QUANTIFYING QUANTITATIVE LITERACY:

AGE HEAPING AND THE HISTORY OF HUMAN CAPITAL

Brian A´Hearn, Franklin and Marshall College, USA
bahearn@fandm.edu

Joerg Baten, Univ. Tuebingen and Univ. Pompeu Fabra, Barcelona, Spain
Joerg.baten@uni-tuebingen.de

Dorothee Crayen, Univ. Tuebingen, Germany
Dorothee.Crayen@uni-tuebingen.de

Version 19 Nov., 2006

Abstract

Age data frequently display excess frequencies at round or attractive ages, such as even numbers and multiples of five. This phenomenon of age heaping has been viewed as a problem in previous research, especially in demography and epidemiology. We see it as an opportunity and propose its use as a measure of human capital that can yield comparable estimates across a wide range of historical contexts. A simulation study yields methodological guidelines for measuring and interpreting differences in age heaping, while analysis of contemporary and historical datasets demonstrates the existence of a robust correlation between age heaping and literacy at both the individual and aggregate level. To illustrate the method, we generate estimates of human capital in Europe over the very long run, which support the hypothesis of a major increase in human capital preceding the industrial revolution.

Keywords: Human Capital, Age Heaping, Growth, Industrial Revolution, Numeracy
JEL codes: I21, N01, N30, 015

I. Introduction

In the twenty-first century, legislatures and educational bureaucracies creak into action at the publication of the latest international league tables of educational attainment.[1] In the U.S., such concerns have a long history and a particular focus on math, dating back to post-Sputnik anxiety about falling behind in a technological race with economic and national security implications. This led to reforms like the "New Math" of the 1960s, measures like the Education for Economic Security Act of 1984, and debates like the current "Math Wars" about classroom pedagogy. But these are only the most recent manifestations of national math anxiety. More than a century ago it was British fear of German technological ascendancy that spurred a reform movement demanding better mathematical, technical and scientific education.[2]

Concern with educational policies is not misplaced, according to a voluminous (if inconclusive) literature on the determinants of economic growth.[3] But growth regressions have yielded few insights into the importance of *different types* of knowledge, education, or skill. Large panel datasets rarely have information more detailed than gender- and level-specific enrollment or attainment rates. An oft-cited analysis indicating that engineering students raise growth while law students lower it is the exception rather than

[1] Influential comparative projects include TIMMS (Trends in International Mathematics and Science Study; International Association for the Evaluation of Educational Achievement), PISA (Programme for International Student Assessment; OECD), and ALL (Adult Literacy and Lifeskills; Statistics Canada and OECD).
[2] This fear was captured in the title of a well-known and pessimistic analysis of the industrial position of the country: "Made in Germany." See Sanderson (1999).
[3] A few influential studies that investigate the connection between schooling and economic growth include Levine and Renelt (1992), Mankiw, Romer and Weil (1992), Barro and Lee (1994), Barro (1997), Sala-i-Martin (1997), Bils and Klenow (2001), and Hanushek and Kimko (2000).

the rule.[4] In studies of individual labor market outcomes, the consistent and powerful predictive capacity of human capital variables is not in question. Debate concerns only the precise mechanisms at work, for example whether education primarily screens ability or creates it, or whether skill- or college-premia are increasing in an era of rapid technological change. In this context, achievement and skill test results are often introduced as measures of ability alongside educational attainment and experience variables. Invariably, ability measured in this way matters; occasionally, it is broken down by specific skill areas such as quantitative and verbal reasoning.[5] In a study of the U.S. labor market, Murnane, Willett, and Levy (1995) found that cognitive ability had greater predictive power than educational attainment, with mathematics the ability most highly correlated with wages. Similarly, Rivera-Batiz (1992) found that "quantitative literacy" significantly (in both senses) raised the probability of full-time employment among U.S. workers. Outside the U.S. recent studies have found numeracy to be positively associated with labor force participation, full-time employment, annual weeks worked, and income in Britain, Canada, and Australia.[6] Often numeracy dominates literacy as an explanatory factor, particularly for women and among the less educated.

Interest in the historical evolution of human capital has been given fresh stimulus by the recent development of very-long-run growth models. These models attempt a unified explanation of pre-modern Malthusian dynamics and modern economic growth, typically assigning a key role to a fertility transition in which families switch from

[4] Murphy, Shleifer and Vishny (1991).
[5] Studies with evidence on general cognitive ability include Neal and Johnson (1996) and Heckman (1995).
[6] See Chiswick, Lee, and Miller (2003) for Australia; Charette and Meng (1998) and Finnie and Meng (2001) for Canada; Parsons and Bynner (2005) for Britain.

quantity to quality of children, investing more in human capital.[7] The empirical basis of such theorizing is shaky, however. For the nineteenth century, measures of human capital investment, such as school enrollments and attainment, or census data on literacy, are available – at least for some of today's rich countries. O'Rourke and Williamson (1997) were able to include schooling in European convergence regressions for the 1870-1913 period, for example, concluding that globalization forces were in fact a much more important influence on comparative development.[8] No data whatsoever are available on individual cognitive ability or on different types of ability. Pushing back into the early nineteenth century and before, schooling data dry up and literacy must generally be inferred from a proxy: ability to sign one's name on marriage registers and legal documents.[9] For the years around 1800, Reis (2005) is able to assemble such data for 15 European regions. The figures indicate that male literacy varied widely, from over 60% in northwestern Europe to below 20% in parts of Italy and under 10% in eastern Europe. Pushing back still further into the early modern era, it becomes increasingly difficult to find systematic, comparable data. Relatively plentiful data on reading ability in Scandinavia and signature ability in the Netherlands, Britain, France, and Spain allow Graff (1987) to document a considerable improvement in literacy in the seventeenth and

---

[7] In Galor and Weil (2000), population size eventually causes technological progress to speed up, raising the return to human capital, causing families to increase investment in human capital, lowering fertility and increasing the rate of growth. Thus is the Malthusian cycle broken. Becker, Murphy and Tamura (1990), Lucas (2002), and Cervellati and Sunde (2005) are further examples of models explicitly built on such human capital foundations.

[8] Tortella (1994), using literacy data, offers a different interpretation, at least for southwest Europe.

[9] The limitations of signature ability as a measure of functional literacy are obvious, but can also be raised with respect to self-reported "ability to read". In practice the two measures are well correlated where both can be observed. In historical curricula, reading instruction was largely completed before writing was started.

eighteenth centuries. But there is little information available on the rest of Europe. Allen’s (2003) conclusion that human capital has no ability to explain progress and poverty in Europe between 1300 and 1800 may result more from his use of urbanization as a proxy than from literacy’s actual irrelevance. Indeed, Baten and van Zanden (2006) reach the opposite conclusion using book production to measure human capital.

What of numeracy as a historical measure of human capital? For Weber, Sombart, and Schumpeter, numeracy was at the very heart of modern, rational capitalism. They traced the roots of both to the invention of double-entry bookkeeping in late medieval Italy. Carruthers and Espeland (1991) describe in some detail the process of abstraction and organization inherent in compiling a ledger, which made possible the development of concepts like capital, depreciation, and rate of profit.[10] It is no accident that the introduction of Arabic numerals into Europe (by the merchant Leonardo of Pisa, a.k.a. Fibonacci) and the earliest accounts of mathematics education date from this same time and place. Numerous *scuole d’abbaco* thrived in Renaissance Florence according to Goldthwaite (1972), where the young sons of the commercial classes studied for a year or two a mathematics curriculum that would change little before the nineteenth century.[11] Italy remained the European center of publication and instruction in mathematics and accounting until at least 1500, according to Swetz (1987). Emigh (2002) has investigated the numeracy of ordinary Tuscans in this period by analyzing their tax declarations for

[10] The authors argue that although double-entry bookkeeping truly was a superior technology, that its potential was seldom exploited by practicing merchants. It came to have a powerful rhetorical significance, as a symbol of meticulousness and probity.
[11] Routine commercial calculations in the middle ages could include conversions between non-decimal monetary systems with fluctuation exchange rates, estimation of the volume of containers, the reckoning of interest, or the division of profits between partners with different amounts of capital invested at different times.

the famous Florentine *catasto* of 1427. She finds that ordinary citizens and peasants much more often provided too much quantitative information (rents, size of plots, yields, debts, salaries) than too little, with respect to the demands of the tax officials. This implies that causation ran from market activity to numeracy to tax design, rather than in reverse. Cohen (1982), another historian of numeracy, finds that in the expansion of market activity in the early nineteenth century U.S. was only partially responsible for making Americans “a calculating people.” A reform of math pedagogy actually de-emphasized commercial applications in favor of teaching abstract thinking in this period.[12] A sophisticated literature on the history of numeracy certainly exists, but it does not yield statistical measures. Can we quantify quantitative reasoning?

It turns out that we can. As signature ability can proxy for literacy, so accuracy of age awareness can proxy for numeracy, and for human capital more generally. A society in which individuals know their age only approximately is a society in which life is governed not by the calendar and the clock but by the seasonal cycle, in which birth dates are not recorded by families or authorities, in which numerical age is not a criterion for access to privileges (e.g. voting, office-holding, marriage, holy orders) or for the imposition of responsibilities (such as military service or taxation), in which individuals who know their birth year have difficulty accurately calculating their age from the current year. Within a society, the least educated and those with the least interaction with state, religious, or other administrative bureaucracies will be least likely to know their age accurately. Age awareness thus tells us something about both the individual and the

[12] Equally important was the triumph of “political arithmetic” as the favored tool for assessing the experiment of republican America and influencing public opinion. This presupposed a certain basic level of numeracy.

society he or she inhabits. Approximation in age awareness manifests itself in the phenomenon of age heaping in self-reported age data. Individuals lacking certain knowledge of their age rarely state this openly, but choose instead a figure they deem plausible. They do not choose randomly, but have a systematic tendency to prefer "attractive" numbers, such as those ending in 5 or 0, or even numbers, or in some societies numbers with other specific terminal digits. Age heaping can be assessed from any sufficiently numerous source of age data: census returns, tombstones, necrologies, muster lists, legal records, or tax data, for example. While care must be exercised in ascertaining possible biases, such data are in principle available much more widely than signature rates and other proxies for human capital.

