Aging, cognitive abilities and retirement*

Fabrizio Mazzonna
Tor Vergata University

Franco Peracchi
Tor Vergata University and EIEF

August 2009
This version: November 6, 2010

Abstract

We investigate the relationship between aging, cognitive abilities and retirement using the Survey on Health, Aging and Retirement in Europe (SHARE), a household panel that offers the possibility of comparing several European countries using nationally representative samples of the population aged 50+. We use a version of the model proposed by Grossman (1972) as a guide for our empirical specification of the age-profile of cognitive abilities. According to the model, retirement plays a fundamental role in explaining the process of cognitive deterioration. Our empirical results confirm this key prediction. They also indicate that education plays a fundamental role in explaining heterogeneity in the level of cognitive abilities.

Keywords: Aging; cognitive abilities; retirement; education; SHARE.

JEL codes: J14, J24, C23.

* Corresponding author: Franco Peracchi, Faculty of Economics, Tor Vergata University, via Columbia 2, 00133 Rome, Italy (franco.peracchi@uniroma2.it). We thank Josh Angrist, Anne Case, Angus Deaton, Maria Glymour, Giovanni Mastrobuoni, and Chris Paxson for insightful discussions. We also thank participants to the 2009 American Economic Association Meeting, the 2009 Congress of the European Economic Association, the VIII Brucchi-Luchino workshop, and seminar participants at Collegio Carlo Alberto, Princeton University, Universitat Autònoma de Barcelona, and the University of Zurich for useful comments. Finally, Franco Peracchi thanks the Center for Health and Wellbeing at Princeton University for generous hospitality during Fall 2010.

1 Introduction

For many countries, aging is one of the great social and economic challenges of the 21st century. In Europe, for example, the ratio of persons aged 65 and over as a percentage of the population aged 18–65 (the old-age dependency ratio) is expected to increase from its current levels of 25 percent to about 50 percent in 2060 (Eurostat 2008).

A fundamental aspect of the aging process is the decline of cognitive abilities. Schaie (1989) shows that cognitive functions are relatively stable until the fifth decade of life. After this period, the decline becomes apparent and the incidence of cognitive impairments increases sharply with age. At all ages, however, there is large variation across individuals in the level of cognitive performance.

The age-related process of neurodegeneration is complex and its determinants are not yet well understood. One conceptual framework, due to Horn and Cattel (1967) and Salthouse (1985), distinguishes between two types of abilities. The first type, 'fluid intelligence', consists of the basic mechanisms of processing information which are closely related to biological and physical factors. One important aspect of these abilities is the speed with which many operations can be executed (Salthouse 1996). The second type, 'crystallized intelligence', consists of the knowledge acquired during the life with education and other life experiences. Unlike fluid intelligence, which is subject to a clear decline as people get older, crystallized intelligence tends to be maintained at older ages and is subject to a lower rate of age-related decline. As argued by Salthouse (1985), dimensions of cognitive functioning such as orientation, memory, fluency and numeracy, are generally based on different combinations of fluid and crystallized intelligence. This consideration suggests that accounting for the different dimensions of cognitive functioning may be important for the analysis of the process of neurodegeneration associated with aging.

Another conceptual framework, due to Stern (2002), is that individuals have different levels of cognitive reserve, and that higher levels allow them to prevent or slow down the process of neurodegeneration associated with aging. Individual heterogeneity in cognitive performance may reflect both genetic differences in the level of cognitive reserve and life events (individual choices or exogenous shocks) that may affect the cognitive endowment and the rate of age-related decline.

Recent research in neuroscience (see van Praag et al. 2000 for a review) has questioned the idea that age-related cognitive decline is inevitable and fixed. Although neural plasticity is reduced in old age, it remains more substantial than previously recognized. In a comprehensive review of cognitive-enrichment effects at old ages, Hertzog et al. (2008) describe how the age-profiles of cognitive abilities can differ over the life span in response to various type of enrichment behaviors of an individual ('cognitive-enrichment hypothesis'). As shown by several empirical studies, important factors in this process are education (Le Carret 2003), occupational choices (Adam et al. 2006), leisure activities (Scarmeas et al. 2003), home environment and parental influences in both childhood (Kaplan et al. 2001, Cunha and Heckman 2007, Case and Paxson 2009) and adolescence (Richards et al. 2004), social activities (Trouton et al. 2006), lifestyle (Cervilla et al. 2000), and chronic diseases like hypertension or heart disease (Meyer et al. 1999).

Most of this literature is descriptive, with only few efforts at interpreting the empirical evidence within a well-defined model. For instance, the popular 'use-it-or-lose-it hypothesis' (see for example Rohwedder and Willis 2010), by which intellectually engaging activities help buffer individuals against cognitive decline, does not explain individual differences in the time and effort allocated to these intellectually engaging activities (Stine-Morrow 2007). Further, empirical results are often based on small cross-sectional samples and cross-country comparisons are lacking. The few existing longitudinal studies do not account for sample selection due to panel attrition, a potentially serious problem in samples of older people.

There are at least two reasons why understanding the process of age-related decline in cognitive abilities is important to economists. First, cognitive abilities are fundamental for decision making, for they influence individuals' ability to process information and to make the right choices. As many countries have moved more towards systems of individual provision for retirement income, decision making ability is becoming a crucial element for the appropriate formulation of consumption and saving plans (Banks et al. 2007, Christelis et al. 2010).

Second, cognitive abilities may be regarded as one aspect of human capital, along with education, health, and non-cognitive abilities. Economists have focused their attention mainly on human capital accumulation, much less on human capital deterioration. As stressed by McFadden (2008), "natural questions to ask are how human capital at various stages in the life cycle can be measured [...]; the degree to which the depreciation of human capital components is an exogenous consequence of aging or can be controlled through work, study, and behavioral choices; and the degree to which depreciation is predictable or random". Recent attempts in this direction (Adam et al. 2006, Bonsang et al. 2010, Rohwedder and Willis 2010) lack a clear conceptual framework, which is not without consequences for their empirical analysis. For example, all these papers use a simple dummy variable for being retired, which implies no role for the length of time spent in retirement. Further, they all fail to control for differences by gender and education, and tend to adopt a "kitchen sink" approach by including a very long list of controls some of which, like health conditions, can hardly be treated as exogenous.

The standard human capital model (Schultz 1961, Ben-Porath 1967) offers two key insights. One is that the observed age-related decline in cognitive abilities need not be the same as natural deterioration, because people may respond to aging by investing in cognitive-repair activities. The other is that the amount of repair investment depends on market and non-market incentives, relative prices, discount rates, etc. Our paper follows the human capital approach by using a discrete time version of the model originally proposed by Grossman (1972) to generate predictions about the evolution of the stock of cognitive abilities over the life cycle. The simple version of the model that ignores the possible utility value of cognitive investment predicts that, after retirement, the cognitive stock should decline rapidly because there are no longer market incentives to invest in cognitive repair. Although this sharp prediction is lost in the full version of the model, we may still ask whether the data provide evidence of an increase in the rate of decline of the cognitive stock after retirement. A key issue is of course the potential endogeneity of the retirement decision, an issue that we address using an instrumental variables approach.

The empirical part of this paper is based on microdata from the the Survey of Health, Aging, and Retirement in Europe (SHARE), a large household panel which contains data on the individual life circumstances of about 30,000 individuals aged 50+ in eleven European countries, including measures of cognitive function based on simple tests of orientation in time, memory, verbal fluency and numeracy.

The remainder of this paper is organized as follows: Section 2 describes the data used for this study; Section 3 describes the model; Section 4 discusses features of the data that complicate identification of the causal effect of aging on cognitive abilities; Section 5 presents our results; Section 6 contains some robustness checks; finally, Section 7 offers some conclusions.

2 Data

Our data are from Release 2 of the first two waves (2004 and 2006) of the Survey of Health, Ageing and Retirement in Europe (SHARE), a multidisciplinary and cross-national bi-annual household panel survey coordinated by the Mannheim Research Institute for the Economics of Aging (MEA) with the technical support of CentERdata at Tilburg University. The survey collects data on health, socio-economic status, and social and family networks for nationally representative samples of elderly people in the participating countries.

2.1 Description of SHARE

SHARE is designed to be cross-nationally comparable and is harmonized with the U.S. Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA). The base-line (2004) study covers 11 countries, representing different regions of continental Europe, from Scandinavia (Denmark, Sweden) through Central Europe (Austria, Belgium, France, Germany, the

Netherlands, Switzerland) to Mediterranean countries (Greece, Italy, Spain). Four other European countries (Czech Republic, Ireland, Poland and Slovenia) have been subsequently added.

The target population consists of individuals aged 50+ who speak the official language of each country and do not live abroad or in an institution, plus their spouses or partners irrespective of age. The common questionnaire and interview mode, the effort devoted to translation of the questionnaire into the national languages of each country, and the standardization of fieldwork procedures and interviewing protocols are the most important design tools adopted to ensure cross-country comparability (Börsch-Supan et al. 2005).

The interview mode is Computer Assisted Personal Interview (CAPI), supplemented by a self-administered paper-and-pencil questionnaire. The CAPI questionnaire consists of 20 modules covering several aspects of life circumstances: demographics, physical and mental health, behavioral risks, health care, employment and pensions, grip strength and walking speed, children, social support, housing, consumption, household income, assets, financial transfers, social and physical activities, and expectations. The paper-and-pencil questionnaire is instead used to collect more sensitive information, like social and psychological well-being, religiosity and political affiliation.

Programming of the CAPI interviews is done centrally by CentERdata using the Blaise software language. Besides enforcing standardized interview conditions across countries, this system offers an unprecedented amount of information on the time respondents spend on each single question in the CAPI interview. This information, stored in the so-called 'keystroke files', can be used in a number of different ways. For example, it provides a useful diagnostic tool to identify problems occurring during the interview process, or to detect cases where interviewers did not follow the SHARE protocol. It also enables one to compute an accurate measure of duration of the CAPI interview. We use the time spent on cognitive questions in a novel way, namely as a measure of a respondent's processing speed, a second dimension of cognitive abilities evaluation. As argued by Salthouse (1985), aging is associated with a decrease in the speed at which many cognitive operations can be executed. The keystroke files allow us to capture this characteristic of cognitive deterioration.

In this paper, we restrict attention to the countries that contributed to the 2004 baseline study. To separate age and cohort effects we also make use of the refreshment sample in the second wave. In this case, we drop the Austrian sample because no refreshment sample is available. We restrict attention to individuals aged 50–70 at the time of the interview who answered the retrospective question on past employment status and reported being in the labor force at age 50. These selection criteria give a working sample of 13,753 individuals from the first wave and 4,445 from the refreshment sample in the second wave. Table 1 shows the composition of our working sample by country and gender. Due to the lower female employment rate we end up undersampling females, especially in Italy, Greece and Spain. This is also likely to make our female sample somewhat special. Table 2 shows the labor force status of people aged 50–70 in the SHARE sample by country. We distinguish between people who never did any paid work, and are therefore excluded from our working sample, and people who were in the labor force at the age of 50, and are therefore included in our working sample. It is easy to note that a selection issue arises for females from Mediterranean countries, as about one third of them never worked.

