Ageing, cognitive abilities and retirement*

Fabrizio Mazzonna
Munich Center for the Economics of Aging

Franco Peracchi
Tor Vergata University and EIEF

August 2009
This version: January 10, 2012

Abstract

We investigate the relationship between ageing, cognitive abilities and retirement using the Survey on Health, Ageing 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+. The human capital framework suggests that retirement may cause an increase in cognitive decline, since after retirement individuals lose the market incentive to invest in cognitive repair activities. Our empirical results, based on an instrumental variable strategy to deal with the potential endogeneity of retirement, confirm this key prediction. They also indicate that education plays a fundamental role in explaining heterogeneity in the level of cognitive abilities.

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

JEL codes: J14, J24.

* Corresponding author: Franco Peracchi, Faculty of Economics, Tor Vergata University, via Columbia 2, 00133 Rome, Italy (franco.peracchi@uniroma2.it). We thank two anonymous referees for their comments, and Josh Angrist, Anne Case, Angus Deaton, Maria Glymour, Giovanni Mastrobuoni, Chris Paxson and Facundo Piguillem 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. Franco Peracchi thanks the Center for Health and Wellbeing at Princeton University for generous hospitality during Fall 2010. This paper uses data from SHARE release 2. The SHARE data collection has been primarily funded by the European Commission through the 5th, 6th and 7th framework programmes. Additional funding from the U.S. National Institute on Aging as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions).

1 Introduction

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

A fundamental aspect of the ageing process is the decline of cognitive abilities. Schaie (1989) shows that cognitive functioning is 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 process of cognitive ageing is complex and 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. 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 suggests that accounting for the different dimensions of cognitive functioning may be important for an analysis of the process of cognitive ageing.

Another conceptual framework, due to Stern (2002), is that individuals have different levels of cognitive reserve and a higher level allows them to prevent or slow down the process of neurodegeneration associated with ageing. 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 their comprehensive review, Hertzog et al. (2008) describe how the age-profiles of cognitive abilities can differ over the life span in response to various types of behavior ('cognitive-enrichment hypothesis'). As revealed by many empirical studies, important factors in this process are education (Banks and Mazzonna 2011), occupational and retirement choices (Adam et al. 2006, Bonsang et al. 2010, Rohwedder and

Willis 2010), leisure activities (Scarmeas et al. 2003), home environment and parental influences in childhood (Cunha and Heckman 2007, Case and Paxson 2009) and adolescence (Richards et al. 2004), lifestyles (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 (Schaie 1989, Richards et al. 2004, Bonsang et al. 2010) do not account for sample selection due to attrition, a potentially serious problem in panels 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 functioning is fundamental for decision making, for it influences 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 ageing or can be controlled through work, study, and behavioral choices; and the degree to which depreciation is predictable or random".

Following the human capital approach, in this paper we adapt the model of health capital accumulation originally proposed by Grossman (1972) to derive a framework that allows us to better understand the link between cognitive abilities, ageing and retirement. One insight of this model is that the observed age-related decline in cognitive abilities need not be the same as natural deterioration, because people may respond to ageing by investing in cognitive repair activities. Another insight is that the amount of repair investment depends on market and non-market incentives, relative prices, discount rates, etc. In particular, the fact that retired individuals lose the market incentive to invest in repair activities may cause an increase in the rate of cognitive decline after retirement. In the empirical part of the paper, we employ 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 functioning based on simple tests of orientation in time, memory, verbal fluency and numeracy. Based on the predictions of our theoretical framework, our empirical specification accounts for the distance from retirement to capture the increase in the rate of cognitive decline after retirement. We also control for individual differences by gender, education and country of residence. A key issue is, of course, the endogeneity of retirement. We address this issue using an instrumental variables (IV) approach that exploits variation between and within countries in eligibility ages for early and normal retirement.

Recent papers by Bonsang et al. (2010) and Rohwedder and Willis (2010) also employ the SHARE data, along with other data sets, to estimate the causal effect of retirement on cognitive abilities using a somewhat similar IV approach. These papers lack a clear conceptual framework, which has important implications for their empirical strategy. First, they essentially consider retirement as a binary treatment that only causes a one-time shift in the level of cognitive abilities, with no effect on the slope of their age profile. As a result, they model its effects through a simple dummy variable for being retired. Second, the lack of a clear theoretical framework implies that important explanatory variables are omitted, while some of the included regressors are likely to be endogenous. For example, Rohwedder and Willis (2010) ignore important controls such as gender, education and country of residence. On the other hand, Bonsang et al. (2010) 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. Both papers also have specific limitations in their identification strategy. For example, when using the SHARE data, they only rely on the cross-country variation in eligibility ages for early and normal retirement at one point in time. Not only they do not exploit the substantial within-country variation arising from the pension reforms of the 1990s, but their cross-sectional variation is actually quite small, as more than half of the countries in their sample have the same eligibility ages.