Age heaping is a well-known phenomenon among demographers. Already a half-century ago influential studies by Bachi (1951) and Myers (1954) investigated age heaping and its correlation with education levels within and across countries. Myers (1976) demonstrated the correlation at the individual level between age awareness and income. For others, including epidemiologists, age heaping is a problem to be solved, a source of distortion in age-specific vital rates. In this context it remains a standard topic in U.N. analyses of developing country population data.[13] Development economists and anthropologists use age heaping as a measure of data quality and consistency. Meanwhile, historians have studied age heaping as a topic of interest in its own right. A pioneering example is the study by Herlihy and Klapisch-Zuber (1978) of the Florentine tax records from the fourteenth and fifteenth centuries. In a chapter devoted to age

[13] Discussions of age heaping as a problem in the demography literature include Coale and Kisker (1986), Preston, Elo, Rosenwaike and Hill (1996), and Vallin, Meslé, Adamets, and Pyrozhkov (2002). Denic, Khatib and Saadi (2004) discuss the issues for medical research. See also U.N. Statistics Division (2003).

heaping, they documented marked heaping on even numbers for children and on multiples of five for adults, to a degree similar to that reported for Egyptian census data in 1947. Among other findings, they demonstrated that age heaping diminished substantially over successive tax enumerations from 1371 to 1470, and that it was more prevalent among women, in rural areas and small towns, and among the poor. A second well known example is the study of Duncan-Jones (1990), who used ages from tombstones to estimate age heaping for men and women in twelve provinces of the Roman Empire. He found age heaping on multiples of five at levels not dissimilar to those for medieval Tuscany or developing countries of the 1950s and ’60s and higher for women than men.[14]

The first use of age heaping as an indicator of human capital in the economic history literature is relatively recent. Mokyr (1983) tested for positive selection or “brain drain” in pre-famine Irish emigration by comparing age heaping among migrants and in the population at large. Developing original measures of age heaping along the way, he found no support for the conventional wisdom that the best and brightest emigrated. Budd and Guinnane (1991) studied Irish age misreporting in linked samples from the 1901 and 1911 censuses. They found considerable heaping on multiples of five in the 1901 census, which was greater among the illiterate, the poor, and the aged. The introduction of state pensions for individuals 70 and older in 1908 changed incentives regarding age reporting. For those potentially eligible, there was on the one hand an incentive to exaggerate age, on the other a concern to report age accurately, since pension examiners had access to census returns. On balance, age heaping declined significantly in the 1911 census, and its

[14] Other interesting studies of age heaping are Kaiser and Engel (1993) on early modern Russia, and Jowett and Li (1992) on contemporary Chinese ethnic groups.

variation across social groups became much less clear. The longitudinal data enabled the authors to check the consistency of individual age reports. Mean elapsed time was not 10 but approximately 12 years, the positive discrepancy being larger for Catholics, illiterates, and those reporting heaped ages in 1901. In another study of Ireland, O'Grada (2006) used a higher degree of age heaping among Dublin's immigrant Jewish population to show that their lower literacy did not refer only to the English language and that their lower mortality was the result of religious practices rather than education. In a similar linked census sample for Britain in 1851 and 1881, Long (2005, forthcoming) assessed both aggregate age heaping at the county level and, exploiting the repeated observations, individual age discrepancies. Fully a quarter of his sample of 1851 school-aged children reported ages in 1881 with discrepancies of from two to five years. While countywide age heaping had a limited impact on individual outcomes once other county characteristics were controlled for, individual age discrepancy has a significant impact on socio-economic status, wages (10% higher for 0-discrepancy individuals), and the probability of rural-urban migration. What these studies have in common is that all find evidence of significant age heaping, and that it varies across individuals or groups in a way consistent with its interpretation as a measure of human capital.

Researchers in a number of disciplines are familiar with age heaping, and it has been deployed as an indicator of human capital in studies of particular times and places. But several different measures of age heaping have been employed, complicating comparisons. And more generally, awareness of age heaping is somehow less than the sum of these parts. Age heaping has an unexploited potential to yield new insights into the comparative historical evolution of human capital. In this paper we sketch the outlines of an age heaping methodology. Sections II and III use simulations to investigate the reliability of various available measures of heaping for the first time. Sections IV and V establish the connection of age heaping with another measure of human capital – literacy, at both the individual and aggregate level, in both contemporary and historical data. Section VI provides an illustrative example of age heaping's potential to track human capital over a very long span of time and across widely differing societies. Section VII offers concluding thoughts and suggestions for further research.

II. Measuring age-heaping

To deploy age-heaping as a useful indicator of human capital, we require a measure that allows us to track its variation over time and across groups. This is a question of *how much* age heaping is present in the sample, conceptually distinct from the question whether age heaping is present at all. For the second of these questions, the familiar Pearson chi-squared statistic would be suitable:

$$(1) \qquad H_P = \sum_{i=1}^{k} \frac{\left(n_i - \hat{n}_i\right)^2}{\hat{n}_i}$$

where H is chosen to stand for heaping and P for Pearson, i indexes the k ages to be considered, and n_i and $\hat{n}_i$ are the observed and expected frequencies, respectively. Note that the chi-squared statistic does not attempt to estimate a parameter such as, for example, the share of ages incorrectly reported. It can take on quite a wide range of non-negative values, depending on k and the particular pattern of expected frequencies (the null hypothesis). Like the other indices to be considered, H_P is based on comparison of actual and expected frequencies, and like them it must standardize and aggregate these deviations. The chi-squared statistic standardizes by squaring and expressing the result as a percentage of the expected frequency. It aggregates by summing with equal weights. H_P

is asymptotically distributed as a χ^2_{k-1} random variable, facilitating hypothesis testing concerning the presence of heaping.

But our primary interest is in comparing the *degree* of age-heaping in different samples: the “how much” question. And for this purpose the chi-squared statistic has the drawback of scale dependence. For clarity, it is useful to distinguish *mathematical* from *statistical* or probabilistic, scale dependence. The chi-squared statistic is scale dependent in the mathematical sense that inflating both observed and expected frequencies by some common factor causes H_P to grow by the same factor. It is *not* scale dependent in a statistical sense, in that $E\left(\chi^2_{k-1}\right) = k - 1$, regardless of sample size. The two statements are reconciled by noting that, under the null hypothesis, a larger sample is not expected to increase all observed frequencies by the same factor. Rather, the observed and expected distributions will tend to conform more closely on average. Mathematical scale dependence complicates comparisons, since two samples of different sizes, with identical patterns and degrees of heaping, will yield different values of H_P. Similarly problematic is the dependence of H_P on the age range considered (k), which varies across samples. Of course, it is possible to consider only subsamples of like size and age-range, but this either throws away valuable information or necessitates a cumbersome resampling and averaging procedure. Another potential weakness of H_P is that it weights the deviations for all ages equally, even those for which smaller frequencies and greater variability are expected.

In this paper two indices based on the chi-squared statistic are considered. Both are mathematically scale-independent, and both weight deviations by the age’s expected sample share so as to reduce the influence of potentially unreliable observations. They differ in their methods for standardizing the deviations. The first, dubbed the "ABC" Index after the initials of authors, is a weighted sum of the squared percentage deviations from predicted values:

$$(2)\ \ H_{ABC} = \sum_i \frac{\hat{n}_i}{N}\left(\frac{n_i - \hat{n}_i}{\hat{n}_i}\right)^2,$$

where the notation is as before and N denotes the sample size. If possible discrepancies between expected and actual sample size are ignored, H_{ABC} can be rewritten as $\frac{1}{N}H_P$, so that it is in fact a scaled version of the chi-squared statistic.

Squaring the deviations has the effect of heavily weighting outliers. The second index, "Lambda," avoids this by relying on the absolute value of the percentage deviations:

$$(3)\ \ H_\lambda = \frac{\sum |n_i - \hat{n}_i|}{\sum \hat{n}_i} = \sum_i \frac{\hat{n}_i}{N}\frac{|n_i - \hat{n}_i|}{\hat{n}_i}.$$

Neither the ABC nor the Lambda Index embodies particular assumptions about how expected frequencies are generated. Both can take on a wide range of values depending on the number of ages k and the particular distribution of expected frequencies.[15]

Several commonly used indices aggregate observed and expected frequencies over terminal digits before calculating deviations. Bachi's index sums the differences between actual and expected frequencies of each terminal digit, considering only the

[15] The Lambda Index was proposed by Mokyr (1983), who also considered a "Gamma" index, which can be rewritten as $\sum_i \frac{\hat{n}_i}{N^2}\frac{(\hat{n}_i - n_i)^2}{\hat{n}_i}$, showing it to be a weighted chi-squared type index. In simulations, Gamma's performance did not differ in interesting ways from that of H_{ABC} and H_λ.

positive values – intuitively similar to the use of absolute values in (3).[16] The resulting sum is expressed relative to sample size:

$$(4)\ H_B = \frac{1}{N}\sum_{j=0}^{9} \mathrm{I}(j)\cdot(n_j - \hat{n}_j), \quad \mathrm{I}(j)=1 \text{ if } n_j > \hat{n}_j, \text{ else } \mathrm{I}(j)=0,$$

where I(j) is an indicator variable as defined above, and j indexes terminal digits from 0 to 9. The Bachi index is not mathematically scale dependent, weights terminal digits with positive deviations equally, and makes no assumptions about how to derive the expected frequencies for each terminal digit. If all expected terminal digit frequencies are assumed to be 10%, H_B can take values between 0 and 0.90 and can be intuitively interpreted as an approximation of the percentage of the sample reporting an inaccurate age.

Another commonly used measure is Myers' Blended Index, which differs from those considered thus far by making a specific assumption about expected frequencies, and by making an adjustment to observed frequencies. Predicted terminal digit shares are set at 10% (meaning the index must be applied to age intervals that are multiples of 10 years). This is of course not always a realistic assumption; in a sample of individuals aged 60 to 79, one would expect more ages ending in 0 than in 9, for example. Rather than adjust *expected* frequencies accordingly, H_M adjusts the *observed* frequencies using a "blending" procedure.[17] Unlike H_B, which considers only positive deviations and expresses them relative to N, the Myers Index sums the absolute values of all deviations, and expresses them relative to $2N$:

---

[16] This is only one of three indices proposed by Bachi. The others focus on a particular age or a particular terminal digit.

[17] See Shryock and Siegel (1973) for an abbreviated description of Myers' blending procedure.

$$(5)\ H_M = \frac{1}{2}\sum_{j=0}^{9}\left|\frac{\tilde{n}_j}{N} - .10\right| = \frac{1}{2}\sum_{j=0}^{9}\frac{\left|\tilde{n}_j - \hat{n}_j\right|}{N},\ \hat{n}_j = 0.1N\ \forall\, j,$$

where $\tilde{n}_j$ is the "blended" observed frequency of a particular terminal digit *j*. The index is not mathematically scale dependent, and can vary between 0 and 0.90 like the Bachi.

The indices considered so far can be used to detect any type of heaping, and do not rely on particular mechanisms for generating expected frequencies (with the exception of the Myers Index). More specialized measures focus on a particular type of heaping. The well-known Whipple Index is designed to capture heaping on ages ending in 0 or 5. H_W sums the frequencies of all ages ending in 0 or 5 and expresses the result relative to one-fifth the sample size:

$$(6)\ H_W = \frac{\sum(n_{25} + n_{30} + n_{35} ... + n_{60})}{\frac{1}{5}\sum_{i=23}^{62} n_i}.$$

The summation notation in the denominator (rather than the *N* used in Equations 2-5) is meant to emphasize that H_W must be defined over an interval in which each terminal digit occurs an equal number of times, such as 23 to 62. Implicitly, equal terminal digit shares in unheaped data are assumed. This would be correct for a uniform distribution of ages, but can only be approximate for typical samples in which frequency decreases with age. The Whipple index makes no adjustment to correct for this problem. H_W can range from 0 in the case of no observations on 0s and 5s, through 1 in the case of a uniform distribution of terminal digits, to 5 in the case of 100% heaping. In application, H_W is typically multiplied by 100.

