.e. the magnitude of education equalization.⁵ Moreover, they differ with regards to the *gender-gap* in education inequality. In Iran, even if education is more equally distributed among men than among women across all age groups, the gap is most pronounced for cohorts aged between 45 and 70. The polarization between a small segment of highly educated and a broad group with basic education is thus more pronounced among women

⁵See Crespo-Cuaresma *et al.* (2013) for a detailed discussion and analysis of the dynamics of age-group specific education inequality.

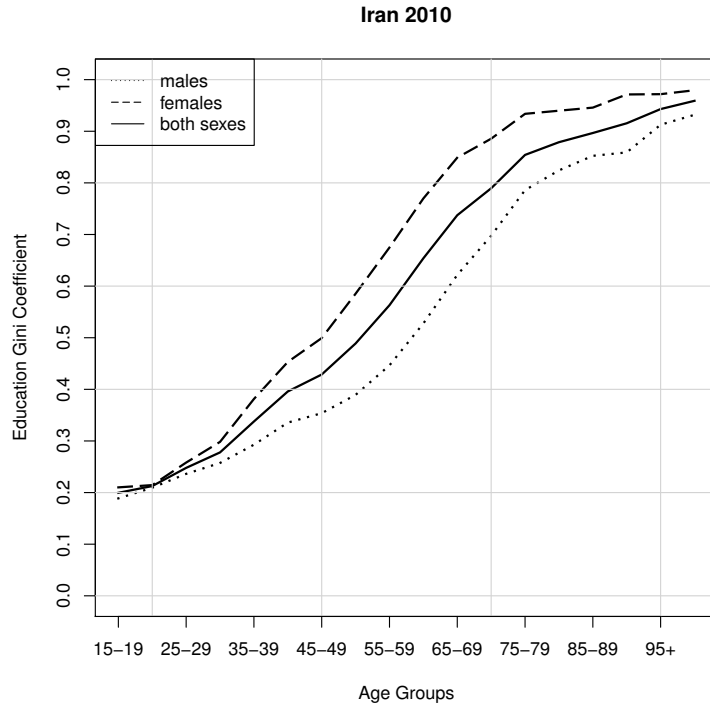


Figure 1: Education Gini Coefficients Across Age and Gender

than among men. At the age of 20, the education Gini indices of males and females are equal to 0.2. However, stronger segmentation in the access to tertiary education among women can explain the slightly higher degree of inequality in subsequent cohorts.

In order to ascertain to measure the degree of inequality in completed educational attainment, I restrict my analysis to cohorts aged 25 and over. The finer the grouping, the more homogeneous groups are. This increases the relative importance of between-versus within-group variation. Considering this impact on the Gini index decomposition, I construct three broader age groups for people aged 25 to 44, 45 to 64 as well as 65 and over. I thus obtain 6 subgroups of the population which I use to examine the relevance of age, gender and within-group differences for the degree of inequality in educational attainment.

5 Decomposing Education Inequality

Globally, the education Gini coefficient of the population aged 25 and over averages at 0.44. With values greater than 0.6, the most unequal distributions of educational attainment come upon in South Asian (SA) and Sub-Sahara African (SSA) countries. On the other hand, high income (HI) OECD as well as Central Asian and Eastern European (CAEE) countries show values below 0.3, on average. Despite their relatively low level, the latter

region was able to reduce the education Gini by 2.5% each year between 1970 and 2010.⁶ The general trend towards more equal education distributions was also pronounced in East Asia and the Pacific (EAP) and in Middle Eastern and North African (MENA) countries; high-dynamic regions in the medium spectrum of global education inequality. The decline over time is not only relatively sluggish in Latin American and Caribbean (LAC) and high income OECD countries, but also in Sub-Saharan Africa. In South Asia too, the education Gini decreased only by 0.8% a year (see Table 2). Applying Equations (8), (12) and (14), I decompose the education Gini coefficient of the total population aged 25 and over of 171 countries from 1970 to 2010 into five components as follows,

$$I_G^E = I_{age}^E + I_{age/sex}^E + I_{sex}^E + I_{within}^E + I_{residual}^E \quad (15)$$

where I_{age}^E captures the contribution of gender-specific differentials within age-groups while $I_{age/sex}^E$ compares the educational attainment of different sexes *and* cohorts. I_{sex}^E is the component of the education Gini which is due to differentials between males and females of the identical age groups. The extent of inequality within population subgroups, in turn, is given by I_{within}^E . Finally, $I_{residual}^E$ is the residual component.

