GLOBALIZATION AND EDUCATIONAL INEQUALITY DURING THE 18TH TO 20TH CENTURIES: LATIN AMERICA IN GLOBAL COMPARISON

JOERG BATEN
(University of Tuebingen and CESifo)[1]
CHRISTINA MUMME
(University of Tuebingen)

This is the last working paper version before this study was submitted and accepted. Please cite as Baten, Joerg and Mumme, Christina (2010). "Globalization and Educational Inequality in Long-Run Development during the 17th to 20th Centuries: Latin America and other Developing World Regions", *Journal of Iberian and Latin American Economic History* 28-2 (2010), 279 -305. Online: doi:10.1017/S021261091000008X

ABSTRACT

This paper explores inequality of numeracy and education by studying school years and numeracy of rich and poor, as well as tall and short individuals. To estimate numeracy the age heaping method is used for the 18th to early-20th century. Testing the hypothesis that globalization might have increased inequality of education, we find evidence that 19th century globalization actually increased inequality in Latin America, but 20th century globalization had positive effects by reducing educational inequality in a broader sample of developing countries. Moreover, we find strong evidence for Kuznets' inverted U hypothesis, i.e. rising educational inequality with GDP per capita in the period until 1913 and the opposite after 1945.

Keywords: Human Capital, Inequality, Age heaping, Numeracy, Globalization
JEL Classification: I21, N30, N36, O57

---

[1] Dept. Economics, Address: Mohlstrasse 36, 72074 Tuebingen, joerg.baten@uni-tuebingen.de. Mumme: mumme@gmx.net Acknowledgements: We are indebted to participants of the conference "A Comparative Approach to Inequality and Development: Latin America and Europe" in Madrid 2009 and seminar participants at Tuebingen for scientific discussions and advice. We thank the ESF-GlobalEuroNet project for financial support. This paper is produced as part of the project Historical Patterns of Development and Underdevelopment: Origins and Persistence of the Great Divergence (HI-POD), a Collaborative Project funded by the European Commission's Seventh Research Framework Programme, Contract number 225342. For important comments and data, we thank an anonymous referee, Lety Arroyo Abad, Pablo Astorga, Luis Bértola, Tarcísio Botelho, Stephen Broadberry, Albert Carreras, Rafael Dobado, Ewout Frankema, Thomas Ginter, Anton Glatz, Bishnu Gupta, Marianne Hock, Christina Jedermann, Dacil Juif, Kerstin Manzel, Adolfo Meisel, Alexandra Mylius, Sarah Nalle, Ricardo Paixao, Javier Puche Gil, Heike Schmutz, Bernd Stein, Yvonne Stolz, Linda Twrdek, Gerrit Ulrichs, Jan Luiten van Zanden and Carolina Vicario.
Appendix: https://www.wiwi.uni-tuebingen.de/cms/fileadmin/Uploads/Schulung/Schulung5/Paper/baten_mumme_appendix.pdf

RESUMEN

En este artículo se estudia la inequidad en la alfabetización matemática –*numeracy*- y en la educación analizando los años de escolarización y alfabetización matemática de ricos y pobres, ya sean individuos altos o bajos. Para estimar la alfabetización matemática utilizamos el método *age heaping* para los siglos XVIII y XIX. Contrastando la hipótesis de que la globalización puede haber incrementado la inequidad educacional, encontramos evidencias de que la globalización del siglo XIX aumentó esta inequidad mientras la globalización del siglo XX tuvo, sin embargo, efectos positivos en la reducción de la inequidad educacional en una amplia muestra de países en desarrollo. Además, hemos encontrado evidencias que confirman la hipótesis de la U invertida de Kuznets: una relación positiva entre la inequidad educacional y el PBI per cápita en el período anterior a 1913 y lo opuesto en el período posterior a 1945.

Palabras clave: capital humano, desigualdad, *age heaping*, globalización

1. INTRODUCTION

Inequality is an important factor in today's globalization of the world economy and one of the underlying causes of income inequality, educational inequality, is at the core of the debate. Educational inequality is in many cases difficult to measure because micro surveys, which are not normally comparable across countries and periods, are necessary. This contribution uses the numeracy difference between occupational groups as a measure of inequality for the period from the 18th to the early-20th century, and the difference in years of schooling between the taller and shorter half of the female population as a measure for late-20th century educational inequality. For the earlier period we concentrate on Latin America, a region where 20th century income inequality is famously high, and for the latter period we compare the countries of this region with less developed countries and medium income countries elsewhere.

The relationship between globalization and educational inequality is one of today's major issues. Is it possible that the current globalization will fail, just as the previous globalization tendency in the period 1850-1914 did, because inequality stimulates anti-integration forces? Timmer and Williamson (1998) found that during the 19th century, inequality in new world countries such as the USA, Brazil, Argentina, Canada and Australia provoked anti-immigration policies that led to the disintegration of Atlantic labor markets. Rising inequality could also decrease the legitimation of international integration, so that those groups which normally benefit from it (for example, the well-educated in rich countries and world inhabitants in general) might not give it their full support.

There are other reasons why the study of educational inequality determinants is important. Firstly, inequality is now often considered as a component of the standard of

living: Being at the bottom of the income distribution is much harder to bear if the distance to the wealthier part of the economy is large, and educational inequality is a determinant of later income inequality. This also applies to the inequality of schooling.[2]

Wood (1997) argued that for the 1980s and 1990s more open trade increased wage inequality in some parts of the world, particularly in Latin America.[3] His studies have focused on the 1980s and 1990s. The question is whether the relationship of globalization and inequality holds before the 1980s. In our study we are not restricted to only two decades. We are able to analyze the relationship for three centuries and will assess the difference between educational inequality in the 'First Era of Globalization Period', as O'Rourke and Williamson (1999) defined it – namely, the 1850-1913 period – with the early phases. For a second study period, 1945-1984, we will use the openness indices of Sachs and Warner as well as trade shares to assess whether openness increased educational inequality. Inequality, though, is a complex phenomenon and many potential determinants should be taken into account as well as globalization. We will therefore control for as many other potential determinants as possible.

Why should openness matter for educational inequality? Most research in this field has focused on income inequality in OECD countries, arguing that imports of goods mainly produced with unskilled labor could decrease the demand for unskilled labor within the rich OECD countries, depressing unskilled wages and increasing inequality. However, factor endowments and relative scarcities in developing countries (LDCs) differ fundamentally.[4]

Our expectation is that openness in fact increases inequality in countries with abundant land, and lowers inequality if unskilled labor, relative to potential trading partners, is the abundant factor. In the absence of unusual complementarities between factor inputs and other counter-acting forces, poor countries will increase their exports of unskilled-labor intensive products in globalization periods because their abundant factors and their comparative advantage are likely to be in this segment. Increasing production with unskilled labor should increase unskilled labor demand and wages, and the opposite should be true for land-rich countries. If labor demand rises (falls), even children of unskilled workers should receive some schooling (or less), although in most cases not enough to move into the upper half of income recipients.

Now, Latin America was clearly a land-rich region between the 18th and early 20th

[2] Castello and Domenech (2002), Thomas et al. (2001)
[3] Wood (1997), UN (1995), Cepal (2004)
[4] Wood (1994, 1997, 1998)

century.[5] Therefore, the expectation would be a rise in inequality during the "First Era of Globalization" (1850-1913).[6] Does the expectation also hold for the broader sample of developing countries during the 1945-84 period? It is somewhat less clear as some were already industrializing during the period and land became less important.

A number of scholars have studied the influence of globalization and de-globalization on within-country income inequality worldwide.[7] Bourguignon and Morrison (2000) found a strong decline of within-country inequality during the de-globalization phase of 1914-45, whereas within-country inequality rises during globalization phases. Prados (2007) finds increasing Latin American income inequality during this period. In the next section we will assess whether educational inequality also increases with growing openness.