The remainder of this paper is organized as follows. Section 2 describes the data used for this study. Section 3 describes our theoretical framework. Section 4 discusses features of the data that complicate identification of the causal effect of ageing 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 Munich Center 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. In Section 6.3 we also use the information from the third (2008) wave of SHARE on schooling achievements and household environment when the respondent was aged 10.

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 baseline (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), ageing 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. Of course, as suggested by one referee, the time it takes to answer a question may also be influenced by other factors, such as personality of the interviewer.

In this paper, we restrict attention to the countries that contributed to the 2004 baseline study. In Section 6.1 we also use the refreshment sample in the second wave, dropping Austria because no refreshment sample is available, while in Section 6.3 we use the additional information on school performance and early-life conditions available in the recently released third (2008) wave of SHARE, called SHARELIFE. Our working samples consist of individuals aged 50–70 at the time of their first interview, who answered the retrospective question on past employment status, reported being in the labour force at age 50, and classified themselves as employed, unemployed or retired.

These selection criteria give a sample of 13,753 individuals from the first wave and 4,445 individuals from the refreshment sample in the second wave. For 10,402 of them (nearly 58%), additional information is available in SHARELIFE. Table 1 shows the composition of our three working samples (baseline, refreshment and SHARELIFE) by country and gender.

Table 2 shows the distribution by labour force status of people aged 50–70 in the first wave or in the refreshment sample of the second wave of SHARE. We distinguish between people who were not selected into our working samples and people who were selected and were employed, unemployed or retired at the time of the interview. Due to the lower female attachment to the labour force, we end up undersampling women, especially in Greece, Italy, the Netherlands and Spain where more than half of them never worked. This is also likely to make our female sample somewhat special.

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 6.2.

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 baseline sample from 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 ageing 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 functioning at older ages (Banks and Mazzonna 2011). 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 for women, possibly due to the small number of employed women 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 as a guiding framework to understand the link between cognitive abilities, treated as unidimensional 'cognitive capital', ageing and retirement. One 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 ageing (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).

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 a sequence $\{(c_t, a_t)\}_{t=0}^{T}$ of consumption and cognitive investment to maximize her lifetime utility

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

where T is life length, assumed to be known, and ρ is the rate of time preference. The period utility function may depend on t and is assumed to be strictly increasing in both arguments with decreasing marginal utilities. Notice that, as in the standard model of consumer choice, preferences are defined over the two goods, consumption and cognitive investment. Although preferences may also depend on the stock K_t of cognitive capital, as in the original model by Grossman (1972), this simpler specification is enough for our purposes for, qualitatively, our main results do not change when we include the cognitive stock in the utility function.

In solving this problem, the individual takes as given her initial stocks of cognitive capital and of other assets, K_0 and 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, δ_t is the natural deterioration rate of cognitive capital, namely the rate at which it would deteriorate in the absence of repair investment, and $K_{T+1} = 0$ is the terminal condition. The concept that individuals must invest in cognitive-repair activities in order to offset the natural deterioration rate 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 second constraint forces repair investment to be nonnegative, which implies that

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

This constraint 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

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

where p_t is the price of cognitive repair investment, r is the real interest rate and $y_t = F_t(K_t)$ is earnings which, by analogy with the standard Mincerian formulation, depend on the cognitive stock at time t. We assume that the earnings production function F_t is strictly increasing, with diminishing marginal product $f_t = F_t'$.