An alternative index of heaping on the terminal digits 0 and 5 makes a less restrictive assumption regarding expected frequencies: that they evolve approximately linearly over any three year age range. The “Multiples of Five” index is a simple, equally weighted, average of the frequencies of terminal digits 0 and 5, each expressed relative to the average frequency of immediately adjacent ages, as in the following example:

$$(7)\ H_{M5} = \frac{1}{3} \cdot \left\{ \frac{2n_{20}}{n_{19}+n_{21}} + \frac{2n_{25}}{n_{24}+n_{26}} + \frac{2n_{30}}{n_{29}+n_{31}} \right\}.$$

III. Evaluating Indices of Age-Heaping

Which of the indices defined in Section II is best? The familiar criteria used to assess estimator performance, bias and variability, are not directly applicable here, as the indices under consideration are not estimators of unknown parameters. Three desirable properties of an age-heaping index are statistical scale-independence, a linear response to the degree of heaping, and the ability to reliably rank samples from populations with different degrees of heaping. Because the distributions of the indices are unknown, these properties must be investigated by simulation.

a. *Simulation details*

Three types and five degrees of heaping are distinguished in the study. The *types* are: heaping on even numbers, heaping on multiples of five, and mixtures of the two. The *degrees* are: for the pure types, 0, 5, 10, 15, and 20 percent of the sample subjected to heaping; and for the mixed type 5 percent even-heaping plus an additional 5, 10, 15, or 20 percent multiples of five heaping. The predominance in the mixed scheme of heaping on multiples of five over even-heaping conforms to the pattern found in a wide range of historical datasets. The mixed schemes are the focus of most discussion here.

The three heaping patterns were imposed on random samples drawn from distributions in which frequencies decrease with age, as is typical of data from military, census, and randomly-sampled sources. More specifically, samples of size 250, 500, 1000, 2000, and 5000 were drawn from the 23-42 range of a *N*(20,10) distribution rounded to integer values, then heaped according to the relevant scheme. The entire exercise was repeated for the wider age range 18-57, yielding results sufficiently close to those reported here as not to merit separate discussion. 1000 repetitions were carried out for each sample size and heaping scheme.

As noted in Section II, the ABC, Lambda, and Bachi indices do not embody specific assumptions about expected frequencies. In the simulation study, expected frequencies were generated using locally weighted regression. The intuition is to use a regression of observed frequencies on age to generate predicted frequencies, allowing the estimated relationship to vary locally. For each age, the slope is estimated using only data in a window around that age (here set to include 80% of all observations), using a kernel (here the tricube) to weight them inversely with the distance from the age being considered. This approach is similar to other smoothing methods, in particular kernel density estimation. It has the advantage of being easily implemented and flexible, making no *a priori* assumptions about the form of the underlying (unheaped) distribution of ages. Its potential disadvantage is oversensitivity to (undersmoothing of) the observed, randomly- or systematically-heaped frequencies. In principle, this problem can be addressed by optimal choice of bandwidth. But this choice depends on the type and degree of heaping and on the underlying distribution of unheaped ages, all of which are unknown in practice. Hence, a uniform (non-optimal) bandwidth was applied for all sample sizes, heaping schemes, and indices. The evaluation of the ABC, Lambda, and Bachi indices is in fact a joint evaluation of the index itself and the lowess method of deriving expected frequencies. Similarly, the Myers Index is evaluated together with its blending method of adjusting observed frequencies.

Figure 1 about here.

As an example of the results, Figure 1 displays kernel density estimates of the distribution of the Lambda Index under the mixed heaping scheme with a sample size of 1000. The leftmost curve corresponds to no heaping, the rightmost to maximum heaping. It is evident that on average the index value increases with the degree of heaping, as it was designed to. The graph also indicates for this specific case that the increase is not linear with the degree of heaping, that the distribution is right-skewed at low degrees of heaping, that the variance of the distribution increases, and that there is substantial overlap in the distributions for successive degrees of heaping.

b. *Scale dependence*

The indices under study are mathematically independent of scale by construction. But they may be *statistically* scale dependent, in the sense that their expected value may be a function of sample size. Random sampling variation always creates deviations between observed and expected frequencies, even when the population is unheaped. Since such deviations are squared, converted to absolute values, or considered only when positive in Equations 2-5, there is no tendency for them to cancel each other out, so that the means of these indices are not zero even in the absence of heaping. When such deviations are expressed relative to expected frequencies or sample size, however, they tend to diminish in larger samples. This creates what is termed here statistical scale dependence, which creates the same difficulties as mathematical scale dependence.

Figure 2 about here.

Figure 2 plots the means of all six indices as a function of sample size under the 05-10 mixed heaping scheme (5 percent even-heaping, 10 percent heaping on multiples of five). Statistical scale dependence is apparent for all but the Whipple Index. The ABC and Lambda indices perform poorly on this criterion. Their decrease with a doubling of sample size (at least when starting from a small sample) is enough to offset the increase that would result from doubling the degree of heaping. Among the "all-purpose" indices, the Bachi seems to do best. The pattern of Figure 2 is representative of results under all types and degrees of heaping, though the severity of scale dependence is somewhat reduced when the degree of heaping is extreme.

c. *Response to heaping*

Figure 3 plots mean index values for increasing degrees of mixed heaping, for sample sizes of 500. All index means rise monotonically with the degree of heaping, as designed, but responsiveness varies considerably. The Multiples of 5 and Whipple Indices increase linearly, while the ABC and Lambda Indices increase at increasing rates. The sharply nonlinear response of the ABC and Lambda indices creates two related problems. Most importantly, it is difficult to distinguish low degrees of heaping from zero and from each other, especially once random sampling variability is taken into account. In addition, even substantial differences in heaping across samples become less easily interpretable. An increase of 0.2 in the Whipple Index always corresponds to an increase of approximately 5% in the share heaped on multiples of five. No analogous simple rule describes the indices with a non-linear response. They do not, for example, double when the share heaped doubles. The Bachi and Myers Indices have an intermediate response pattern, which is reasonably close to linear.

Figure 3 about here

The response patterns in Figure 3 are representative of all sample sizes and of both mixed and pure heaping on multiples of five. Pure even-heaping produces somewhat different results. It is no surprise but deserves emphasis that the Whipple Index does not respond at all, actually declining somewhat as the share subjected to heaping increases in the simulation results. Similarly, the Multiples of 5 Index increases only imperceptibly. The other indices increase monotonically with the degree of heaping, in a distinctly non-linear way in the case of the ABC and Lambda Indices. The rate of increase is less than for types of heaping including multiples of five, making it difficult to distinguish varying degrees of even-heaping. As a practical matter, however, datasets from a wide variety of countries and time periods have failed to yield an example of pure even-heaping. When present, even-heaping is "outweighed" by heaping on multiples of five.

d. *Precision and the probability of ranking errors*

Accuracy in ranking samples depends not only on the response of the index mean to age heaping, but equally on its variation around that mean in repeated sampling. Figure 1 illustrated how the distribution of the Lambda Index varied with the degree of heaping, shifting steadily to the right, decreasing in skewness, and increasing in variance. The substantial overlap of the estimated densities for successive degrees of heaping in Figure 1 signaled likely difficulties in distinguishing small variations in the degree of heaping from random sampling noise. Index distributions are also a function of sample size. This is illustrated for the Bachi Index in Figure 4, which plots kernel density estimates for 05-10 heaping in sample sizes of 250, 500, 1000, 2000, and 5000. The Bachi's previously-noted scale dependence is again evident; the low, wide curve with the rightmost mode corresponds to a sample of 250, while the tall, narrow, leftmost curve is for sample size 5000. Also evident is a quite dramatic decrease in the variance of the distribution as the sample size increases, a characteristic of all the indices.

Figure 4 about here.

A useful gauge of reliability in discerning the degree of heaping turns out to be the probability of incorrectly ranking sample pairs. A ranking error occurs when the sample from a population with a low degree of heaping, say 05-05, yields an index value exceeding that for the sample from a population subject to a greater degree of heaping such as 05-10. We could write this as $H^{05-05} > H^{05-10}$, or alternatively $D = H^{05-10} - H^{05-05} < 0$. The distribution of differences like D can be estimated by recourse to the already simulated index realizations under various types and degrees of heaping. In many cases these differences have an approximately normal distribution. Figure 5 provides an example: the estimated density of the difference $D = H_B^{05-10} - H_B^{05-05}$ for the Bachi Index. The probability of a negative difference, hence a ranking error, is given by the shaded area, which in this case equals 0.15.

Figure 5 about here.

Table 1 presents estimates of the probability of a ranking error for different sample sizes under the mixed heaping type. The differences refer to "adjacent" degrees of heaping; $D1 = H^{05-05} - H^{00-00}$, $D2 = H^{05-10} - H^{05-05}$, and so on. Immediately apparent from the figures in Table 1 is that none of the indices perform reliably in small samples. In samples of 250, the "all-purpose" indices routinely misrank samples from unheaped populations as having more heaping than those from 05-05 populations: 37% of the time. The Whipple Index offers dramatically better performance for low degrees of heaping, and remains the best by a small margin even with more severe heaping. In large samples, the accuracy of all indices is much improved, the error probabilities comfortably small. Again the Whipple Index is clearly the most accurate of the group, especially in distinguishing among relatively low degrees of heaping. It bears emphasis that the Whipple is quite incapable of capturing forms of heaping other than multiples of five. Among the more flexible indices, the Bachi offers the lowest error probabilities.

Table 2 reports the analogous probabilities for samples that are two degrees of heaping apart. In other words, $D20 = H^{05-10} - H^{00-00}$, $D31 = H^{05-15} - H^{05-05}$, and so on. While all indices have trouble reliably distinguishing fine differences in the degree of heaping, the results in Table 2 indicate considerably better accuracy for larger heaping differences. Here again, the Whipple Index is clearly the most reliable – so long heaping on terminal digits 5 and 0 characterizes the data. In large samples, all indices are extremely accurate, with incorrect rankings not a practical problem.

Tables 1, 2 about here.