I analyze the variation of components across countries and over time by calculating its relative contribution (Contr) to the overall education Gini in Table 2. I also test whether these contributions exhibit significant region-specific trends over time and relate to the total level of inequality (see Figure 2). Moreover, I calculate the ratio between each components' time trend and that of the total Gini, estimated from fixed effects regressions of the respective factors against time, to evaluate the relative contribution of components to the declining trend of education inequality over time (Δ Contr). In general, I find components to play differential and varying roles in the process towards more equal education distributions. The vast majority of components exhibit a significant decreasing trend over time which corresponds to the secular decline of the total Gini index. The respective ratios in Table 2 are thus positive while zeros indicate insignificance.

Differentials across age groups of same sexes are intimately connected to processes of educational improvements. The contribution of the age component is thus lowest in the high-inequality countries of South Asia (12%) and Sub-Saharan Africa (14%) where younger cohorts tend to be as similarly low educated as their predecessors. Its relative importance increases slightly to a maximum impact of 22% in CAEE countries as the overall Gini decreases. Higher educational attainment of the youth compared to that of the elderly is thus an important factor contributing to *low* levels of education inequality. Declining differences between cohorts also play a significant role in the reduction of overall inequality. Besides in CAEE, the impact is, however, relatively large in North Africa and the Middle East. In these two regions, the relative contribution of age is significantly decreasing over time. This is due to the attainment share of young cohorts to strongly decrease relative to the two older groups, indicating increasingly stagnant education levels

⁶The group of CAEE countries consists, however, of countries in Eastern Europe (e.g. Latvia and Ukraine), in Central Asia (e.g. Armenia and Kazakhstan) as well as of countries in South-Eastern Europe (e.g. Albania, Macedonia and Turkey). While education inequality is historically relatively low in the former two groups of countries, the latter group showed relatively high levels of education inequality in the 1970s and 1980s and was able to reduce it significantly thereafter.

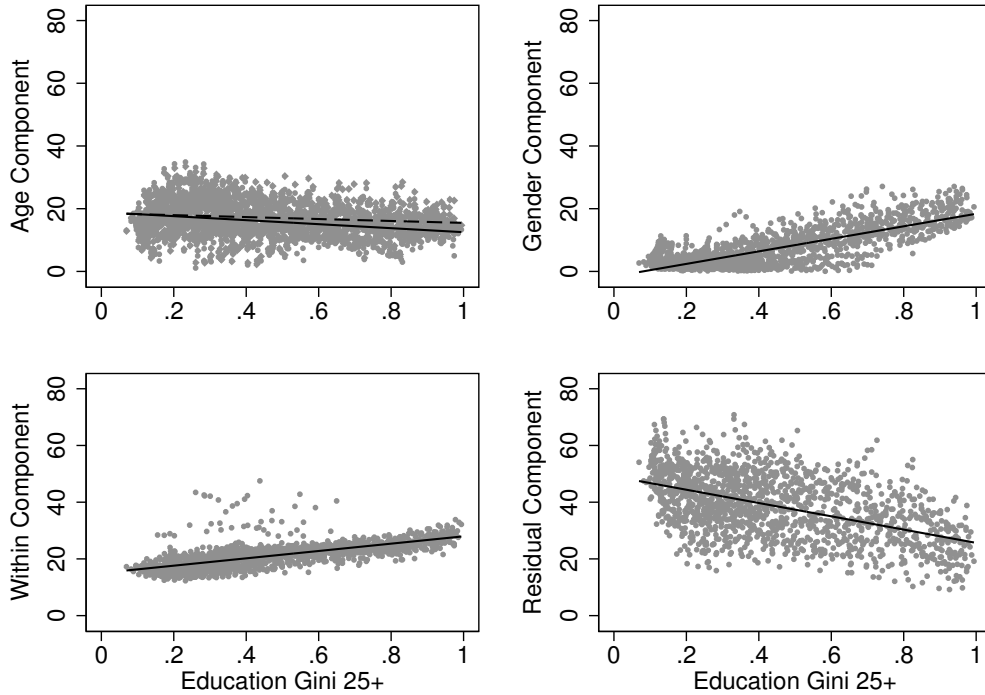


Figure 2: Contribution of Components (%) & Overall Inequality

of successive generations. While this is true for both sexes in CAEE, the education share of young females is still increasing relative to their counterparts in the Middle East and North Africa. Yet, the development of young males predominates. In countries of EAP, SA and SSA, on the other hand, the two opposing processes cancel each other out. While the education share of female cohorts is increasingly higher than that of their predecessors, males aged 25 to 44 continuously lose compared to the older groups. Gaps between middle and old groups exist in LAC and in high income OECD countries, where convergence of age groups dominates since 1990.

