Please Note:

This is the last working paper version before this study was submitted and accepted. Please cite as Baten, Joerg and Tollnek, Franziska: “Age-Heaping-Based Human Capital Estimates,”, in Diebolt, Claude, Haupert, Michael (Eds.): *Handbook of Cliometrics*, Berlin/Heidelberg: Springer, 2016, 131-154.

**Age-Heaping Based Human Capital Estimates**

Franziska Tollnek* and Joerg Baten**

January 2014

**Abstract:**
In this article, we provide comprehensive insights into the implementation and the use of the age-heaping method. Age-heaping can be applied to approximate basic numerical skills and hence basic education. We discuss the advantages and potential issues of different indicators and we show the relationship of those indicators with literacy and schooling. The application of age-heaping based indicators enables us to explore various topics on basic education such as the gender gap and the divergence of countries in the very long run. This well-established technique has been used by a great variety of authors who also show that numeracy has a large impact on growth.

*University of Tuebingen, **University of Tuebingen, CEPR and CESifo

## 1. Introduction

Education is one ofthe driving factors for the development and long-term economic growth of countries. Many projects in development aid are set up to increase school enrollment rates or years of schooling to improve education and thus the prospects of future generations. Nowadays, there are plenty of measures and indexes at hand to quantify different levels of education among children, adolescents and adults. Through various tests and methods, the levels of education or human capitalare comparable on an international basis. In the famous Programme for International Student Assessment(PISA), scholars compare cognitive skills of students fromvarious countries around the world. On the one hand, the impact of such a programme is enormous: the countries with lower scores invest financial means or restructure their schedules to push forward in the range. On the other hand, the results build one of the largest databases on students' education worldwide with which scholars are able to conduct analyses and draw conclusions for the future.

However, if we go some decades further back in time, we have to rely on other measures of human capital such as years of schooling, enrollment rates orliteracy because we simply lack other indicators. The differentiation between different years of schooling, for example, is slightly less exact than that of the cognitive skills tests of the PISA study. Moreover, there are other issues that might occur with these indicators: if a kid is enrolled in school, it does not mean necessarily that he or she acquires a certain level of reading or mathematical skills before dropping out potentially. Literacy rates are often self-reported or even have to be constructed from people's ability to append theirsignatures to documentsas marriage registers or wills, for example, which does not necessarily mean that the person is able to read and write. Reis (2005) reports such estimated literacy rates for a number of European countries around 1800. The English database implemented by Schofield (1973)reaches back to the middle of the 18$^{th}$ century. By analyzing wills, Gregory Clark (2007) also constructed a large database on English literacy that even dates back to 1585.

The construction of data bases onliteracy reaching back in time as early as the 16$^{th}$ century is, of course, an exceptional case and only possible for a country like Englandwhere the availability of sources is much better than in most of the other countries in the world. For most of the countries, data sources are scarce and do not provide literacy or enrollment rates until after the Industrial Revolution. For some less developed countries or world regions, we do not even find comprehensive enrollment rates for the past fifty years because schooling was not obligatory or there were no schools nearby the children could go to. But how can we measure human capital in times in which education was only available for the rich or in regions where data sources are very scarce until today?

In numerous surveys, churchregisters or census lists, people reported information from which scholars are able to derive a basic indicator of human capital: their age. The underlying concept for calculating such an indicator is the so called "age-heaping": in earlier times, when people did not have birth certificates or passports, they were often not aware of their true age or they did simply not know it because no one kept record of their exact date of birth. As a consequence, when people were asked for their age and they did not know it, they stated a rather "popular" number. For instance, they

claimed to be 35 when they were in reality 34 and 36. Hence, the age distribution shows "heaps" or "spikes" at these popular digitsthat are mostly multiples of five. Why does this clearly not reflect the true distribution of ages? We can explore that with a small example: If in the year 1935, for example, one hundred people stated to be 35 years old but only fifty persons reported being 34 or 36 years of age, this would mean that in 1900 twice as many children were born as in the years 1901 and 1899, respectively. This is a very unlikely scenario and most probably due to age non-awareness. This phenomenon causes problems for demographers because they have difficulties estimating the true distributionof males and females in certain age groups or the life expectancyof a population (see, for example, A'Hearn, Baten and Crayen, 2009).

But, while being a disadvantage to the accuracy of demographic research, this pattern is actually a benefit for the research on basic education: By implementing an indicator as the Whipple, for example, we can calculate the ratio of the individuals who were able to report their own ages exactly in contrast to those who stated rounded numbers. Consequently, an indicator based on age-heaping enables us to conduct studies on basic numeracy or human capitalfor a great variety of countries and in the very long run.

Many authors used the well-establishedage-heaping method on various topics related to basic education: A'Hearn, Baten and Crayen (2009) demonstrated the strong relationship between age-heaping based indicators and literacy or mathematical skills. Myers (1954), Zelnik (1961), Duncan-Jones (1990), Budd and Guinnane (1991), O'Grada (2006), Manzel, Baten and Stolz (2012)as well asCrayen and Baten (2010a, 2010b), among others, studieddifferences in numeracy of various countries, world regions and time periods. De Moor and Van Zanden (2010), Manzel and Baten (2009) as well as Friesen, Baten and Prayon (2013) assessed gender inequalities in numeracy in different world regions, whereas Juif and Baten (2013) compared the numeracy levels of Inca indios before and after the Spanish conquest. Stolz and Baten(2012) analyzed the effects of migration on human capital selectivity –hence, they measured the extent of "brain drain" or "brain gain"of countries through migration.[1]Charette and Meng (1998), for instance, assessed the impact of literacy and numeracy on labor market outcomes.

In the following section we will explain in greater detail the advantages and potential caveats of the age-heaping method.We also discuss the indicators that are commonly used to approximate basic numeracy and we describe in which way the calculation has to be done. Furthermore, we explore the relationship between age-heaping based indicators and other measures such as literacy and schooling. In section three, we describe different research topics that have been assessed by implementing the age-heaping method, while in section four we discuss studies that explore differences in numeracy levels acrossvarious world regions. In section five we provide concluding remarks about the impact of basic numeracy.

[1]Brain drain means that highly educated people emigrate from their country of origin to another. Brain gain means the opposite effect.

## 2. Age-heaping based indicators: advantages, potential biases and indexes

### Advantages, potential issues and heaping patterns

The requirement for employing numeracy as an indicator for human capital is that a certain share of people in earlier times – especially before the Industrial Revolution – were not aware of their actual age because they did not know their date of birth or they were not able to calculate the number of years from their date of birth to the actual year.[2] Consequently, when individuals were asked for their age and could not stateit exactly, they did not choose any number randomly, but they tipicallytended to report a number divisible by five such as 35, 40, 45, and so on (see, for example, Duncan-Jones 1990, A'Hearn et al. 2009).

While the afore-mentioned is the most commonly detected heaping pattern, there are alsosocieties in which heaping on multiples of two – hence, even numbers – is more popular. The age awareness of a population with this type of heaping can be consideredslightly better than that of people rounding on multiples of five. As Duncan-Jones (1990) outlines, populations can show rounding behavior on any digit, probably caused by cultural differences: Roman Egyptians preferred the number six, while Roman Africans rather rounded on digits ending in one.De Moor and Van Zanden (2010) report even a preference for multiples of twelve in different medieval and early modern sources, among them a census from Tuscany in 1427 and another from Reims in 1422. This phenomenon could be the result of religious orientations and the underlying usage of the number twelve as a holy number (De Moor and Van Zanden 2010, p. 192). Interestingly, this heaping pattern was more often adopted by women than by men, especially during early modern times in the South Netherlands.[3]

In some special cases such as the Chinese culture, one might think of a different heaping pattern, for example the avoidance of the number four which pronounced sounds similar to the word for 'death' or the preference of the number eight which can beassociated with fortune (Crayen and Baten 2010a). However, Baten, Ma, Morgan and Wang (2009)found that Chinese migrants to the US heaped considerably more on multiples of five than, for instance, on the birth year of the dragon, which is a very popular animal sign in China.