2.2 Cognitive measures

The measures of cognitive ability in SHARE are the outcomes of simple tests of orientation in time, memory, verbal fluency and numeracy. These tests are administered to all respondents and are carried out after the first four modules (Cover Screen, Demographics and Networks, Physical Health, and Behavioral Risks) of the questionnaire. The tests are comparable with similar tests implemented in the HRS and ELSA, and follow a protocol aimed at minimizing the potential influences of the interviewer and the interview process. An important drawback of SHARE is that the exact same tests were administered to all respondents of the same household and to the same individual over time. Repeated exposure to the same tests may induce learning effects which are likely to improve the cognitive scores of some respondents. The potential impact of these effects is analyzed later in Section 4.3.

The test format adopted by SHARE is based on the Telephone Interview of Cognitive Status-Modified (TICS-M) test which utilizes a format for the assessment of cognitive functions that can be administered in person or by telephone and is highly correlated with the Mini-Mental State Exam (MMSE) (Folstein et al. 1975), a screening tool frequently used by health-care providers to assess overall brain function. While the MMSE is limited by a ceiling effect, and therefore is relatively insensitive to early evidence of cognitive impairment (de Jager et al. 2003), the TICS-M test allows more discrimination in the range of cognitive performance because it uses 10-word recall instead of 3-word as in the MMSE.

The test of orientation in time consists of four questions about the interview date (day, month, year) and day of the week. This test shows very little variability across respondents. Almost 87 percent of the baseline sample answered correctly all four questions, with 86 percent of the errors concerning the question about the day of the month. Thus, to better discriminate between respondents we make use of the time spent to answer these four questions to construct an adjusted test score that combines the raw score of the original test with a measure of processing speed. In practice, we proceed as follows: we first group respondents by their raw score, which ranges between 0 and 4 depending on the number of correct answers; for all respondents with a positive score, we then group respondents in each group by quintile of the time length distribution. In this way, we obtain an adjusted score with 21 different values.

The test of memory consists of verbal registration and recall of a list of 10 words (butter, arm, letter, queen, ticket, grass, corner, stone, book, stick). The speed at which these words are displayed to the interviewer and then read out to the respondent is automatically controlled by the CAPI system. The respondent hears the complete list only once and the test is carried out two times, immediately after the encoding phase (immediate recall) and at the end of the cognitive function module (delayed recall). The raw total scores of both tests correspond to the number of words that the respondent recalls. As for the test of orientation in time, we again use the keystroke files to combine the raw score with the time needed to answer to the corresponding recall question. Following a procedure similar to that described above, we obtain an adjusted score with 51 different values.

The test of verbal fluency consists of counting how many distinct elements from a particular category the respondent can name in a specific time interval. The specific category used in SHARE is members of the animal kingdom (real or mythical, except repetitions or proper nouns) and the time interval is one minute for all respondents. Notice that, because of the fixed time interval, we cannot use processing speed in this case.

Finally, the test of numeracy consists of a few questions involving simple arithmetical calculations based on real life situations. Respondents who correctly answer the first question are asked a more difficult one, while those who make a mistake are asked an easier one. The last question is about compound interest, testing basic financial literacy. The resulting raw total score ranges from 0 to 4. A full description of the sequence of questions used for this test is given in Appendix numeracy. Here again we use the keystroke files to combine the raw total score with the time needed to provide all correct answers. As for the test of orientation in time, we obtain an adjusted score with 21 different values.

Table 3 provides the mean and the standard deviation of raw and adjusted scores, along with the correlation between the scores on the various domains. To interpret the table, notice that the maximum score for orientation in time is 4, for recall is 10, and for numeracy is 4. As for the correlations between test scores, orientation in time is only weakly correlated with the other domains (about .20), immediate and delayed recall have the highest correlation (close to .70), and the correlations between all other domains is about .40.

2.3 Summary statistics

This section presents a few summaries of the distribution of adjusted test scores in the first wave of SHARE. These summaries have been constructed by smoothing average test scores by age using a 3-year centered running mean. Averaging of the individual observations is based on the cross-sectional survey weights provided by the public-use data files.

Figure 1 plots the age-profiles of average test scores separately for men and women. The figure shows substantial gender differences in the outcome of the various tests. Women tend to do better than men in the tests of recall (both immediate and delayed), especially at younger ages, whereas men tend to do better than women in the test of numeracy. In the other two domains the confidence bands of the two curves overlap. The figure shows clear evidence of falling average test scores with age. Although suggestive, we cannot conclude from this evidence that aging causes a decline of cognitive abilities because the observed pattern combines both age and cohort effects. Due to the cross-sectional nature of the data, we cannot distinguish between these two different effects.

Figure 2 plots the age-profiles of average test scores separately for people with and without a high-school degree ('HS graduates' and 'HS dropouts' respectively). This figure is consistent with the hypothesis that education is an important determinant of heterogeneity in cognitive functions at older ages (Le Carret 2003). Higher education corresponds to better scores in all cognitive tests at all ages. However, education differences are mainly in the level of test scores, not in their rate of decline with age. Notice that education is particularly important in the case of numeracy but does not seem to matter much in the case of orientation in time.

Figure 3 plots the age-profiles of average test scores by macro-region. Our macro-regions correspond to the classical geographical aggregation into Scandinavia (Denmark and Sweden), Central Europe (Belgium, France, Germany, the Netherlands and Switzerland) and Mediterranean countries (Greece, Italy, Spain). The figure shows large differences in average test scores between Mediterranean countries and the other countries of continental Europe. Differences between Scandinavia and Central Europe are instead much less marked.

Figure 4 plots the age-profiles of average test scores by employment status, distinguishing between employed and retired people. The latter include the unemployed because, in many European countries, unemployment programs provide early retirement benefits well before the Social Security early retirement age (Gruber and Wise 2004). We do not report average test scores after age 65 for those who are employed because the employment rate is very small after that age. The figure shows large differences in average test scores between employed and retired people, particularly for the numeracy and fluency tests. Employed people have higher average test scores at all ages. Differences are mainly in the intercept and there is no clear evidence of systematic differences in the rate of decline with age.

Finally, Table 4 shows average cognitive scores by gender, retirement status and age group (60–65 and 66–70). The table distinguishes between people who are still employed, are retired by 5 years or less, and are retired by more than 5 years. In the 60–65 age range not only retired people show lower test scores, but also the distance from retirement seem to matter. People retired by more than 5 years, in fact, shows on average lower test scores than people retired by 5 years or less. Moreover, these differences are in most cases statistically significant at 5% level. If we look instead at the 66–70 age range, no clear pattern emerges, possibly due to the small number of employed people after the age of 65.

3 Theoretical framework

In this section, we present a discrete-time version of the model proposed by Grossman (1972), which we use to generate explicit predictions about the life-time profile of the individual stock of cognitive abilities, treated as unidimensional 'cognitive capital'. The key insight of this model is that individuals can to some extent control the level of their cognitive capital by investing in cognitive-repair activities to partly offset exogenous age-related deterioration. In psychology, this has been called the 'Dumbledore hypothesis' of cognitive aging (Stine-Morrow 2007). By cognitive-repair investment we mean all types of cognitive-promoting behavior, including extensive reading, as well as cultural and other intellectually stimulating activities (Adam et al. 2006, Hertzog et al. 2008). In our version of the model, cognitive repair investment enters the utility function along with consumption. This is a key difference from the Grossman model, where instead the cognitive stock enters directly into the utility function. Our point is that there is no reason to think that cognitive abilities directly contribute to an individual's utility, while cognitive repair activities may be enjoyed as any other good. As customary in this literature, we disregard educational choices and assume that an individual's education is determined outside the model.

Formally, we consider an individual who at the age when planning starts ($t = 0$) chooses sequences of consumption $c_0, \ldots, c_T$ and cognitive investments $a_0, \ldots, a_T$ in order to maximize lifetime utility

$$U = \sum_{t=0}^{T} \frac{u(c_t, a_t)}{(1+\rho)^t},$$

where T is life length, assumed to be known with certainty, and ρ is the rate of time preference. We assume that u is strictly increasing in both arguments with decreasing marginal utilities $u_{c_t} = \partial u(c_t, a_t)/\partial c_t$ and $u_{a_t} = \partial u(c_t, a_t)/\partial a_t$.

In solving this problem, the individual takes as given her initial stock of cognitive capital k_0

and of other assets A_0, and faces three constraints. The first constraint is the law of motion for the stock of cognitive capital

$$k_{t+1} = \gamma_t a_t + (1 - \delta_t) k_t, \qquad t = 0, \ldots, T-1, \tag{1}$$

where γ_t is the efficiency of cognitive repair and δ_t is the natural deterioration rate of cognitive capital, defined as the rate at which cognitive capital would deteriorate in the absence of repair investment. The concept that individuals must invest in cognitive-promoting activities in order to offset the natural rate of cognitive deterioration accords well with the experimental training literature, which shows that training effects often dissipate within a few years unless there are additional attempts to provide reinforcement to maintain the intervened behavior (Willis et al. 2006). The parameters γ_t and δ_t in (1) may depend on education and other individual characteristics. Unlike Grossman original specification, they are also allowed to vary over time. This is mainly intended to capture alternative hypothesis about the efficiency of cognitive repair before and after retirement. For example, greater efficiency while working is consistent with the idea that people who work face an environment that is more challenging and stimulating (Rohwedder and Willis 2010). On the other hand, lower efficiency while working is consistent with the hypothesis that retired people may be able to devote more time to cognitive repair activities.

The second constraint forces cognitive-repair investment to be nonnegative,

$$a_t \geq 0, \qquad t = 1, \ldots, T-1,$$

with the terminal condition that $a_T = 0$, or equivalently

$$k_{t+1} \geq (1 - \delta_t) k_t, \qquad t = 0, \ldots, T-1, \tag{2}$$

with $k_T = (1-\delta_{T-1}) k_{T-1}$. We assume that the period utility function is such that the nonnegativity constraint on consumption can be ignored. Constraint (2) is important because it places an upper bound on the rate of decline of cognitive capital, which cannot exceed the natural deterioration rate δ_t. Notice that cognitive-damaging behavior enters the model not through a_t but by increasing the rate δ_t at which cognitive capital depreciates (Muurinen 1982). As Stine-Morrow (2007) puts it, 'losses come "for free"; gains are hard won'.

The third constraint is the life-time budget constraint

$$\sum_{t=0}^{T} \frac{c_t}{(1+r)^t} + \sum_{t=0}^{T} \frac{p\, a_t}{(1+r)^t} = A_0 + \sum_{t=0}^{T} \frac{y_t}{(1+r)^t}, \tag{3}$$

where p is the price of cognitive-repair investment, r is the real interest rate, and $y_t = F_t(k_t)$ is earnings, which depend among other things on the cognitive stock at time t. This specification is consistent with the traditional Mincerian earning function that allows for individual heterogeneity in earnings due to unobserved ability. We assume that the earnings production function F_t is strictly increasing with diminishing marginal productivity $f_t = F_t'$.

Letting $u_{c_t} = \partial u/\partial c_t$ and $u_{a_t} = \partial u/\partial a_t$, the first order conditions for this problem are

$$u_{c_t} = \lambda \left(\frac{1+\rho}{1+r}\right)^t, \qquad t = 0, \ldots, T, \tag{4}$$

$$u_{a_t} = \lambda \left(\frac{1+\rho}{1+r}\right)^t \left(p - \frac{\partial Y_t}{\partial a_t}\right), \qquad t = 0, \ldots, T-1, \tag{5}$$

and

$$0 = \nu_t[k_{t+1} - (1-\delta_t)k_t], \qquad t = 0, \ldots, T-1, \tag{6}$$

where λ is the Lagrange multiplier associated with the budget constraint (3), Y_t is the discounted value at time t of all subsequent earnings,

$$Y_t = \sum_{s=1}^{T-t} \frac{y_{t+s}}{(1+r)^s} = \sum_{s=1}^{T-t} \frac{F_{t+s}(k_{t+s})}{(1+r)^s},$$

and the ν_t are the Lagrange multipliers associated with the non negativity constraints (2).