Lower trend shares coincide with no significant development of the age contribution over time in high income OECD countries as well as in SA and SSA. Further inspection of these regions reveals a structural break in the age-component time-trend, being significantly positive before 1990 and negative thereafter.⁷ The gender and time-specific patterns of the age component highlight the varying role of the distribution of education between cohorts in the process of education expansion, with an increasing divide if younger generations become higher educated than the older and a narrowing gap as these cohorts age. Beyond that, these findings indicate that before including females, education expansion started among young males.

⁷For 33% of total observations, the 5-year difference in the age component is positive. 90% of these observations are before 1990.

As opposed to the gender-specific age component, the contribution of the factor which compares age groups across sexes does not reveal any link to the overall education Gini. In general, the contribution of differences between age groups as well as sexes is larger than that of the pure age component. This applies also to the component's share in declining overall inequality⁸ and is especially true for South Asia and Sub-Saharan Africa. The divide in educational attainment is thus larger, comparing, for example, males and females aged 30 and 50 respectively, than comparing females of the concerning age groups. In Niger, the country with the largest degree of education inequality, the time-average share of women aged 65 and over in total educational attainment is not only lower than that of younger women, but also significantly lower than that of males in each age group. In contrast, the education share of women in the youngest age group is higher than that of males in the two older cohorts. There is thus an additional factor of gender to age-group inequalities. The two age-related component contributions exhibit the same trend over time in CAEE, LAC, EAP and MENA. In general, this implies that cohort effects apply to both sexes equally. However, in MENA and EAP, young women increased their education share relative to older male cohorts at least until the 1990, while young males lose relative to older females over the whole time span. This holds even more in Sub-Saharan Africa and South Asia, resulting in an insignificant trend of the respective contribution over time.

Closing the divide between males and females of the same cohorts has significantly contributed to declining education inequality around the globe. Not only do I observe a positive relation between the total Gini index and the relative contribution of the sex component, but also a decreasing trend of the component as well as its contribution over time for all world regions. In MENA, SA and EAP, this is mainly due to continuously increasing education shares of women aged 25 to 64. In some Eastern Asian countries, e.g. Hongkong, the education share of young females is even higher than that of young males. In contrast, the education gap between males and females is increasing in the middle age group in some Sub-Saharan African countries. Education gender gaps are negligible in high income OECD countries. Compared to its low overall education Gini, CAEE shows a relatively high contribution of gender inequality which stems from relatively large differentials in South Eastern Europe. On the other hand, overall education inequality can be relatively large even if gender differences are marginal. Many Latin American countries exceed a total Gini value of 0.4 but the education shares of both sexes evolve almost simultaneously in all age groups. From this it follows that differential institutions are a relevant factor in shaping social inequalities.

Educational divides within subgroups of the population arise from segmentation along various social lines, including ethnicity, geography and divisions between individuals from different socio-economic backgrounds, among other things. The distribution of education within population subgroups thus plays a significant role in shaping the overall level of inequality. Besides the degree of within-group inequality, the within-group component depends on the each group's share in the population as well as in total educational attainment. The within-group component is thus generally larger for females than for males; and

⁸I only find a structural break in the time trend of the age-sex component in LAC countries, which is significantly increasing before 1990 and decreasing thereafter.

for older⁹ than for younger cohorts. Across the globe, it contributes 21% to the total Gini index of the population aged 25 and over. Its relevance is generally increasing with the level of education inequality and decreasing over time. The relative within-group component contribution is therefore large and greater than the three between-group components in MENA, Sub-Saharan Africa as well as in South Asia. In the latter region, the within component has added a relatively high share of 33% to declining education inequality on average per year. Notably, with a relative contribution of 26%, the within-component plays a considerable role in countries of the Middle East and North Africa, compared to the level of education inequality in the region, but adds lesser to the declining trend than in South Asia. In contrast, the within component is almost equally relevant as the age as well as the age-sex component in high-income OECD economies while it is dominated by both age-related factors in CAEE.

⁹This effect can be negligible if life expectancy is very low, so is the population share of the elderly.