Another potential heaping scheme is the heaping on rounded figures of birth years (see, for example, Myers 1976). For example, when people were asked for their birth year, and they were not able to report it exactly, they tended to state an even year. In case the people were asked for both, their actual age and their birth year, the results of numeracy indexes could be biased because of the different heaping schemes that appear. Consequently, it is more reliable to work with sources in which the individuals had to make either of the statements (Crayen and Baten 2010a). When applying age-heaping methods to any kind of source, it is generally very important to be aware of different heaping

[2]However, we have to keep in mind that there are individuals still living today, predominantly inthe least developed countries, which are not aware of their true age when they are asked for it (see, for example, Juif and Baten 2013a).

[3] This could be due to a stricter adherence of religious practices or beliefs by women than by men, though this is not scientifically proven so far (see also De Moor and Van Zanden 2010).

schemes or preferences for certain numbers that might be of cultural or religious origin. If necessary, special indexes to capture those biases have to be implemented.

One great advantage of an age-heaping based indicator is that it enables us to assess basic numeracy for a large number of countries, covering a very long period of time, because this phenomenon appeared presumably in most of the societies of the past until a certain point in time (see, for example, Duncan-Jones 1990). The second advantage is that there exists a large number of sources that can be employed to calculate numeracy indexes.Generally, we can use any list for which people had to report their ageincluding census lists, ecclesiastical surveys, tax lists, marriage registers, death registers and shipping lists, just to name a few. The probably earliest census in the history of mankind that we are aware of is the population census decreed by Emperor Augustus – around the birth of Christ – for which Maria and Joseph were heading to their place of birth to be enumerated. Admittedly, we do neither know if the persons back then were asked for their age nor are there any documents preserved. However, Duncan-Jones (1990) reveals another way to measure age awareness in ancient times: the inscriptions in tombstones in the Roman world. Age-heaping on multiples of five was very common in the first centuries after Christ, with levels of age-misreporting of up to 60 percent.

Although all of these sources are of quite different origin or carried out for distinct reasons, the indicator estimates derived from their age information are comparable both across countries and time periods. This is due to the fact that the principle of heaping ages because of ignoranceor missing number discipline always works the same way, independent of space and time – provided that an analysis based on one indicator with certain prerequisites does not include mixed types of heaping patterns. If we observe two types of age-misreporting such as rounding on multiples of five and on birth years in one census, we should not proceed with only one indicator measuring a certain type of heaping because this could bias the results (Crayen and Baten 2008a).

The most important factor when calculating age-heaping levels derived from suchlists is that the ages of the individuals are self-reported and not counterchecked.[4]In some cases, related especially to church survey data such as marriage registers, it is possible that an ambitious priest counterchecked the ages of the bride and groom by their respective birth dates in a birth or baptism register. In case thatthe ages are counterchecked, we can mostly not detect any age-heaping at all. Hence, if numeracy levels are extremely high, especially in the case of very early samples of rural parishes, we should either eliminate the regarding sample from the data set or check the possibility of high numeracy levels.We could, for example,compare the numeracy levels to the corresponding literacy rates of the parish or to the numeracy levels of regions or villages with a similar infrastructure, education system, and so on (A'Hearn et al 2009, p. 795). Generally we can say that the further back in time the period

[4]Self-reporting is, of course, not an option if we consider inscriptions in tombstones or death registers. The ages provided in these sources reflect the heaping pattern of the individual who reported the age in place of the respective person. But even in such cases there are gender- or social group-specific differences observable (Duncan-Jones 1990, pp. 83). It is most likely that the persons providing the ages for the tombstones were related to the person they reported it for or were at least of similar social or educational status.

of interest lies and the higher the age-heaping is, the more likely is it that ages are not counterchecked. In censuses executed by governmental authorities and in times in which obligatory identification did not exist, we can assume with certainty that ages are not counterchecked.

Another possible objection could be the question: whose age-heaping do we measureafter all? Do the statements truly reflect the pattern of the respondents or is the observed age-heaping actually caused by the census taker?Critics could argue that the census taker might have estimated the ages of the people by himself or corrected those that seemed implausible to him. It has to be examined carefully for each data source, if this could be an issue. However, there are various hints that this is not the case in the studies under discussion. According to Manzel and Baten (#) some of the executive authorities explicitly required the census takers to interrogate the people individually.[5] Moreover, if the age-heaping results were influenced by the individual numeracy level of the census taker, the results of different censuses should vary within one region or country for the same birth cohorts. The authors, however, find that the results of different censuses display very similar levels of age-heaping for the respective birth decades.

Another strong argument in favor of the self-reporting of surveyed individuals is the difference in numeracy levels that we find between occupational or social groups. Baten and Mumme (2010) as well as Tollnek and Baten (2013) reveal that better educated groups of professionals – such as merchants – show significantly higher levels of basic numeracy than unskilled or partly skilled individuals. Furthermore, A'Hearn et al (2009) show that the correlation between literacy and numeracy rates is very strong on a regional or country-wide basis. Clearly, we are only able to detect such considerable region or occupation specific differences if people stated their ages by themselves.

Related to information about households or married couples, there is a further possible question to discuss: Did women report their ages themselves or did their husband help them – or even answer for them? How reliable are comparisons between male and female numeracy originating from the same source? In various studies, scholars suggest that we can rely on the age statements made by or assigned to women: According to De Moor and Van Zanden (2010), the indexes of women and men in a Belgian census, for example, were actually not that different so that it seems plausible that the individuals responded themselves. Furthermore, they find that women sometimes displayed preferences for different numbers than men – such asmultiples of the number twelve – which could not occur if the husband answered for them.

Manzel, Baten and Stolz (2012) find also evidence in favor of the self-reporting of household members, which they prove with results from the 1744 census of Buenos Aires: in case the household head stated the ages in place of the other family members, there should be substantial differences in the numeracy levels because one might assume that the heads were better educated than the other members, given that he or she provided the family income and in most cases had an occupation. However, the difference is almost negligible.Moreover, the authors report sources in which the

[5] They found information on censuses from which it becomes clear that the authorities required the census takers of surveying each person individually.

interviewer made complementary remarks. Related to a certain person who reported to be 30 years old, he noted "[…] but looked considerably older" (Manzel et al. 2012). Such statements strengthen the assumption that census takers asked the people individually for their ages and did not accept someone else to answer in their place. With all the results of the afore-mentioned studies and the information provided on the procedure of various censuses, we can most certainly assume that the studies discussed in this paper deliver reliable information on the basic education of the respective population.

**Whipple, ABCC and other indexes**

There are various indexes we can adopt for measuringage-heaping. In some cases the employed scheme varies from one study to another, depending on the author. What many of the indexes have in common, though, is the assumption that ages, stated as integers, follow a discrete uniform distribution. For example, ten percent of the people in the ten-year age group from 30 to 39 are expected to report their age as 31, i.e. with '1' as the terminal digit since it is the only number ending with this digit in this ten-number interval.Applied to heaping on multiples of five, this implies that 1/5 (two out of ten) or 20 percent of the ages in this age group end in the digit '0' or '5'. O'Grada (2006), for example, implements a simple index by observing the frequency of the numbers divisible by ten in the age groups 30-34, 40-44 etc. Observing five ages in each group should, in the simplest case, deliver the same frequency for each digit. A value greater than 0.2 (which equals 1/5), indicates a rounding pattern of the respondents. Consequently,we expect each age to be reported by about the same amount of individuals. We have to be careful about the assumptions of age distributions in general, though. Especially in older age groups, it is most likely that a higher share of people is alive at age 60 in contrast to those aged 69 (see, for example, A'Hearn et al. 2009).