At an interior solution ($\nu = 0$), the marginal rate of substitution (MRS) between cognitive investment and consumption is

$$\frac{u_{a_t}}{u_{c_t}} = MRS_t = p - \frac{\partial Y_t}{\partial a_t}, \tag{7}$$

where the right-hand side is the effective price of cognitive investment in terms of foregone consumption. If Y_t does not depend on a_t, then we have the familiar condition $MRS_t = p$. An example when this may happen is retirement. If $\partial Y_t/\partial a_t >$ before retirement while $\partial Y_t/\partial a_t = 0$ after retirement, for example because of a substantial lump-sum component in pension benefits as in Galama et al. (2008), and if the MRS between cognitive investment and consumption is decreasing in a_t, then the model predicts a decrease in cognitive repair after retirement. When the effective rate of investment $\gamma_t a_t/k_t$ falls below the natural deterioration rate δ_t, the cognitive stock declines.

In the special case when cognitive investment does not enter the utility function (the pure investment model), $u_{a_t} = 0$ for all t and so (5) becomes

$$0 = \lambda \left(\frac{1+\rho}{1+r}\right)^t \left(p - \frac{\partial Y_t}{\partial a_t}\right), \qquad t = 0, \ldots, T-1,$$

which is satisfied by setting $p = \partial Y_t/\partial a_t$. We show in Appendix A that, when γ_t takes a constant value γ, this condition is equivalent to the condition

$$\pi_t = f_t(k_t) \tag{8}$$

where $\pi_t = p(\delta_t + r)/\gamma$ is the user cost of cognitive capital. If f_t is strictly decreasing, then an increase in π_t (for example because of an increase in the natural deterioration rate δ_t) causes k_t to fall.

On the other hand, at a boundary solution ($\nu_t > 0$), the nonnegativity constraint (2) implies that the rate of decline of the cognitive stock must equal the natural deterioration rate δ_t, a point stressed by Case and Deaton (2005). We henceforth assume that (2) is not binding at $t = 1$. So, an important question is whether it may be binding later in life.

Consider first the pure investment model. If post-retirement income does not depend on the current level of cognitive abilities then, in a model without (2), the cognitive stock should drop to zero immediately after the age R at which retirement occurs. Since (2) does not allow this, the cognitive stock can only decline at its maximal rate δ_t. Thus, for the pure investment model with strictly decreasing f_t,

$$k_t = \begin{cases} f_t^{-1}(\pi_t), & t \leq R, \\ (1 - \delta_{t-1})k^*_{t-1}, & t > R. \end{cases} \tag{9}$$

This may be viewed as one way of formalizing the use-it-or-lose-it hypothesis (Rohwedder and Willis 2010). As an illustration, Figure 5 shows the optimal path of the stock of cognitive capital in a simple version of the pure investment model where post-retirement income is a lump-sum unrelated to previous earnings and the natural deterioration rate is constant. We consider four otherwise identical individuals: one retiring at age 50 (orange line), one retiring at age 60 (green line), one retiring at age 70 (blue line), and one who never retires (black line). The figure illustrates clearly two sharp conclusions of the model. First, the optimal stock of cognitive capital drops rapidly after retirement. Second, the 'cognitive gap' between initially identical individuals who only differ in their retirement pattern widens rapidly with age.

Although the simplicity of (9) is lost when a_t enters the utility function, condition (7) does not rule out the possibility of an increase in the rate of decline of the cognitive stock after retirement. This rate of decline, however, will be less than the pure deterioration rate δ_t because there also non-market incentives to cognitive investment. Also notice that heterogeneity in the parameters of the utility function, in particular a preference for more cognitive stimulating activities may play a role in determining the effect of retirement on the cognitive stock and its rate of decline.

4 Identification issues

This section discusses several important identification issues that arise when using the SHARE data to estimate a model based on the arguments in Section 3. The first issue is the potential endogeneity of the retirement decision. The second issue is cohort heterogeneity in cognitive abilities, which complicates the interpretation of estimates from a single cross-section as we cannot easily distinguish between pure aging and cohort effects. The other issues is learning effects, due to the fact that exactly the same cognitive tests were submitted to all eligible respondents within a household and to the same individual over time, and panel attrition. Both problems severely limit the usefulness of the panel dimension of SHARE.

4.1 Endogeneity of retirement

Endogeneity of retirement poses the main empirical challenge when trying to identify the effect of retirement on cognitive performance. On the one hand, simple OLS estimation may be biased because of potential reverse causality (people with lower cognitive abilities may decide to retire earlier) or correlation between the retirement choice and unobservable factors (i.e. health). On the other hand, the available empirical evidence (such as the country studies in Gruber and Wise 2004) indicates that, for most workers in Europe, the retirement decision is simply to retire at the earliest possible date, which is determined by exogenous laws and Social Security regulations.

We approach the problem using a standard instrumental variables (IV) strategy. Key to our approach is the availability of instruments that are both relevant, i.e. directly related to the retirement decision, and exogenous, i.e. they affect cognitive abilities only indirectly through their effects on the age of retirement. Our instruments are the legislated early and normal ages of eligibility for a public old-age pension, two variables that are easily shown to be relevant (Section 5.2) and are arguably exogenous.

Figures 8 and 9 show the histogram of the retirement age by country, respectively for men and women. The vertical blue and red lines respectively denote the eligibility ages for early and normal retirement, while the blue and red areas indicate changes in the eligibility rules for the cohorts in our sample. Major changes occurred Italy, while smaller changes occurred in Austria, Germany and Sweden. While other countries also modified Social Security rules during the 1990's, these changes mostly restricted other criteria for early retirement (e.g. year of contribution or definition of invalidity status) and eliminated financial incentives to retire, but did not change the ages of eligibility. Eligibility ages differ substantially by country and gender. For instance, the early retirement age ranges from 52 in Italy before 1994 to 61 in Sweden after 1999. Smaller cross-country and gender differences are observed for the normal retirement age. This is 65 years in many countries, but varies for both males and females from a minimum of 60 to a maximum of 65 years. We refer to Appendix C for further detail about pension eligibility rules.

Notice that, unlike other papers using Social Security laws to construct instruments (e.g. Bonsang et al. 2010 and Rohwedder and Willis 2010), we do not use the early and normal eligibility ages at the time of the interview (2004 in our case), but rather the eligibility ages at the time when individuals faced their retirement decisions. Thus, we explicitly account for changes in eligibility rules that differently affect the cohorts in the SHARE countries.

4.2 Cohort heterogeneity

Cohort heterogeneity in cognitive abilities may reflect differences in both initial conditions and mortality across cohorts. The role of differences in initial cognitive endowment and early life-environment has recently been stressed by Richards et al. (2004), Cuhna and Heckman (2007), Case and Paxson (2009) and Currie (2009). If cohort heterogeneity is only a fixed effect, reflecting different initial conditions, then one solution is to difference it out by exploiting the panel dimension of SHARE. The problem with this approach is that, along with the fixed effects, all time-invariant personal characteristics are also differenced out. Further, nonrandom attrition and retest effects (see below), due to repeated exposure to the same tests, introduce different and perhaps bigger problems.

Differences in mortality, cumulated over time between birth and the age at which a cohort is observed, may also induce substantial cohort heterogeneity. The problem may not be so important for the younger cohorts, but it is very relevant for the older ones, especially those that survived the Second World War. The survivors from these cohorts are in fact expected to have better physical and cognitive health than average.

As a consequence of cohort heterogeneity, the coefficient on age from a cross-sectional regression may be affected by two different sources of bias, one due to differences in initial conditions, the other due to differences in mortality. These biases are likely to have opposite sign. Cohort differences in initial conditions may cause overestimation of the age effect because of the dramatic improvements in childhood conditions in all European countries after the Second World War. Cohort differences in mortality may cause underestimation of the age effect because mortality rates are typically higher for people with poor health and poor cognitive abilities (Glymour 2007).

The debate on the direction and magnitude of cohort differences in cognitive abilities is still ongoing. On the one hand, there is an extensive literature, stimulated by the analysis of Flynn (1987), arguing that important IQ gains have occurred across generations in several countries, including many European ones. On the other hand, Alwin (1991) reports a decline in education-adjusted verbal test scores. The contrasting evidence from this literature may be due to differences in the type of measured abilities. Flynn's IQ test measures principally fluid intelligence, while Alwin focuses on cohort differences in verbal abilities, usually defined as part of crystallized intelligence.

In Section 6.1 we control for cohort differences by using the refreshment sample from the second wave. There are two main reasons for this. First, using data from the second wave allow us to distinguish between age and cohort effects, because for each cohort we now have two different ages. Second, the refreshment sample does not show problems of attrition and learning effects that characterize the longitudinal sample.

4.3 Learning effects and panel attrition

SHARE submits the exact same cognitive tests to all eligible respondents within a household, and to the same individual over time. This feature of the survey, which was meant to guarantee testing equivalence across individuals and over time, may cause two types of learning effects. The first is intra-household learning, namely the fact that respondents may learn from the response given by other household members. The second is retest effects, namely the fact that respondents in a given wave may learn from their own test experience in a previous wave. It is reasonable to conjecture that both these effects may bias test scores upwards.

Intra-household learning may bias cognitive test scores in both waves, but is only be a problem for respondents in households with at least two respondents. In principle, it should be prevented by the SHARE interviewing protocol as the cognitive tests should be administered without third persons, in a separate room, as free as possible from interruptions, and without proxy respondents. In practice, these conditions have not always been satisfied. Specifically, for about 20 percent of the respondents in wave 1, other persons were present during the cognitive module of the interview. In Section 6.2 we control for this learning effect by adding as an extra regressor a dummy variable that is equal to one for individuals who witnessed the interview of another household member and is equal to zero otherwise.

A problem that complicates a longitudinal analysis is retest effects due to the fact that, in SHARE, individuals are repeatedly exposed to exactly the same tests. Unlike intra-household learning effects, that can be identified from a single cross section, an analysis of retest effects must be based on the longitudinal sample and cannot ignore the potential selectivity effects associated with nonrandom attrition.

Panel attrition in SHARE is nonnegligible, as about one third of the original sample is lost between the first and the second wave of the survey. Loss rates also vary substantially by country, ranging from about 19 percent in Greece to about 47 percent in Germany, and are typically higher for men than for women. While aspects of the survey design and of the fieldwork may be important determinants of attrition probabilities, Zamarro et al. (2008) also find that people in poor health and with poor cognitive abilities are more likely to drop out of the panel. Given the high attrition rate, and the fact that those who are lost seem to be those with low cognitive skills, we cannot exclude that this selectivity effect is driven by unobservable factors. Thus, ignoring attrition or assuming random attrition may lead to invalid inference.

As sample attrition and retest effects attrition are likely to operate in the same direction, ignoring one may lead to overestimating the other. Taking all this into account, it is safer to confine attention to the cross-sectional sample, possibly trying to control for cohort heterogeneity and intra-household learning effects.