Table 2: Education Gini Components by Region

Region	Comp	Contr	Trend ^a	Δ Contr ^b	Comp	Contr	Trend	Δ Contr
<i>HI OECD</i>	I_G^E ^c	0.207	↓	-0.002	Sex	3.55	↓	10.47
	Age	16.46	None	14.86	Within	16.59	↓	20.39
	Age/Sex	17.66	↓	21.14	Resid	45.75	↑	33.14
<i>CAE</i>	I_G^E	0.270	↓	-0.007	Sex	5.02	↓	10.89
	Age	22.45	↓	24.67	Within	15.87	↓	18.66
	Age/Sex	22.74	↓	26.98	Resid	33.90	↑	18.80
<i>LAC</i>	I_G^E	0.386	↓	-0.005	Sex	2.61	↓	4.43
	Age	14.34	↑	9.75	Within	20.93	↓	25.85
	Age/Sex	14.28	↑	10.71	Resid	47.84	None	49.27
<i>EAP</i>	I_G^E	0.419	↓	-0.007	Sex	6.50	↓	19.13
	Age	16.95	None	16.04	Within	21.23	↓	25.57
	Age/Sex	18.15	None	19.95	Resid	37.18	↓	19.31
<i>SA</i>	I_G^E	0.734	↓	-0.006	Sex	13.46	↓	37.5
	Age	11.99	↑	3.27	Within	24.27	↓	33.16
	Age/Sex	15.63	None	18.34	Resid	34.64	↑	0
<i>MENA</i>	I_G^E	0.565	↓	-0.01	Sex	10.53	↓	26.3
	Age	15.3	↓	18.58	Within	25.81	↓	27.78
	Age/Sex	15.61	↓	18.87	Resid	32.75	↑	8.46
<i>SSA</i>	I_G^E	0.684	↓	-0.007	Sex	13.17	↓	33.1
	Age	14.15	None	16.00	Within	23.94	↓	26.88
	Age/Sex	16.49	↓	21.73	Resid	32.25	↑	0

^aStatistically significant time trend from a fixed effects regression of component contributions (%) against time.

^bTime trend of component relative to region specific trend of I_G^E , estimated from a fixed effects regression. A value equal to zero indicates insignificance.

^cEducation of the total population 25+, Contr is equal to the Gini value while Trend and Δ Contr correspond to the time trend of the overall Gini.