Letting $u_{ct} = \partial u_t / \partial c_t$ and $u_{at} = \partial u_t / \partial a_t$, the first order conditions for an interior solution require the marginal rate of substitution (MRS) between cognitive investment and consumption to be equal to the effective price of cognitive investment in terms of foregone consumption

$$\frac{u_{at}}{u_{ct}} = MRS_t = p_t - \frac{\partial Y_t}{\partial a_t}, \tag{4}$$

where Y_t is the discounted value at time t of all subsequent earnings. If Y_t does not depend on a_t, then we have the familiar condition $MRS_t = p_t$. An example when this may happen is retirement. If $\partial Y_t / \partial a_t > 0$ 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. (2009), 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'), $u_{at} = 0$ for all t and so (4) is satisfied by setting $p_t = \partial Y_t / \partial a_t$. It can be shown that this condition is equivalent to the condition

$$\pi_t = f_t(K_t), \tag{5}$$

where $\pi_t = (1+r)p^*_{t-1} - (1-\delta_t)p^*_t$, with $p^*_t = p_t/\gamma_t$, may be interpreted as the user cost of cognitive capital. If p^*_t takes the constant value p^*, then $\pi_t = (r + \delta_t)p^*$. When f_t is strictly decreasing, an increase in π_t (due for example to an increase in the natural deterioration rate δ_t) causes K_t to fall.

On the other hand, at a boundary solution, 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) and Galama and Kapteyn (2011). We henceforth assume that (2) is not binding at $t = 0$. So, an important question is whether it may be binding later in life.

Consider first the pure investment model. If income depends on the level of cognitive abilities up to retirement but not afterwards 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,

$$K_t = \begin{cases} f_t^{-1}(\pi_t), & t \leq R, \\ (1 - \delta_{t-1})K_{t-1}, & t > R. \end{cases} \tag{6}$$

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, retiring respectively at age 50 (orange line), age 60 (green line), age 70 (blue line), and never retiring (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. An important implication of (6) is that simply introducing a retirement dummy is an inadequate way of modeling the effect of retirement on cognitive abilities.

Although the simplicity of (6) is lost when a_t enters the utility function, condition (4) allows for the possibility of an increase in the rate of decline of the cognitive stock after retirement. This rate of decline will be less than the pure deterioration rate δ_t, because there also non-market incentives to cognitive investment, and may also depend on γ_t, the efficiency of cognitive repair. 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.

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. For example, if education is associated with different preferences for cognitive investment, then this effect may vary across individuals depending on their educational attainments. Education may also increase the efficiency of cognitive investment, γ_t, and lower the rate of deterioration of cognitive abilities, δ_t, so more educated people may have a higher stock of cognitive capital throughout their life (Muurinen 1982).

4 Identification issues

This section discusses several important identification issues that arise when using the SHARE data to estimate models motivated by the theoretical framework in Section 3. The first issue is potential endogeneity of retirement (Section 4.1). The second is cohort heterogeneity in cognitive abilities, which complicates the interpretation of estimates from a single cross-section as we cannot easily separate pure ageing from cohort effects (Section 4.2). Other important issues are 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 (Section 4.3). Although these two issues severely limit the usefulness of the panel dimension of SHARE, they are completely ignored by Bonsang et al. (2010). Another neglected issue in the literature is heterogeneity in the effect of retirement and its dependence on educational and occupational choices (Section 4.4).

4.1 Endogeneity of retirement

Endogeneity of retirement represents 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 by using a standard 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 in Italy, while smaller changes occurred in most other countries, in particular for women. 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 men and women from a minimum of 60 to a maximum of 65 years. Together with these changes, during the 90's most of these countries also restricted other criteria for early retirement (e.g. year of contribution or definition of invalidity status) and eliminated financial incentives to retire. We refer to Appendix B 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 across cohorts in both initial conditions and mortality. The role of differences in initial cognitive endowment and early life-environment has recently stressed by Richards et al. (2004), Cuhna and Heckman (2007), Case and Paxson (2009) and Currie (2009). The role of difference in schooling quality has also been discussed (Alwin 1991). 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.

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 exactly the 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 respondents in the first wave, 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 the 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 baseline 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 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.

4.4 Heterogeneity of retirement effects

The effect of ageing and retirement on cognitive abilities may be heterogeneous across individuals. As suggested by both the available literature and the descriptive evidence in Section 2.3, this heterogeneity may depend not only on gender but also on the educational and occupational choices of an individual. In turn, this creates two kinds of problems. One is how to model the dependence of the effect of ageing and retirement on educational and occupational choices. The other is potential endogeneity of educational and occupational choices, that is, the fact that they may depend on unobservables that also affect test scores.