5 Empirical results

In this section we present the results obtained by estimating a class of statistical models motivated by the discussion in Sections 3 and 4. All models in this class represent the age-profile of test scores for the ith individual in our sample as a continuous piecewise-polynomial function of age with a single knot at the reported retirement age R_i, defined as the age at which the last job ended.

After experimenting with polynomials of various order, we find that a continuous piecewise-linear function of age (i.e. a linear spline) is systematically preferred by standard model selection criteria. Our class of statistical models is therefore of the form

$$Y_i = \beta_{0i} + \beta_{1i} Age_i + \beta_{2i} DistR_i + \beta_3^\top X_i + U_i, \tag{10}$$

where β_{0i}, β_{1i} and β_{2i} are possibly heterogeneous parameters, Age_i is the individual's current age, $DistR_i = \max\{0, Age_i - R_i\}$ is the number of years spent in retirement (equal to zero if the individual is not yet retired), X_i is a vector of exogenous regressors, and U_i is a regression error. Model (10) is always estimated separately for males and females.

We begin by presenting the results obtained using ordinary least squares (OLS). These results may be interpreted as purely descriptive statistics or, under the unlikely assumption of exogeneous retirement, as estimates of the causal effect of retirement on cognitive abilities. Then, in Section 5.2, we compare these results with those obtained using two-stage least squares (2SLS) to control for potential endogeneity of retirement.

5.1 OLS

Table 5 shows our OLS estimates separately by gender and cognitive domain. To facilitate comparisons, cognitive test scores are standardized by subtracting off their sample mean and dividing by their sample standard deviation. Estimated standard errors are robust to clustering at the country and cohort level.

Our first specification (Model A) includes as regressors only the linear age spline with a single knot at retirement and a low-education dummy ($LowEd$). This specification assumes that β_{1i}

and β_{2i} do not vary in the population and that $\beta_3 = 0$. For all domains, the age effects are statistically significant, except for orientation in time for males, and have the expected negative sign. Consistently with the prediction of the Grossman model, the coefficient on $DistR$ is negative and statistically significant for all domains, indicating a sharp negative change in the slope of the age-profile after retirement. Moreover, we find that the low-education dummy is always statistically significant and has the expected negative sign. To give an idea of the size of the estimated effect of education, people with a high-school degree recall on average almost one word more in both recall tests than people without a high-school degree. They can also name 4 more animals in the fluency test and get half point more in the numeracy test on the original 0–4 scale. These results are consistent with the theory of fluid and crystallized intelligence (Horn and Cattel 1967), as age appears to be relatively more important for the two recall tests but less important for fluency and numeracy that are more influenced by formal education.

Model B adds to Model A a set of country dummies (with Belgium as the reference country), thus removing the restriction that $\beta_3 = 0$. Including the country dummies does not change much the estimated coefficients but increases dramatically the fit of the model as measured by the adjusted R^2.

To evaluate the presence of heterogeneity across educational groups in the linear spline, Model C allows β_{1i} and β_{2i} to differ depending on educational attainments. However, the interaction between the low-education dummy and the age spline is rarely statistically significant. Figure 6 illustrates our results by showing the estimated age-profile of test scores for people who retire at age 60. The negative change in the slope of the age-profile after retirement is evident for all domains. The effect of having completed high school is also clearly positive for all cognitive domains. However, its magnitude is different across domains and is strongest for numeracy.

Finally, to further investigate the issue of cross-country heterogeneity, we estimate Model C, without country dummies, separately for our three macro-regions: Scandinavia, Central Europe, and Mediterranean countries. Figure 7 shows the age-profiles of predicted test scores by education level and macro-region implied by the estimated model. Compared with Figure 3, the differences between the various regions are now smaller, especially in the case of numeracy, delayed recall, and for more educated people. However, regional differences persist and are sizeable, with people living in Mediterranean countries showing lower test scores in all cognitive domains except orientation in time. These differences may be due to differences in the quality of schooling system or, according to the model, to differences in the wage premium on cognitive skills.

5.2 2SLS

Our instruments for the potentially endogenous variable $DistR$ are $DistE = \max\{0, Age - Early\}$ and $DistN = \max\{0, Age - Normal\}$, namely the positive part of the difference between the actual age and the legislated age of eligibility for early and normal retirement ($DistE$ and $DistN$ respectively). Our instruments for the interaction of $DistR$ with $LowEd$ in Models B and C are the interactions of $DistE$ and $DistN$ with $LowEd$.

Table 6 shows the estimated coefficients from the first stage regression of $DistR$ on the two instruments and the exogenous variables in the various models, along with the R^2 of the regression and the F-test statistic on the exogenous instruments. Model A, B and C contain exactly the same covariates as in Table 5. Our results confirm that eligibility rules are important for retirement in the sense that, for both genders and across all models, distances from the eligibility ages are strong predictors of the distance from retirement ($DistR$). They are also consistent with the evidence in Gruber and Wise (2004) on the importance of early retirement incentives, especially for males.

Table 7 presents the estimated coefficient on $DistR$ from the second stage of the 2SLS procedure. These estimates confirm the negative effect of retirement on the age-profile of cognitive test scores. The coefficients for Model A, without country dummies, are large compared with OLS, especially for women, which is in line with the results in Rohwedder and Willis (2010). The absence of country dummies is the main reason for these large differences, as countries with higher level of cognitive abilities (such as Denmark and Sweden) also have higher ages of eligibility for retirement. This is clearly seen in Model B where the size of the coefficients dramatically decreases, although they remain negative and statistically significant in most cases. Finally, Model C adds to Model B controls for heterogeneity across educational levels in the effect of retirement. For males, the results are quite similar across educational groups except for fluency. For females instead, we observe a large heterogeneity across educational groups, with a large and statistically significant coefficient only for the people with lower education.

In general, 2SLS coefficients are bigger than OLS, although this difference is sizeable only for females. In fact, a standard test of the null hypothesis of exogenous retirement based on the difference between OLS and 2SLS estimates rejects the exogeneity assumption at conventional significance level only in the case of delayed recall for males, but in all domain for females. Selection issues are likely to be the explanation of the big differences between males and females, as our sample selection criteria exclude a large fraction of females of the original SHARE sample, especially in Mediterranean countries. This also implies that other papers that do not fully control for gender differences (such as Bonsang et al. 2010 and Rohwedder and Willis 2010) may also offer biased estimates of the causal effect of retirement on cognitive abilities.

6 Robustness checks

We now present the results of robustness checks for the presence of cohort effects (Section 6.1) and intra-household learning (Section 6.2).

6.1 Cohort effects

Table 8 shows the results from estimating three versions of Model C using a pooled sample that combines the first (2004) wave of SHARE and the refreshment sample from the second (2006) wave. The refreshment sample allows us to separate the effect of cohort differences from that of aging.

Our basic model (Model C1) is just Model C, now estimated on the pooled sample. Model C2 adds to Model C1 a set of dummies for the 1934–38, 1946–50, and 1951–56 birth cohorts. The reference cohort is people born during World War II (1939–1945). Model C3 adds a time dummy to Model C2. The results show that controlling for cohort effects only slightly modifies the age-profiles of our test scores. In particular, with the exception of orientation in time for females, the coefficient on $DistR$ is now negative and statistically significant for all domains. The only noticeable difference is that, after controlling for cohort effects, the coefficient on age in the fluency and numeracy tests for males are no longer significant.

As a final check for the importance of cohort effects we also included in the model interactions between age and country dummies. The results, not reported in our paper, do not show important differences in the estimated age-profiles of cognitive abilities.

6.2 Learning effects

To control for intra-household learning effects, we estimate Model C with an added indicator which is equal to one for individuals who witnessed the interview of another household member and is equal to zero otherwise. Assuming that, given the regressors in Model C, the added indicator is conditionally independent of the unobservables that affect the test scores, its associated coefficient measures the average intra-household learning effect for our baseline individual. Because this effect may be expected to be greater for more educated people, we also estimate the model separately by educational attainment.

Table 9 shows the IV coefficients on the added indicator for the model estimated on the pooled data, the subsample of people without a high-school degree ('HS dropouts'), and the subsample of people with a high-school degree ('HS graduates'). As expected, the intra-household learning effect is positive, significant, and increasing in the level of education. The effect seems to be stronger for orientation in time and for the two memory tests. For example, respondents with a high-school degree recall, on average, half a word more on the delayed recall test if they witnessed the interview of another household member.

7 Conclusions

In this paper we investigate the relation between age and cognitive abilities using a version of the human capital model and data from SHARE, a survey that has the unique feature of providing measures of cognitive functions for a representative sample of people aged 50+ in Europe.

Our findings reveal an increase in the rate of decline of cognitive abilities after retirement. In the light of the model, this reflects the reduced incentives to invest in cognitive-repair activities after retirement. An implication of our result is that incentives to early retirement and mandatory retirement rules cause important losses of human capital because they make it less attractive to preserve the level of human capital inherited from the past, a point also made by Rohwedder and Willis (2010).

Equally important is the role of education in explaining heterogeneity in the level of cognitive abilities and, to a lesser extent, in their age-related decline.

Finally, even after controlling for education, age and length of the retirement spell, we find large and systematic differences in measured cognitive functions across European regions, with lower educated people in Mediterranean countries showing lower test scores in all cognitive domains except orientation in time.

References

Adam S., Bay C., Bonsang E., Germain S., and Perelman S. (2006), "Occupational activities and cognitive reserve: A frontier approach applied to the Survey of Health, Ageing and Retirement in Europe (SHARE)." CREPP WP 2006/05, University of Liege.

Bonsang E., Adam S., and Perelman S. (2010), "Does Retirement Affect Cognitive Functioning?" RM/10/005 Maastricht University.

Alwin D.F. (1991), "Family of origin and cohort differences in verbal ability." *American Sociological Review*, 56: 625–638.

Banks J., and Oldfield Z. (2007), "Understanding pensions: Cognitive functions, numerical ability and retirement saving." *Fiscal Studies*, 28: 143–170.

Ben Porath Y. (1967), "The production of human capital and the life cycle of earnings." *Journal of Political Economy*, 75: 352–365.

Börsch-Supan A., and Jürges H. (2005), *The Survey of Health, Aging, and Retirement in Europe. Methodology*, Mannheim Research Institute for the Economics of Aging (MEA).

Case A., and Deaton A. (2005), "Broken down by work and sex: How our health declines." In Wise D.A. (ed.), *Analyses in the Economics of Aging*. University of Chicago Press, 185–212.

Case A., and Paxson C.H. (2009), "Early life health and cognitive function in old age." *American Economic Review*, 99: 104–109.

Cervilla J.A., Prince M., Joels S., Lovestone S., and Mann A. (2000), "Long-term predictors of cognitive outcome in a cohort of older people with hypertension." *British Journal of Psychiatry*, 177: 66–71.

Christelis D., Jappelli T., and Padula M. (2010), "Cognitive abilities and portfolio choice." *European Economic Review*, 54: 18–38.

Cunha F., and Heckman J. (2007), "The technology of skill formation." *American Economic Review*, 97: 31–47.

Currie J. (2009), "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development." *Journal of Economic Literature*, 47: 87–122.

de Jager C.A., Budge M.M., and Clarke R. (2003), "Utility of TICS-M for the assessment of cognitive function in older adults." *International Journal of Geriatric Psychiatry*, 18: 318–324.

Flynn, J.R. (1987), "Massive IQ gains in 14 nations: What IQ tests really measure." *Psychological Bulletin*, 101, 171-191.