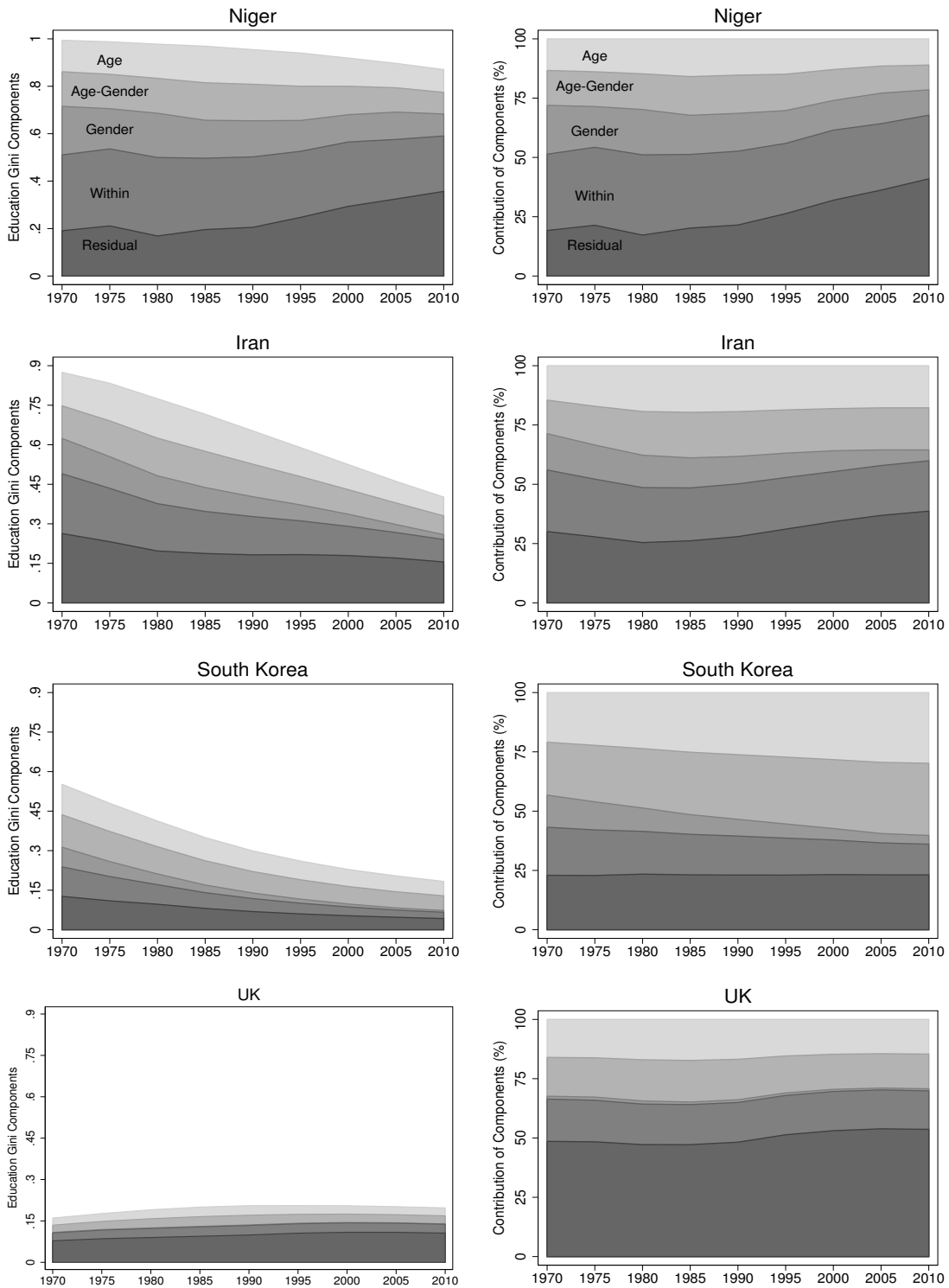


Figure 3: Components of Education Inequality, Selected Countries

The within-group component compares individuals of one group while the between component compares population groups by their average educational attainment. The residual component, in turn, interacts with both factors. First, the lower between-group inequality, the more are subgroups of the population overlapping. Or, the more permutations are necessary to rank individuals first, by the average educational attainment of their group and second, by the relative position within their group. Therefore, the relative contribution of the residual component is not only declining in the overall level of education inequality as well as increasing over time, but also strongly decreasing with the relative contribution of the between-group component (see Figure 4). Even if the residual component explains a large part of education inequality in each world region, its contribution of slightly above 30% is relatively low in SA and SSA, where it did not add to the declining trend in the total Gini. Compared to the overall degree of inequality, countries in the Middle East and North Africa stand out with their low level of residual contribution. This is due to outliers in the Middle East,¹⁰ characterized by untypical low between-group-inequality contribution in conjunction with particularly high within-group inequality.¹¹ On the other hand, the residual component adds above 45% to education inequality in high income OECD and LAC countries.

The negative relation is still present when the relative contribution of the between-group component is related to the *level* of the residual component. It is weaker, though, as the variation of the residual substantially increases with declining relevance of between-group inequality. This is due to countries differing with respect to the relevance of within-group inequality. Comparing individuals from different groups solely by group averages abstracts from comparisons of outliers not represented by their group's mean attainment. The more polarized groups are, i.e. the greater the spread within groups, the more relevant these comparisons become. Hence, the higher the residual component. Even if I do not observe any relation between the respective relative contributions, I find the *level* of the residual component to increase with rising relative relevance of within-group inequality. This relation is even stronger if I exclude MENA countries with particularly high within-group contributions but low shares of between-group components.

To sum up, the analyzed components of education inequality exhibit differential roles across regions and vary over time in the process of education improvements. Even if the sample period I observe is dominated by equalizing processes between cohorts, differing trends between males and females as well as over time indicate that the gap between age groups fluctuates in conjunction with the level of educational attainment. This is also true for differences between sexes. My findings indicate that education expansion processes have predominantly started among young males before including females. Yet, closing the gap between sexes of equal age groups has significantly contributed to the declining trend of overall education inequality. The contribution of within-group inequality, in turn, is relatively large but decreasing over time. Beyond that, polarization within groups impairs between-group comparisons, resulting in larger residual components of education inequality.

¹⁰Especially Qatar and the United Arab Emirates.

¹¹The low mean residual contribution in CAEE countries is driven by South-eastern countries.