IV. Age heaping and illiteracy today

To assess age heaping's potential as a measure of human capital, we begin by using contemporary data to investigate its range of variation and its correlation with illiteracy, the most commonly available alternative indicator. The Demographic Health Surveys (DHS) sponsored by USAID are a rich source of individual-level data on the health, nutrition, and household demographics of women in developing countries. Indices of age heaping and rates of illiteracy were calculated from the DHS data for women aged 23-42 from 415 regions in 52 countries. A *functional* definition of illiteracy was used in the surveys: the inability to read "easily" or to read "a whole sentence." Based on the results of Section III, the Whipple and Bachi indices were chosen as measures of age heaping, the Bachi being calculated from sample sizes of 500 (averaging repeated samples where possible). Age heaping ranges widely at the national level: the Whipple from below 100 in Kyrgyzstan to a peak of 236 in Sudan; the Bachi by almost a factor of five between Trinidad & Tobago (0.06) and Sudan (0.28). Female illiteracy also ranges widely in the DHS data, from under 10% in Kyrgyzstan or the Philippines to over 90% in Chad, Mali, and Senegal.

As documented in Table 3, both measures of age heaping are well correlated with illiteracy. The regressions use the natural logarithm of age heaping to accommodate the moderate degree of nonlinearity in the relationship with illiteracy found in preliminary testing; age heaping increases at an increasing rate with illiteracy. Robust standard errors are reported due to the presence of heteroskedasticity. While the age heaping – illiteracy relationship is strong and statistically significant for both the Bachi and Whipple indices, the R^2's indicate that it is considerably less noisy in the case of the Whipple. (In these bivariate regressions, R^2 is just the square of the simple correlation coefficient, hence comparable even though the dependent variables differ.) The results are similar for either the 52 countries or their 415 separate regions. Adding country dummies to the regional regression yields the estimates headed "country fixed-effects."[18] The "R^2 within" figures indicate that within-country variation in illiteracy explains 19% of within-country variation in (ln) age-heaping, using either index. The robustness of the slope estimates to the inclusion of country dummies shows that the illiteracy-age heaping relationship is not

[18] Estimated country effects are not reported in Table 3 for reasons of space, but several are large. Chad, Pakistan, Sudan, and Yemen all have unexpectedly high levels of age heaping given their levels of illiteracy, for example.

driven solely by a few outlying countries with high illiteracy and age heaping. In light of these results, it seems reasonable to conclude that age heaping can function not only as a direct measure of numeracy but also as a proxy for illiteracy.

Table 3 about here.

Analysis of individual data permits further insight into whether age heaping more reflects the characteristics of individuals or of the society they inhabit. From 97 DHS surveys a sample of over 700,000 adult (20-49) women from 47 countries was extracted, with information on age, illiteracy, and region of residence.[19] Given the results of the aggregate analysis, in which the Whipple Index was more highly correlated with illiteracy than the Bachi Index, we suspect that the predominant form of heaping is multiples of five. In fact, the sample share of ages ending in 0 and 5 is 26%, considerably in excess of a naïve prediction of 1/5 or 20%. For this reason, heaping is defined in terms of terminal digits 0 and 5 in what follows.

The probability of reporting a multiple of five age is modeled as a logistic function of both individual characteristics and regional illiteracy. Table 4 reports the estimated marginal effects of each explanatory variable, evaluated at the mean of all variables. In the most basic model, personal illiteracy raises the probability of reporting a multiple of five age by 5.7 percentage points, a large effect relative to the six points of excess frequency observed at these ages, and the estimate is highly statistically significant. When regional illiteracy is added to the equation, its estimated effect is significant in both senses of the word. Increasing the regional illiteracy rate by one standard deviation (26.8) from its mean (47.7) increases the predicted probability of an

[19] These surveys overlap the regional data sources and include repeated surveys of the same country, but not the same households in the same country except by chance.

individual reporting a heaped age by about three percentage points, equivalent to the effect of individual illiteracy in this equation. In columns three and four the effect of the regional illiteracy rate is estimated separately for literate and illiterate individuals. The constants (not reported) of course differ across the two groups. But so do the slope coefficients; among illiterates, the marginal effect of regional illiteracy is more than twice that among literates. Illiterates are in this sense more dependent on their neighbors.

Table 4 about here.

We can probe the determinants of age heaping more deeply by adding age-group dummies to the explanatory variables. The probability of reporting a heaped age can vary with age itself, even controlling for personal and social illiteracy. One reason is that older individuals are more likely to forget their ages.[20] (The extremely old, who sometimes deliberately exaggerate their age, have been excluded from the sample.) Secondly, society may provide different incentives to different groups to distort their true age. For example women still single in their thirties or beyond might feel pressure to report a low age to increase "marriageability", and this distortion could be associated with some form of heaping (though not necessarily multiples of five).[21] Finally, the degree of regional illiteracy may have been greater when today's older individuals were young, or alternatively the quality and quantity of education may have changed in ways not captured by personal illiteracy but relevant for calculating or knowing one's age. In cross sections like the DHS data, the time trend that might capture such effects is indistinguishable from an aging effect. The third model in Table 4 therefore adds

[20] See the discussions in Kaiser and Engle (1993) and Ewbank (1981).
[21] See Retherford and Mishra (2001) and Narasimhan et al. (1997).

dummies for age below 30 and age greater than or equal to 40 (thereby taking the 30s as the reference group).

Both personal and regional literacy continue to exert a substantial impact in the expanded model. Personal illiteracy raises the probability of reporting a "heaped" age by 2.1%. Regional literacy has a similarly powerful effect; starting from the sample mean, increasing the illiteracy rate by 20 percentage points raises the probability of any individual (literate or not) reporting a heaped age by some 2.4%. Finally, the age-group effects indicate that the probability of reporting a heaped age rises with age itself, by 2.5% from the 20s to the 30s, by a further 1.4% in the 40s and above. Experimentation with alternative specifications confirms the impression that the largest increase in age-heaping occurs in the 30s, which might suggest cultural influences or a recent expansion of schooling, rather than age-induced forgetfulness. The individual-level analysis thus confirms that age-heaping reflects a number of different factors, all of interest to economists: personal human capital, social capital in the form of record keeping, awareness of the calendar, age-specific rights and responsibilities, or generalized education, as well as cultural factors specific to particular societies.

V. Age heaping and illiteracy in the past

Are the patterns identified in Section IV for contemporary developing countries also to be found in historical data? We seek an answer to this question in U.S. census data from the nineteenth century, specifically the Integrated Public Use Micro Samples (IPUMS) of the censuses of 1850, 1870, and 1900.[22] Records for nearly 650,000 men and

[22] The individual records for 1890 do not survive, while the age data from 1880 and 1910 are given in ten-year age intervals for some states and nationalities. The IPUMS data are available online from the Minnesota Population Center at the University of Minnesota.

women aged 20-69, with data on race, age, literacy, and birthplace were extracted from the IPUMS databases. Given the difficulty of precisely estimating age heaping in small samples, only the 83 American states and mostly-European foreign birthplaces with at least 100 observations were used. In the early censuses, literacy was defined simply as the ability to read and write any language. The absence of any reference to proficiency makes illiteracy rates lower than in the DHS data.

As shown in Figure 6, age heaping and illiteracy vary quite widely across birth regions with at least 100 observations, from less than one percent among the white populations of several Northeastern states to values in excess of 75% among the former slaves of the American South in the 1870 census. Age-heaping measured by the Whipple Index ranges from lows in the vicinity of 100 to above 220 for the black populations of Southern U.S. states and for Ireland, Mexico, and parts of the American Southwest. Also evident in Figure 6 is the clear positive correlation of the two human capital measures among native whites, blacks and immigrants.

Figure 6 about here.

Regression analysis reveals patterns that vary only slightly over time and between groups. The estimated effects of birth region illiteracy on age heaping are strong, positive, and statistically significant in all cases (Table 5). Indeed, the relationship is stronger than in the DHS data. For comparability, the semilog specification was retained although there is little evidence of nonlinearity in the IPUMS data. With two exceptions, point estimates of the slope cluster around 0.008, a value clearly greater than the figures near 0.006 for the Whipple Index in Table 3. Also greater are the regression R^2s in the census data, indicating a less noisy relationship in historical than in contemporary data.

The estimated slopes imply that a one standard deviation increase in illiteracy predicts an increase in age heaping of from .7 to .9 standard deviations, depending on the sample in question. In two cases, coefficient estimates are lower: blacks in 1870 and native whites in 1900. In both cases this can be attributed to sample composition.[23] It is worth noting that the correlation of illiteracy and age heaping is quite robust, emerging also in pooled and region fixed-effects models.

Table 5 about here.

Individual data permit a more detailed analysis of the determinants of age heaping. Table 6 reports for the IPUMS data the marginal effects estimated from regressions like those underlying Table 4 (DHS data), in which the probability of reporting a multiple of five age is modeled as a logistic function of personal illiteracy, the regional illiteracy rate, and age group. Dummy variables are added for women (both genders being present in this sample), and for the Irish. Irish immigrants constitute 6% of the sample, trailing only New York, Pennsylvania, and Ohio among birthplaces. They present high values of illiteracy, but *very* high values of age heaping relative to the rest of the sample. (African-American populations display the reverse pattern after the Civil War: high values of age heaping, but *extremely* high illiteracy rates, relative to the sample average.)

---

[23] In the case of African-Americans, the slope declines as former slaves in the Southern states enter the sample in 1870. (In the 1850 census, slaves were not individually enumerated, so the IPUMS data include only free blacks.) Values of both illiteracy and age heaping among the former slaves are initially bunched around extremely high values, without displaying a very systematic relationship. By 1900, significant differentiation among freed black populations emerges, conforming well to the typical log-linear relationship. As for native whites, one third of the decrease in slope in 1900 is due to the addition of South Dakota to the sample, another third to other Midwestern and Western states.

The estimates indicate that the probability of reporting a heaped age depends on both personal illiteracy and the illiteracy rate in the region of birth, in all censuses and for all groups. And the order of magnitude of the estimated effects is quite similar to that in the contemporary DHS data. For native whites, regional illiteracy is initially the more powerful factor; in 1850 personal illiteracy raises the probability of reporting a heaped age by less than 1%, while a 20 point increase in regional illiteracy raises the same probability by 4%. Even a literate person in a highly illiterate society is not so unlikely to report a heaped age, then. By 1900, it is instead personal illiteracy that dominates, raising heaping probabilities by over 3% (exactly the excess frequency observed for ages divisible by five), while a 20 percentage point increase in illiteracy raises this probability by only 2%. African-Americans show a similar trend in the relative strength of the two effects, though both are more powerful than among whites in 1900. Among native-born Americans, then, age heaping comes more and more to indicate personal than social circumstances. Among immigrants, by contrast, the evolution is in the opposite direction, making generalization difficult. In some other respects there is a measure of convergence, both within and across groups. Among both native- and foreign-born whites, initial male-female differences diminish to statistical and meaningful insignificance by 1900. And in all three groups, clear differences in heaping propensity by age diminish markedly. In this respect convergence between groups is also evident; the ageing – age heaping effect is initially much stronger in the immigrant and African-American populations, and among them it diminishes most strongly. This is again consistent with age heaping reflecting less and less the individual's environment and demographic characteristics, more and more his or her individual human capital.