When it comes to measuring the actual degree of age-heaping, there are some desired properties that can improve the results of the indicator, as described by A'Hearn et al. (2009): First,the index should be mathematically scale independent, which means that it delivers comparable results for two samples with the same heaping patterns but different sample sizes. The second valuable feature is the linear response to the degree of heaping, which implies that the indicator increases linearly when heaping rises. Finally, the coefficient of variation should be as small as possible across different random samples.[6]

There are several established measures with at least some of the desired properties such as the indexes suggested by Mokyr, Bachi and Myers.[7]A'Hearn et al. (2006) state that the indicators proposed by Mokyr and Bachi are not calculated on the basis of specific expected frequencies. Hence, they do not rely on a particular assumption about which terminal digit appears with a certain frequency. However, there is a common procedure also discussed by Myers (1954)thatimplies the expected proportion of each terminal digit to be 10 percent. For this procedure it is necessary to sum

[6] Please see A'Hearn et al. (2006, pp. 11) for more detailed information on the properties.
[7]The Mokyr Index we refer to in this section, is also called the Lambda Index (A'Hearn et al. 2006).

up all of the ages ending in zero, then those ending in one and so on, starting at age 20, for example.In the next step, the share of the population stating the respective terminal digit (0 to 9) relative to the whole population is calculated.[8] Each percentage share greater than ten means consequently an overrepresentation of the ages with the respective digit.The "Blended" index proposed by Myers (1954) works in a similar wayas this procedure but with some adjustments: Instead of starting the aggregation at age 20, he uses everyof the terminaldigitsat each age between 23 and 32, for example, as the starting point. Hethen proceeds with the aggregation of the ages with each terminal digit (0 to 9) but instead of counting each unit digit once, it is counted several times, according to the "leading" digit.[9] The result of this procedure represents the relative share of people thatreported ages with the respective last digit. If there is no age-heaping in the data, the percentage share of each figure should not differ largely from 10 percent (Myers 1954, p. 827).[10]

While the Bachi and Myers Indexes are scale independent at least in the mathematical sense, none of the indexes turns out to be scale independent in the statistical sense, meaning that the mathematical scale independency does not hold in random sample settings, as A'Hearn et al. (2006, p. 18) show.[11] However, each of the three indexes discussed in this sectioncan be adopted to reveal any kind of heaping, be it rounding on multiples of five or the preference for any other of the ten digits. This might be a small advantage in contrast to indicators that can only detect a preference for multiples of five.But yet, there is an indicator that exceeds all of the others in terms of its properties: the Whipple Index. The Whipple is statistically scale independent, its expected value rises linearly with the degree of heaping and its coefficient of variation is lower than for the other indicators discussed (A'Hearnet al. 2009, p. 788). The Whipple Indexis calculated aspresented in the following formula (1):

$$(1)\quad Wh=\frac{\sum\left(n_{25}+n_{30}\ldots+n_{65}+n_{70}\right)}{\frac{1}{5}\sum_{i=23}^{72}n_i}\times 100$$

$$Wh=\left(\sum\left[\left(n_{25}+n_{30}\ldots+n_{65}+n_{70}\right)\right]\right)/\left(1/5\sum_{(i=23)}^{72}n_i\right)\times 100$$

$$Wh=\frac{\sum\left(n_{25}+n_{30}\ldots+n_{65}+n_{70}\right)}{\frac{1}{5}\sum_{i=23}^{72}n_i}\times 100$$

[8]The digit '0' includes all ages ending in zero, hence, 30, 40, 50 etc. The digit '1' includes all ages ending in one, hence, 31, 41, 51, and so on.

[9]Myers criticizes that starting the aggregation at a certain age, for example 20, increases the share of people with a digit ending in zero because: "... the 'leading' digits naturally occur more frequently among the persons counted than the 'following' ones" (Myers 1954, p. 826).

[10]For a more detailed description of the "blended" method, see Myers (1954).

[11]Statistical scale dependency means that the assumed mathematical scale independency can change when applying an indicator to random samples of different size. For more information on this topic see (A'Hearn et al. 1990, pp. 11)

In the enumerator, the number of people reporting ages ending in a zero or a five is aggregated. This is divided by all of the reported ages in the age range 23 to 72. Subsequently, we multiply the sum of the reported ages by 1/5 in the denominator. This is based on the assumption that 20 percent of all the people report an age ending with zero or five correctly. The whole term is then multiplied by 100 for convenience. Hence, the Whipple can take on values ranging usually between 100 and 500. If exactly 1/5 of all the individuals state an age ending in a multiple of five, the Whipple takes on the value 100. In the case that all of the persons report a multiple of five, the Whipple increases to 500. We have to be careful by interpreting this figure, though: A value of 500 still would mean that 1/5 of the individuals who state a rounded age, were doing so correctly. Admittedly, with an age-heaping effect of this size, we might as well assume that these individuals did not report their correct age because of age-awareness. In theory, the Whipple can also take on the value zero, if no person reports a multiple of five – this would be the case of perfect "anti-heaping" (see, for example, A'Hearn et al. 2009). The Whipple increases linearly, which means that it rises by 50 percent, whenever the proportion of people reporting a multiple of five increasesby 50 percent (Crayen and Baten 2008a, p. 7)

Because of its design the Whipple Indexobviously does not account for the fact that less people are alive at higher ages. Thus, there naturally is a higher number of people reporting the age of 60 than the age of 69 even if there was no age-heaping in the population otherwise. We are able to reduce this potential bias by calculating the Whipple for age groups of ten-year-steps. Additionally, we arrange the age groups such that the multiples of five, and especially the numbers ending with zero, are more evenly distributed within the age groups: the first age group starts at age 23 and ends with the age 32. The other age groups are arranged accordingly: 33 to 42, 53 to 62, and so on. It is more reliable to exclude individuals older than 72 years because they tend to exaggerate their age. The survivor bias effect could also play a role because people with a higher basic education might have a higher life expectancy due to a higher expected income, for example(Crayen and Baten 2010a#).

It is also common to exclude the individuals younger than 23 years of age from the analysis because of two reasons: First, young people often married around the age of 20 or entered military service at that time. As they often had to report their ages at such occasions, their age awareness is expected to be better than at older ages. Second, younger people tended to round their ages to a much greater degree on multiples of two than of five.Additionally, for children or adolescents still living with their parents, we do not know if they reported their ages themselves or if their parents answered for them (Manzel and Baten 2009).To account for a higher degree of heaping on multiples of two among this group which is not captured directly by the Whipple,Crayen and Baten (2008a#, Appendix)propose an upward adjustmentof the Whipple Index. With this adjustment, the value of the youngest age group increases, hence the estimated numeracy decreases. [12]

[12]If the Whipple indicator is larger than 100, they suggest adding 0.2 units to the value of the age group 33 to 42 for every Whipple unit. The resulting value is the new estimate for those aged 23 to 32. For example, if the value of the age group 33 to 42 is 150, then the digit above 100 has to be multiplied by 0.2 (50*0.2=10). The result is added to the original value (150+10), which delivers the new estimate for the age group 23 to 32 (160) (Crayen

The Whipple Index combines a number of desired properties and is – after making some adjustments – a reliable measure for the degree of age-heaping. However, the adopted scale and the interpretation of its outcomes are not particularly intuitive. A'Hearn, Baten and Crayen (2009) solved this issue by introducing another indicator which they called the "ABCC".[13] The calculation works as shown in the following formula (2):

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

$$Wh = \left(\sum \left[\!\left[ (n_{25} + n_{30} \ldots + n_{65} + n_{70}) \right]\!\right]\right) / \left(1/5 \sum_{(i=23)}^{72} n_i\right) \times 100$$

$$Wh = \frac{\sum (n_{25} + n_{30} \ldots + n_{65} + n_{70})}{\frac{1}{5} \sum_{i=23}^{72} n_i} \times 100$$

The ABCC is a simple linear transformation of the Whipple and ranges between 0 and 100. For the case of "perfect" heaping and thus a Whipple of 500, the ABCC takes on the value 0. If every person states their age correctly, the ABCC value increases to 100. Hence, the ABCC can intuitively be interpreted as the share of people reporting their age correctly. This measure has been successfully used in a variety of studies so far (Manzel and Baten 2009, Mumme and Baten 2010, Manzel et al. 2012, Stolz and Baten 2012, Juif and Baten 2013 as well as Baten and Juif 2014, forthcoming).

Because age-heaping indicators such as the Whipple and the ABCC Indexare employed to approximate basic education if other indicators are not available, it is very important that these indexes correlate with other measures. It turns out that there is a strong correlation between the share of people reporting their correct age and indicators such as literacy or schooling. Myers (1954, p. 830) finds a correlation of high literacy rates and low levels of age misreporting for Australia, Canada and Great Britain. Duncan-Jones (1990) also reports a significant correlation between age-heaping and illiteracy in a number of developing countries in the 20th century, among them Egypt (1947), Morocco (1960) and Mexico (1970). Furthermore, A'Hearn et al. (2009) perform analyses on the relationship between the ABCCIndex and literacy in various countries. They detect a very strong, significant and robust correlation between the two indicators for almost all of the 72# countries in their data set. In the very detailed analysis for the USthe correlation is particularly strong, still when controlling for birth place, ethnic group and gender balance and it is evident for bothpooled and regional fixed effects regressions.