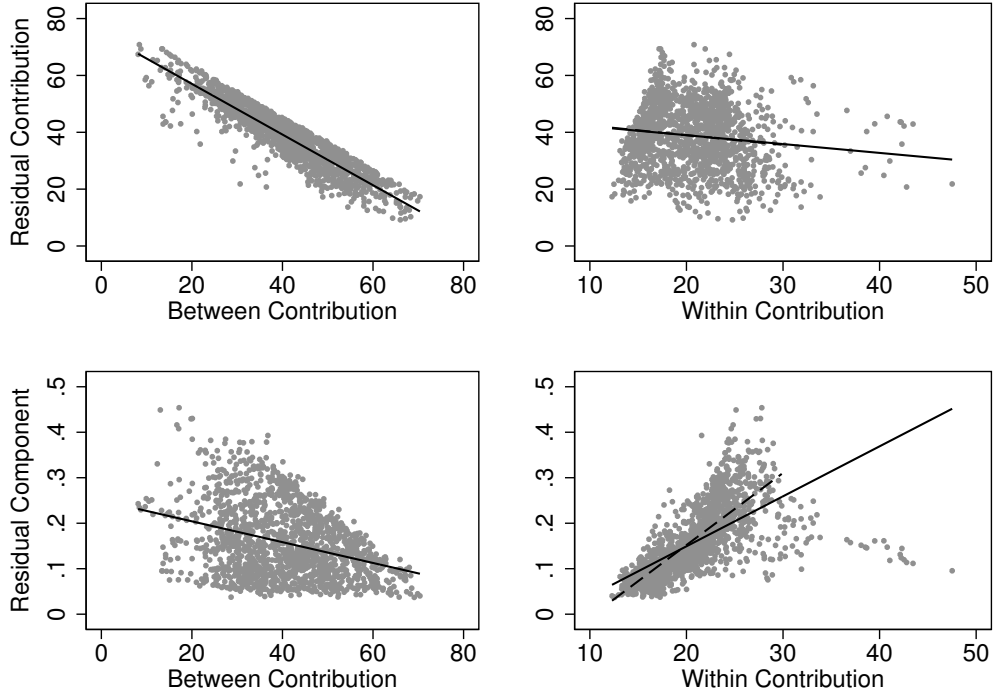


Figure 4: Between - Within & Residual Component

6 Distributional Impacts of Education Expansion

Convergence of educational attainment within and between age groups as well as sexes has accounted for declining inequality in the distribution of education. However, to what extent have education expansion processes around the globe contributed to this trend? Moreover, in how far have components of education inequality have been affected differently?

Previous work on the distribution of education found evidence for a strong negative relation between the degree of education inequality and average educational attainment, mostly measured by mean years of schooling in the concerning population. Yet, Sauer & Zagler (2012) observed that even if this relation holds across countries, it need not be strictly negative within countries over time. Also my findings in Section 5 suggest a non-linear relation between education inequality and average attainment which is driven by the dynamics of between- and within-group inequality. I test whether a non-linear relation between each component and average educational attainment exists in a panel regression framework. Beyond that, I investigate the distributional impacts of variations in the formal duration of schooling cycles. A simple model which aims to explain the level as well as the relative contribution of each component can be written as,

$$\text{Comp}_{i,t} = \alpha_i + \beta_1 \bar{Y}_{i,t} + \beta_2 \bar{Y}_{i,t}^2 + \beta_3 \text{dur}_{i,t}^P + \beta_4 \text{dur}_{i,t}^C + \beta_5 \text{dur}_{i,t}^S + \lambda_t + \varepsilon_{i,t} \quad (16)$$

Table 3: Regression Results: Level of Components

	I_G^E	I_{age}^E	$I_{age/sex}^E$	I_{sex}^E	I_{within}^E	$I_{residual}^E$
\bar{Y}	-0.107*** (0.002)	-0.014*** (0.001)	-0.014*** (0.001)	-0.031*** (0.001)	-0.030*** (0.001)	-0.016*** (0.002)
\bar{Y}^2	0.003*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.002*** (0.000)	0.001*** (0.000)	-0.000*** (0.000)
dur^P	0.152*** (0.038)	0.037* (0.020)	0.097*** (0.022)	0.033 (0.023)	0.018 (0.015)	-0.032 (0.040)
dur^C	-0.152*** (0.041)	-0.052** (0.022)	-0.077*** (0.024)	-0.034 (0.025)	-0.039** (0.016)	0.051 (0.044)
dur^S	0.040** (0.019)	0.018* (0.010)	0.036*** (0.011)	0.025** (0.012)	0.018** (0.007)	-0.057*** (0.020)
Obs	1,539	1,539	1,539	1,539	1,539	1,539
N	171	171	171	171	171	171
T	9	9	9	9	9	9

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

where $Comp_{i,t}$ is the concerning component of education inequality or its relative contribution. $\bar{Y}_{i,t}$ is a measure of mean years of schooling in the population aged 25+ and its square accounts for the presumed non-linearity in the relation of interest. $dur_{i,t}^P$, $dur_{i,t}^C$ and $dur_{i,t}^S$ are the formal durations necessary to complete primary, compulsory and upper secondary education respectively. This data is taken from UIS (see Section 4) and measured as an average over the years each of 16 cohorts above 25 went to school to complete the concerning education level. Thereby, the time lag between the education policy and educational outcomes is accounted for. I include country- (α_i) and time-specific (λ_t) intercepts. Finally, $\varepsilon_{i,t}$ captures the time varying component of the error term. I estimate the parameters in 16 using the fixed-effects estimator as the expected value of the time-invariant country-specific effects conditional on the explanatory variables cannot be assumed to be zero. The results are provided in Tables 3 and 4.