Given the close relationship between educational and occupational choices, which reflects the fact that they largely depend on the same set of unobservables ("latent ability"), it is unlikely that one needs to control for both of them. Thus, in Section 6.3, we control for latent ability by using the information from the third wave of SHARE on schooling achievements and household environment when the respondent was aged 10. Exploiting the available information on the early-life environment may also help address the concerns arising because of cohort heterogeneity.

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, for the reason already given in Section 2.3, as the age at which the last job ended or the current spell of unemployment began.

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. Some robustness checks are presented in Section 6.

5.1 OLS

After experimenting with polynomials of various order (linear, quadratic and cubic), we find that a linear age spline, namely a continuous piecewise-linear function of age with a single knot at the reported retirement age R_i, is systematically preferred by standard model selection criteria, such as AIC and BIC. Our basic model (Model A), fitted separately be gender and cognitive domain, is therefore of the form

$$Y_i = \alpha_0 + \alpha_1 Age_i + \alpha_2 DistR_i + \beta^\top X_i + U_i, \tag{7}$$

where Y_i is the standardized test score for the ith individual in the sample, 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 set of country dummies (with Belgium as the reference country), U_i is a regression error uncorrelated with Age_i and X_i but potentially correlated with $DistR_i$, and α_0, α_1, α_2 and β are parameters to be estimated. Notice that, conditional on U_i, the effect of one additional year of age on test scores is equal to α_1 up to retirement and to $\alpha_1 + \alpha_2$ after retirement.

Column A of Table 5 shows the OLS estimates of model (7). Estimated standard errors are robust to clustering at the country and cohort level. We find that the linear age term is statistically significant for all domains except orientation in time for men, and has the expected negative sign. Consistently with the prediction of the model in Section 3, the coefficient on $DistR_i$ is negative and statistically significant for all domains, indicating a sharp negative change in the slope of the age-profile after retirement.

Model B allows for more heterogeneity by adding a dummy for educational attainments, $LowEd_i$, treated here as an exogenous regressor. The dummy is equal to one for people with less than secondary education, and is equal to zero otherwise. Including the education dummy changes the estimated coefficients only little.

To evaluate the presence of heterogeneity across educational groups in the linear spline, Model C allows both the intercept and the slope parameters in (7) to differ depending on educational attainments. It turns out that the interaction between the 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 increased negative slope of the age-profile after retirement is evident for all domains. Although the effect of having completed high school is clearly positive for all cognitive domains, its magnitude differs 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, regional differences are now smaller, especially in the case of numeracy, delayed recall, and for more educated people. However, they persist and are still 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 to differences in the wage premium on cognitive skills.

5.2 2SLS

In this section we address the issue of potential endogeneity of the number of years spent in retirement, $DistR_i$, which arises if the age of retirement depends on unobservable variables that also affect the test scores. Our instruments for $DistR_i$ are $DistE_i = \max\{0, Age_i - E_i\}$ and $DistN_i = \max\{0, Age_i - N_i\}$, namely the positive part of the difference between the actual and the legislated ages of eligibility for early and normal retirement, E_i and N_i respectively. Similar instruments were also used by Angelini et al. (2009) to study the causal effect of early retirement on financial hardship. We treat educational attainments as exogeneous, and leave the issue of their potential endogeneity to Section 6.

Table 6 shows the estimated coefficients from various first-stage regressions for $DistR_i$. The table also shows the regression R^2 and the F statistic for the significance of the excluded instruments. Model 1A includes as regressors Age_i, $DistE_i$, $DistN_i$ and the country dummies, Model 1B adds the education dummy, while Model 1C further adds the interactions of Age_i, $DistE_i$, and $DistN_i$ with the education dummy. 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 $DistR_i$, the distance from retirement. They are also consistent with the evidence in Gruber and Wise (2004) on the importance of early retirement incentives, especially for men. Notice that introducing the education dummy in Model 1B does not affect the estimated coefficients on our two instruments. This indicates that, conditional on the other exogenous variables, there is very little correlation between education and our instruments.

Table 7 shows the estimated coefficient on $DistR_i$ from the second stage of the 2SLS procedure, along with the Sargan-Hansen J statistic for testing the overidentifying restrictions implied by the model. The estimates of Model 2A, which only includes the country dummies, confirm the negative effect of retirement on the age-profile of cognitive test scores. Further, the coefficients on $DistR$ are usually larger than those from OLS, except for delayed recall and fluency in the case of men.