Table 6 about here.

VI. Human capital accumulation in the long run

The evidence indicates that age heaping is well correlated with literacy, at both the aggregate and individual level, in both contemporary and recent historical periods, within countries around the world and between them. In addition to its inherent interest as an indicator of numeracy, age heaping can therefore also be taken as a proxy for literacy. In both senses it is a measure of human capital. But age data are available for many times and places where other measures such as literacy are not. Age heaping can be estimated from any documentary or archaeological source yielding a sufficiently large sample, from tombstones to marriage registers, necrologies, legal records, muster lists, and of course census data. To illustrate age heaping's potential as culture- and time-spanning index of human capital, we calculate values of the Whipple Index from published and archival data from several countries, over a period running from classical antiquity up through the nineteenth century (Figure 7).

Figure 7 about here.

The data for Roman times are drawn from Duncan-Jones (1990), who calculated the prevalence of multiples of five among tombstone ages, separately for men and women and for twelve provinces of the Empire. The archaeological sources date primarily from the Principate (roughly 0-200 C.E.), to a lesser extent from the subsequent period. Tombstone data may be subject to bias in age heaping, but the direction is unclear. On the one hand, primarily wealthier families would have been able to afford funerary monuments; on the other, ages may have been estimated by relatives or religious authorities. In Figure 7, Duncan-Jones' index of rounding has been converted to a

Whipple Index by a simple linear transformation and averaged over men and women.[24] The figures for Italy, France (Gaul) and Germany are tightly clustered in the 250-300 range, while that for the Alpine region is something of an outlier (with respect to other provinces not shown in Figure 7 as well). The next available estimate is for one of medieval Europe's most advanced regions, Tuscany. Our calculation covers both sexes and ages 23-62, and is based on data from the celebrated Florentine cadastre of 1427 published by Herlihy and Klapisch-Zuber (1978). Within Tuscany the Whipple Index varies from 189 for males in the city of Florence to 314 for women in the smaller towns and rural areas. The overall average of 287 shows no meaningful change from Italian values in the Roman Empire a millennium before.

Moving forward a century to the years around 1500 reveals continued stagnation in human capital levels. The apparent increase in Italian age heaping is in fact probably a regional difference, for the later figure is an average for the southern towns Pozzuoli and Sorrento (modern Campania), based on census data reported in Duncan-Jones.[25] Though data specific to Roman Campania are lacking, the Renaissance value of 378 exceeds that of all Roman provinces but the Alpine region, making any sort of improvement since the earlier period highly unlikely. The German estimates for this period refer to the Southwest region (modern Württemberg). They are based on new archival data exploited

[24] The figures for Italy and the Alpine region are simple averages of constituent provinces: Rome and Italy outside Rome in the first case, the provinces of Noricum and Pannonia in the second. Aggregation is advisable where possible due to likely small sample sizes underlying Duncan-Jones' figures. Noricum is in modern Austria. No data are available for Raetia to its west in modern Switzerland. Pannonia, to its east, is in modern Hungary, hence no longer geographically in the Alps. In Roman times, however, the two provinces were ethnically similar and economically integrated.

[25] The Pozzuoli census was conducted in 1489, that of Sorrento in 1561. Geography rather than chronology dictated averaging Pozzuoli with Sorrento rather than Tuscany (1427). The original source for the Southern data is Beloch (1937).

here for the first time: the population muster rolls (an early type of census) of 1523 from the town of Balingen.[26] Whipple Index values in the range of 310-330 are on the high side of Roman provincial figures – about the same as the index for males in Roman Germany.[27] In central Europe, too, human capital levels of Roman times were at best maintained, or perhaps recovered after an intervening decline.

At some point in the early modern period, age heaping in Western Europe begins to diminish. In France and in the neighboring Alpine region, the decline is early and gradual. The Alps are represented here by Geneva, for which estimates based on death reports are given in Duncan-Jones (1990). Between the mid-sixteenth and the mid-eighteenth centuries, the Whipple Index value falls steadily from over 300 to under 200, a level decisively below anything seen in medieval Europe or Roman antiquity. Early eighteenth century France has an even lower degree of age heaping, according to Duncan-Jones' Paris death report data. And this low incidence of age heaping seems to have been characteristic of France already in the late seventeenth century, judging by preliminary results from other datasets.[28] Census data from the nineteenth century show

[26] The Württemberg data are for ages 23-52. The authors gratefully acknowledge Lisbeth Zahawi for providing these data from the Balingen City Archive.

[27] Figure 7's relatively low Whipple Index for Roman Germany (256) results from averaging the male figure of 329 with an astonishingly low female estimate of 183, based on a sample of unknown size. The apparent increase of the index may therefore be spurious.

[28] Among 3,700 French soldiers born between 1650 and 1700, and aged from 23 to 52, the Whipple Index is a very low 153. The data are a military sample made available by John Komlos at www.uni-tuebingen.de/uni/wwl/dhheight.html. For a full analysis of age heaping in this dataset, see Crayen and Baten (2006).

age heaping in France to have diminished even further among the birth cohorts around 1800, and to have fallen to negligible levels by the mid-1800s.[29]

Other countries in the sample display somewhat different patterns, though sharing the general downward trend. In Germany, the decrease in age heaping is later and more rapid. German estimates for the decades around 1700 are for the town of Lippborg in western Germany (near the Ruhr). The data from the family and soul registers preserved in the city archive are analyzed using age heaping techniques for the first time here.[30] In the space of a few decades, age heaping in Lippborg fell from values around 300, typical of both late medieval and Roman Germany, to decisively lower levels near 200. This rapid fall must have continued, for by the time census data become available for the birth cohorts of the mid-nineteenth century, age heaping has all but disappeared.

In the U.S. and Italy, the decrease in age heaping appears to have followed a two-step path. Both countries arrive in the early nineteenth century with Whipple Indices clearly lower than anything observed in medieval or Renaissance times, but clearly higher than figures for France and Germany. In Italy we cannot observe the transition; for the U.S. a Whipple Index in excess of 200 can be conjectured for the mid-seventeenth century, which would put it in the same range as Germany, the Alpine region, and perhaps France.[31] Both Italy and the U.S. then display a sharp reduction in age heaping

[29] Census data for France, Germany, and Italy are all drawn from Rothenbacher (2002) and the accompanying CD. Whipple Indices are calculated for ages 23-62, where possible averaging values for the same birth cohort observed in different censuses to avoid confounding cohort effects with aging effects.

[30] The Whipple Index for Lippborg is based on ages 23-72. The authors express their thanks to M.Thérèse Haemers-Van Roey Neerpelt for kindly making these data available.

[31] Fischer (1977), p. 85 reports ratios of frequencies at multiples of ten to a 10-year moving average frequency, for a sample of some 4,000 individuals in Essex County,

during the mid-nineteenth century, converging on Whipple Index values already achieved in France and Germany.

The latest and most rapid decrease in age heaping is found in Russia. Age distributions covering birth cohorts of the late 1600s have been published by Kaiser and Engel (1993) for two cities: Tula in central Russia and Viatka (Kirov today) in the Northeast. Clear differences are evident between long-settled Tula, where the Whipple Index averages 257 over two enumerations, and frontier Viatka, where the analogous figure is 342. The average for the two cities reported in Figure 7 is 299, which is similar to medieval and Renaissance values found in other sample countries, clearly above early-declining France and the Alpine region, and in line with only the earliest Lippborg figures. By the nineteenth century Russia appears to lag behind the rest of the sample by two centuries.[32] For birth cohorts of the 1820s and ’30s, among whom age-heaping has been largely eliminated in France and Germany, and is low (values around 150) in Italy and the U.S., it is still in excess of 200 in Russia. Improvement, when it does come to Russia, is revolutionary. By the birth cohort of the 1860s, the Whipple Index falls to 129, much closer to (if still above) other sample countries (where the Whipple Index varies from 100 in France to 115 in the U.S.).

This survey only scratches the surface of historical age data available in published sources and archives. And space precludes us here from investigating age heaping’s relationship with the few alternative measures of human capital available before modern times. But even this illustrative example suggests some fascinating hypotheses. Medieval

---

Massachusetts, in the years 1636-72. The same index calculated for the Tula data of 1715 yields very similar values. Tula in 1715 has a Whipple Index of 223.

[32] All-Russian figures were calculated on the basis of materials from the first modern census of the Empire, conducted in 1897, located in the INION Library, Moscow.

and Renaissance Europe appears to have had the same level of numeracy as in Roman antiquity over a millennium before. The lack of improvement is startling. But if we hypothesize some decline in numeracy skills during the Dark Ages, the implication is that the medieval economic recovery was preceded by or contemporaneous with a parallel recovery in human capital levels. It is worth noting that seventeenth century Russia does not look out of place in European company, falling clearly behind only after this point. Well documented by age heaping is the considerable improvement in human capital levels *preceding* the onset of industrialization and modern economic growth. With different timing and intensity, age heaping levels fell dramatically over the early modern period in all sample countries but Russia. This lends strong support to accounts that assign human capital the role of cause, rather than effect, of the industrial revolution.[33]

VII. Summary and directions for future research

Age heaping has the potential to broaden and deepen the empirical foundation for research on human capital and its role in economic development. It is both a complement and a substitute for literacy-based measures. It complements literacy by revealing different cognitive abilities – abilities that are equally important in determining individual labor market outcomes today, and which arose in tandem with the market economy historically. It is also complementary in the sense of yielding insights not only about individuals but also about the societies in which they live. Age heaping can be a substitute or proxy for (il)literacy, as the two are well correlated between and within countries, in aggregate and individual data, in contemporary and historical contexts.

[33] It parallels developments in literacy in Britain, where signature rates among brides and grooms improved slowly but steadily from 1500 to 1750, reaching about 50%. They then increased by only about another 10 percentage points over the following century of radical economic transformation. See Mitch (1991), pp. 1-4.

Though tenuous at very low levels of illiteracy and age heaping, this correlation is remarkably robust over the full range of observed variation in contemporary and historical data. The type of age heaping that predominates in historical data on European adults (on multiples of five) can be accurately measured using a computationally simple, intuitive indicator: the Whipple Index. With large samples (1,000 or more), even small differences in the Whipple Index reliably reflect differences in the underlying populations. Even with samples as small as 250, differences of the magnitude frequently observed in the historical record (e.g. 250 vs. 150) very rarely result from random variation and can be considered reliable indicators of population differences.

Our preliminary exploration of age heaping as a measure of human capital illustrates the breadth and power of the approach. It yields comparable estimates from data that range widely in sources (tombstones, tax records, death records, and census data), over time (from antiquity to the turn of the twentieth century), and across space (from Russia to North America). It offers decisive support for the hypothesis that an increase in human capital levels preceded the industrial revolution. The next stages in the development and implementation of the method include: a more comprehensive survey of potential data sources and expansion of the database; an investigation of whether age-heaping or literacy is more closely correlated with measures of economic development; and the extension of the study to contexts in which literacy measures are completely absent or inherently more problematic, as with the ideographic writing systems of East Asia.