2. YEARS OF SCHOOLING, AGE HEAPING, OCCUPATIONS AND INEQUALITY

We will first discuss the general idea of the age heaping method in this section, before discussing our measures for inequality of human capital. Age heaping is often used nowadays as a basic numeracy indicator. The share of people who are able to report their exact age rather than report a rounded age has been found to be strongly correlated with numerical abilities (Crayen and Baten 2010a). A widely-used measure for age heaping is the ABCC index as suggested by A'Hearn, Baten and Crayen (2010a), which divides the number of people who reported non-preferred ages (that is, ages which are not a multiple of five) by the total number of people. The index ranges from 0 to 100. If everybody reports the correct age, ABCC has a value of 100. Here we restrict the evidence to the age groups 23-32, 33-42, 43-52 and 53-62, because ABCCs of younger and older individuals might be biased. Only units which have at least 30 cases per skill and age group are studied. The ABCC index can best be understood by considering an example. If we have 100 people with unskilled occupations of a specific age group reporting age in the census of, say, 1870, we would expect 20 of them to report an age ending in 0 or 5 (because 2 of 10 ages end in 0 or 5). For the remaining 80 individuals the question is: do they report other ages or do they also choose an age ending in 0 or 5? If 20 do the latter, then one-quarter (of the 80 remaining persons) report probably a wrong age, the ABCC is then 75 percent (one minus one-quarter). If the people with skilled occupations in the same country and birth decade have an ABCC of 85, then the social difference of numeracy between those occupational groups is 10 (=85-75).[8]

[5] Prados (2007)
[6] O'Rourke and Williamson (1999)
[7] Lindert and Williamson (2001)
[8] This holds only if age distributions are relatively smooth. Crayen and Baten (2010a) studied the influence of famines, epidemics, wars and civil wars and found that the effect was randomly distributed and in the vast

It is important, however, to counter-check whether census-takers or recruitment officers did explicitly ask for the age (and did not 'correct' the reported ages afterwards). In the case of the samples studied here we have good reasons to believe that the people were actually asked for their age, and the number of corrections made afterwards was not large. Otherwise, the relatively high level of age heaping that we observe in the data would probably not have occurred.[9]

How close is the relationship between age heaping and other human capital indicators such as literacy and schooling? A'Hearn, Baten and Crayen (2009) used the large U.S. census sample to perform a very detailed analysis of this relationship. They subdivided by race, gender, high and low educational status and other criteria. In each case, they obtained a statistically significant relationship. The fact that the coefficients are relatively stable between samples is also or noteworthy, i.e. a unit change in age heaping is associated with similar changes in literacy across the various tests. The correlation was both statistically and economically significant for any country studied so far which had substantial age heaping.[10]

Some uncertainty remains about whether age heaping in the sources contains information about the numeracy of the responding individual or, rather, about the diligence of the reporting personnel who wrote down the statements. A potential bias always exists if more than one person is involved in the creation of a historical source. For example, if literacy is measured by analysing the share of signatures in marriage contracts, there might have been priests who were more or less interested in obtaining real signatures, as opposed to just crosses or other symbols. We find it reinforcing that we estimate generally much more age heaping (and less numeracy) for the lower social strata, and among the half of the sample population which had lower anthropometric values. Moreover, the regional differences of age heaping are similar to regional differences in illiteracy.

We conclude that the age heaping method is now a well-established indicator for numeracy of groups, but the problem regarding how upper and lower group members can be distinguished from each other for historical populations for which we typically have no individual income data remains. Occupations have often been used to classify upper versus lower income group individuals and we will apply this criterion to Latin American data until

---

majority of cases not influential for individual age groups. The method also assumes that ages ending in 0 and 5 are the most clearly preferred ages. This is least clear for the age group 23-32 because heaping also frequently takes place on multiples of 2. Crayen and Baten (2010a) suggest reducing the ABCC for this age group, a recommendation which we also follow.

[9] Even if the precise birthday (often related to a saint's day or a holiday) is known to the individual, it might well be the case that the exact number of years since birth means little to an individual although the annual event is celebrated again and again.

[10] On the regions of Argentina, see for example Manzel and Baten (2009).

the 20th century (similarly to Crayen and Baten, 2010b). Of course, occupations such as 'day-laborer' or 'agricultural worker' typically yielded a low income, whereas professionals, noblemen, factory owners, and skilled craftsmen had higher incomes. As a caveat to this method it should be noted that some occupations represent a wide income range (farmers, for example).

For the study of the 1945-1984 period, we also use an alternative, similarly rough proxy to distinguish social groups, based on human stature as Crayen and Baten (2010b) suggested. This involves contrasting the number of years of schooling of the taller and shorter 50 percent of the sample. Almost all anthropometric studies that considered occupational or income groupings found that the well-off strata of society were taller.[11] A second very interesting aspect to this strategy is that tall individuals are much less likely than short individuals to have suffered from infant protein deficiency syndrome (IPDS) which reduces learning abilities to a certain extent. The syndrome was widespread during the 1945 to 1984 period in the poorest countries of the world, when malnutrition was so common that most populations were severely stunted (with adult males being shorter than 170 cm on average). Support for this claim comes from biologists and psychologists who have conducted experiments on the influence of protein malnutrition in childhood and intellectual ability later in life.[12]

One caveat to the proposed anthropometric method is clearly that there is also genetic height variation (especially on the individual level, see Magnusson, Rasmusson and Gyllenstein, 2006). Nonetheless, we are confident that most individual variation can be averaged out by means of sufficiently large sample sizes.

Why do we use such a special method to measure inequality of education? Are there not other data sets available that contain some social classification criterion (such as occupation or income), as well as educational measures for the 1945-1985 period? To the best of our knowledge, for such a large number of countries as used here (42 countries), consistent data sets of this type do not exist. [13]

---

[11] For recent collections of anthropometric studies see Steckel and Floud (1997) and Baten and Komlos (2004).
[12] Paxson and Schady (2007).
[13] Another interesting measure of educational inequality was proposed by Frankema (2008) who uses a 'comparative grade enrolment distribution' to determine educational inequality. His idea is that the higher the secondary school completion shares of the attainment distribution, the larger the educational 'middle class' which might imply less educational inequality. Frankema (2008) finds that grade distribution in Latin America is skewed towards lower grades during the mid-to-late 20th century with almost 43% of the pupils leaving school without passing the first grade and more than 70% dropping out of school with less than 4 years of school attendance. So despite the fact that Latin America reached almost full primary school enrolment rates, levels of school completion were very low. Unfortunately, this alternative measure is not available for the early period studied here and it is not available by birth cohort for the later period. Yet another measure of human capital inequality uses skill premia, as studied for long-term periods by van Zanden (2009). He found, for example, that less developed countries such as Indonesia and India had quite high skill premia.

[TABLE A and B AROUND HERE]

3. DATA SOURCES, SELECTIVITIES AND REPRESENTATIVENESS

Many population counts were carried out in colonial Latin America, aiming at an overview of the population, taxpayers and the military potential. Most early counts were focused on limited regions or cities within a country. Larger censuses were carried out after the mid-18th century, covering a higher share of the national population (Table A).[14] For the post-colonial period, censuses of the republics were carried out mostly after the mid-19th century, while the early-19th century is clearly less documented. The Latin American countries currently have the best source situation of historical population enumerations among today's developing countries.[15] Our samples cover Argentina, Brazil, Colombia, Ecuador, Mexico, Uruguay and Venezuela[16] and represent therefore a large part of this world region. All in all, the countries under study represent today around 80% of the Latin American population. An important question is whether our various sources are representative of the whole society during the period under study. This issue has been studied intensively by Manzel and Baten (2009) who used mostly the same sources to study long-run trends. The population enumerations were supposed to have universal coverage in the whole area considered as well as in all social strata. Manzel and Baten have assessed many potential weaknesses of the data, such as social and regional biases. For example, one potential criticism of the padrones of the 18th century is underenumeration. The government wanted to know the population number and age structure in order to learn about the potential of taxpayers. One could imagine that this stimulated avoidance behaviour among the richer part of the population. However, it was not easy for the rich and well-educated strata to avoid being included in the census. We find them in large numbers in our census lists, as is evident from the occupations listed.

Another potential caveat is the problem regarding who really answered the question about the age. Is it possible that perhaps only the head of the household answered for the whole house? Manzel and Baten (2009) applied an indirect method by calculating the age heaping indices for household heads and other members of the household. The expectation was that the head knew his exact age more often than the age of other household members.

[14] Manzel and Baten (2009)
[15] Platt (1998, p.7)
[16] While borders changed during the colonial and post-colonial period, we always refer to today's borders, as far as possible.

The difference between the two groups, however, was not very large. Other scattered evidence comes from remarks of the census officials about heads and other household members. In both cases, there were statements such as 'she did not know her age', combined with an age statement of a preferred age. This can be interpreted as evidence that other household members were also actually asked.

The ethnic composition is important for Latin America. Were Indios or slaves of African origin sometimes omitted from the padrones? The direct comparison of population structure by ethnic group given in the literature and the composition of 18th century padrones[17] revealed that the bias was limited.[18] In some Mexican censuses there was some under-representation of American Indios and in one of the Buenos Aires padrones there was some under-representation of African Americans, but in general the samples were quite representative in terms of ethnicity.

Regional bias is another issue that we need to address. Clearly, the early samples in particular were more often concentrated on the population of the capital (Table A and B). Large cities tend to have higher levels of inequality,[19] hence we expect higher inequality values for the 18th century. We will assess this effect with appropriate dummy variables below.