Folstein M.F., Folstein S.E., and McHugh P.R. (1975), "Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician." *Journal of Psychiatric Research*, 12(3): 189–198.

Galama T., Kapteyn A., Fonseca R., and Michaud P.-C. (2008), "Grossman's health threshold and retirement." Rand Corporation Working Paper WR-658, Santa Monica.

Glymour M.M. (2007), "When bad genes look good: APOE-e4, cognitive decline, and diagnostic thresholds." *American Journal of Epidemiology*, 165: 1239–1246.

Grossman M. (1972), "On the concept of health capital and the demand for health." *The Journal of Political Economy*, 80: 223–255.

Gruber J., and Wise D. (2004), *Social Security Programs and Retirement around the World: Micro-Estimation*. University of Chicago Press, Chicago.

Hertzog C., Kramer A.F., Wilson R.B., and Lindenberger U. (2008), "Enrichment effects on adult cognitive developments. Can the functional capacity of older adults be preserved and enhanced?" *Psychological Science in the Public Interest*, 9: 1–65.

Horn J.L., and Cattell R.B. (1967), "Age differences in fluid and crystallized intelligence." *Acta Psychologica*, 26: 107–129.

Kaplan G.A., Turrell G., Lynch J.W., Everson S.A., et al. (2001), "Childhood socioeconomic position and cognitive function in adulthood." *International Journal of Epidemiology*, 30: 256–263.

Le Carret N., Lafont S., Letenneur L., Dartigues F., et al. (2003), "The effect of education on cognitive performances and its implication for the constitution of the cognitive reserve." *Developmental Neuropsychology*, 23: 317–337.

McFadden D. (2008) "Human capital accumulation and depreciation." *Review of Agricultural Economics*, 30: 379–385.

Meyer J.S., Rauch G.M., Crawford K., Rauch R.A., et al. (1999), "Risk factors accelerating cerebral degenerative changes, cognitive decline, and dementia." *International Journal of Geriatric Psychiatry*, 14: 1050–1061.

Muurinen J.M. (1982), "Demand for health: A generalized Grossman model." *Journal of Health Economics*, 1: 5–28.

Richards M., Shipley B., Furher R., and Wadsworth M.E. (2004), "Cognitive ability in childhood and cognitive decline in mid-life: longitudinal birth cohort study." *British Medical Journal*, 328: 552-554.

Rohwedder S., and Willis R. (2010), "Mental retirement." *Journal of Economic Perspectives*, 24: 1–20.

Salthouse T.A. (1985), *A Theory of Cognitive Aging*. Amsterdam, North-Holland.

Salthouse T.A. (1996), "The processing-speed theory of adult age differences in cognition." *Psychological Review*, 103: 403–428.

Scarmeas, N., and Stern, Y. (2003), "Cognitive reserve and lifestyle." *Journal of Clinical & Experimental Neuropsychology*, 25: 625–663.

Schaie K.W. (1989), "The hazards of cognitive aging." *The Gerontologist*, 29: 484–493.

Schultz T.W. (1961), "Investment in human capital." *American Economic Review*, 51: 1–17.

Stern Y. (2002), "What is cognitive reserve? Theory and research application of the reserve concept." *Journal of the International Neuropsychological Society*, 8: 448–460.

Stine-Morrow E.A.L. (2007), "The Dumbledore hypothesis of cognitive aging." *Current Directions in Psychological Science*, 6: 295–299.

Trouton A., Stewart R., and Prince M. (2006), "Does social activity influence the accuracy of subjective memory deficit? Findings from a British community survey." *Journal of the American Geriatrics Society*, 54: 1108–1113.

van Praag H., Kempermann G., and Gage F.H. (2000), "Neural consequences of environmental enrichment." *Nature Review Neuroscience*, 1: 191–198.

Willis S.L., Tennstedt S.L., Marsiske M., Ball K., Elias J., Koepke K.M., et al. (2006), "Long-term effects of cognitive training on everyday functional outcomes in older adults." *Journal of the American Medical Association*, 296: 2805–2814.

Zamarro G., Meijer E., and Fernandes M. (2008), "Mental health and cognitive ability." In Börsch-Supan A., Brugiavini A., Jürges H., Kapteyn A., et al. (eds.), *First Results from the Survey of Health, Ageing and Retirement in Europe (2004-2007). Starting the Longitudinal Dimension*. Mannheim Research Institute for the Economics of Aging (MEA).

Table 1: Sample size by country and gender.

		Main sample			Refreshment sample		
		Males	Females	Total	Males	Females	Total
AT	Austria	534	503	1037	.	.	.
BE	Belgium	1043	674	1717	96	72	168
CH	Switzerland	282	230	512	222	217	439
DE	Germany	863	713	1576	304	252	556
DK	Denmark	504	466	970	427	449	876
ES	Spain	547	284	831	145	75	220
FR	France	830	731	1561	253	253	506
GR	Greece	846	413	1259	307	158	465
IT	Italy	716	413	1129	333	209	542
NL	Netherlands	822	453	1275	257	178	435
SE	Sweden	887	999	1886	110	128	238
Total		7874	5879	13753	2454	1991	4445

Table 2: Labor force status of people aged 50–70.

	Male		Female	
Country	Never worked	Selected	Never worked	Selected
AT	0.00	90.34	9.56	68.28
BE	0.57	84.03	12.62	48.69
CH	0.94	91.25	4.13	67.85
DE	0.10	89.41	4.04	67.50
DK	0.18	90.61	0.71	81.06
ES	1.28	87.20	36.46	33.49
FR	0.22	89.96	7.00	69.35
GR	0.44	91.20	29.26	41.58
IT	0.50	85.96	29.44	39.94
NL	0.21	84.31	7.85	40.34
SW	0.10	92.26	0.35	87.19
Total	0.37	88.39	13.34	57.22

Table 3: Mean and standard deviation of raw and adjusted cognitive scores.

Raw scores	Mean	Std.	Correlations				
Orientation	3.88	0.37	1.000				
Recall imm.	5.21	1.67	.106	1.000			
recall del.	3.80	1.90	.118	.663	1.000		
Fluency	20.56	7.29	.081	.374	.345	1.000	
Numeracy	2.61	1.02	.125	.326	.297	.315	1.000
Adjusted scores	Mean	Std.	Correlations				
Orientation	3.50	0.45	1.000				
Recall imm.	4.89	1.69	.112	1.000			
Recall del.	3.43	1.87	.129	.635	1.000		
Fluency	20.56	7.29	.110	.386	.350	1.000	
Numeracy	2.25	1.04	.152	.331	.302	.327	1.000

Table 4: Average cognitive scores by gender and retirement status.

	Aged 60-65			Aged 66-70		
	Empl.	Ret.≤ 5 yr	Ret.> 5 yr	Empl.	Ret.≤5 yr	Ret.> 5 yr
Males						
Orientation	3.53	3.45	3.41	3.38	3.47	3.41
	(0.01)	(0.02)	(0.02)	(0.06)	(0.02)	(0.02)
Recall imm.	4.79	4.48	4.22	4.96	4.16	4.13
	(0.05)	(0.06)	(0.07)	(0.17)	(0.08)	(0.05)
Recall del.	3.28	2.75	2.64	3.33	2.72	2.50
	(0.06)	(0.06)	(0.08)	(0.21)	(0.08)	(0.05)
Fluency	19.93	19.27	18.01	20.73	17.99	18.11
	(0.25)	(0.24)	(0.29)	(0.76)	(0.32)	(0.20)
Numeracy	2.42	2.20	2.07	2.53	1.96	2.07
	(0.03)	(0.03)	(0.04)	(0.12)	(0.05)	(0.03)
Females						
Orientation	3.55	3.48	3.55	3.51	3.50	3.45
	(0.02)	(0.02)	(0.02)	(0.06)	(0.03)	(0.02)
Recall imm.	5.18	4.78	4.73	4.19	4.15	4.45
	(0.07)	(0.07)	(0.08)	(0.26)	(0.11)	(0.06)
Recall del.	3.63	3.40	3.16	2.71	2.99	2.83
	(0.08)	(0.08)	(0.09)	(0.26)	(0.12)	(0.07)
Fluency	20.58	19.68	18.74	18.73	16.62	17.62
	(0.31)	(0.28)	(0.35)	(1.05)	(0.44)	(0.27)
Numeracy	2.04	2.02	1.92	1.56	1.71	1.77
	(0.04)	(0.04)	(0.05)	(0.15)	(0.06)	(0.04)

Table 5: OLS estimates for the level of the test scores in the baseline sample (Model A: linear spline in age and low-education dummy; Model B: Model A plus a full set of country dummies; Model C: Model B plus interaction between the linear spline and the low-education dummy. Significance levels: (*) p-values between 10 and 5 percent; (**) p-values between 5 and 1 percent; (***) p-values below 1 percent.)

	Males			Females		
Orientation	A	B	C	A	B	C
Age	-.003	-.004	-.005 *	-.010 ***	-.010 ***	-.007 **
DistR	-.011 ***	-.013 ***	-.010 **	-.004	-.007 *	-.011 **
LowEd	-.069 ***	-.102 ***	-.082 ***	-.084 ***	-.086 ***	-.117 ***
Age*LowEd			.003			-.005
DistR*LowEd			-.007			.009
R^2	.006	.035	.035	.013	.039	.039
Recall imm.	A	B	C	A	B	C
Age	-.015 ***	-.019 ***	-.016 ***	-.009 **	-.015 ***	-.013 ***
DistR	-.017 ***	-.012 ***	-.017 ***	-.022 ***	-.016 ***	-.016 **
LowEd	-.516 ***	-.413 ***	-.447 ***	-.484 ***	-.390 ***	-.395 ***
Age*LowEd			-.006			-.005
DistR*LowEd			.012 **			.001
R^2	.116	.150	.150	.098	.157	.157
Recall del.	A	B	C	A	B	C
Age	-.016 ***	-.021 ***	-.021 ***	-.016 ***	-.024 ***	-.023 ***
DistR	-.021 ***	-.012 ***	-.015 ***	-.024 ***	-.011 ***	-.010 *
LowEd	-.411 ***	-.345 ***	-.360 ***	-.387 ***	-.333 ***	-.329 ***
Age*LowEd			.002			-.002
DistR*LowEd			.005			-.002
R^2	.092	.121	.121	.083	.131	.131
Fluency	A	B	C	A	B	C
Age	-.007	-.012 ***	-.010 ***	-.003	-.012 ***	-.013 ***
DistR	-.019 ***	-.011 ***	-.019 ***	-.027 ***	-.017 ***	-.012 **
LowEd	-.523 ***	-.386 ***	-.434 ***	-.541 ***	-.424 ***	-.393 ***
Age*LowEd			-.005			.003
DistR*LowEd			.017 ***			-.010
R^2	.090	.218	.218	.102	.238	.238
Numeracy	A	B	C	A	B	C
Age	-.007 **	-.009 ***	-.006 **	-.008 *	-.011 ***	-.008 **
DistR	-.022 ***	-.016 ***	-.021 ***	-.014 ***	-.014 ***	-.014 **
LowEd	-.629 ***	-.565 ***	-.593 ***	-.615 ***	-.558 ***	-.568 ***
Age*LowEd			-.008 *			-.006
DistR*LowEd			.009			.002
R^2	.136	.185	.185	.121	.169	.169

Table 6: Estimated coefficients from the first stage regression for $DistR = \max\{0, Age - R\}$ (Models A, B and C are as in Table 5).