References

Allen, R.C., 2003. Progress and poverty in early modern Europe. Explorations in Economic History 56 (3), 403-443.

Bachi, R., 1951. The tendency to round off age returns: measurement and correction. Bulletin of the International Statistical Institute 33, 195-221.

Barro, R.J., 1997. Determinants of Economic Growth : A Cross-country Empirical Study. MIT Press, Cambridge MA.

Barro, R.J., Lee, J.-W., 1994. Sources of economic growth (with commentary). Carnegie-Rochester Conference Series on Public Policy 40, 1-57.

Baten, J., van Zanden, J.L., 2006. The impact of human capital on early modern welfare growth, 1450-1913: book production as a human capital indicator. University of Tübingen Working Paper.

Becker, G.S., Murphy, K.M., Tamura, R., 1990. Human capital, fertility, and economic growth. Journal of Political Economy 98, 12-37.

Beloch, K.J., 1937. Bevölkerungsgeschichte Italiens I. Walter De Gruyter, Berlin.

Bils, M., Klenow, P.J., 2001. Does schooling cause growth? American Economic Review 90 (5), 1160-1183.

Budd, J.W., Guinnane, T., 1991. Intentional age misreporting, age heaping and the 1908 old age pensions act in Ireland. Population Studies 45 (3), 497-518.

Carruthers, B.G., Espeland, W., 1991. Accounting for rationality: double-entry bookkeeping and the emergence of economic rationality. American Journal of Sociology 97 (1), 31-96.

Cervellati, M., Sunde, U., 2005. Human capital formation, life expectancy and the process of development. American Economic Review 95 (5), 1653-1672.

Charette, M.F., Meng, R.,1998. The determinants of literacy and numeracy, and the effect of literacy and numeracy on labour market outcomes. Canadian Journal of Economics 31 (3), 495-517.

Chiswick, B., Lee, Y.L., Miller, P.W., 2003. Schooling, literacy, numeracy and labour market success. The Economic Record 79 (245), 165-181.

Coale, A., Kisker, E.E., 1986. Mortality crossovers: reality or bad data? Population Studies 40, 389-401.

Cohen, P.C., 1982. A Calculating People: The Spread of Numeracy in Early America. University of Chicago Press, Chicago.

Crayen, D., Baten, J., 2006. Numeracy, inequality, age heaping, and economic growth: new estimation strategies for eestern Europe and the U.S. (17th - 19th centuries). University of Tübingen Working Paper.

Denic, S., Khatib, F., Saadi, H., 2004. Quality of age data in patients from developing countries. Journal of Public Health 26 (2), 168-171.

Duncan-Jones, R., 1990. Structure and Scale in the Roman Economy. Cambridge University Press, Cambridge.

Emigh, R.J., 2002. Numeracy or enumeration? The uses of numbers by states and societies. Social Science History 26 (4), 653-698.

Ewbank, D.C., 1981. Age Misreporting and Age-selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis. National Academic Press, Washington, D.C.

Finnie, R., Meng, R., 2001. Cognitive skills and the youth labour market. Applied Economic Letters 8, 675-679.

Fischer, D.H., 1977. Growing old in America. Oxford University Press, New York.

Galor, O., Weil, D.N., 2000. Population, technology and growth: from the Malthusian regime to the demographic transition. American Economic Review 110, 806-828.

Goldthwaite, R.A., 1972. Schools and teachers of commercial arithmetic in Renaissance Florence. Journal of European Economic History 1, 418-433.

Graff, H., 1987. The Legacies of Literacy: Continuities and Contradictions in Western Culture and Society. Indiana University Press, Bloomington IN.

Hanushek, E.A., Kimko, D.D., 2000. Schooling, labor-force quality, and the growth of nations. American Economic Review 90(5), 1184-1208.

Heckman, J.J., 1995. Lessons from the Bell Curve. Journal of Political Economy 103 (5), 1091-1120.

Herlihy, D., Klapisch-Zuber, D.,1978. Les Toscans et leurs familles: une étude du catasto florentin di 1427. Editions de l'Ecole des Hautes Etudes en Sciences Sociales / Presses de la Fondation Nationale des Sciences Politiques, Paris. Chap. XIII.

Jowett, A.J., Li, Y-Q., 1992. Age heaping: contrasting patterns from China. GeoJournal 28 (4), 427-442.

Kaiser, D., Engel, P., 1993. Time and age-awareness in early modern Russia. Comparative Studies in Society and History 35 (4), 824-839.

Levine, R., Renelt, D., 1992. A sensitivity analysis of cross-country growth regressions. American Economic Review 82, 942-963.

Long, J., (2005). Rural-urban migration and socioeconomic mobility in Victorian Britain. Journal of Economic History 65 (1), 1-35.

Long, J., (forthcoming). The economic return to primary schooling in Victorian England. Journal of Economic History.

Lucas, R., 2002. Lectures on Economic Growth. Harvard University Press: Cambridge, MA. Ch. 5.

Mankiw, N.G., Romer, D., Weil, D.N., 1992. A contribution to the empirics of economic growth. Quarterly Journal of Economics 107 (2), 19-36.

Mitch, D., 1991. The Rise of Popular Literacy in Victorian England. University of Pennsylvania Press, Philadelphia.

Mokyr, J., 1983. Why Ireland Starved. George Allen & Unwin, London. Chap. 8.

Murnane, R.J., Willet, J.B., Levy, F., 1995. The Growing importance of cognitive skills in wage determination. Review of Economics and Statistics 77(2), 251-266.

Murphy, K. M., Shleifer, A.,Vishny, R.W., 1991.The allocation of talent: implications for growth. Quarterly Journal of Economics 106 (2), 503-530.

Myers, R., 1976. An instance of reverse heaping of ages. Demography 13 (4), 577-580.

Myers, R., 1954. Accuracy of age reporting in the 1950 United States census. Journal of the American Statistical Association XLIX, 826-831.

Narasimhan, R. L., Retherford, R.D., Mishra, V., Arnold, F., Roy, T.K., 1997. Comparison of fertility estimates from India's Sample Registration System and Family National Health Survey. National Family Health Survey Subject Reports 4. IPPS, Mumbai.

Neal, D.J., Johnson, W.R., 1996. The role of pre-market factors in black-white wage differences. Journal of Political Economy 104 (5), 869-895.

O'Grada, C., 2006. Dublin Jewish demography a century ago. UCD Centre for Economic Research Working Paper Series WP 06/01.

O'Rourke, K., Williamson, J.G., 1997. Around the European periphery 1870-1913: globalization, schooling and growth. European Review of Economic History 1, 153-190.

Parsons, S., Bynner, J., 2005. Does Numeracy Matter More? National Research and Development Centre for Adult Literacy and Numeracy, London.

Preston, S.H., Elo, I.T., Rosenwaike, I., Hill, M., 1996. African-American Mortality at Older Ages: Results of a Matching Study. Demography 33 (2), 193-209.

Reis, J., 2005. Economic growth, human capital formation and consumption in western Europe before 1800. In: Allen, R.C., Bengtsson, T., Dribe, M. (Eds.), Living Standards in the Past. Oxford University Press, Oxford, pp. 195-225.

Retherford, R.D., Mishra, V., 2001. An evaluation of recent estimates of fertility trends in India. National Family Health Survey Subject Reports 19. IPPS, Mumbai.

Rivera-Batiz, F., 1992. Quantitative literacy and the likelihood of employment among young adults in the Unites States. Journal of Human Resources 27, 313-328.

Rothenbacher, F., 2002. The European Population, 1850-1945. Palgrave Macmillan, New York.

Sala-i-Martin, X., 1997. I just ran 4 million regressions. National Bureau of Economic Research Working Paper no. 6252.

Sanderson, M., 1999. Education and Economic Decline in Britain, 1870 to the 1990s. Cambridge University Press, Cambridge.

Shryock , H.S., Siegel, J.S., 1973. The Methods and Materials of Demography. U.S. Bureau of the Census, Washington, D.C. Chap. 8.

Swetz, F.J., 1987. Capitalism and Arithmetic: The New Math of the 15th Century. Open Court Publishing Company, La Salle.

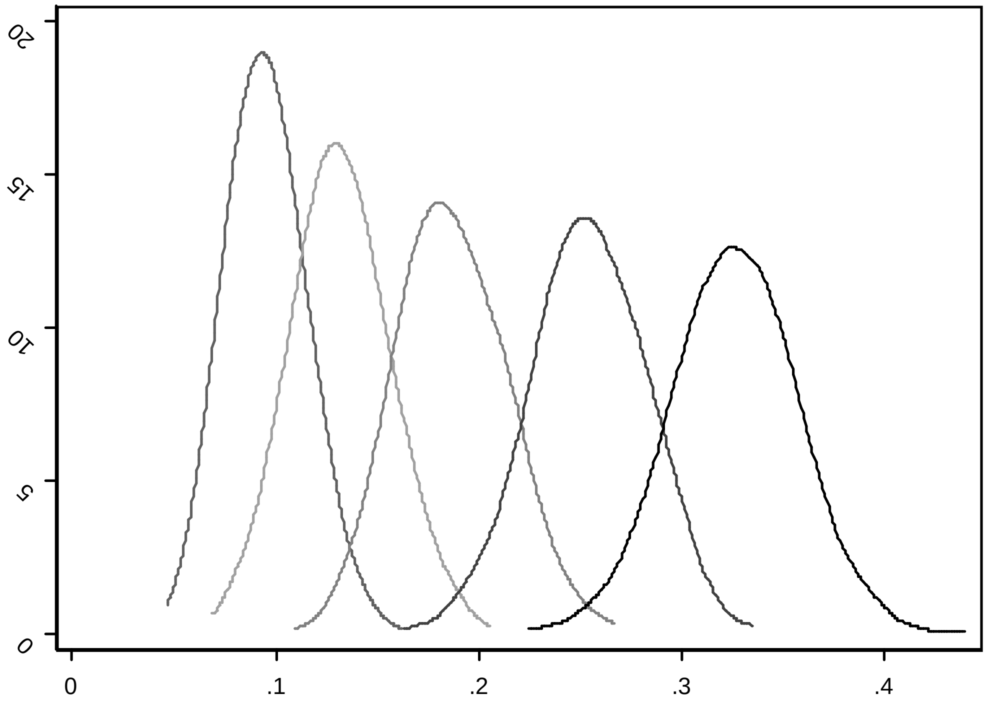
Tortella, G., 1994. Patterns of economic retardation and recovery in south-western Europe in the nineteenth and twentieth centuries. Economic History Review 47 (1), 1-21.

United Nations Statistics Division, 2003. Nonsampling Errors in Surveys. United Nations Secretariat, New York.