After introducing the education dummy in Model 2B, we only observe a small and statistically insignificant decrease in the estimated effect of retirement, largely because of the small correlation of our two instruments with education. Finally, Model 2C fully interacts the linear spline in age with the education dummy. In addition to the linear age term and the country dummies, the set of instruments now includes $DistE_i$, $DistN_i$ and their interaction with the education dummy. For men, the results are quite similar across educational groups except for fluency. For women, 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 women. In fact, a standard test based on the difference between OLS and 2SLS estimates rejects the hypothesis of exogeneity of retirement at conventional significance levels only in the case of delayed recall for men, but always for women. Selection issues are likely to be the explanation of the big differences between men and women, as our sample selection criteria exclude a large fraction of women, especially in Mediterranean countries. This also implies that other papers that do not fully control for gender differences (e.g. Bonsang et al. 2010 and Rohwedder and Willis 2010) may offer biased estimates of the causal effect of retirement on cognitive abilities. Finally, it is worth noting that, except in the case of fluency and delayed recall for women, the J test never rejects the hypothesis that the instruments are valid.

6 Robustness checks

We now present the results of robustness checks for the presence of cohort effects (Section 6.1), intra-household learning (Section 6.2), and endogeneity of education (Section 6.3). We also carried out robustness checks for other aspects of model specification such as the use of raw rather than adjusted test scores, the use of quadratic rather than linear splines, and the inclusion of the unemployed with the employed rather than the retired, but did not find any significant departure from the basic results in Table 7. The results of these additional checks are available upon request.

6.1 Cohort effects

Table 8 shows the results of 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 ageing.

Our basic model is Model 2C, now estimated on the pooled sample. Model 2D adds to Model 2C 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 2E adds to Model 2D a time dummy. We find that controlling for cohort effects only slightly modifies the age-profiles of test scores. In particular, the coefficient on $DistR_i$ is now negative and statistically significant for all domains except orientation in time for women. The only noticeable difference is that, after controlling for cohort effects, the coefficient on age in the fluency and numeracy tests for men are no longer significant.

As a final check for the importance of cohort effects we also included in the model interactions between the age and the country dummies. The results, not reported to save space, 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 2C 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. If we assume that, given the regressors in Model 2C, the added indicator is conditionally independent of the unobservables that affect the test scores, then 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 higher for HS graduates. The effect seems to be stronger for orientation in time and the two memory tests. For example, HS graduates recall, on average, half a word more on the delayed recall test if they witnessed the interview of another household member.

6.3 Endogenous education

Endogeneity of education does not cause biases in the estimates from Model 2A, provided that one is willing to assume that heterogeneity in the effect of retirement, if there is any, is unrelated to education. The problem, in this case, is that 2SLS estimate a "local average treatment effect" (Imbens and Angrist 1994) which may or may not be an interesting parameter to consider.

Allowing for endogenous education in Model 2B, and especially in Model 2C where education is fully interacted with the age spline, is not so simple. We consider two alternative strategies. One strategy exploits variation in compulsory schooling laws across countries and cohorts. A similar strategy was used by Brunello, Fort and Weber (2009) to estimate the returns to education. We tried this strategy but, in our case, the weak first stage for education leads to unreliable estimates of the effect of retirement in the second stage.

The other strategy controls for latent ability and early-life conditions by using the additional information available in SHARELIFE. Specifically, we use the answers to the questions on the performance in math and language relative to other children in the same class, the number of books at home, and the occupation of the main breadwinner in the household. All these variables refer to the time when the person was 10 years old. The main problem with this strategy is the decreased sample size due to the large attrition between the first and the third wave, with an overall loss of about 40 percent of the baseline sample. To reduce the resulting loss of precision, we merge the SHARELIFE information with both the baseline sample from the first wave and the refreshment sample from the second wave, dropping individuals who could not be interviewed in SHARELIFE. Table 10 shows our results. Because the sample for this table differs from that employed for Table 7, we compare the estimates obtained when we do not use the information on initial conditions (Models 2B and 2C) and when we use it (Models 3B and 3C). Although the two sets of estimates are qualitatively very similar, the coefficients on $LowEd$ and its interaction with $DistR$ are somewhat smaller when we control for latent ability and early-life conditions.