Moreover, Crayen and Baten (2010a) tested the impact of several factors such as primary schooling, height and state antiquity, among other factors, on age-heaping.[14]For a global dataset, they

and Baten 2010a#, Appendix 1, p. 35).

[13] The name of the index is constructed by the initials of the last names of the three authors plus Gregory Clark's.

[14]Height is employed as a proxy indicator for infant malnutrition, because the smaller a person is, the more likely

found that school enrollment is one of the driving factors for the development of numerical abilitiesamong societies. In all of the modifications and independent of the factors controlled for, it is always highly and significantly correlated with age-heaping. Consequently, we assume that age-heaping based indicators are valid estimators for basic education.

## 3. Applied age-heaping indicatorsin various research topics

**Reconstructing very early numeracy differences: the example of Inca indios**

Acemoglu, Johnson and Robinson (2001, 2002) studied the differences between former European colonies. They compare the former colonies that are rich today to those that are poor. Acemoglu et al. argue that the Europeans created exploitative institutions in colonies that had an adverse disease environment for Europeans. In contrast, they implemented a growth-promoting type of institutions in colonies where Europeans settled. Examples for the latter would be the USA, Australia, Argentina, partly South Africa etc., whereas classical examples for the former would be West Africa. The more or less growth-promoting nature of colonial institutions translated into better or worse institutions during the late 20th century, which had animpact on today's difference in real income per capita because institutions tend to remain similar for a longer period. Applying the age-heaping technique to this topic is particularly useful because alternative views suggest a strong role of human capital channels (Glaeser et al. 2004). A related question is, for example, whether there were 'pre-colonial legacies': how much did the ancient economies and societies invest before the colonialists arrived?

A paper by Juif and Baten (2013) employs an early Spanish census that was directly taken after the invasion of the Inca Empire. It makes use of the fact that basic numeracy is usually attained during the first decade of life. Clearly, the question needs to be considered whether such a birth cohort-specific analysis could be distorted by later learning processes. However, the numeracy values of the cohort born before the invasion are close to zero, hence they cannot be upwardly biased. The cohorts born after the invasion in contrast were slowly rising. Hence, the most important result of this study was that in fact some pre-colonial legacy – or burden – existed in Andean America. This legacy has not been reduced during the colonial times, as colonial institutions such as the Peruvian 'Mita' reinforced educational inequality (Dell 2010). During the early period, it is interesting that some Indio groups that were alliedwith the Spanish during the invasion (and received tax exemptions and a slightly less terrible standard of living after the invasion in return) also displayed a better numeracy. A likely interpretation would be that their slightly higher net income allowed more investments into the basic numeracy of their children. This observation stands also in contrast to the suspicion that cultural attitudes could have implied a different number rounding behavior. Another problem considered by the authors could be, whether colonial officials did not ask the Indios for their age but rather estimated it without asking (if they estimated after asking, this would not be a problem for the age-heaping procedure because in this case the respondent did most likely not know his age either). Juif and Baten

is it that he or she did not have access to protein-rich nutrition which also hinders the development of numerical skills. State antiquity approximates the quality of institutions.

rejected these doubts in their studywith arguments based on the effect that the social difference of numeracy within the Indio groups was substantial. In addition, the colonial officials sometimes explicitly noted thoughts such as "She says she is 30, but she looks more like 40", which indicate that the Indios were in fact asked for their age. As a result, this earliest numeracy study for a Non-european country found that a pre-colonial legacy was in fact very likely.

**Religion and numeracy**

A number of scholars have recently studied potential religious determinants of human capital formation (See Becker and Woessmann 2009 for a widely cited study and a good overview). The relative exogenous character of religious rules has been stressed by this literature, because beliefs about the necessity to read religious texts are considered to be less influenced by economic factors and profit-maximizing educational investment decisions. Botticini and Eckstein have explained how religious rules to provide education to one's (male) offspring appeared in the Jewish faith: in the first century BCE, a conflict between two influential religious fractions of Judaism took place. One of these fractions, the Pharisees, stressed the religious duty to educate and they gained stronger influence on Judaism than other group. Botticini and Eckstein emphasize that the education rule was not economically motivated, because the large majority of the Jewish were farmers and rural day-laborers, for whom a substantial educational investment would not yield sufficient returns during this period. Only with the substantial urban growth in Mesopotamia during the 8$^{th}$ and 9$^{th}$ centuries CE, the Jewish population living there could use their religiously determined education in order to achieve profitable positions as merchants and, later on, as bankers. Medieval Western Europe actually first tried to attract this religious and occupational group because the kings of England and France assumed correctly that government revenues might increase. The famous restriction of Jewish population groups to being exclusively merchants, bankers, and other trades that were forbidden to the Christian population were only created later, during the High Middle Ages. Botticini and Eckstein therefore reject the hypothesis that this restriction caused high Jewish educational levels.

The debate on religious differences of education and numeracy in particular has important implications for history and for our understanding of human capital formation. For that reason, Juif and Baten (2014) studied the differences between the average population and persons who were accused by the inquisition to execute Jewish practices in Iberia and Latin America. The period under study runs from the 15$^{th}$ to the 18$^{th}$ centuries. The sources that are available for this early period were primarily created by the inquisition. A question about the age of the accused was included for identification purposes. Besides the evidence from the inquisition lists, we also included a number of census-based numeracy evidenceto compare the average population in the same regional units.We studied potential selectivities and biases intensively and dismissed them ultimately. The most important result of this study of religion and numeracy is that persons who were accused of being Jewish had a substantially higher numeracy than the average population. If we accept the working hypothesis that most persons accused of Judaism came from families of a different educational

behavior (and a different educational self-selection), the religious factor appears to be of important influence. However, the authors find that also the catholic elites (such as priests) had a substantially higher numeracy compared to the average Iberian and Latin American population.

**Path Dependency of early numeracy and land inequality as a determinant of modern math and science skills?**

Within the framework of Unified Growth Theory, Galor, Moav and Vollrath (2009) have focused on land inequality as one of the crucial obstacles to human capital formation. They describe the political economy of regions and countries with higher and lower land inequality, assuming an influential role of two different elite groups: large land-owners and industrial capitalists. In regions with lower land inequality, industrialists obtained a larger relative power in the decision process about educational investments. In contrast, in regions with high land inequality large land-owners remained in power and were not particularly interested in spending their taxed income for primary schooling: first of all, their agricultural day-laborers did not need much education to fulfil their manual tasks (at least that is the traditional view). Secondly, additional primary schooling would have increased their burden of taxation. Thirdly, educated workers might have moved to cities or even have initiated land reforms. In a study of this land inequality effect on modern math and science skills, Juif and Baten (2014) also include early numeracy (around 1820) as the second main determinant. They find that early numeracy has a large explanatory share, even after controlling for land inequality and a number of other explanatory variables. It seems that this path dependency worked via economic specialization: if an economy specialized early on the production of human-capital-intensive products, the relatively high income allowed investing into education for the next generation. In addition, such human-capital-intensive production methods probably resulted in substantial switching costs – hence,the countries specialized in this type of production and developed a branding and reputation for their products. As a consequence, they most likelywere entering a high degree of path-dependency.

**Numeracy Differences between Occupational Groups in Preindustrial Times**

When it comes to the question of who stated the ages written down on a census list – the enumerator or the respondent – the analysis of numeracy between occupational groups is crucial. If the age-heaping levels between the occupational groups vary significantly, this might indicate that the respondents stated their age themselves. De Moor and Van Zanden (2010, p. 204) were able to verify differences in numeracy between three occupational groups, professionals, craftsmen and unskilled laborers, for the 17$^{th}$ century in Amsterdam. While highly skilled professionals had relatively low age-heaping levels (with a Whipple index below 100), the opposite was the case for non-skilled individuals and slaves who displayed a high degree of age-heaping. The craftsmen had slightly better values than the unskilled group.