The estimated relationship between average educational attainment and overall educational inequality as well as its components hints at a U-shape. The total education Gini of the population aged 25+ is predicted to significantly decrease in the process of education expansion until eight years of schooling, but increases thereafter. This non-linear relation is slightly less pronounced for the two age components, but even stronger for the gender as well as the within-group component. The residual component, on the other hand, is increasingly negative related to mean years of schooling.

Looking at relative contributions of education inequality components reveals a somewhat different picture. Most importantly, I find evidence for an education Kuznets curve,

Table 4: Regression Results: Contribution of Components

	I_G^E	I_{age}^E	$I_{age/sex}^E$	I_{sex}^E	I_{within}^E	$I_{residual}^E$
\bar{Y}	-0.107*** (0.002)	1.302*** (0.161)	2.007*** (0.159)	-2.562*** (0.125)	-0.771*** (0.084)	0.023 (0.316)
\bar{Y}^2	0.003*** (0.000)	-0.064*** (0.009)	-0.071*** (0.009)	0.141*** (0.007)	0.009** (0.005)	-0.017 (0.018)
dur^P	0.152*** (0.038)	14.478*** (3.856)	20.661*** (3.792)	1.843 (2.994)	-5.532*** (2.001)	-31.451*** (7.549)
dur^C	-0.152*** (0.041)	-19.674*** (4.214)	-20.684*** (4.144)	3.784 (3.272)	2.549 (2.187)	34.025*** (8.250)
dur^S	0.040** (0.019)	9.457*** (1.936)	10.685*** (1.904)	-0.259 (1.503)	0.803 (1.005)	-20.686*** (3.791)
Obs	1,539	1,539	1,539	1,539	1,539	1,539
N	171	171	171	171	171	171
T	9	9	9	9	9	9

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

i.e. an inverted U-shape relation between the relative contribution of both age related components and mean years of schooling. Thus, as education expands, relative differentials between age groups first increase but vanish if educational attainment of successive young cohorts stagnates. In contrast, the relative relevance of inequality between sexes as well as within groups increases with higher average educational attainment. While the turning point is not yet observable for the within-group component, it is relatively low at 5 years of schooling for the gender component. This result, on the one hand, supports the previous finding that education expansion processes have started among young men and, on the other hand, reveals the tendency of women to attain higher education levels than men not only in high income OECD economies but also some countries of East Asia, Eastern Europe and Latin America. In general, these results provide indication for inequality to change in waves with educational attainment. As long as specific groups within societies are the first to benefit from improvements, inequality will first rise but decrease as larger parts of the population take part in education and remain constant until the next impulse for advancements.

Variations in the formal duration of schooling can have large distributional effects. Depending on the education level in which changes take place, particular population groups benefit from such policies. Holding the duration of higher levels constant, an increase in the duration of primary education significantly widens the gaps between age groups, thereby increasing the relative contribution of both age components. This effect is, however, almost entirely compensated by an expansion of compulsory education via rising the formal duration of lower secondary schooling. In contrast, a longer duration of

primary education significantly contributes to reduce the relative contribution of within-group inequality. Finally, an increase in the formal duration of upper secondary education significantly increases the relative relevance of the gap between the youth and the elderly. This is also true for the corresponding levels of age-related components. Beyond that, a longer duration of higher education levels tends to benefit the sex or other group of the population which is relatively high represented.

7 Conclusions

Investigating the trends of education inequality between as well as within population subgroups sheds light into the distributional implications of different education expansion trajectories. The analyzed components exhibit differential roles across regions and vary over time as societies become educated. In general, as long as specific groups within societies are the first to benefit from improvements, inequality will first rise but decrease as larger parts of the population take part in education; and remain constant until the next impulse for advancements. This is especially true for divides between age groups, which increase as younger generations become higher educated and decrease as these cohorts age. However, I find evidence on gender specific developments which indicate that education expansion processes have predominantly started among young males before including females. Yet, closing the gap between sexes of equal age groups has significantly contributed to the declining trend of overall education inequality throughout the observed sample period. Indeed, young women are already higher educated than their male counterparts in some middle and highly educated countries. Inequality in the distribution of education within population subgroups accounts for a large part in overall education inequality. Thus, even its relevance is decreasing over time, the ethnic background or the social and economic status of people continues to determine the educational prospects of people. Beyond that, the larger the polarization within groups, the less meaningful are group comparisons, resulting in an underestimation of between-group inequality. Cross-regional variations in the roles of divides between sexes as well as within population groups indicate that differential institutions are a relevant factor in shaping social inequalities.