Males	A	B	C
DistE	.316 ***	.395 ***	.496 ***
DistN	.283 ***	.246 ***	.210 **
DistE*LowEd			-.156 ***
DistN*LowEd			.020
R^2	.549	.570	.572
F-stat.	122.87 ***	180.19 ***	108.66 ***
Females	A	B	C
DistE	.210 ***	.270 ***	.298 ***
DistN	.444 ***	.388 ***	.488 ***
DistE*LowEd			-.067
DistN*LowEd			-.181 **
R^2	.600	.611	.614
F-stat.	120.20 ***	187.65 ***	97.76 ***
Country effects	No	Yes	Yes

Table 7: Second step coefficient on $DistR$ (Models A, B and C are as in Table 5).

Males	A	B	C
Orientation	-.021	-.027 **	-.028 **
(LowEd)			.002
Recall imm.	-.073 ***	-.020 **	-.021 **
(LowEd)			.004
Recall del.	-.051 ***	.013	.014
(LowEd)			-.008
Fluency	-.062 ***	-.007	.003
(LowEd)			-.027 *
Numeracy	-.092 ***	-.030 ***	-.030 ***
(LowEd)			.003
Females			
Orientation	-.004	-.005	-.006
(LowEd)			.004
Recall imm.	-.102 ***	-.055 ***	-.017
(LowEd)			-.083 ***
Recall del.	-.098 ***	-.029 ***	-.005
(LowEd)			-.051 ***
Fluency	-.111 ***	-.023 **	.003
(LowEd)			-.055 ***
Numeracy	-.049 ***	-.038 ***	-.012
(LowEd)			-.055 ***
Country effects	No	Yes	Yes

Table 8: OLS estimates for the level of the test scores in the pooled sample that combines the first wave and the refreshment sample from the second wave (Model C1: linear spline in age fully interacted with the low-education dummy, plus a full set of country dummies; Model C2: Model C1 plus cohort dummies; Model C3: Model C2 plus a time dummy).

	Males			Females		
Orientation	C1	C2	C3	C1	C2	C3
Age	-.002	-.009 *	-.006	-.004 ***	.003	.003
DistR	-.018 ***	-.017 ***	-.018 ***	-.001	-.001	-.001
LowEd	-.107 ***	-.107 ***	-.100 ***	-.096 ***	-.096 ***	-.096 ***
Age*LowEd	.001	.002	.002	-.004	-.004	-.004
DistR*LowEd	-.103	-.107	-.161 **	.019	.019	.016
R^2	.020	.020	.021	.015	.015	.015
Recall imm.	C1	C2	C3	C1	C2	C3
Age	-.015 ***	-.017 ***	-.015 ***	-.017 ***	-.018 ***	-.014 **
DistR	-.016 ***	-.015 ***	-.015 ***	-.012 ***	-.010 ***	-.011 ***
LowEd	-.420 ***	-.419 ***	-.415 ***	-.428 ***	-.425 ***	-.417 ***
Age*LowEd	-.002	-.001	-.001	.001	.002	.002
DistR*LowEd	.138 **	.131 **	.100	.048	.037	-.024
R^2	.145	.145	.145	.164	.164	.165
Recall del.	C1	C2	C3	C1	C2	C3
Age	-.018 ***	-.016 ***	-.015 ***	-.023 ***	-.023 ***	-.021 ***
DistR	-.013 ***	-.014 ***	-.014 ***	-.009 ***	-.008 ***	-.009 ***
LowEd	-.362 ***	-.362 ***	-.359 ***	-.353 ***	-.352 ***	-.349 ***
Age* LowEd	-.003	-.003	-.003	.001	.001	.001
DistR* LowEd	.095	.099	.074	.122 **	.114 **	.088
R^2	.125	.125	.125	.145	.145	.145
Fluency	C1	C2	C3	C1	C2	C3
Age	-.017 ***	-.008	-.004	-.017 ***	-.020 ***	-.019 ***
DistR	-.010 ***	-.010 ***	-.011 ***	-.011 ***	-.010 ***	-.010 ***
LowEd	-.390 ***	-.390 ***	-.382 ***	-.432 ***	-.428 ***	-.426 ***
Age* LowEd	.006 **	.006 **	.006 **	.001	.002	.002
DistR*LowEd	.036	.037	-.034	.011	-.005	-.022
R^2	.217	.217	.218	.240	.240	.240
Numeracy	C1	C2	C3	C1	C2	C3
Age	-.007 ***	-.009	-.009	-.009 ***	-.013 *	-.013 *
DistR	-.016 ***	-.016 ***	-.016 ***	-.010 ***	-.007 ***	-.007 ***
LowEd	-.532 ***	-.531 ***	-.530 ***	-.572 ***	-.566 ***	-.566 ***
Age*LowEd	-.003	-.002	-.002	-.003	-.001	-.001
DistR*LowEd	.062	.060	.054	.069	.046	.046
R^2	.171	.171	.171	.153	.154	.153

Table 9: Intra-household learning effects.

	Pooled	HS dropouts	HS graduates
Orientation	.323 ***	.325 ***	.349 ***
Recall imm.	.276 ***	.261 ***	.294 ***
Recall del.	.306 ***	.277 ***	.343 ***
Fluency	.096 **	.038	.171 ***
Numeracy	.135 ***	.096	.192 ***

Figure 1: Age-profiles of average test scores by gender.

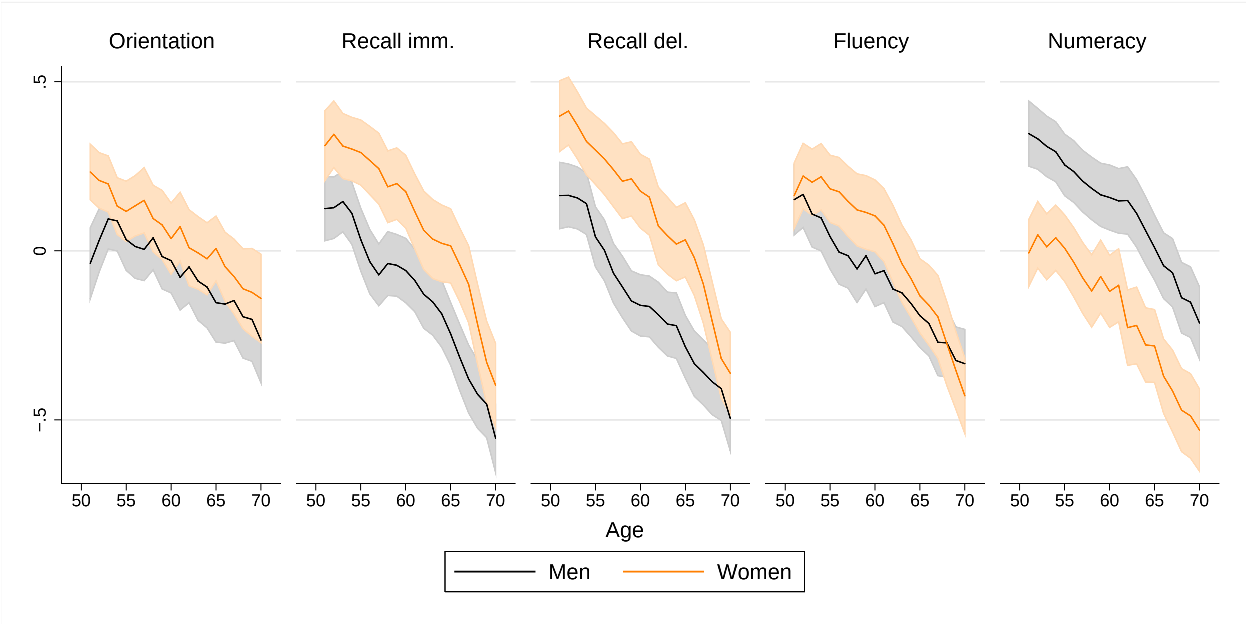


Figure 2: Age-profiles of average test scores by education level.

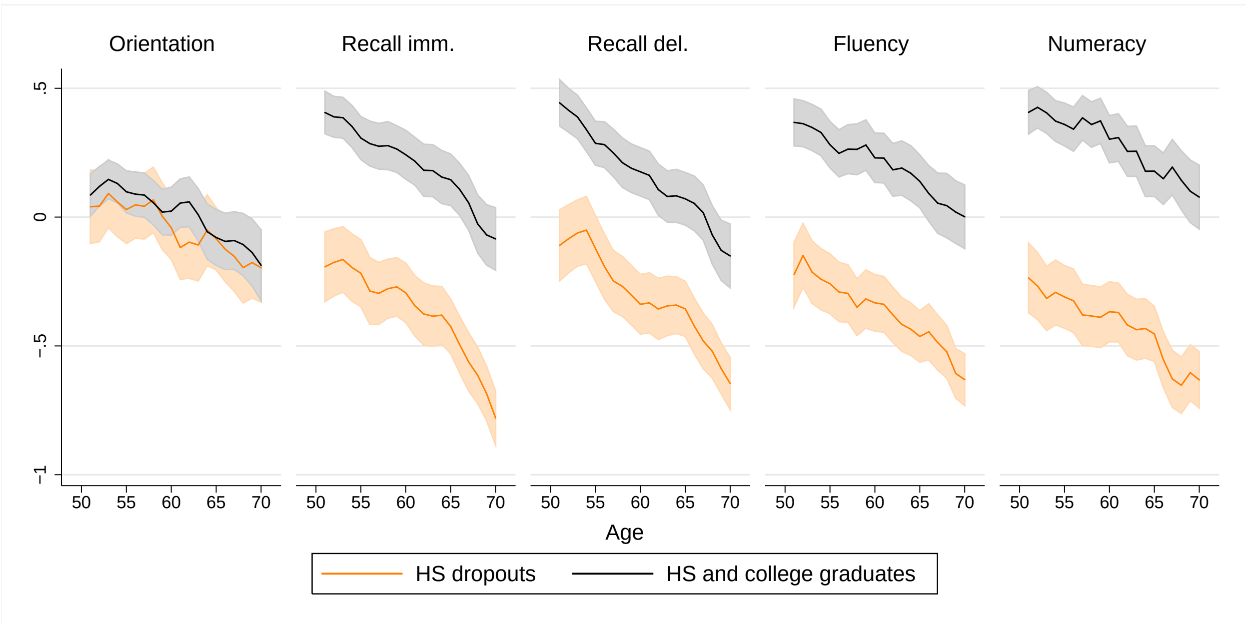


Figure 3: Age-profiles of average test scores by macro-region.