Tollnek and Baten (2013) assess the numeracy of occupational groups for four countries in early modern Europe (Austria, Spain, Southern Italy and Germany) as well as for Uruguay. Additional

information is provided by literacy data from Switzerland. In total, the comprehensive dataset includes nearly 30.000 observations with information on age, sex and occupation of individuals. The authors distinguish between six occupational groups, adapting the Armstrong-scheme (Armstrong 1972): the professionals (doctors, lawyers etc.), the intermediate (administers and higher clerks), the skilled group (craftsmen and shop keepers), the partly skilled (herdsmen and carriage drivers), the unskilled group (day laborers) and the farmers (smallholders and farmers with medium-sized or larger farms).[15] The descriptive analysis already reveals large differences between the groups: in all of the European countries, the professionals have the highest numeracy values (ABCC index between 86 and 96), followed by the intermediate and skilled groups that reflect lower numerical abilities (Tollnek and Baten 2013, p. 33). The two least skilled groups, partly skilled and unskilled, have the lowest values of numeracy in all of the countries.[16] Interestingly, the farmers have considerably low age-heaping levels, with numeracy values similar to the skilled group. In Germany and Uruguay, the farmers' values even are close to the groups with the highest levels, which are the professionals in Germany and the skilled in Uruguay.

The authors assess these differences also in a logit regression with 'numeracy' as the dependent variable: it takes on the value one if the individual stated an exact age and zero otherwise.[17] They control for birth half century, country, age (because younger people might know their age more exactly) and most important the occupational groups. The regression results strongly confirm the decriptive results for all of the countries in their sample: the three upper groups and the farmers have a significantly higher probability of being numerate than the two lowest groups, partly skilled and unskilled, the pooled reference category (Tollnek and Baten 2013, p. 28). The values of the coefficients range between roughly 18 for the professional groups and 8 for the skilled. The farmers have the third highest chance for success (hence, 'numerate' takes on the value one) with a coefficient of nearly 9, which can be translated into a higher probability of being numerate of about 9 percent in contrast to the two lowest groups. These results are also confirmed by regression results using literacy evidence from Switzerland.

## 4. The Development of numerical skillsin different world regions and time periods

### A human capital revolution in Europe

A'Hearn et al. (2009) discuss the development of numeracy all over Europe from the late middle ages to the early modern period. The European countries experienced a striking increase in numeracy during this time period, which can be identified as a "human capital revolution". While the numeracy values rose in all of the European countries, there was variation between the different parts of Europe. The Western European countries showedan exceptional development. As early as around 1450, the

[15] The occupations in brackets are only examples. In total, there are hundreds of occupations in the dataset that were arranged according to the Armstrong-scheme.

[16] Germany is an exceptional case: the values of the intermediate, skilled, partly skilled and unskilled groups differ only slightly.

[17] The coefficients are subsequently multiplied by 125 to correct for the 20 percent of people who state a multiple of fivecorrectly. For further information, please see Appendix B in Tollnek and Baten (2013).

Netherlands represented numeracy values (ABCC) of roughly 70 percent (A'Hearn et al. 2009, pp. 801/804). Britain and France surpassed this value at around 1600 and 1650, respectively. Britain and Denmark, on the other hand, already experienced numeracy rates of 90 percent or more in the period of 1700. While Denmark's rates grew continuously until the end of the period, Britain's values remained at the same level. The picture looks similar if we look at Central Europe. Austria and Protestant Germany already had high numeracy levels of between 78 and 87 percent around the period of 1600. Catholic Germany had somewhat lower values (68 percent in circa 1700), but it converged strongly thereafter. The Eastern European countries, in contrast to the rest of Europe,lagged slightly behind:around 1600, Bohemia represented numeracy values of only 44 percent. One period later, around 1650, Russia and Hungary showed levels of 43 and 32 percent, respectively. However, towards the end of the early modern era at approximately 1800, the overwhelming majority of the European countries managed to increase their human capital values significantly: even those regions that lagged behind such as Bohemia and Russia reached numeracy levels well above 80 or close to 90.

**Numeracy levels in Latin America**

Manzel, Baten and Stolz (#) analyze long-term trends in numeracy for a number of Latin American countries from the 17th to the beginning of the 20th century. Some of the countries experienced strong increases of human capitaltroughout the whole time period such as Argentina and Uruguay that are comparable to those of some European countries. While Argentina started with an ABCC value of less than 20 percent in the birth decade 1680, it reached values of almost 70 percent around 1800 (Manzel et al. #, p. 44).[18] With an exceptional increase during the 19th century, Argentina reached almost full numeracy yet at the beginning of the 20th century. The development of Uruguay is similar, showing even higher numeracy levels than Argentina in parts of the 19th century. Despite such great examples of convergence, some of the Latin American countries underwent a process of divergence during the 19thcentury: in Colombia, Mexico and Ecuador the ABCC levels stagnated. While Mexico started off well with continuously growing numeracy levels from 1680 to 1790, there was almost no improvement throughout the 19th century. Ecuador's levelseven worsened slightly. Brazil was a particular case because it experienced a short period of stagnation at the first half of the 19th century but increased human capital afterwards again. Towards the beginning of the 20th century, numeracy levels rose considerably in all of the observed countries.

**Industrialized countries vs. the rest of the world?**

Crayen and Baten (2010a#) assess long-term trends of numeracyimplementing the Whipple index in 165 countries all over the world. The development of some industrialized countries not discussed so faris of interest:the United States (US), Australia and New Zealand started with Whipple values around 150 at the beginning of the 19th century, which are among the highest numbers compared to the

[18]The values for Argentina and Mexico are estimates based on regression results. They are controlled for capital effects and male share. For further information, please see Manzel et al. (#)

other industrialized countries in the same period (Crayen and Baten 2010a#, p. 8#). Towards the middle of the 19th century, the values of all three countries decreased significantly to around 120. They converged continuously in the following decades and reached values of circa 100 at the end of the 19th century. Spain also had values above 150 around 1830. The increase of Spain's numeracy developed more slowly, but it also reached levels close to 100 at the beginning of the 20th century. Exceptional cases are also Greece and Cyprus, which had values above 200 and 250, respectively, still at the end of the 19th century. However, their rates decreased dramatically throughout the 20th century. Ireland is one of the few industrialized countries, in which the Whipple index increases slightly around 1870, which is probably due to the Great Famine that took placea few decades earlier. An exceptional case in the positive sense is Japan, which had values of 100 already at the beginning of the period under study.

The comparison of world regional numeracy trends reveals some crucial differences: South Asian countries had the highest age-heaping levels with Whipple values of more than 450 around 1840 (Crayen and Baten 2010a#, p. 12#). The numbers declined steadily throughout the following decades, reaching a Whipple index of roughly 300 towards the 1940s. The Middle East and North Africa had the second lowest levels of numeracy with values higher than 400 in the 1820s. Egypt had probably the highest age-heaping level in this region with a Whipple of almost 500 (the case of "perfect" heaping) (Crayen and Baten 2010a#, p. 9#). But similar to South Asia, the Middle Eastern and North African countries managed to increase their numeracy levels continuously (Crayen and Baten 2010a#, p. 12#). The industrialized countries were on the upper range of the strata with the highest numeracy levels. East Asia, dominated by Chinese data (as Japan being part of the industrialized countries), still had age-heaping levels of above 150 at the beginning and in the middle of the 19th century. In only a few decades, though, age-heapingin China decreased strongly and vanished around 1880. South East Asia and Latin America ranged in the middle between the groups with quite high and relatively low levels of age-heaping.

**Numeracy trends of women and the gender gap in different world regions**

Equality between women and men related to education or wages is a controversely discussed topic until today. Even in countries with relatively high levels of income and education, such as the European countries or the US, there is an ongoing debate about wage differentials between men and women. Women with the same degree of education and experience often receive considerably lower wages than their male counterparts working in the same field or position (Quelle?#).

But what about educational differences between men and women before formal schooling became accessible for most people? When did the gender gap open and did it worsen or improve over time? Duncan-Jones's (1990, p. 86) analysis of inscriptions on tombstones reveals a numeracy difference between men and women in Roman times which is probably the earliest gender gap measurable. Although the age reported on the tombstone rather reflects the numerical abilities of a relative, supposedly, the ages of women show a higher heaping pattern than those of men. The indicator implemented by Duncan-Jones represents the percentage share of people who report a

rounded age, relative to those who state their age correctly.[19] While in some regions such as Moesia or Pannonia the women had considerably higher heaping levels than the men (28.1 and 17.1 percent), the differences were relatively small in most of the other regions: in Mauretania, for instance, the womens' index was only 4.8 percent higher than the mens' and in Rome 6.8 percent. However, there were also regions in which women had lower heaping values such as Italy outside Rome (-1.9 percent).