Finally, an important point for Latin America in particular is whether migrants should be included in the individual samples. Here we are mainly interested in the educational inequality of countries and migrants contributed to this inequality. Therefore, we decided to include migrants as well.[20]

While the sources for the study of Latin American educational inequality during the 18th to early-20th century deserve the most scrutiny, the later 20th century evidence, which we are employing for our second study period, is easier to use. The company Macro International Inc. performs surveys of child health and health-related behaviour in order to create a solid and representative database for improving child health (among other aims). They recorded years of schooling and heights of women mostly born between 1945 and 1984 in many developing countries. We included only those aged 20 to 50 in many developing countries. As the height of adults is mostly determined in the three years after birth, the height of the mother can shed light on the development of status differences in this period after birth. One potential lacuna in our data is the environmental influence on growth at later ages, especially during the

[17] During the 19th century, the New Republican governments forbade statements about ethnicity
[18] Manzel and Baten (2009, Table 3)
[19] Baten (1999)
[20] See notes to Table C.

adolescent growth spurt. However, Baten (2000b) finds that this effect is negligible compared to the impact of the first three years, as long as individuals have reached their final height when measured. Secondly, there could be survivor bias effects, but Moradi and Baten (2005) and Guntupalli and Baten (2006) rejected this possibility in detailed studies.[21]

We consider here the difference in years of schooling of the taller 50 percent compared with the shorter half, and organize the data by individual country and birth decade. It is remarkable that in most cases taller women had more years of schooling (Table C). Small differences refer to cases such as Ghana, Madagascar or Tanzania, in which the urban centers of education differed from the regions of tallest heights, which were sometimes characterized by a specialization in cattle farming.[22] In Latin America, some of the strong educational inequalities by height group are partly determined by the Indio versus European ancestor difference. It is difficult to disentangle socioeconomic from nutritional habit differences (and perhaps genetic ones) here. Hence we will rely on fixed effects regressions below, which control for country-specific characteristics.

[TABLE C AROUND HERE]

4. DEVELOPMENT OF EDUCATIONAL INEQUALITY

We would expect levels of educational inequality in the various countries of Latin America to be quite different because the institutional and economic structures were so varied. For example, Mesoamerica and the Andes had large shares of indigenous or mestizo populations, who received less schooling and other public goods compared with the middle and upper strata of European origin. In these regions, the inequality heritage of land distribution in favour of the Spanish conquerors and later European immigrants might have been strongest.[23] Similarly strong differences might have prevailed in the countries that kept slavery until the late-19th century and whose population component of African origin was disadvantaged. In contrast, the population of the Southern Cone was more homogenous in ethnicity-related aspects, because the Indio population share was smaller and slavery was abolished earlier.

[21] The data set refers mostly to mothers. Moradi (2002) explored the potential difference between mothers and non-mothers. He finds a very moderate selectivity among young mothers. Mothers at age 20-25 were slightly less educated than the reference population. By employing usual height elasticities for education levels, Moradi estimates about one millimeter shorter height of mothers, compared with the reference population of all women aged 20-25. There was no significant selectivity among older women. This result suggests that selectivity of mothers might not be a major problem.

[22] Moradi and Baten (2005)

[23] Lambert (1968, p. 581)

However, ethnicity, slavery and colonial heritage were not the only factors at work. O'Rourke and Williamson (1999) argued convincingly that the Southern Cone countries had strongly increasing inequality during the late-19th century globalization movement.

Going further back in time, how might inequality have differed between colonial times and post-independence Latin America? Unfortunately, today we have little evidence for the pre-independence inequality history of Latin America. Williamson (2009) and Dobado and Garcia (2009) have recently raised some doubts about the early colonial heritage hypothesis (of continuously high inequality). Dobado and Garcia argued that real wages were quite high in some parts of Bourbon Latin America, whereas average income was lower than in Europe. Hence, inequality might actually have been lower than in Europe. The question is, of course, whether this wage evidence is representative and can inform us about the situation of other poorer strata, which did not earn wages (such as the population majority of peasants). Williamson (2009) considered the fact that especially the low population density of the 17th century and early-18th century might have generated relatively low inequality, again compared with Europe. In times of labor scarcity, wages tend to be higher and even the nutrition and general treatment of slaves and indigenous bound labor might be slightly less terrible. Recent work on Uruguay suggests that in the Southern Cone during the 18th century, inequality might also not have been very pronounced.[24]

Economic policy differences certainly also mattered for country-specific differences. For example, the famous Mexican dictatorship era of Porfirio Díaz (1877-1911) increased average income and education, but had a reputation for bringing about sharp increases in inequality.[25] In sum, during the 19th century the variation of inequality between countries was probably large. Pre-independence inequality history is largely unexplored, but some authors recently argued for a modest inequality level relative to Europe.

In the following, we firstly compare numeracy of the upper and lower occupational strata in Latin American countries for the early period. For Argentina, we have some data for Buenos Aires for the birth decades until 1740 and some representative national data after this date. The ABCC Index increased from 24 to 56 for the lower income groups in Buenos Aires from the late-17th to the mid-18th century (Table B). The upper income groups started with a level above 40 percent age numeracy in the 1680s, but grew to just 64 percent in the 1740s. The fact that the early evidence on Argentina covers only Buenos Aires is certainly a caveat. The gap for the 19th century was large and relatively constant, declining only slightly from a 14 percent difference in the 1820s to a 9 percent difference in the 1860s.

[24] Vicario (2010)
[25] Tutino (2001, p. 700-701)

Secondly, we have long-term data on Mexico covering sufficient observations to study both social groups between the 1730s and 1760s, and in the 1880s to 1900s. The latter evidence is nationally representative, while the former relates to some Mexican regions, including Mexico City. In short, social differences in age numeracy were small and in one case even negative in 18th century Mexico. This is consistent with the observation of Tanck de Estrada (1999) that the Bourbon reforms of the 18th century had some positive impact even on school building in Indio villages, i.e. for the poorest Mexicans. However, the late-19th century saw a highly stratified society with large differences. This fits with available evidence on the Porfiriato.

We have more scattered data for a number of other Latin American countries (Table B). In Brazil, the highest educational inequality levels were reached in the 1840s, which is the latest value we have.[26] In Uruguay, inequality was declining but this might have been caused by the fact that the later birth cohorts were based on data from prisons. Venezuela and Colombia had quite high inequality, whereas in poor Ecuador even the skilled groups had low levels of numeracy. Summing up, we can document a number of Latin American countries but the cases for which we have sufficient numbers of observations are somewhat distributed over the different centuries. Most striking in Table B is the fact that, of the 40 cases studied only two have negative values for the difference between the skilled and unskilled groups.

Now we move to the birth cohorts of the mid-to-late 20th century. We studied the difference of school years, subtracting the figure for the taller half from that of the shorter half (Figure A). The years of schooling are a better measure for the mid-to-late 20th century than age heaping which had already disappeared in many countries by then. It should be noted that there are many African countries in the sample, a few Latin American ones, but fewer countries from other regions. One important result from these samples is that the taller half of the population (those who probably came from more advantaged family backgrounds) always had higher or equal school year values, whereas school year values tended to be lower among the shorter half of the population.[27] Latin America had the highest difference (which means the largest educational inequality) whereas Soviet Central Asia had the lowest value (Figure A).

[FIGURE A AND TABLE D AROUND HERE]

[26] We thank Yvonne Stolz, who plans to study the Brazilian case in more detail, for providing the 18th century evidence.

[27] See also the Appendix referenced in footnote 1. In a separate Appendix (available from the authors) we show that the inequality of literacy between the taller and shorter half correlates with the inequality of numeracy measured in the same way.

5. OPENNESS AND OTHER POTENTIAL DETERMINANTS OF INEQUALITY

We first describe the explanatory variables in Table D that we include in the regressions and then discuss the results.

Openness. How can we measure 'openness'? Given the importance of this variable, much work has been done in this field. Most economists agree that simple trade shares of GDP are insufficient by themselves to capture the degree of openness of an economy. If two neighbouring free-trade countries have exactly the same factor endowments, it is possible that their trade is relatively low in spite of their openness simply because production is so similar. On the other hand, two highly protected countries can experience high trade shares if their endowments are sufficiently different. One alternative measure is the openness index of Sachs and Warner (1995) for 79 countries. These authors consider high tariffs, important tariff barriers, plus state monopolies of major commodity exports, a high black market premium for national currencies, and a socialist economic system. This variable is coded as a binary variable. Rodriguez and Rodrik (1999) criticized the fact that two factors in particular, the state monopoly and currency black market premium, might measure other economic characteristics rather than just a lack of openness. The currency value distortion also indicates other macroeconomic problems. However, if there is no perfect measure of openness, it is a promising strategy to use both this one and the trade share. In spite of our conceptual skepticism against the trade share of GDP as a measure of 'openness', we will also test this variable below.