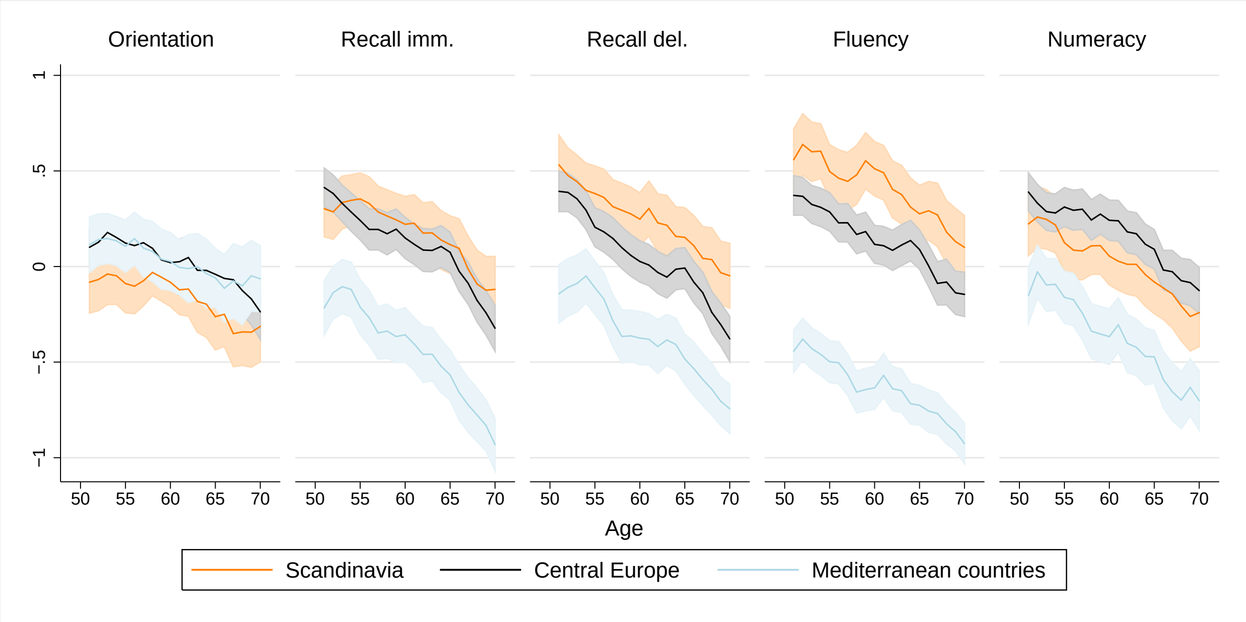


Figure 4: Age-profiles of average test scores by employment status.

Figure 5: Optimal path of cognitive stock over the life cycle implied by the model in Section 3.

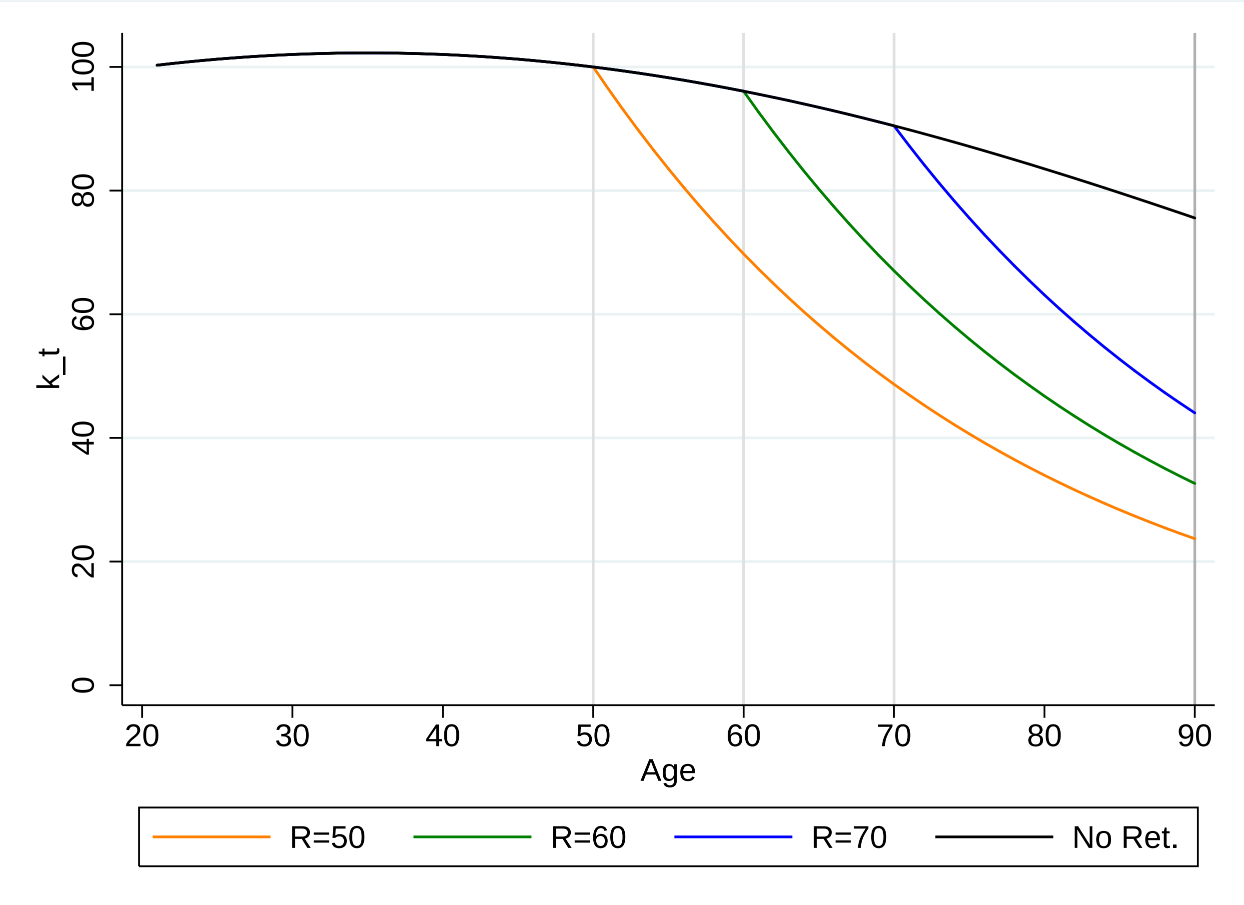


Figure 6: Predicted test scores at baseline by age, education and retirement at the age of 60.

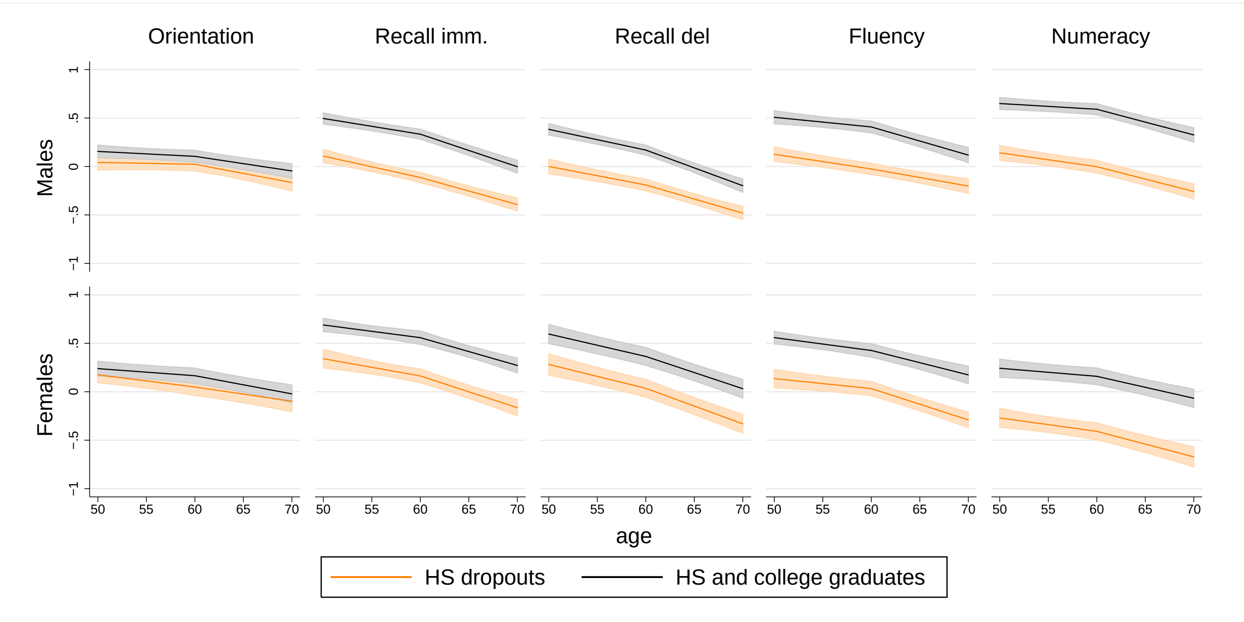


Figure 7: Age-profiles of predicted test scores by education level and macro-region and retirement at the age of 60.

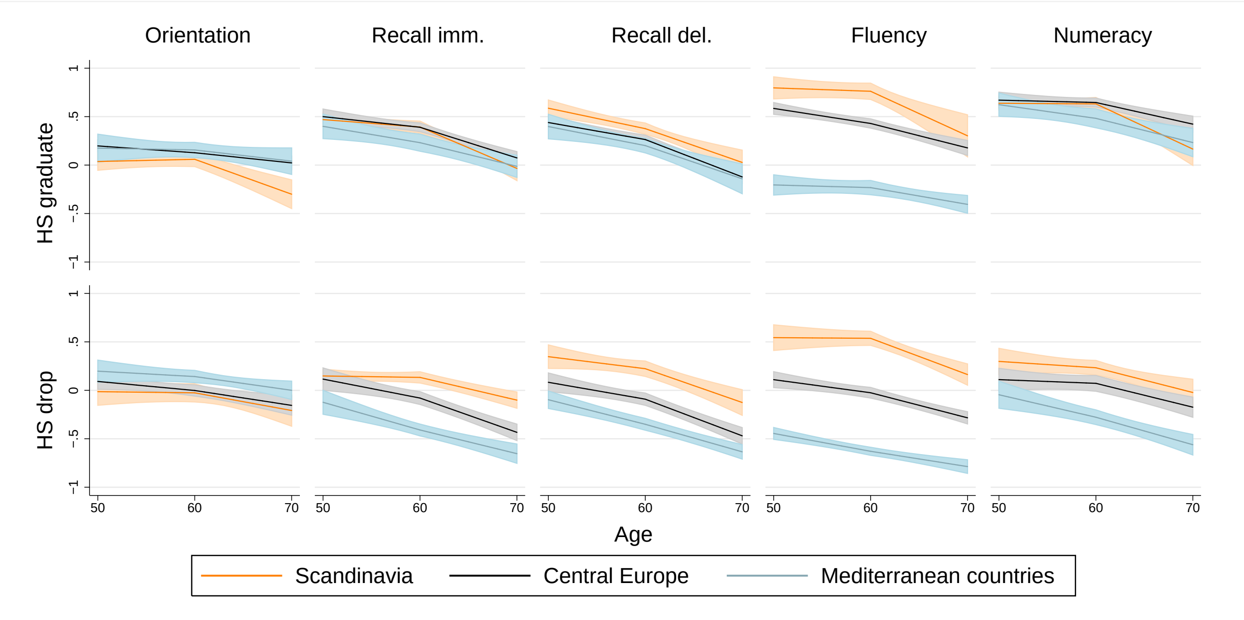


Figure 8: Early and normal eligibility ages for pension benefits, by country (males).