Vallin, J., Meslé, F., Adamets, S., Pyrozhkov, S., 2002. A new estimate of Ukrainian population losses during the crises of the 1930s and 1940s. Population Studies 56 (3), 249 – 264.

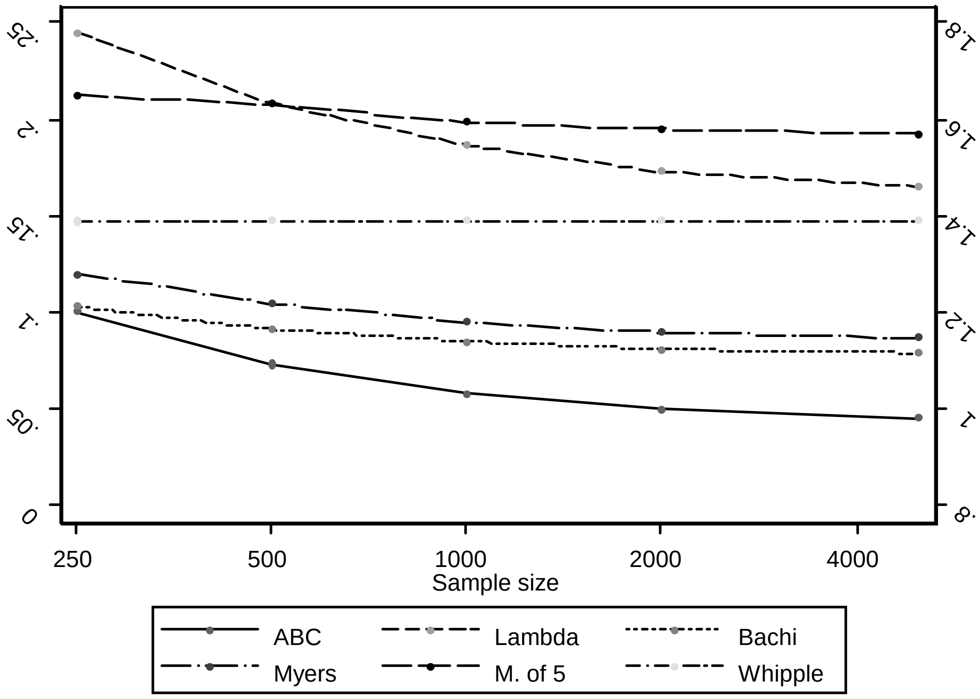
FIGURES

Figure 1. Distribution of the Lambda Index under increasing degrees of mixed heaping



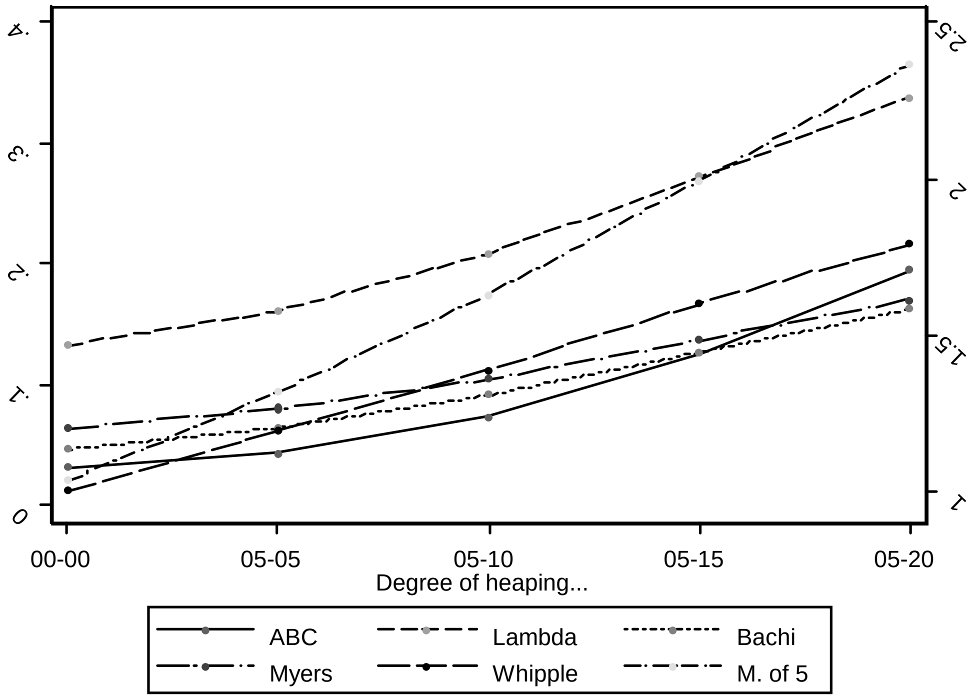
Kernel density estimates (Epanechnikov kernel, bandwidth .01); mixed heaping schemes as described in text; sample size 1000; repetitions 1000.

Figure 2. Scale dependence under mixed heaping



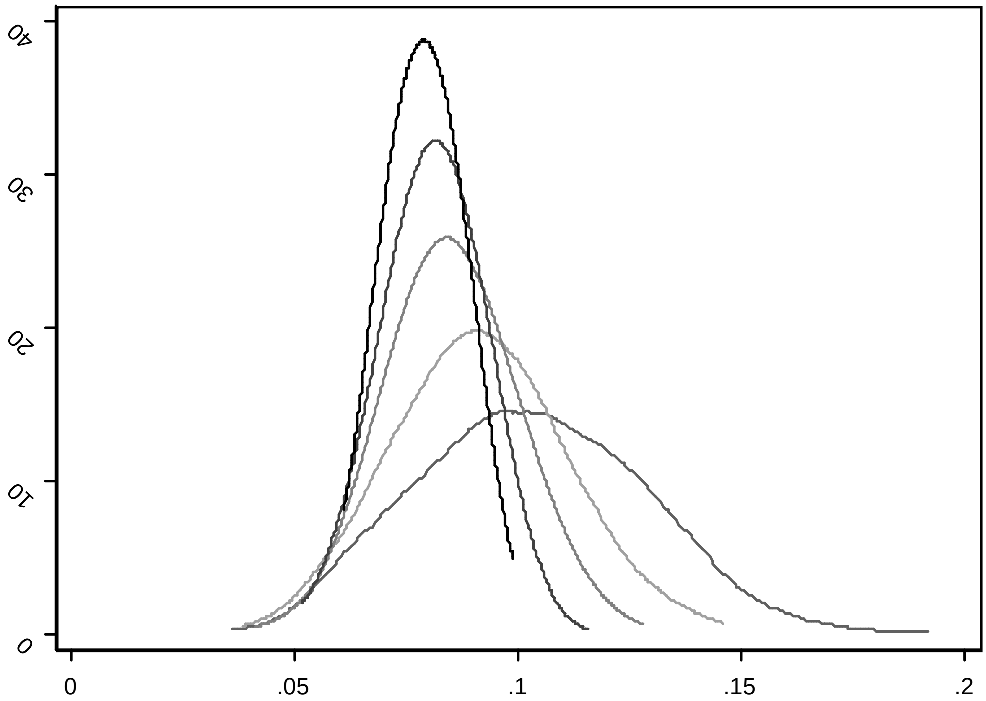
Mean values of indices under mixed heaping scheme 05-10 for indicated sample sizes;
Multiples of 5 and Whipple indices plotted against right axis; log scale x-axis; 1000 repetitions.

Figure 3. Index response to degree of heaping



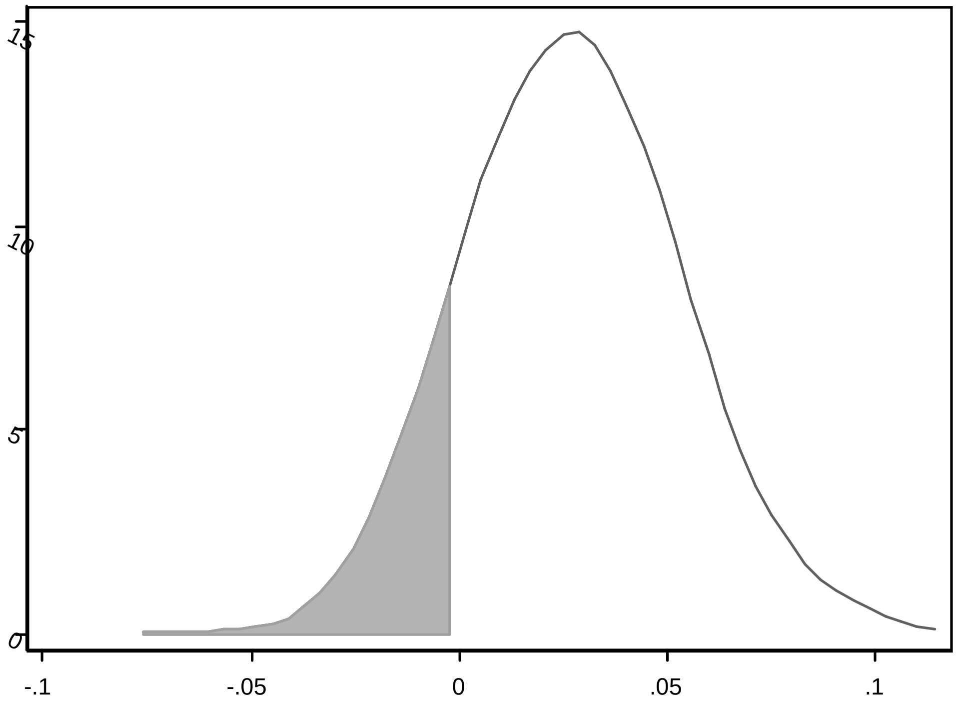
Mean values of indices under mixed heaping schemes as indicated on horizontal axis and described in text; sample size 500; Multiples of 5 and Whipple Indices plotted on right axis; 1000 repetitions.