De Moor and Van Zanden (2010) assess human capital levels in the medieval and early modern Low Countries. The results of the Whipple Indexes of Bruges in Belgium (1474-1524) suggest that the differences between women and men were relatively small in total: the men have an index of 159 and the women 168 (De Moor and Van Zanden 2010, p. 192). In the city of Bruges, the women even surpassed the men slightly by 4 Whipple points.[20] The authors also found similar results for Holland during the 16th to 18th centuries. The gender gaps were small then and sometimes the Whipple Index was lower for women, indicating a higher numeracy than for men.

Myers (1954 p. 830) reports for the (US) that women in the 1950s showed significantly higher levels of age-heaping than men. For the other countries included in his study – Australia, Canada and Great Britain – he detects only very slight differences in age misreporting between women and men in the late 1940s or early 1950s. In Great Britain, women reported their ages even more precisely than men.

These examples suggest that in particular regions and time periods womens' access to basic education was not as limited as one might have expected. However, we have to keep in mind that the Low Countries, for example, are different to many other countries with respect to the position of women. Men and women seemed to have had a relatively equal standing in the household, already in early modern times (De Moor and Van Zanden 2010). But what about womens' basic education in the rest of the world?

Manzel and Baten (2009) as well as Friesen, Baten and Prayon (2013) assess the development of womens' basic education for a large number of countries in Latin America and Asia via age-heaping based indicators. They perform their analyses following a fundamental theory about labor force participation developed by Goldin (1995). Goldin argues that female labor force participation follows a U-shaped pattern over time: in societies with low income and rather low levels of education, at least among women, they engaged to a large extent in home production of agricultural goods and work on family farms. At this stage of the process, labor force participation shares are high for both men and women. With increasing levels of income and market integration, more women are tied to household activities and child care, while men work in factories, for example, where new production techniques overcome the traditional home production. Hence, womens' level of labor market

[19] He subtracts the 20 percent of people who report a multiple of five correctly from the total number of people who state a rounded age. Hence, the reported percentage share contains those who incorrectly state a rounded age.

[20] The women, however, represent higher values at the 'Dozen Index' that detects rounding behaviour on multiples of twelve. This is probably based on religious practices and rather among Catholics (De Moor and Van Zanden 2010).

participation decreases. One possible reason for that development could be that womens' work in factories is socially stigmatized. The third stage of the processis observable in countries that have reached a high level of income and education. Women are able to achieve higher degrees of education and enter white-collar occupations that are less stigmatized than manufactual work. In this last phase of the U-shape, women participate actively in the labor force again.

Manzel and Baten (2009) were able to prove this pattern based on numeracy estimates for 28 countries in Latin America and the Caribbean during 1880 to 1940.[21] Instead of testing the relative labor force participation of women, they implement "the relationship between average education and the ratio between female and male education" as an indicator to demonstrate the U-shape development (Manzel and Baten 2009, p. 3, WP Version #). As a general measure of educational equality between men and women, they subtract the Whipple Index of men from that of women and divide the result by the Whipple of men. This is subsequently multiplied by -100 for convenience. If the outcome is positive, the women have a numeracy advantage over the men (and the other way round, if the index is negative). The positive index is defined as "gender equality" in basic education. It turns out that the equality index is negative for most of the countries. However, for some countries with high levels of basic numeracy throughout the time period, the equality is relatively high as well, indicating the last stage of the U-shape hypothesis. This is the case for Argentina, Uruguay, Guyana and Suriname, meaning that gender equality increases if basic education is well-established in the society in general. The ABCC values for Argentina, to state an example, reach from about 95 to 100 percent and the equality index is slightly above 0 (Manzel and Baten 2009, Appendix p. 41, WP #).[22] In Guatemala and the Dominican Republic, for example, the authors find the opposite effect, namely low levels of basic numeracy and low equality indexes. Colombia, however, has ABCC levels between 79 and 89, while the equality index ranges approximately between -49 and -2, meaning that women have large educational disadvantages in Colombia at the beginning of the period, which decrease over time (Manzel and Baten 2009, Appendix p. 41, WP #). But there are also cases such as Haiti, where numeracy is low, whereas gender inequality is not observable, indicating the first stage of the U-shape hypothesis. In general, the Non-hispanic parts of the Caribbean represent considerably higher equality indexes as well as higher ABCC levels than the Latin American countries during the whole time period.[23] Towards the end of the period equality rises with increasing levels of basic numeracy in all of the countries: in Latin America, the ABCC values increase from roughly 78 percent in 1880 to about 93 percent in the 1940s and in the Non-hispanic Caribbean from about 90 percent to 99 percent (Manzel and Baten 2009, p. 31, WP #). The equality values increase from less than -12 to about -5 in

[21]The data are arranged in age groups and then transferred into birth decades, with the census year as the starting point, minus the value of the respective age group. If the census year was 1940, for instance, then the age group 23 to 32 was born in the decade 1910 (1940-23).

[22]All of the numbers in this passage are taken from figures 3 to 6 (pp. 31-34) and from the appendix table A.1 (p. 41) in the working paper version of Manzel and Baten (2009#).

[23] The low inequality of Non-hispanic countries might be due to the institutional framework created by slavery. As both men and women were torn away from their home countries and had to work equally, the 'traditional' gender roles did not evolve as it did in other countries. Besides, Caribbean women tended to work outside the household more often than Latin American women did (Manzel and Baten 2009).

Latin America and from roughly -3 to slightly above 0 in the Non-hispanic Caribbean (Manzel and Baten 2009, p. 34, WP #). The ABCC and equality values of the Hispanic Caribbean lie mostly below the values of Latin America.

To test the U-shaped hypothesis, Manzel and Baten perform a regression analysis with the equality index as the dependent variable, controlling for a number of other factors such as female voting rights and a democracy index. The most important factors for the U-shape are the ABCC values to approximate basic education: they are included as a linear parameter to control for initial levels of education and they are added as squared values to test for higher levels of education. As expected, the linear (and hence lower) ABCC values have a significant negative impact on equality, while higher levels of education (squared ABCCs) have a significant and positive impact on gender equality. The authors also plot the estimated values as to illustrate the U-shape: the downward slope tends to be rather smooth, whereas the upward sloping part is strongly observable in the data. Hence, they could demonstrate that Goldin's hypothesis also applies for basic education in Latin America and the Caribbean.

Friesen, Baten and Prayon (2013) test the U-shape hypothesis for 14 countries in Asia from 1900 to the 1960s.[24] They use the ABCC index to approximate basic numeracy. Furthermore, they employ the educational gender equality index based on the Whipple Index in the same way as Manzel and Baten (2009) did. Besides the age-heaping based indicators, Friesen et al. (2013, p. 9) discuss literacy and school enrollment rates in the Asian countries in the dataset that clearly indicate high levels of inequality between men and women. In Afghanistan, for example, the equality indicator for enrollment reaches a value below -95 in 1955 as only 4 percent of the girls are enrolled.

The analysis of the ABCC values provides further information on basic education between the sexes, especially when enrollment rates are not available for some of the regions. The authors find quite different results for the womens' ABCC indexes among the observed regions: the vast majority of the Southeast Asian women was already quite numerate around 1900, especially in Hong Kong and Thailand, while Indonesia lagged slightly behind. However, the picture looks quite different for women in South and West Asia: While Sri Lanka started off with ABCC values of around 59 percent in 1900 and reached almost full numeracy in the 1950s, all of the other countries in this region reflect values far below. Women from Pakistan and Bangladesh represent the lowest levels, reaching not even values of 50 percent towards the end of the period (Friesen et al. 2013, p. 14).