Kuznets curve effects. Kuznets (1955) found that inequality first rises and then declines with economic development. He explained his inverted U curve by labor-market disequilibria: Since technological progress initially favours the rewards for some specialized skills, demand for unskilled labor decreases and its wage falls. Therefore, inequality rises initially. The diffusion of skills and economic policies serve as egalitarian forces which reduce inequality. We therefore add Kuznets variables by adding real GDP per worker in linear and quadratic form and expect a positive coefficient of the former and a negative effect for the latter.[28]

Democracy. Li, Squire and Zou (1998) also emphasized the importance of political freedom for income equality. If dictatorship provides privileges to certain groups in society,

[28] Kuznets effects have recently been studied by Morrisson and Murtin (2007) for educational inequality. They construct a within-country indicator on the basis of primary, secondary, tertiary enrollment rates and confirm the existence of an educational Kuznets curve by studying this indicator for the 1870 to 2000 period. Kuznets curves have also been a traditional field of study for income inequality research. For example, Prados (2008) recently found a Kuznets curve for Spain, 1850-2000.

this might lead to higher inequality. The research project 'POLITY IV', at the University of Maryland, created comprehensive surveys on the democratic or autocratic behaviour of governments in recent history, approximating democracy with a numerical score.

Demographic effects ('Mature'). Did competition reduce the wages of baby boomers? Demographic effects could have an influence on inequality. According to the normal life-cycle effect of income, people receive their highest income in their 40s and 50s. Teenagers and young adults between age 15 and 40 earn less on average, and beyond the age of 60 income starts to decline again. If 'fat' cohorts (for example, the 'baby boom' generation of the 1960s) enter the labor market, we would expect a rise in inequality because the supply of 'young' labor is very large, whereas the share of the richer 'mature' age group is relatively smaller. Higgins and Williamson (1999) found a robust influence of both cohort sizes of the mature age groups. We include cohort size effects by taking the share of the mature population (aged 40-59), relative to the total population of the age groups 15-69 (working age), using the same specification as Higgins and Williamson (1999).

Speed of structural change. How much did agricultural productivity lag? Agricultural productivity, and therefore agricultural incomes, might lag behind industry and services and this could lead to rising inequality.[29]

Civil War. Civil war is one of the strongest determinants of welfare and educational development in developing countries. Civil war has a very destructive effect on average schooling levels, but it is less clear whether this terrible military nightmare increases or decreases the inequality of schooling. In some cases, the better-off population might be able to flee to quiet parts of the country and their children might continue attending school. On the other hand, the destruction of expected human capital returns in the future might particularly affect those strata which otherwise would have invested a lot in the schooling of their children. Hence, it is an empirical question whether this variable increased or decreased educational inequality.[30]

Results for the early period. All regression models are estimated as fixed effects in order to control for unobservable characteristics, such as cultural or geographic factors. In the regressions for the early period, we employed a dummy variable for the 'First Era of Globalization' (1850-1913) as an indicator for openness. For the 18th to early-20th century, there is insufficient evidence to reconstruct the trade share or political protectionism, except

[29] Baten and Fraunholz (2004)

[30] Descriptive statistics of the explanatory variables are available in the Appendix, see footnote 1. Some variables are slightly skewed, but in most cases the skewness is only modest. Given that some of the values of the dependent variables are negative, we decided not to use a logarithmic transformation.

perhaps for the last few decades of our study period (and for this period alone the number of observations would be too small). We find the regressions to have a positive coefficient of the First Era of Globalization dummy variable (Table D).

The absolute level of numeracy might reduce the inequalities in this period slightly (only significant in column 2). The 'baby boom effect' of the 'mature' variable is not visible in this period. Civil war did not have strong and significant effects on early inequality.

In the second regression we also included GDP per capita as well as its squared term. The value of the former was large and positive while the latter had a large negative value. The former term is larger than the latter and has a greater effect on the observed values.[31] Hence, for this early period the Kuznets curve was on the rise.

Results for the 1945-1984 period. We compare three regression models for the later period (Table D, Column 1 to 3). Openness actually reduced educational inequality in this period. The coefficients of Sachs/Warner openness are statistically significant although the values of the coefficients are not very large. In column 5 of Table D, we also included an alternative measure of openness, the trade share as reported in the Penn World Tables. However, this alternative measure is not significant (also after adjustment for population size, it remains insignificant).

In contrast, GDP per capita and its squared term are significant, except in column 4 which records a much smaller number of cases. The coefficient of the non-squared GDP term is much larger than the squared term. This implies that educational inequality declines with increasing income during this period. At very high levels the decline stops. The results are also quite robust over different specifications.

The level of average education increases the gap in less developed countries, which is quite the contrary to what we might have expected. In contrast, civil war mostly reduces educational inequality – it appears that the richer strata do not send their children to school during a civil war either. Finally, there are no obvious effects of democracy, productivity lags, or the 'baby boom' effect (that is, the 'mature' variable). The explanatory power of these models is in general quite large.

Is openness endogenous here? The question is whether the lower educational inequality of children born in a specific birth decade would cause more openness. One could imagine that – based on the general Stolper-Samuelson view – labor-abundant countries with high inequality would open their economies during a globalization period to profit from more demand and therefore higher wages in that sector or vice versa. This might at least be the case

[31] Predicted values are available from the authors.

if unskilled workers have sufficient political power. On the other hand, especially among richer countries, there might be an economic or psychological effect of higher inequality leading to less openness. Baltzer and Baten (2008) tested these hypotheses and provided evidence that low inequality in Latin American countries in the mid-to-late 20th century did not lead to more openness. Hence, we tentatively conclude that endogeneity is not a major problem here.

6. CONCLUSION

We explored inequality of numeracy and education by studying school years and numeracy of rich and poor, as well as tall and short individuals. To estimate numeracy, the age heaping method was used. In this study, we mobilized a large body of new evidence on inequality, going back to the 18th century and covering a number of Latin American countries, namely Argentina, Brazil, Colombia, Ecuador, Mexico, Uruguay, and Venezuela. Looking at the time trend of educational inequality, Mexico displays only modest numeracy advantage for the skilled groups in the 18th century, but the gaps between the upper and lower strata increased strongly until the 19th century. Similarly, Argentina suffered substantial educational inequality during the 19th century. In a regression analysis, the "First Era of Globalization" was mostly confirmed as having higher inequality than earlier periods.

We studied many developing countries in the period from the 1940s to the 1980s, looking at the schooling difference between the taller half of the population and the shorter half. One remarkable finding was that the taller half always had more years of schooling. This applied to 42 different countries without exception. Latin America had the greatest educational inequality in this period, which is certainly one of the reasons for its high income inequality today.

Testing the hypothesis that globalization might have increased inequality of education, we found evidence that 20th century globalization had positive effects by reducing educational inequality. Moreover, we found strong evidence for Kuznet's inverted U hypothesis, which was on the rise during the 18th and 19th century in Latin America and tended to fall in the second half of the 20th century in the developing world.

REFERENCES

A'HEARN, B., CRAYEN, D. and BATEN, J. (2009): «Quantifying Quantitative Literacy: Age heaping and the History of Human Capital». *Journal of Economic History* 68-3, p.783-808.

BALTZER, M. and BATEN, J. (2008): «Heights, Inequality, and Trade in the Latin American Periphery, 1950-2004». *Economics and Human Biology* 6 (2), pp. 191-203.

BATEN, J. (2000b): «Height and Real Wages: An International Comparison**».** *Jahrbuch fuer Wirtschaftsgeschichte* 2000-1, pp. 17-32.

BATEN, J. and FRAUNHOLZ, U. (2004): «Did Partial Globalization Increase Inequality? The Case of the Latin American Periphery, 1950-2000». *CESifo Economic Studies* 50 (1), pp. 45-84.

BATEN, J. and KOMLOS, J. (2004): «Looking Backward and Looking Forward: Anthropometric Research and the Development of Social Science History». *Social Science History* 28 (2), pp. 1-24.

BATEN, J., PELGER, I. and TWRDEK, L. (2009): «The Anthropometric History of Brazil, Lima (Peru), and Argentina during the 19th – 20th Centuries». *Economics and Human Biology 7-4*, pp. 319-333.

CAILLET-BOIS, R.R. (1919): *Documentos para la historia argentina. Vol. 11: Territorio y poblacion. Padrones de las ciudades y campanas de Buenos Aires y Montevideo*. Buenos Aires, Compania Sud-Americana de Billetes de Banco.