7 Conclusions

In this paper we investigated 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 show an increase in the rate of decline of cognitive abilities after retirement. In the light of our theoretical framework, this reflects the reduced incentives to invest in cognitive repair activities after retirement. Our result has two important implications. First, incentives to early retirement and mandatory retirement rules both cause important losses of human capital because they make it less attractive to preserve the level of human capital inherited from the past. Second, the loss caused by retirement is not one-time, but increases with the length of the retirement spell.

We also find that education plays an important role 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.

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

Angelini V., Brugiavini A., and Weber G. (2009), "Ageing and unused capacity in Europe: is there an early retirement trap?" *Economic Policy*, 24: 463–508.

Banks J., and Mazzonna F. (2011), "The effect of education on old age cognitive abilities: Evidence from a regression discontinuity design." *Economic Journal*, forthcoming.

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

Bonsang E., Adam S., and Perelman S. (2010), "Does retirement affect cognitive functioning?" Netspar Discussion Paper 11/2010-069.

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).

Brunello G., Fort M., and Weber G. (2009), "Changes in compulsory schooling, education and the distribution of wages in Europe", *Economic Journal*, 119: 516–539.

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.

Eurostat (2008), "Ageing characterises the demographic perspectives of the European societies." *Statics in Focus*, 72/2008.

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: 189–198.

Galama T., and Kapteyn A. (2011), "Grossman's missing health threshold." *Journal of Health Economics*, 30: 1044–1056.

Galama T., Kapteyn A., Fonseca R., and Michaud P.-C. (2009), "Grossman's health threshold and retirement." RAND 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.

Imbens G.W., and Angrist J.D. (1994), "Identification and estimation of local average treatment effects." *Econometrica*, 62: 467-475.

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.

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.

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.

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: Baseline, refreshment and SHARELIFE sample sizes by country and gender.

		Baseline		Refreshment		SHARELIFE	
		Men	Women	Men	Women	Men	Women
AT	Austria	534	503	-	-	-	-
BE	Belgium	1,043	674	96	72	769	518
CH	Switzerland	282	230	222	217	342	333
DE	Germany	863	713	304	252	524	464
DK	Denmark	504	466	427	449	653	637
ES	Spain	547	284	145	75	354	194
FR	France	830	731	253	253	608	579
GR	Greece	846	413	307	158	855	418
IT	Italy	716	413	333	209	679	397
NL	Netherlands	822	453	257	178	606	359
SE	Sweden	887	999	110	128	511	602
Total		7,874	5,879	2,454	1,991	5,901	4,501

Table 2: Distribution by labour force status of people aged 50–70 in the first wave or in the refreshment sample of the second wave of SHARE (percent by gender).

	Men				Women			
	Not	Selected			Not	Selected		
	Selected	Empl.	Unempl.	Retired	Selected	Empl.	Unempl.	Retired
AT	6.6	32.6	4.0	56.9	28.8	19.7	3.1	48.5
BE	11.7	39.4	5.4	43.5	44.1	26.0	7.8	22.1
CH	7.5	64.1	2.2	26.3	30.1	47.8	2.4	19.8
DE	6.3	41.1	8.4	44.2	28.9	34.4	6.0	30.7
DK	6.7	56.3	6.9	30.1	15.8	49.2	5.7	29.4
ES	10.2	42.9	5.3	41.6	65.2	22.0	5.0	7.8
FR	6.7	43.3	5.8	44.2	29.1	39.0	4.8	27.6
GR	5.7	55.7	2.3	36.3	55.1	22.3	2.1	20.4
IT	8.0	33.8	3.6	54.5	57.3	17.1	1.5	24.2
NL	13.4	47.2	3.4	36.0	58.0	31.4	2.1	8.5
SW	6.1	57.6	3.5	32.8	11.2	55.1	2.8	30.9
Total	8.3	45.8	4.7	41.2	39.8	32.0	4.1	24.2

Table 3: Mean and standard deviation (S.D.) of raw and adjusted cognitive scores.

Raw scores	Mean	S.D.	Correlations				
Orientation	3.88	.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**	**S.D.**	**Correlations**				
Orientation	3.50	.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. For each gender and age group, the reference is people retired by 5 years or less (*: p-values between 10 and 5 percent, **: p-values between 5 and 1 percent, ***: p-values below 1 percent).