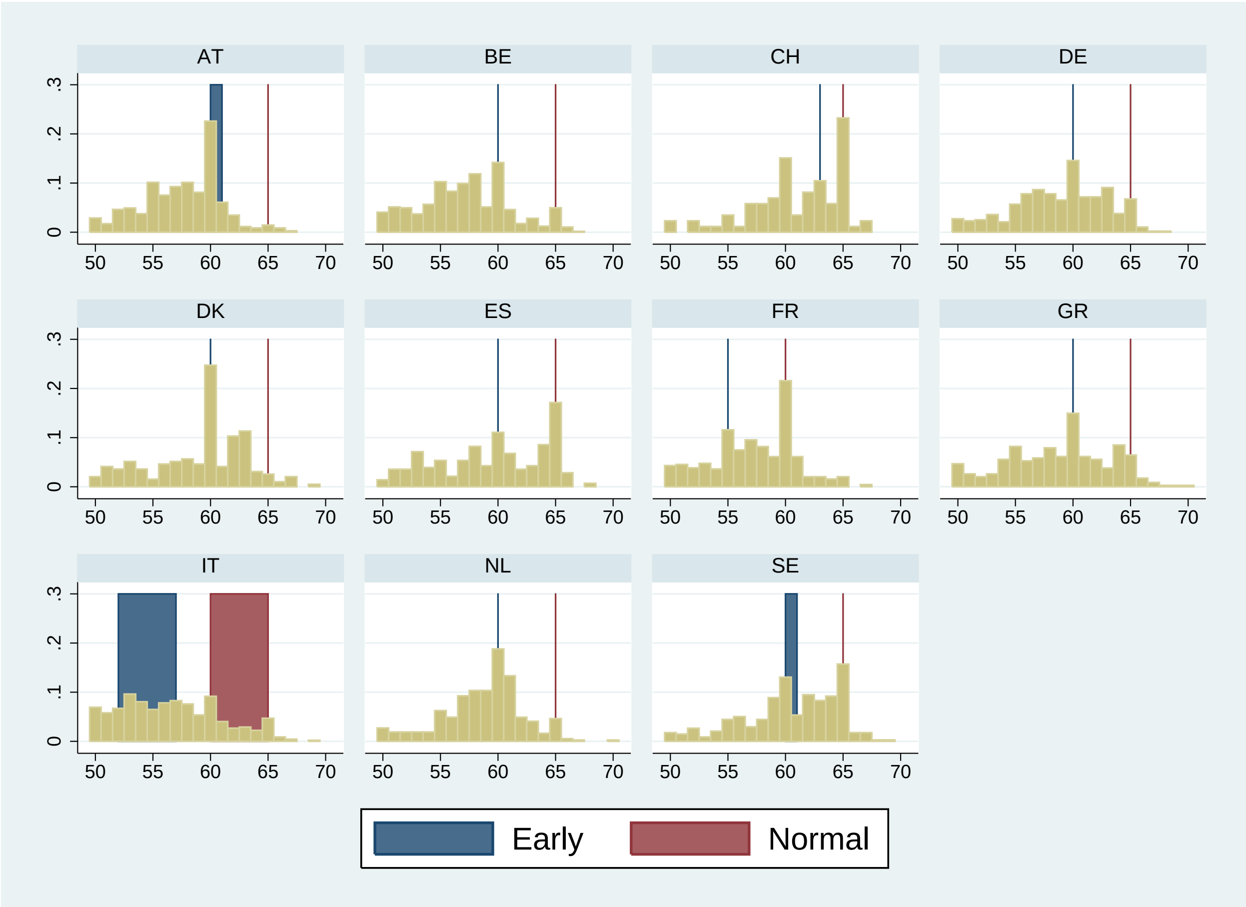
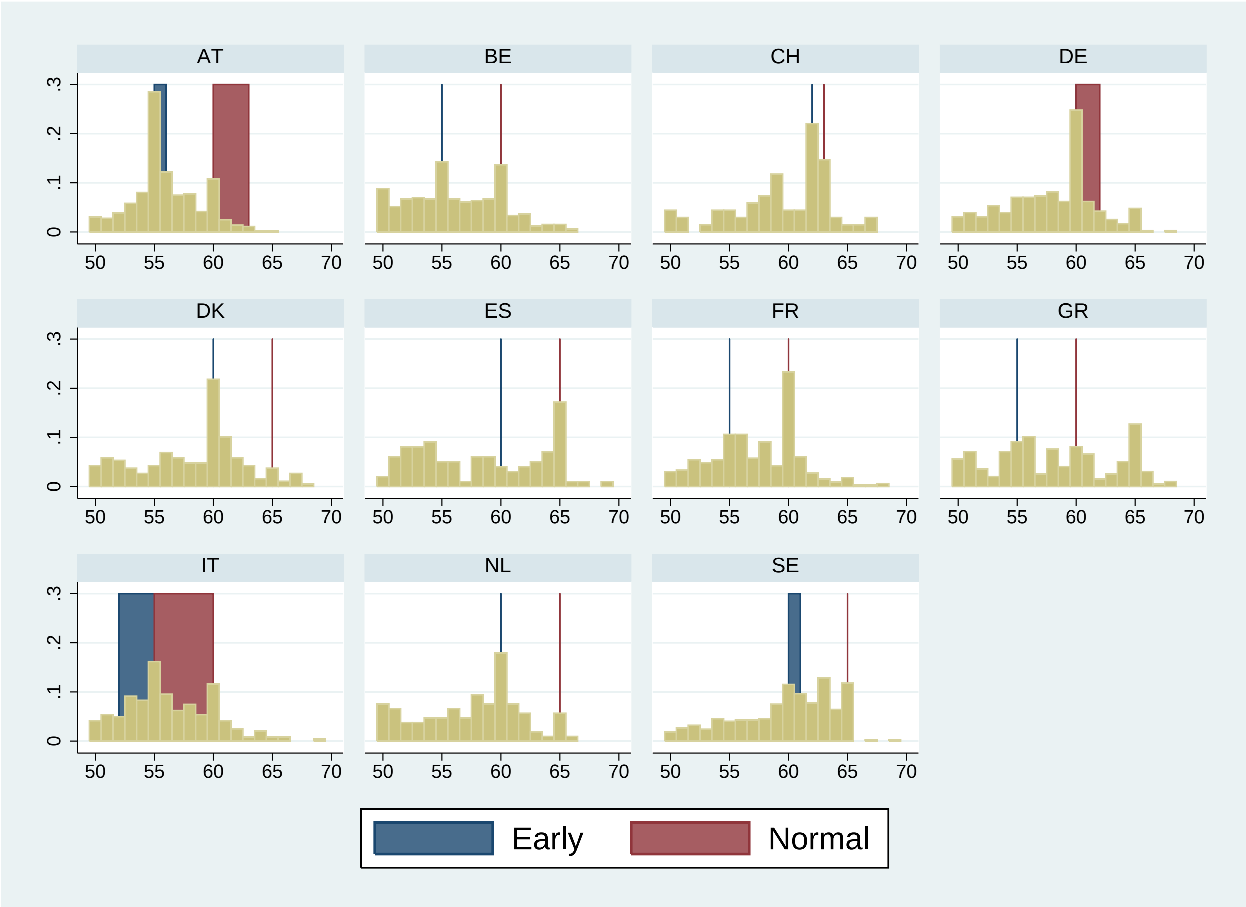


Figure 9: Early and normal eligibility ages for pension benefits, by country (females).



A Evolution of cognitive investment over time

The marginal productivity of cognitive investment at time t is

$$\frac{\partial Y_t}{\partial a_t} = \sum_{s=1}^{T-t} \frac{f_{t+s}}{(1+r)^s} \frac{\partial k_{t+s}}{\partial a_t}$$

with $f_{t+s} = F'_{t+s}(k_{t+s})$ and

$$\frac{\partial k_{t+s}}{\partial a_t} = \gamma_t \prod_{j=1}^{s-1} (1 - \delta_{t+j}),$$

where we used the fact that, from (1),

$$k_{t+s} = \sum_{h=0}^{s-1} \gamma_{t+h} a_{t+h} \left[\prod_{j=1+h}^{s-1} (1 - \delta_{t+j}) \right] + k_t \prod_{j=0}^{s-1} (1 - \delta_{t+j}).$$

It is not difficult to show that the following recursive relationship links $\partial Y_{t-1}/\partial a_{t-1}$ to $\partial Y_t/\partial a_t$

$$\frac{\partial Y_{t-1}}{\partial a_{t-1}} = \frac{\gamma_{t-1}}{1+r} \left[f_t + \left(\frac{1-\delta_t}{\gamma_t} \right) \frac{\partial Y_t}{\partial a_t} \right].$$

Substituting in the expression for $u_{a_{t-1}}$ obtained from (5) gives the following recursive relationship for the evolution of cognitive investment over time

$$u_{a_t} = \frac{\gamma_t}{\gamma_{t-1}} \frac{1+\rho}{1-\delta_t} u_{a_{t-1}} + \lambda \frac{\gamma_t}{1-\delta_t} (\pi_t - f_t) \left(\frac{1+\rho}{1+r} \right)^t, \tag{11}$$

where

$$\pi_t = \frac{p}{\gamma_{t-1}} \left(\frac{\gamma_t - \gamma_{t-1}}{\gamma_t} + r + \frac{\gamma_{t-1}}{\gamma_t} \delta_t \right)$$

Given diminishing marginal utility of cognitive investment, if γ_t is constant over time and $\pi_t = f_t$, then we should observe a steady decline of cognitive investment because $u_{a_t} > u_{a_{t-1}}$ since ρ and δ_t are both positive. This is also true after retirement, when $f_t = 0$. However, if γ_t is not constant over time, then we cannot make predictions about the evolution of a_t and, in particular, about the effect of retirement. In this case, the effect of retirement depends on the shape of the utility function and could be different for individuals with different preferences.

B The SHARE numeracy test

The set of questions asked in the SHARE numeracy test are:

1. *"If the chance of getting a disease is 10 percent, how many people out of one thousand would be expected to get the disease?"*

2. *"In a sale, a shop is selling all items at half price. Before the sale a sofa costs 300 Euro. How much will it cost in the sale?"*

3. *" A second hand car dealer is selling a car for 6,000 Euro. This is two-thirds of what it costs new. How much did the car cost new?"*

4. *" Let's say you have 2,000 Euro in a saving account. The account earns ten percent interest each year. How much would you have in the account at the end of two years?"*

All respondents start from question 1. If a respondent answers this question correctly, then she is asked 3. Otherwise, she is asked 2 and the test ends. If the respondent answers 3 correctly, then she is asked 4 and the test ends. Otherwise, the test ends with 3. For each question, interviewers are asked to code the answers provided by respondents on a grid of possible answers which always includes "other" as a category. The grid of possible answers is never shown to the respondent. The raw total score of this test is computed as follow. Answering 2 incorrectly gives a score of 0, while answering correctly gives a score of 1. Answering 3 incorrectly gives a score of 2, answering 4 incorrectly gives a score of 3, while answering 4 correctly gives a score of 4.

C Pension eligibility rules in the SHARE countries

The main source of information on early and normal ages of eligibility for public old-age pensions in the SHARE countries is the Mutual Information System on Social Protection (MISSOC) database. The MISSOC collects information on social protection for the member states of the European Union and other countries, including Switzerland. This source was supplemented with information from Gruber and Wise (2004).

For each country, we consider the different rules that affect the different cohorts of respondents. For Greece, we assume that the individuals in our sample first started working before 1992. Under this assumption, the early retirement age for Greece corresponds to eligibility for pension benefits with at least 35 years of contribution. In Italy, a sequence of pension reforms (Amato 1992, Dini 1995, Prodi 1995) changed repeatedly the criteria for eligibility. We conventionally assume that high-school dropouts started working at age 17 while people with at least a high-school degree started at age of 19, and both contributed continuosly to the public pension system until retirement.