My findings involve some suggestions for education policies, which are able to affect the distributional implications of education expansion. As long as societies are segmented along various lines, policies benefit particular groups. For example, I find increasing the formal duration of lower education levels to exert equalizing effects while longer durations of higher levels tend to widen educational gaps between both sexes and socio-economic groups. From this it follows that a comprehensive understanding of the distribution of education between and within various population groups in conjunction with the identification of target groups should help to improve the distributional effects of policies.

Not only can the new indicators of between- and within-group inequality in educational attainment provide novel insights into comparative education expansion trajectories around the globe, but also into the macro-economic consequences and broader societal effects of education inequality. For example, the age component measures the extent of education expansion and can be used to examine the consequences for economic growth. Relating my measures of gender inequality in educational attainment to democratization,

female labor force participation or fertility can contribute to the understanding of the role of women in development processes. Moreover, they can be used to analyze whether closing education gender gaps have contributed to narrowing gender wage gaps at the aggregate level.

References

- BARRO, ROBERT J., & LEE, JONG-WHA. 2010. *A New Dataset of Educational Attainment in the World, 1950-2010*. NBER Working Paper 15902.
- BAUER, RAMON, POTANCOKOVÁ, MICHAELA, GOUJON, ANNE, & KC, SAMIR. 2012. *Populations for 171 Countries by Age, Sex, and Level of Education around 2010: Harmonized Estimates of the Baseline Data for the Wittgenstein Centre Projections*. IIASA Interim Report IR-12-016.
- BENAABDELAALI, WAIL, HANCHANE, SAID, & KAMAL, ABDELHAK. 2012. Chapter 13 Educational Inequality in the World, 1950-2010: Estimates from a New Dataset. *In: BISHOP, JOHN A., & SALAS, RAFAEL (eds), Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber (Research on Economic Inequality, Volume 20*. Emerald Group Publishing Limited.
- BIRDSALL, NANCY, & LONDONO, 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.
- CRESPO-CUARESMA, JESUS, SAUER, PETRA, & KC, SAMIR. 2013. *Gini Coefficients of Educational Attainment: Age Group Specific Trends in Educational (In)equality*. WWWforEurope Working Paper no 6.
- 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.
- INTER-AMERICAN DEVELOPMENT BANK. 1999. *Facing up to Inequality in Latin America: Economic and Social Progress in Latin America, 1998-99 Report*. Johns Hopkins University Press.

- 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.
- KC, SAMIR, POTANCOKOVÁ, MICHAELA, BAUER, RAMON, GOUJON, ANNE, & STRIESSNIG, ERICH. 2013. *Summary of Data, Assumptions and Methods for New Wittgenstein Centre for Demography and Global Human Capital (WIC) Population Projections by Age, Sex and Level of Education for 195 Countries to 2100*. IIASA Interim Report IR-13-018.
- 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.
- LONDONO, 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.
- MAAS, JACOB, & CRIEL, CEERT. 1982. *Distribution of Primary School Enrollments in Eastern Africa*. World Bank Staff Working Paper 511.
- POTANCOKOVÁ, MICHAELA, KC, SAMIR, & GOUJON, ANNE. 2014. *Global Estimates of Mean Years of Schooling: A New Methodology*. IIASA Interim Report IR-14-005.
- 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*.
- SAUER, PETRA, & ZAGLER, MARTIN. 2012. Economic Growth and the Quantity and Distribution of Education: A Survey. *Journal of Economic Surveys*, **26**(5), 933–951.
- SAUER, PETRA, & ZAGLER, MARTIN. 2014. (In)equality in Education and Economic Development. *Review of Income and Wealth*, **60**, 353–79.
- SHERET, MICHAEL. 1988. Equality Trends and Comparisons for the Education System of Papua New Guinea. *Studies in Educational Evaluation*, **14**(1), 91–112.
- SILBER, JACQUES. 1989. Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality. *The Review of Economics and Statistics*, **71**(1), 107–115.

A Between-group Inequality

The contribution of inequality between any two categories or groups p, q can be written as follows (Silber, 1989),

$$\begin{aligned} I_{Bpq}^E &= -\left(\frac{n_q}{n}\right)n_p\bar{s}_p + \left(\frac{n_p}{n}\right)n_q\bar{s}_q \\ &= \frac{n_p + n_q}{n}(n_p\bar{s}_p + n_q\bar{s}_q)I_{pq}^E \end{aligned} \quad (17)$$

where

$$I_{pq}^E = \frac{n_q}{n_p + n_q} \frac{n_p\bar{s}_p}{n_p\bar{s}_p + n_q\bar{s}_q} - \frac{n_p}{n_p + n_q} \frac{n_q\bar{s}_q}{n_p\bar{s}_p + n_q\bar{s}_q} \quad (18)$$