CASTELLO, A. and DOMENECH, R. (2002): «Human Capital Inequality and Economic Growth: Some New Evidence». *Economic Journal* 112, pp. C187-C200.

CEPAL (2004): *Social Panorama of Latin America*. Santiago.

CLODFELTER, M. (2002): W*arfare and Armed Conflicts. A Statistical Reference to Casualty and Other Figures, 1500-2000*. Jefferson, North Carolina, and London: McFarland&Company, Inc.

CRAYEN, D. and BATEN, J. (2010a): «Global Trends in Numeracy 1820-1949 and its Implications for Long-Run Growth». *Explorations in Economic History 47-1, pp. 82-99*.

CRAYEN, D. and BATEN, J. (2010b): «New Evidence and New Methods to Measure Human Capital Inequality before and during the Industrial Revolution: France and the U.S. in the Seventeenth to Nineteenth Centuries». *Economic History Review* 53-2, pp. 452-478.

DOBADO, R. / GARCIA, (2009). Neither so low nor so short! Wages and Heights in Eighteenth and Early Nineteenth Centuries Colonial Hispanic America. *Working Paper of the Universidad Complutense, Instituto Complutense de Estudios Internacionales Madrid* 14-09.

DUNCAN-JONES, R. (1990): *Structure and Scale in the Roman Economy*. Cambridge: Cambridge University Press.

FRANKEMA, E. (2008): *The historical evolution of inequality in Latin America: a comparative analysis, 1870-2000*. Dissertation, Groningen: Rijksuniversiteit.

GUNTUPALLI, A. and BATEN, J. (2006): «The Development and Inequality of Heights in North, West and East India, 1915-44**».** *Explorations in Economic History* 43 (4), pp. 578-608.

HIGGINS, M. and WILLIAMSON, J. G. (1999): «Explaining Inequality the World Round: Cohort Size, Kuznets Curves, and Openness**».** *NBER working paper* 7224.

KUZNETS, S. (1955): «Economic Growth and Income Inequality». *American Economic Review* 45 (1), pp. 1-28.

LAMBERT, D.C (1968). «La reparticion des los ingresos y las desigualdades sociales en la America Latina. » *Revista Mexicana de Sociologia* 30-3, pp. 571-592.

LI, H., SQUIRE, L. and ZOU, H. (1998): «Explaining International and Intertemporal Variations in Income Inequality». *Economic Journal* 108, pp. 26-43.

LINDERT, P. and WILLIAMSON, J. G. (2001): «Does Globalization Make the World More Unequal?». *NBER-Working Paper* No. 8228.

MAGNUSSON, P., RASMUSSEN, F. and GYLLENSTEIN, U. (2006): «Height at age 18 years is a strong predictor of attained education later in life: Cohort study of over 950,000 Swedish men**».** *International Journal of Epidemiology* 35, pp. 658-663.

MANZEL, K. and BATEN, J. (2010): «Convergence and Divergence of Numeracy: The Development of Age Heaping in Latin America, 17th to 20th Century. » University of Tübingen, Working Paper.

MANZEL, K. and BATEN, J. (2009): «Gender Equality and Inequality in Numeracy – the Case of Latin America and the Caribbean, 1880-1949». *Revista de Historia Económica – Journal of Latin American and Iberian Economic History (1)* (forthcoming).

MORADI, A. (2002): «Height and Health of Women in Sub Saharan Africa and South Asia 1950-80**».** *Conference Proceedings of the International Economic History Association Conference, Buenos Aires* (on CDROM).

MORADI, A. and BATEN, J. (2005): «Inequality in Sub-Saharan Africa: New Data and New Insights from Anthropometric Estimates**».** *World Development* 33 (8), pp. 1233-1265.

MORRISON, C. and MURTIN, F. (2007): «Education Inequalities and the Kuznets Curves: a Global Perspective since 1870**».** *Paris School of Economics Working Paper* 2007-12.

O'ROURKE, K. and WILLIAMSON, J. G. (1999): *Globalization and History*. Cambridge: Cambridge University Press.

OROZCO y BERRA, M. (1980): *História de la ciudad de México desde su fundacion hasta 1854.* Mexico: SepDiana.

PAXSON, C. and SCHADY, N. (2007): «Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health and Parenting». *Journal of Human Resources* 42 (1), pp. 49-84.

PLATT, L.D. (1998): *Census Records for Latin America and the Hispanic United States.* Baltimore, MD: Genealogical Publishing Com.

PRADOS DE LA ESCOSURA, L. (2007): «Inequality and Poverty in Latin America: a long-run Exploration**»,** in T.J. Hatton, K.H. O'Rourke and A.M. Taylor (eds.), *New Comparative Economic History*. Cambridge: MA, pp.291-315.

PRADOS DE LA ESCOSURA, L. (2008): «Inequality, Poverty and the Kuznets Curve in Spain, 1850-2000**».** *European Review of Economic History* 12, pp. 287-324.

RODRIGUEZ, F. and RODRIK, D. (1999): «Trade Policy and Economic Growth: A Sceptic's Guide to Cross-National Evidence**».** *NBER Working Paper* 7081.

SACHS, J. and WARNER, A. (1995): «Economic Reform and the Process of Global Integration». *Brookings Papers on Economic Activity*, 1995, pp. 1-118.

SOMOZA, J. AND LATTES, A. (1967): Muestras de los dos primeros censos nacionales de población, 1869 y 1895. *Documento de Trabajo No 46, Instituto T. Di Tella*, CIS, Buenos Aires.

STECKEL, R. and FLOUD, R. (1997) (eds.): *Health and Welfare during Industrialization*. Chicago and London: Chicago University Press.

STOLZ, Y., BATEN, J., and BOTELHO, T. (2010): Growth Effects of Brazilian Immigration. Working

Paper Tuebingen/Belo Horizonte.

TANCK DE ESTRADA, D. (1999) : *Pueblos de Indios y Educación en el México Colonial, 1750-1821*. Mexico City: El Colegio de México.

THOMAS, V., WANG, Y. and FAN, X (2001): «Measuring Education Inequality: Gini Coefficients of Education, 1960-2000». World Bank Research Paper No. 2525.

TIMMER, A. and WILLIAMSON, J. G. (1998): «Immigration Policy Prior to the 1930s: Labor Markets, Policy Interactions, and Globalization Backlash». *Population and Development Review* 24, pp. 739-771.

TUTINO, J. (2001). Rural Economy and Society: 1821-1910. In: Werner, M.S. Concise Encyclopedia of Mexico. Chicago: Fitzroy Dearborn Publishers, pp. 697-705.

UNITED NATIONS (1995): *Economic Survey of Latin America and the Caribbean 1994-1995.* Santiago.

VAN ZANDEN, J. L. (1995): «Tracing the Beginning of the Kuznets Curve: Western Europe during the Early Modern Period». *Economic History Review* 48, pp. 643-664.

VAN ZANDEN, J.L. (2009): «The Skill Premium and the 'Great Divergence'». *European Review of Economic History* 13, pp. 121-153.

VICARIO, C. (2010). Estructura social y ocupacional en el Sur Atlántico. Montevideo 1760-1860. Master Thesis, Univ. de la Republica de Uruguay.

WILLIAMSON, J. (2009) History without Evidence: Latin American Inequality since 1491. Inequality Conference Madrid, May 2009.

WOOD, A. (1994): *North-South Trade, Employment and Inequality.* Oxford: Clarendon Press.

WOOD, A. (1997): «Openness and Wage Inequality in Developing Countries: The Latin American Challenge to East Asian Conventional Wisdom». *World Bank Economic Review* 11, pp. 33-57.

WOOD, A. (1998): «Globalisation and the Rise in Labour Market Inequalities». *Economic Journal* 108, pp. 1463-1482.

WORLD BANK (1987): *World Development Report.* Washington: Oxford University Press and Johns Hopkins University Press for the World Bank.

WORLD BANK (1999): *World Development Indicators on CD-ROM*. Washington: World Bank Publications Office.