The equality index reflects mostly the different stages of the U-hypothesis: in the countries with very low human capital values for both women and men such as Pakistan, Bangladesh and India, equality values are only slightly below 0, indicating relative equality between women and men(Friesen et al. 2013, p. 23). This is also the case for the countries with high numeracy values, for example Hong Kong and Thailand, for which the equality values range slightly below or above the zero line (Friesen et al. 2013, p. 25). The equality indexes of the other countries lie mostly considerably below 0 (for

[24] Included countries are: Afghanistan, Bangladesh, India, Iran, Sri Lanka, Nepal, Pakistan, Hong Kong, Indonesia, Cambodia, Federation of Malaya, Sarawak, Philippines and Thailand.

example in Indonesia or Sri Lanka). Most of the countries with negative values experienced an increase towards the end of the period, which in some cases even turn into a positive index such as in the Philippines. The opposite effect takes place in Afghanistan, for instance: while the inequality is not as high around 1910 (about -12), it decreases continuously until reaching a value below -60 in the 1950s (Friesen et al. 2013, p. 23).

In the following, the authors test the U-hypothesis in different regression models, in which the equality index is the dependent variable. They control for factors such as female voting rights and religion. The most important determinant, the ABCC index, is included as a linear and a squared parameter (as in Manzel and Baten 2009). The results for the ABCCs are always highly significant and the correlation is, as expected, negative for the linear ABCCs and positive for the squared ones. Furthermore, Friesen et al. (2013, p. 37) plot the regression results to illustrate the fitted values. The scatter plot shows an exact U-shaped pattern. Hence, the assumption of low gender inequality at low levels of human capital, rising inequality at increasing levels of education and, in the last phase, high levels of education and equality is fulfilled in the analysis of the 14 Asian countries under study.

## 5. Conclusion: The Impact of Numerical Abilities on Growth

In this article, we showed that the age-heaping technique provides a unique opportunity to approximate basic education, especially in pre-industrial times. One might argue, though, that the mere knowledge of numeracy levels between different countries, for example, does not contribute to achieve a higher goal. However,although numeracy correlates strongly with literacy, number discipline might even have a larger impact on the development of market exchange (see, for example, De Moor and Van Zanden 2010). In many cases, we do not even know what literacy measures exactly: a broad range reaching from ‘is able to read and write’ to ‘is only able to sign with his name’ is possible. Numeracy or the ability to count, on the contrary, is the basis for participating actively in market mechanisms and for the emergence of capitalism. Crayen and Baten (2010a) prove that numerical skills in fact have a strong impact on growth patterns across different world regions. In their analysis, the authors regress GDP growth rates on various factors, “growth capabilities”, such as initial GDP levels and numeracy, approximated by the Whipple index, as well as number of other control variables. It turns out that numeracy has a not even a significant, but also an economically meaningful impact on the growth rates of the included countries. Hence, the economy of those countries displaying higher levels of numeracy also grows at a faster pace than of the countries with lower numeracy. All in all we showed that age-heaping based human capital estimates provide the opportunity to track potential reasons for the divergence of countries or world regions in the very long run.

**References**

Schofield (1973)
Duncan, Jones (1990)
Gregory Clark (2007)
Baten and Juif (2013b): Land Inequality (WP)

A'Hearn, B& Baten, J& Crayen, D, (2009) Quantifying Quantitative Literacy: Age Heaping and the History of Human CapitalThe Journal of Economic History, 69(03), 783-808

Acemoglu, D, Johnson, S, and Robinson, JA(2001) The Colonial Origins of Comparative Development: An Empirical InvestigationAmerican Economic Review, 91 (5),1369-1401

Acemoglu, D, Johnson, S, and Robinson, JA(2002) Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income DistributionThe Quarterly Journal of Economics, 117(4), 1231-1294

Baten, J, Juif, D(2014) A Story of Large Land-Owners and Math Skills: Inequality and Human Capital Formation in Long-Run Development, 1820-2000", Journal of Comparative Economics (forthcoming)

Baten, J, Ma, D, Morgan, S, Wang, Q(2010) Evolution of living standards and human capital in China in the 18-20th centuries: Evidences from real wages, age-heaping, and anthropometricsExplorations in Economic History, 47(3), 347-359

Baten, J, Mumme, C(2010) Globalization and educational inequality during the 18th to 20th centuries: Latin America in global comparison, Revista de Historia Económica, 28(02), 279-305

Becker, SO, and Woessmann, L(2009)Was Weber Wrong? A Human Capital Theory of Protestant Economic HistoryQuarterly Journal of Economics, 124, 531-596

Budd JW, Guinanne T(1991) Intentional Age-Misreporting, Age-Heaping, and the 1908 Old Age Pensions ActIreland Population Studies, 45(3), 497-518

Crayen, D, Baten, J(2010) Global trends in numeracy 1820-1949 and its implications for long-term growthExplorations in Economic History, 47(1), 82-99

Crayen, D, Baten, J(2010) New evidence and new methods to measure human capital inequality before and during the industrial revolution: France and the US in the seventeenth to nineteenth centuriesEconomic History Review, Economic History Society, 63(2), 452-478

Charette, MF, Meng R(1998) The Determinants of Literacy and Numeracy, and the Effect of Literay and Numeracy on Labour market OutcomesThe Canadian Journal of Economics, 31(3), 495-517

De Moor T, van Zanden J (2010) "Every Woman Counts": A Gender-Analysis of Numeracy in the Low Countries during the Early Modern PeriodJournal of Interdisciplinary History, 41(2), 179-208

Dell, M(2010) The Persistent Effects of Peru's Mining MitaEconometrica, 78(6), 1863–1903

Friesen, J, Baten, J, Prayon, V(2012) Women Count: Gender (in-)equalities in the human capital development in Asia, 1900-60University of Tuebingen Working Papers in Economics and Finance 29, University of Tuebingen, Faculty of Economics and Social Sciences

Galor, O, Moav, O, and Vollrath, D(2009) Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great DivergenceReview of Economic Studies, 76(1), 143-179

Glaeser, EL, La Porta, R, Lopez-de-Silanes, F, and Shleifer, A(2004) Do Institutions Cause Growth? Journal of Economic Growth, 9(3), 271-303

Goldin, C(1995) The U-Shaped Female Labor Force Function in Economic Development and Economic HistoryIn: TPSchultz, edInvestment in Women's Human Capital (Chicago: The University of Chicago Press), p 61-90

Juif, DT, Baten, J(2013)On the Human Capital of 'Inca' Indios before and after the Spanish ConquestWas there a "Pre-Colonial Legacy"?, Explorations in Economic History 50(2), 227-41

Juif, DT, Baten, J(2014)Dangerous Education? The Human Capital of Iberian and Latin American Jews and Other Minorities during the InquisitionWorking Paper UnivTuebingen

Manzel, K, Baten, J, Stolz,Y(2012)Convergence and divergence of numeracy: the development of age heaping in Latin America from the seventeenth to the twentieth centuryEconomic History Review, Economic History Society, 65(3), 932-960

Manzel, K, Baten, J(2009)Gender Equality and Inequality in Numeracy: The Case of Latin America and the Caribbean, 1880–1949Revista de Historia Económica, 27(01), 37-73

Myers RJ(1954)Accuracy of Age Reporting in the 1950 United States CensusJournal of the American Statistical Association, 49(268), 826-831

Myers RJ(1976)An Instance of Reverse Heaping of AgesDemography, Population Association of America, 13(4), 577-580

Ó Gráda, C(2006)Dublin Jewish Demography a Century AgoThe Economic and Social Review, Economic and Social Studies, 37(2), 123-147

Reis J (2005) Economic Grouwth, Human Capital Formation and Consumption in Western Europe before 1800In: Allen R C, Bengtsson T, Dribe M (ed) Living Standards in the Past, Oxford Univrsity Press Inc, New York, p 218-249

Stolz, Y, Baten, J(2012)Brain drain in the age of mass migration: Does relative inequality explain migrant selectivity? Explorations in Economic History, 49(2), 205-220

Tollnek, F, Baten, J(2012)Farmer Families at the Heart of the Educational Revolution: Which Occupational Group Inherited Human Capital in the Early Modern Era?, CEH Discussion Papers 008, Centre for Economic History, Research School of Economics, Australian National University

Zelnik, M(1961) Age Heaping in the United States Census: 1880-1950The Milbank Memorial Fund Quarterly, 39(3), 540-573

**Not for publication: Appendix B: Methodology and basic concepts of age heaping (Internet Appendix)**

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 will apply the age heaping methodology.[1] The idea is that in less developed countries of the past, only a certain share of the population was able to report the 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.[2]

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. Demographers see this age misreporting as a problem when calculating life expectancies and other population statistics. But exactly this misreporting 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.[3] To calculate the Whipple index of age heaping, the number of persons reporting a rounded age ending with 0

[1] For more detailed surveys on the age heaping methodology see A'Hearn, Baten and Crayen (2009).
[2] 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.
[3] 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.

or 5 is divided by the total number of people, and this is subsequently multiplied by 500. Thus, the index measures the proportion of 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.[4]

$$(1)\ Wh = \left( \frac{(Age25 + Age30 + Age35 + ... + Age60)}{1/5 * (Age23 + Age24 + Age25 + ... + Age62)} \right) \times 100$$

For an easier interpretation, A'Hearn, Baten, and Crayen (2009) suggested another index, which we call the ABCC index.[5] 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 persons 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.[6] 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

[4] 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.