Figure 4. Distribution of the Bachi Index for various sample sizes



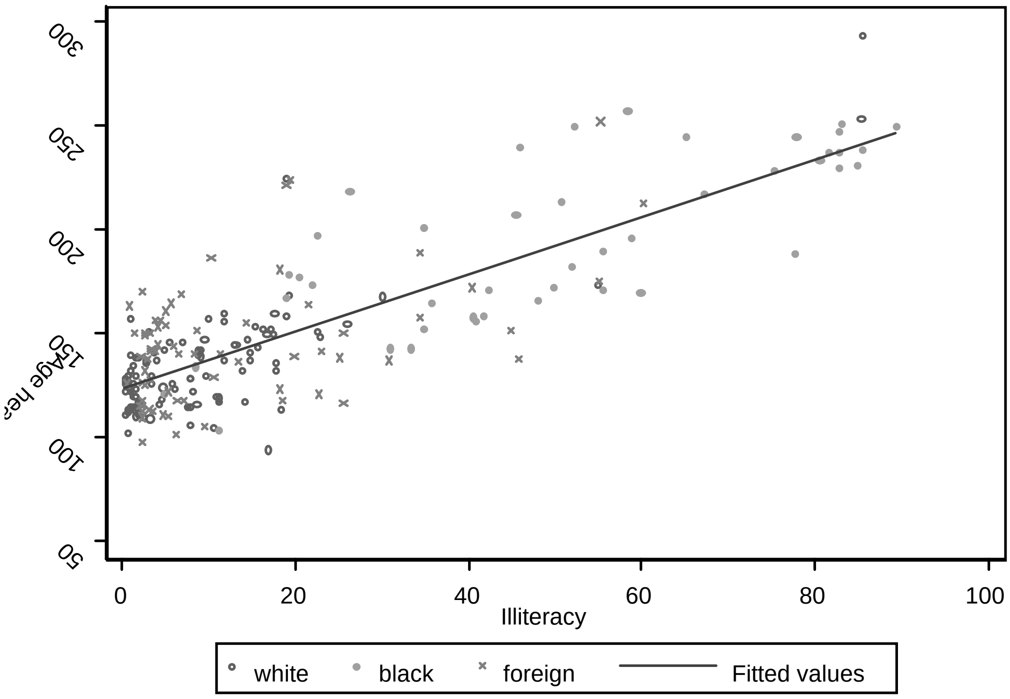
Kernel density estimates (Epanechnikov kernel, bandwidth .075); Sample sizes 250, 500, 1000, 2000, 5000; Mixed heaping type 05-10.

Figure 5. Bachi Index probability of ranking error



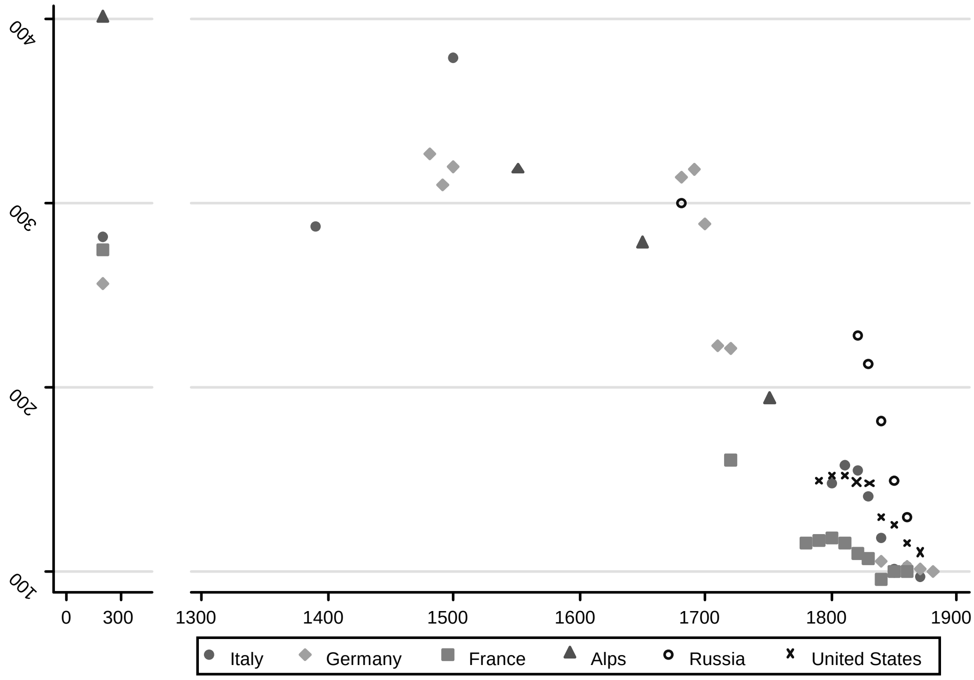
Kernel density estimate of the distribution of difference between Bachi Indices for random samples from 05-10 and 05-05 populations (Epanechnikov kernel, bandwidth .01); Sample size 500.

Figure 6. Age heaping and illiteracy in three U.S. censuses



Whipple Index and illiteracy rate by birth regions (US states and territories, foreign countries and provinces) with at least 100 observations; IPUMS data.

Figure 7. Age heaping in the long run



Notes: Whipple Index values, combined male and female data; time refers to approximate birth decade. *Sources*: Duncan-Jones (1990) pp. 86, 90; Herlihy and Klapisch-Zuber (1978) pp. 656-59; City archive of Balingen, Musterungsliste 1523, A 28a M21; City archive of Lippborg, Catalogus Familiarum ... parochiae Libborgensis de dato 20. Martii 1750; IPUMS; Rothenbacher (2002); Kaiser and Engel (1993) pp. 829-33; Census of the Russian Empire, 1897.

TABLES

Table 1. Probabilities of ranking errors, adjacent degrees of mixed heaping

	Sample size 250					*Sample size* 1000			
Index	*D1*	*D2*	*D3*	*D4*		*D1*	*D2*	*D3*	*D4*
ABC	.37	.27	.18	.20		.11	.04	.03	.04
Lambda	.36	.27	.22	.24		.11	.06	.04	.04
Bachi	.37	.24	.22	.22		.08	.05	.03	.03
Myers	.37	.31	.24	.24		.14	.10	.05	.07
M. of 5	.23	.24	.24	.30		.05	.09	.09	.12
Whipple	.16	.16	.17	.19		.02	.02	.02	.04

Estimated probabilities that *D1* etc. < 0; $D1 = H^{05-05} - H^{00-00}$, $D2 = H^{05-10} - H^{05-05}$ etc.

Table 2. Probabilities of ranking errors, two degrees of mixed heaping apart

	Sample size 250					*Sample size* 1000			
Index		*D20*	*D31*	*D42*			*D20*	*D31*	*D42*
ABC		.17	.06	.04			.00	.00	.00
Lambda		.17	.08	.05			.00	.00	.00
Bachi		.15	.08	.04			.00	.00	.00
Myers		.20	.10	.07			.00	.00	.00
M. of 5		.09	.08	.10			.00	.01	.01
Whipple		.03	.02	.03			.00	.00	.00

Estimated probabilities that *D20* etc. < 0; $D20 = H^{05-10} - H^{00-00}$, $D31 = H^{05-15} - H^{05-05}$ etc.

Table 3. Regressions of (ln) age heaping on illiteracy

	National data		*Regional data*		*Country fixed-effects*	
	Bachi	Whipple	Bachi	Whipple	Bachi	Whipple
Regional	0.0082	0.0061	0.0071	0.0055	0.0075	0.0044
illiteracy	(0.002)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Constant	-2.731	4.499	-2.681	4.515	-2.700	4.568
	(0.068)	(0.026)	(0.025)	(0.013)	(0.042)	(0.024)
N	52	52	415	415	415	415
R^2	0.35	0.49	0.31	0.44	0.74	0.79
R^2 "within"					0.19	0.19

Notes: Robust standard errors in parentheses (ordinary s.e.'s in fixed-effects model); estimation by OLS; dependent variable is logarithm of age heaping; USAID/DHS national and regional data for women aged 23-49.

Table 4. Marginal effects on the probability of reporting a heaped age

			Illiterates	*Literates*	
Personal	0.0572	0.0271			0.0208
illiteracy	(0.005)	(0.006)			(0.001)
Regional		0.0011	0.0016	0.0006	0.0012
illiteracy		(0.000)	(0.000)	(0.000)	(0.000)
Age <30					-0.0253
					(0.001)
Age >= 40					0.0139
					(0.001)
mean dep. var.	0.257	0.257	0.286	0.230	0.257
N	701,104	701,104	352,854	348,250	701,104
χ^2	2980.6	5404.0	2592.7	316.2	6240.7

Notes: marginal effects of coefficient estimates from logistic regression, evaluated at means of all variables; robust standard errors; χ^2 statistic refers to Wald test of model significance with degrees of freedom equal to the number of explanatory variables (p-values are 0.000 in all cases); USAID/DHS individual data on adult (20-49) women.

Table 5. Regressions of (ln) age heaping on illiteracy

	1850			1870			1900		
	native white	*for.*	*native black*	*native white*	*for.*	*native black*	*native white*	*for.*	*native black*
Regional illiteracy	.0089 (.001)	.0085 (.001)	.0081 (.003)	.0083 (.001)	.0082 (.001)	.0043 (.001)	.0052 (.002)	.0078 (.001)	.0083 (.002)
Constant	4.867 (.014)	5.061 (.027)	5.058 (.124)	4.882 (.016)	4.934 (.019)	5.105 (.035)	4.735 (.015)	4.704 (.024)	4.726 (.068)
N	25	9	7	34	30	20	42	29	17
R^2	0.867	0.741	.682	0.650	0.458	.726	0.253	0.638	.804

OLS regression of ln(age heaping) on illiteracy for birthplaces with samples of at least 100; robust standard errors in parentheses; IPUMS data, ages 20-69.

Table 6. Marginal effects on the probability of reporting a heaped age

	1850			1870			1900		
	Native white	*For.*	*Native black*	*Native white*	*For.*	*Native black*	*Native white*	*For.*	*Native black*
Personal illiteracy	.0081 (.006)	.0584 (.013)	.0070 (.024)	.0368 (.005)	.0573 (.008)	.0272 (.007)	.0321 (.006)	.0285 (.008)	.0624 (.007)
Regional illiteracy	.0021 (.000)	.0019 (.001)	.0020 (.001)	.0021 (.000)	.0021 (.000)	.0024 (.000)	.0010 (.000)	.0021 (.000)	.0015 (.000)
Female	.0088 (.003)	-.0134 (.008)	-.0130 (.023)	.0153 (.003)	-.0051 (.005)	.0627 (.005)	.0023 (.002)	.0026 (.004)	.0150 (.007)
Age < 30	-.0631 (.004)	-.1655 (.009)	-.1782 (.027)	-.0653 (.003)	-.1388 (.006)	-.2108 (.006)	-.0165 (.003)	-.0573 (.006)	-.0860 (.008)
Age $\geq$ 40	.0315 (.004)	.0732 (.010)	.0982 (.028)	.0073 (.003)	.0727 (.005)	.1048 (.006)	.0174 (.003)	.0287 (.005)	.0739 (.009)
Irish		.1132 (.012)			.1231 (.007)			.0617 (.006)	
Mean dpvr.	.282	.369	.431	.271	.341	.473	.231	.247	.329
N	73,381	15,485	2,076	124,304	45,172	43,145	136,341	42,241	20,623
Pseudo-R^2	0.009	0.048	0.047	0.007	0.046	0.061	0.002	0.011	0.027
χ^2 LR	768.1	987.6	134.7	1059.1	2717.6	3632.3	252.6	535.3	701.0

Marginal effects of estimated coefficients of logistic regression, evaluated at means of all variables; IPUMS data on adults aged 20-69; degrees of freedom for the chi-squared likelihood ratio test are 5 or 6 depending on the model, but in all cases the p-value is zero to three decimal places.