Table A

DATA SOURCES FOR THE EARLY PERIOD

Country/Region	Year	No. cases (age 23-62)	Potential bias relative to total population	Source
Buenos Aires, AR	1744	1,146	urban, military, incl. slaves	Military Census in the *Documentos para la Historia Argentina* (Caillet-Bois 1919)
Buenos Aires, AR	1771	4,756	urban, incl. slaves	Archivo Nacional de Argentina, Census 1771
Argentina	1869	43,781	no	Somoza and Lattes (1967)
Santa Fé, AR	1887	808	regional	http://www.digitalmicrofilm.com.ar/censos/geografico.php
Argentina	1895	51,715	no	Somoza and Lattes (1967)
São Paulo, BR	1772	1,665	Household heads,servants	Arquivo Histórico Ultramarino, Cód 1270, 2096, see for a description Stolz, Baten and Botelho (2010).
Floresta, BR	1859	1,283	Household heads,slaves	Arquivo Público do Estado de Pernambuco – Depositum Floresta 1859.
São Cristovão, BR	1870	456	regional	Biblioteca do IBGE.
Colombia	1870	2,362	various regions	Archivo Nacional de Bogotá: CE Cauca, Magdalena, Quibdo, Quindio, MF 2, 4, 6, 15,19.
West Ecuador	1870	19,109	various regions	Archivo Nacional de Ecuador, Censo 1870
Hidalgo/Guanajuato/Oaxaca, MX	1740-43	1,383	regional	AGI: CE Ixmiquilpan 1740, Ind, 107; CE Pozos 1743, Ind, 107; CE Southern central Mexico 1743, Ind, 108; CE Chichihualtepec 1743, Ind, 108.
Central and West Mexiko	1777	3,998	regional	AGI, Mex 2578/9.
Mexico - City , MX	1790	3,079	capital, only Spanish and mestizo household heads	Instituto Nacional de Estadística, Geografía e Informática: CE Revillagigedo (2003).
Coahuila, MX	1823	1,598	regional	Grupo Explaroradores Coahuiltecos

Mexico	1930	7,007	various regions, but nationally representative	FSI: CE Guanajuato, Minas de Luz, Mineral de los Llamitos, Ahualuco, Benitez, Tepoztlán, Mezquital, Tetecala, Tlaltizapan: MF 4107114, 4107751, 4107265, 4107065.
Soriano/Maldonado	1834/36	588	regional	Archivo Nacional de R.O. Uruguay, CE Soriano/Maldonado
Montevideo, UY	1846	1,569	capital, prisoners	Archivo Nacional de R.O. Uruguay
Cumarebo /Quisque, VE	1818/20	1,476	regional	AGI: Cuba 759B

Sources: adapted from Manzel and Baten (2009), plus other sources as explained in column 'source'.
Abbreviations: AGI: Archivo General de Indias; AR:Argentina; BR:Brazil; CE:Census of; FSI: Family Search Indexing Project; MF:microfilm; MX:Mexico, UY:Uruguay, VE: Venezuela.

Table B

ABCC INDICES IN SEVERAL LATIN AMERICAN COUNTRIES BY BIRTH DECADE AND OCCUPATIONAL GROUPS

Country	Birth decade	Unskilled	Skilled	Difference
Argentina	1680	24	41	17
	1690	24	43	19
	1700	38	47	10
	1710	44	58	14
	1720	41	56	15
	1730	51	59	8
	1740	56	64	8
	1810	63	77	14
	1820	68	80	12
	1830	71	84	14
	1840	72	84	11
	1850	77	89	12
	1860	81	90	9
Brazil	1710	63	76	12
	1720	63	63	-1
	1730	60	76	16
	1740	53	67	15
	1810	72	92	20
	1820	79	88	9
	1830	70	83	13
	1840	60	82	22
Colombia	1830	56	74	17
	1840	55	65	10
Ecuador	1810	58	63	4
	1820	62	68	7
	1830	64	68	4
	1840	60	68	8

Mexico	1730	56	63	6
	1740	66	61	-5
	1750	70	70	0
	1760	70	75	5
	1880	61	78	16
	1890	62	85	23
	1900	72	75	4
Uruguay	1780	55	71	16
	1790	62	75	12
	1800	79	85	6
	1810	83	83	0
Venezuela	1780	55	71	16
	1790	62	75	12

Notes: "Skilled" refers to occupational groups that were skilled, or professionals, "Unskilled" refers to those with unskilled or only semi-skilled occupations.

Table C

DIFFERENCES OF SCHOOL YEARS BY HEIGHT (BIRTH YEARS 1945-84)

Country	Difference of school years	School years (tall)	School years (short)	Height(tall)	Height(short)
Bangladesh	0.9	3.3	2.5	1547	1460
Burkina Faso	0.4	1.2	0.7	1663	1568
Benin	0.8	2.1	1.4	1636	1537
Bolivia	1.9	7.1	5.2	1557	1467
Brazil	1.4	6.3	4.9	1607	1508
Central African Republic	0.7	2.2	1.5	1639	1533
Ivory Coast	0.5	2.5	2.0	1641	1545
Cameroon	0.5	5.5	4.9	1651	1553
Colombia	1.3	7.5	6.2	1592	1496
Dominican Republic	0.9	7.3	6.5	1614	1516
Egypt	1.2	5.5	4.3	1621	1532
Ethiopia	0.3	1.7	1.4	1619	1521
Gabon	0.7	6.2	5.6	1631	1533
Ghana	0.3	4.9	4.6	1639	1541
Guinea	0.4	1.2	0.8	1638	1541
Guatemala	1.6	2.9	1.3	1518	1423
Haiti	0.9	3.5	2.6	1632	1532
India	0.9	4.2	3.3	1561	1470
Kenya	0.7	6.2	5.5	1646	1544
Kyrgyztan	0.6	11.3	10.7	1628	1538
Comoros	0.7	2.6	1.9	1592	1504
Kazakhstan	0.5	11.2	10.7	1637	1541
Morocco	1.1	3.5	2.4	1630	1539
Madagascar	0.1	3.4	3.3	1578	1487
Mali	0.5	1.2	0.8	1664	1568
Malawi	0.9	3.9	3.0	1607	1515
Mozambique	0.8	2.6	1.8	1609	1512
Namibia	0.7	5.4	4.7	1660	1561
Niger	0.4	0.9	0.5	1654	1560
Nigeria	1.3	5.0	3.7	1644	1529
Nicaragua	1.5	6.0	4.6	1586	1492
Peru	1.9	8.0	6.0	1549	1461
Ruanda	1.0	4.2	3.2	1631	1530

Senegal	0.4	1.3	0.9	1670	1574
Chad	0.5	1.3	0.7	1678	1578
Togo	0.4	1.6	1.2	1638	1542
Turkey	0.8	4.7	3.9	1600	1511
Tanzania	0.2	3.5	3.2	1606	1508
Uganda	0.5	4.6	4.0	1635	1534
Uzbekistan	0.4	11.0	10.6	1649	1553
Zambia	1.0	5.4	4.4	1628	1531
Zimbabwe	0.8	7.0	6.1	1649	1552

Note: “Tall” is defined here as the tallest 50%, “short” as the shortest 50%. The difference in school years is the number of school years of the taller minus the shorter 50%. Female height is reported in millimetres.