[5] 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.

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

educational status, and other criteria. In each case, they obtained a statistically significant relationship. Remarkable is also 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. The results are not only valid for the U.S.: 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 which could be drawn from the less developed countries of the late 20th century, as mentioned in the introduction to this study, A'Hearn et al. (2009) also assessed age-heaping and literacy in 16 different 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 induced the lower statistical significance.[7]

The possibly widest geographical sample studied so far has been created by Crayen and Baten (2010), 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 in a series of cross-sections between the 1880s and 1940s that primary schooling and age-heaping were closely correlated, with R-squares between 0.55 and 0.76 (including other control variables, see below). Again, the coefficients were shown to be relatively stable over time. This large sample also allowed for the examination of various other potential determinants of age-heaping. To assess whether the degree of bureaucracy, birth registration, and government interaction with

[7] The experience of historical demographers shows that data from premodern times were often very rough, imprecise, or fragmentary. Even the18th century statistical materials are still a sheer jungle of uncertainties and traps, as they were not seldom collected haphazardly and analyzed without sill; as a result, they often embrace just part of the phenomenon which they refer to, are thus incomplete (Szołtysek 2011). This refers in particular to the quality of data on age.

citizens is likely to influence the knowledge of one's exact age, independent of personal education, Crayen and Baten used the number of censuses performed for each individual country up to 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 of the variations of this variable turned out to be insignificant, which would suggest that such an independent bureaucracy effect was rather weak. In other words, it appears to be the case that societies with a high number of censuses and an early introduction of birth registers had a high degree of age awareness. But those societies also introduced schooling early, and this was the variable that clearly had more explanatory power 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 as welfare indicators), and found a varying influence: in some decades, there was a statistically significant correlation, while in others there was none. 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. Baten, Ma, Morgan and Wang (2011) also assessed the cultural factor of dragon year heaping among the Chinese, but found that it did not play a large role.

Was this correlation between numeracy and literacy also visible in Eastern Europe? When comparing the log literacy in the Russian Imperial census of 1897 for the individuals born between 1825 and 1884 on the vertical axis, and their numeracy levels (see Figure 3, expressed by the ABCC index) on the horizontal axis, we can see that there is a clear correlation. The Baltic governments of Estland and Livland, as well as the capital region of St. Petersburg, featured very positively, whereas the 'serfdom'-intensive regions around Belarus had quite low values of both literacy and numeracy.[8] Interestingly, the northeastern districts of European Russia—such as Archangelsk, Wologda, and Perm—were much better in

[8] Please note that, as serfdom was abolished on all these territories in 1864, a potential impact must have stemmed from the first four decades. An outlier is the district of Kowno.

numeracy than they were in literacy. The previous literature has noted that, for literacy development, the existence of schools is even more important than for basic numeracy. For the latter, education in the family contributes more in relative terms. In the thinly populated regions of the northeast, gaining access to schools was much more difficult than in the more densely populated areas farther south.

In conclusion, the correlation between age-heaping and other human capital indicators is quite well established, and the 'bureaucratic' factor does not invalidate this relationship. A caveat relates to other forms of heaping (apart from the heaping on multiples of five), such as heaping on multiples of two, which is quite widespread among children and teenagers, and to a lesser extent among young adults in their twenties.[9] This shows that most individuals knew their age as teenagers, but that only in well-educated societies were they are able to remember or calculate their exact age later in life. At higher ages, this heaping pattern was mostly negligible, but it was, interestingly, somewhat stronger among populations who were numerate enough not to round on multiples of five. We will exclude those below age 23 and above 72 since a number of possible distortions affect those specific age groups, leading to age reporting behavior different from that of the adult group in between. Many young males and females married in their early twenties or late teens, when they also had to register as voters, military conscripts, etc. On such occasions, they were sometimes subject to minimum age requirements, a condition which gave rise to increased age awareness. Moreover, individuals in this age group were physically growing, which makes it easier to determine their age with a relatively high accuracy. All of these factors tend to deflate age-heaping levels for children and young adults, compared with the age reporting of the same individuals at

[9] It has been shown that, in some societies, in addition to the usual overrepresentation of five and zero, there was also a decided preference for figures ending on other digits, whereas avoidance of some numbers was likely to occur in a patterned way as well (Stockwell 1966; Nagi, Stockwell and Snavley 1973).

higher ages. Because the age-heaping pattern of very old individuals is subject to upward as well as downward bias for the reasons mentioned above, the very old should also be excluded.

There remains some uncertainty 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. The age data of the relevant age groups of 23-72 were normally derived from statements from the person himself or herself. However, it is possible that a second party, especially the husband, may have made or influenced the age statement, or even that the enumerator estimated the age without asking the individual. If the latter occurred, we would not be able to measure the numeracy of the person interviewed. In contrast, if the enumerator asked and obtained no response, a round age estimated by him would still measure basic numeracy correctly. A large body of literature has investigated the issue of other persons reporting. Foldvari et al. (2011) speculated, for example, that wives may appear to have been more numerate than they actually were because they improved their age statement with the help of their husbands. They compared the numeracy of married and unmarried women and found that the latter had significantly lower numeracy in some of their samples. However, de Moor (2011) recently rejected this view with a number of good arguments. Moreover, in the early modern period and the 19$^{th}$ century, marriage was often associated with higher educational and social status, as a number of studies have found (for example, Baten and Murray 1998). Baten and Szoltysek (2012) compared male and female numeracy in their sample, and found that women were sometimes more numerate than men, which would support the hypothesis that they reported their age themselves. On the other hand, there was a correlation between the male and female numeracy of different households. Recently, Friesen et al. (2011) compared systematically the evidence of a gender gap in numeracy and in literacy for the late 19$^{th}$ and early 20$^{th}$ centuries, and found a strong correlation. They argued that there is no reason why the misreporting of literacy and age should have yielded exactly the same gap between genders. A more likely

explanation is that the well-known correlation between numeracy and literacy also applies to gender differences. For our study, the question of whether the women answered themselves is slightly less important, because we mainly aim to estimate numeracy of different occupational groups.

Moreover, there is sometimes direct evidence in the sources that the wives themselves were asked. Manzel et al. (2011) reported finding sources on Latin American Indio women in which statements like this one were included: “She says that she is 30, but she looks more like 40.” Even for black female and male slaves in the Cape Colony in South Africa who were accused of crimes, the legal personnel created a separate column that indicated whether the person was guessing her age, or whether she actually knew. It is possible that, if those Indio and African women, who probably were not shown much respect by colonial officers, were asked for their age; then European women might also have been asked for their age, as the level respect shown to them might have been somewhat greater.

Of course, 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 analyzing 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 previous studies have generally found much more age-heaping (and less numeracy) among the lower social strata, and among the half of the sample population who had lower anthropometric values (Baten and Mumme 2010). Moreover, the regional differences of age-heaping are similar to the regional differences in illiteracy. It can be concluded that the method of age-heaping is a useful and innovative tool for assessing human capital.

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.[10] we omitted the age range from 63 to 72, as this age group offers too few observations, especially for the 17th and 18th centuries, when mortality was relatively high.[11]

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. The age heaping measure is a very basic measure of human capital. Therefore, it is 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.

[10] An advantage of this method is to spread the preferred ages, such as 25 or 30, more evenly within the age groups and it adjusts also 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 2009).

[11] 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 (2009) to increase the Whipple value (minus 100) by 24 percent, before calculating the ABCC measure.