Table D

DETERMINANTS OF EDUCATIONAL INEQUALITY

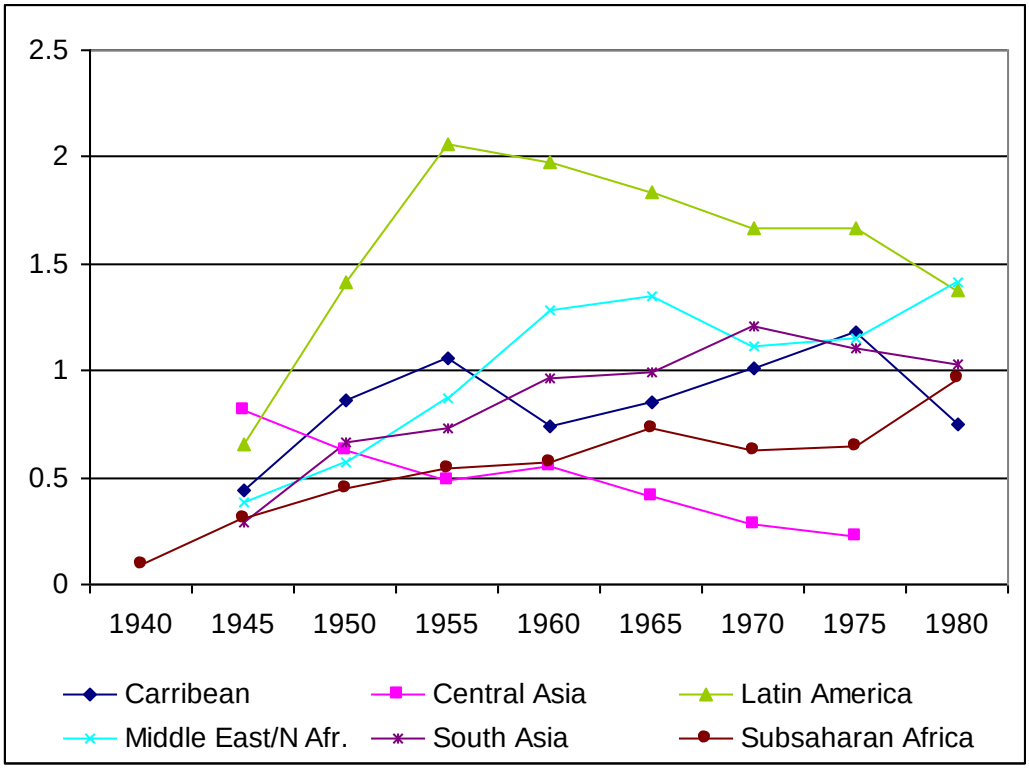
Estimation method	FE	FE	FE	FE	FE
Birth decades	1680s-1900s	1680s-1900s	1940s-1980s	1940s-1980s	1940s-1980s
World regions	Latin America	Latin America	LDCs	LDCs	LDCs
Openness concept	Era of Glob.	Era of Glob.	Sachs/Warner	Sachs/Warner	Penn WT
Openness	7.63**	8.79**	-0.35***	-0.41**	0.01
	(0.041)	(0.042)	(0.004)	(0.040)	(0.94)
Education average	-0.12	-0.33*	0.09***	0.13	0.11***
	(0.26)	(0.059)	(0.0078)	(0.19)	(0.002)
Mature	3.12	0.29		-0.01	
	(0.29)	(0.92)		(0.50)	
Civil War	0.02	-0.96	-0.17*		-0.19**
	(1.00)	(0.77)	(0.088)		(0.042)
GDP p.c.		82.16**	-0.60***	-0.47	-0.62***
		(0.012)	(0.000)	(0.59)	(0.000)
GDP p.c. squared		-43.08**	0.03**	-0.01	0.03***
		(0.011)	(0.020)	(0.95)	(0.009)
Democracy (Polity2)			0.00	0.00	-0.00
			(0.97)	(0.95)	(0.76)
Productivity lag					-0.01
					(0.12)
Constant	15.62**	-0.78	1.22***	1.54	2.06**
	(0.030)	(0.93)	(0.001)	(0.20)	(0.023)
Observations	40	40	174	63	145
R-squared	0.18	0.36	0.35	0.62	0.33

Notes: p-values in parentheses. *, **, *** refer to significance levels of 1, 5, and 10 percent. Dependent variable in columns 1 and 2: difference in numeracy, skilled versus unskilled occupations. Dependent variable in columns 3 to 5: difference of school years, tallest 50% versus shortest 50% (calculated as the difference between the two

groups). As usual with fixed effects regressions, we reported the ‘R-sq within’. The models were also estimated including a prison dummy but there was almost no difference in the other coefficients. We also checked regressions with a capital city dummy, with almost no change in the other coefficients. In columns 3 to 5, we included time fixed effects. Abbreviations: FE: Fixed Effects. LDCs: less developed countries. Glob.: Globalization. GDP per capita is expressed in units of 1000 $ (Geary Khamis $), source: Maddison (2001). Where GDP was lacking, it was linearly interpolated. Ecuador was assumed to have had the average GDP/c of Peru and Brazil. In columns 1 and 2, civil war data come from Clodfelter (2002), the share of mature was calculated from the age distributions in the censuses, see Table A, and the same applies to the education average (using age heaping). For the explanatory variables in columns 3 to 5, see the Appendix (see footnote 1).

Figure A

SCHOOL YEAR DIFFERENCE (VERTICAL AXIS) IN SEVERAL WORLD REGIONS BY BIRTH YEARS



Note: Difference of years of schooling, value of tallest 50 percent minus the shortest 50 percent (females). Years refer to the beginning year of a five-year-birth cohort. Central Asia includes only the formerly Soviet Republics.

Not to be included in the published article:

Appendix

Descriptive statistics, earlier sample:

Variable	Obs	Mean	Std. Dev.	Min	Max
edudiffabs	40	10.14753	6.615123	-5.284439	22.51859
open	40	.125	.3349321	0	1
ab	40	66.25964	12.488	32.40078	85.53741
mature	40	.432292	.3583845	0	1
cw	40	.125	.3349321	0	1
gdpc	40	.83715	.4062134	.359	1.88
gdpsq	40	.8617042	.9123649	.128881	3.5344

Descriptive statistics, later sample:

Variable	Obs	Mean	Std. Dev.	Min	Max
edudiff	174	.9965857	.6043107	-.1482408	2.577521
open	174	.0988506	.2858894	0	1
gdpc1000	174	1.59892	1.282989	.352	8.932
gdpsq1000	174	4.193145	8.693076	.123904	79.78062
eduavg	174	6.608366	3.316112	1.225174	14.57045
cw	174	.1034483	.3054224	0	1
polity2	174	-3.822605	5.497617	-9.8	9

Definitions and Sources of explanatory variables for the later sample:

opensw (Sachs/Warner): openness-dummy by Sachs and Warner (1995), 0=closed; 1=open; The SW openness indicator is a zero-one dummy which takes the value 0 if the economy is closed according to one of the following criteria:

1. average tariff rates = higher than 40%
2. barriers cover on average more than 40% of imports
3. socialist economic system
4. state monopoly of major exports
5. black market premium higher than 20% during the 1970s or 1980s

Source: http://www.nuff.ox.ac.uk/Economics/Growth/datasets/sachs/sachs.htm

openness 'Penn WT' (M+X/Y): The trade share of GDP, unadjusted or adjusted for population size, with the regression given in Gylfason 1999.

Source: Heston and Summers: Penn World Tables 5.6, http://datacentre2.chass.utoronto.ca/pwt/

productivity lag: productivity ratio of industry and services to agriculture = Real GDP per worker in industry and services / Real GDP per worker in agriculture; calculated on basis of 'Agriculture value added per Worker, constant 1995 USD' (Source: 1999 World Development Indicators CD-ROM), 'Labor Force in agriculture, % of total' (Source: 1999 World Development Indicators CD-ROM), 'Real GDP per Worker, 1985 international prices'

Source: Heston and Summers: Penn World Tables 5.6, http://datacentre2.chass.utoronto.ca/pwt/

polity 2: democracy score, -10 (=autocratic) to 10 (=very democratic)

Source: Marshall and Jaggers: Polity IV-data set; http://www.bsos.umd.edu/cidcm/inscr/polity

Civil war: this variable is coded as a dichotomous variable adopting the value 1 if civil war broke out in five-year period. It is defined as sustained combat between the armed forces of a government and forces of another entity for central control or for local issues. 1,000 battle-related deaths per year. Military and civilian deaths are counted. Sources: Correlates of War

Project and Uppsala Conflict Data Project. We recorded all civil wars matching these criteria mentioned in Clodfelter (2002).

rgdpw: Real GDP per Worker, 1990 international prices

Sources: Maddison (2001), see also Alan Heston and Robert Summers: Penn World Tables 5.6, http://datacentre2.chass.utoronto.ca/pwt/

rgdpsq: Real GDP per Worker, squared

Source: see above

mature: proportion of the adult population 15-69 who are 40-59 old, a measure of cohort size effects.

Source: US Census Bureau: International Data Base (IDB); http://www.census.gov/ipc/www/idbnew.html; The Latin American data were adapted from Baten and Fraunholz (2004).

Data Sources

Sources for Argentina

We were able to use a large number of primary sources on the development of numeracy in Argentina, the earliest source being the military census of 1744 for Buenos Aires reproduced in the *Documentos para la Historia Argentina* (Caillet-Bois 1919). The early data for the capital were completed by the census of Buenos Aires for 1771. Later information on the capital is provided by the census of Santa Fé 1887. Moreover, we can incorporate into our analysis the samples of the first two national population censuses of the years 1869 and 1895 that contain extensive information on a representative sample of the Argentinean population and were collected by Somoza and Lattes (1967).

Sources for Brazil

The Brazilian sample consists of early surviving censuses for a number of places in the regions of São Paulo, Floresta (a city in the province of Pernambuco), and São Cristovão (in the province of Rio de Janeiro). We might expect the latter data to be slightly biased as São Cristovão was populated by rather rich people.

Sources for Ecuador

In the case of Ecuador, we could include the western provinces of the country based on the census of 1870. Unfortunately, evidence for the Amazonas region did not survive. We thank Dacil-Tania Juif and Heike Schmutz for their data collection work.

Sources for Mexico

The *Archivo General de Indias* in Seville offers a considerable amount of primary sources for Mexico. For 1740-1743, population enumerations for Hidalgo, Guadalajara, and Oaxaca are included in our sample. For 1777, age data for Mexico City, Durango, Chihuahua, Baja California, Oaxaca, Puebla, and Veracruz are available. We also use a sample on the capital from the *Censo de Revillagigedo* carried out between 1790 and 1794 in Mexico. This was the first enumeration to use a standard format for listing the population by name, age, sex, and family status. Not all census forms have survived but those remaining still provide information on 15 quarters of Mexico City.

For the later period, data from various places in Mexico are available for the year 1930, which were taken to create a nationally representative sample of the country.

Sources for Uruguay

The National Archive in Montevideo offers interesting sources on the development of numeracy in Uruguay: the prison records (1846), the census of Soriano (1834), and the census of Maldonado (1836) that we used in this paper.

Sources for Colombia

The Colombian census data originates from the National Archive in Bogotá. For the 19th century, the census of 1870 offers information on Cauca, Magdalena, Chocó, Quindio, and other departments. The great variety of Colombian provinces in our data enables us to analyze the development of basic numerical abilities in Colombia comprehensively.

Sources for Venezuela

The evidence on Venezuela comes from the Archivo General de Indias. We thank Christina Jedermann and Gerrit Ulrichs for their help. The data contains 1,476 observations about age, occupation and sex of the individuals.

Methodology and basic concepts of age heaping

We study numerical abilities in this article, which are an important component of overall human capital. In order to provide estimates of very basic components of numeracy, we apply the age heaping methodology.[32] The idea is that in less developed countries in the past, only a certain share of the population was able to report their own age exactly when census-takers, army recruitment officers or prison officials asked for it. The remaining population reported a rounded age, for example, 40, when they were in fact 39 or 41. In today's world of obligatory schooling, passports, universities, birth documents, and bureaucracy, it is hard to imagine that people did not know their exact age. But in early and less organized societies this was clearly different. The typical result is an age distribution with spikes at ages ending in a five or a zero and an underrepresentation of other ages, which does not reflect the true age distribution. There was also some heaping on multiples of two, which was quite widespread among children and teenagers and to a lesser extent among young adults in their twenties. This shows that most individuals actually knew their age as teenagers, but only in well-educated societies were they able to remember or calculate their exact age again later in life.[33]

To give an example of rounding on multiples of five, the census of Mexico City 1790 reports 410 people aged 40, but only 42 aged 41. This was clearly caused by age heaping. Apolant (1975, p. 333) gives individual examples of age misreporting: Joseph Milan, who appeared in February 1747 as a witness in a Uruguayan court should have been 48 years old, according to one judicial record. However, in the same year, but in another judicial record, he declared his age to be '45 years'. Demographers see this age misreporting as a problem when calculating life expectancies and other population statistics. However, it is precisely this misreporting that enables us to approximate numerical abilities of historical populations. The ratio between the preferred ages and the others can be calculated by using several indices, one of them being the Whipple index.[34] To calculate the Whipple index of age heaping, the number of people reporting a rounded age ending with 0 or 5 is divided by the total number of people and this is subsequently multiplied by 500. Thus, the index measures the proportion of

[32] For more detailed surveys on the age heaping methodology see A'Hearn, Baten and Crayen (2009).
[33] At higher ages this heaping pattern is mostly negligible, but interestingly somewhat stronger among populations who are numerate enough not to round on multiples of five.
[34] A'Hearn, Baten and Crayen (2009) found that this index is the only one that fulfils the desired properties of scale independence (a linear response to the degree of heaping), and that it ranks samples with different degrees of heaping reliably.

people who state an age ending in a five or zero, assuming that each terminal digit should appear with the same frequency in the 'true' age distribution.[35]

$$(1)\ Wh = \left(\frac{\sum (Age25 + Age30 + ...Age60)}{1/5 \times \sum Age23 + Age24 + Age25 + .. + Age62)} \right) \times 100$$

For an easier interpretation, A'Hearn, Baten, and Crayen (2009) suggested another index, which we call the ABCC index.[36] It is a simple linear transformation of the Whipple index and yields an estimate of the share of individuals who correctly report their age:

$$(2)\ ABCC = \left(1 - \frac{(Wh - 100)}{400}\right) \times 100 \text{ if } Wh \geq 100 \text{; else } ABCC = 100 .$$

The share of people able to report an exact age turns out to be highly correlated with other measures of human capital, like literacy and schooling, both across countries, individuals and over time (Bachi 1951, Myers 1954, Mokyr 1983, A'Hearn, Baten, and Crayen 2009). A'Hearn, Baten, and Crayen (2009) found that the relationship between illiteracy and age heaping for less developed countries (LDCs) after 1950 is very close. They calculated age heaping and illiteracy for not less than 270,000 individuals who were organized by 416 regions, ranging from Latin America to Oceania.[37] The correlation coefficient with illiteracy was as high as 0.7. The correlation with the PISA results for numerical skills was even as high as 0.85, hence the Whipple index is more strongly correlated with numerical skills. They also used a large U.S. census sample to perform a very detailed analysis of this relationship. They subdivided by race, gender, high and low educational status and other criteria. In each case, they obtained a statistically significant relationship. The fact that the coefficients are relatively stable between samples, i.e., a unit change in age heaping is associated with similar changes in literacy across the various tests, is also remarkable. The results are not only valid for the U.S and in any country with substantial age heaping that has been studied so far, the correlation was both statistically and economically significant.

In order to assess the robustness of those U.S. census results and the similar conclusions drawn from late-20th century LDCs, A'Hearn, Baten, and Crayen (2009) also

[35] A value of 500 means an age distribution with ages ending only on multiples of five, whereas 100 indicates no heaping patterns on multiples of five, that is exactly 20 percent of the population reported an age ending in a multiple of five.

[36] The name results from the initials of the authors' last names plus Greg Clark's, who suggested this in a comment on their paper. Whipple indexes below 100 are normally caused by random variation of birth rates in the 20th century rich countries. They are not carrying important information, hence normally set to 100 in the ABCC index.

[37] See A'Hearn, Baten and Crayen (2009), Appendix available from the authors.

assessed age heaping and literacy in 16 European countries between the Middle Ages and the early-19th century. Again, they found a positive correlation between age heaping and literacy, although the relationship was somewhat weaker than for the 19th or 20th century data. It is likely that the unavoidable measurement error when using early-modern data caused the lower statistical significance.

Age heaping has also been compared to other human capital indicators, for example primary schooling rates. The widest geographical sample studied so far was created by Crayen and Baten (2010a), who were able to include 70 countries for which both age heaping and schooling data (as well as other explanatory variables) were available. They found that primary schooling and age heaping were closely correlated in a series of cross-sections between the 1880s and 1940s, with R-squares between 0.55 and 0.76 (including other control variables; see below). Again, the coefficients were relatively stable over time. This large sample also allowed the examination of various other potential determinants of age heaping. To assess whether the degree of bureaucracy, birth registration, and government interaction with citizens is likely to influence the knowledge of one's exact age, independently of personal education, the authors used the number of censuses performed for each individual country for the period under study as an explanatory variable for their age heaping measure. Except for countries with a very long history of census-taking, all variations of this variable turned out to be insignificant, which would suggest that an independent bureaucracy effect was rather weak. In other words, it is sometimes the case that societies with a high number of censuses had high age awareness. But, at the same time, these societies were also early in introducing schooling and this variable clearly had more explanatory power in a joint regression than the independent bureaucracy effect. Crayen and Baten also tested whether the general standard of living had an influence on age heaping tendencies (using height as well as GDP per capita to serve as a proxy for welfare) and found a varying influence: in some decades there was a statistically significant correlation, but in others this was not the case. Cultural determinants of age heaping were also observable, but their strongest influence was visible in East Asia, not in the Latin American countries under study in this article.

In this article, we employ the ABCC measure of age heaping, computing indexes for different countries and birth decades. In order to do so, we use the age groups 23-32, 33-42,

etc.[38] The age range from 63 to 72 was omitted as this age group offers too few observations, especially for the 17th and 18th centuries when mortality was relatively high.[39]

An advantage of the age heaping methodology is that age statements are more widely available than other human capital proxies like signature ability or school attendance. As Reis (2008) argues, the age heaping measure is a very basic measure of human capital. It is, therefore, especially valid to study human capital development in Latin America in the 17th and 18th centuries when more advanced human capital indicators were quite scarce and reflected only the skills of the elite.

[38] An advantage of this method is to spread the preferred ages, such as 25 or 30, more evenly within the age groups and it also adjusts for the fact that more people will be alive at age 50 than at age 54 or at age 55 than at age 59 (Crayen and Baten 2010a).

[39] Given that young adults aged 23 to 32 round partly on multiples of two rather than five, we use the adjustment method suggested by Crayen and Baten (2010a) to increase the Whipple value (minus 100) by 24 percent before calculating the ABCC measure.