	Aged 60–65			Aged 66–70		
Men	Empl.	Ret.$\leq$ 5 yr	Ret.$>$ 5 yr	Empl.	Ret.$\leq$5 yr	Ret.$>$ 5 yr
Orientation	3.53 ***	3.45	3.41	3.38	3.47	3.41 **
Recall imm.	4.79 ***	4.48	4.22 ***	4.96 ***	4.16	4.13
Recall del.	2.42 ***	2.20	2.07 **	2.53 ***	1.96	2.07 *
Fluency	19.93 *	19.27	18.01 ***	20.73 ***	17.99	18.11
Numeracy	3.28 ***	2.75	2.64	3.33 ***	2.72	2.50 ***
Women	**Empl.**	**Ret.$\leq$ 5 yr**	**Ret.$>$ 5 yr**	**Empl.**	**Ret.$\leq$5 yr**	**Ret.$>$ 5 yr**
Orientation	3.55 ***	3.48	3.55 ***	3.51	3.50	3.45
Recall imm.	5.18 ***	4.78	4.73	4.19	4.15	4.45 ***
Recall del.	2.04	2.02	1.92 *	1.56	1.71	1.77
Fluency	20.58 **	19.68	18.74 **	18.73 *	16.62	17.62 *
Numeracy	3.63 **	3.40	3.16 **	2.71	2.99	2.83

Table 5: OLS estimates for the level of the test scores in the baseline sample (p-values are as in Table 4).

	Men			Women		
	A	B	C	A	B	C
Orientation						
Age	-.005 *	-.004	-.005 *	-.011 ***	-.010 ***	-.007 **
$DistR$	-.014 ***	-.013 ***	-.010 **	-.007 *	-.007 *	-.011 **
$LowEd$		-.102 ***	-.082 ***		-.086 ***	-.117 ***
$Age * Educ$			.003			-.005
$DistR * LowEd$			-.007			.009
R^2	.033	.035	.035	.037	.039	.039
Recall imm.						
Age	-.023 ***	-.019 ***	-.016 ***	-.020 ***	-.015 ***	-.013 ***
$DistR$	-.015 ***	-.012 ***	-.017 ***	-.018 ***	-.016 ***	-.016 **
$LowEd$		-.413 ***	-.447 ***		-.390 ***	-.395 ***
$Age * Educ$			-.006			-.005
$DistR * LowEd$			.012 **			.001
R^2	.110	.150	.150	.121	.157	.157
Recall del.						
Age	-.024 ***	-.021 ***	-.021 ***	-.028 ***	-.024 ***	-.023 ***
$DistR$	-.015 ***	-.012 ***	-.015 ***	-.014 ***	-.011 ***	-.010 *
$LowEd$		-.345 ***	-.360 ***		-.333 ***	-.329 ***
$Age * Educ$			.002			-.002
$DistR * LowEd$			.005			-.002
R^2	.095	.121	.121	.107	.131	.131
Fluency						
Age	-.016 ***	-.012 ***	-.010 ***	-.017 ***	-.012 ***	-.013 ***
$DistR$	-.014 ***	-.011 ***	-.019 ***	-.020 ***	-.017 ***	-.012 **
$LowEd$		-.386 ***	-.434 ***		-.424 ***	-.393 ***
$Age * Educ$			-.005			.003
$DistR * LowEd$			.017 ***			-.010
R^2	.188	.218	.218	.198	.238	.238
Numeracy						
Age	-.015 ***	-.009 ***	-.006 **	-.018 ***	-.011 ***	-.008 **
$DistR$	-.020 ***	-.016 ***	-.021 ***	-.018 ***	-.014 ***	-.014 **
$LowEd$		-.565 ***	-.593 ***		-.558 ***	-.568 ***
$Age * Educ$			-.008 *			-.006
$DistR * LowEd$			.009			.002
R^2	.115	.185	.185	.098	.169	.169
N	7,874	7,874	7,874	5,879	5,879	5,879

Table 6: OLS estimates from the first stage regression for $DistR = \max\{0, Age - R\}$ (p-values are as in Table 4).

Men	1A	1B	1C
$DistE$	.476 ***	.474 ***	.467 ***
$DistN$	.166 ***	.166 ***	.248 ***
$DistE * LowEd$			-.011
$DistN * LowEd$			-.140
R^2	.571	.571	.572
F stat.	201.30 ***	197.45 ***	74.88 ***
Women	**1A**	**1B**	**1C**
$DistE$	.352 ***	.354 ***	.353 ***
$DistN$	.322 ***	.318 ***	.420 ***
$DistE * LowEd$			-.033
$DistN * LowEd$			-.171 **
R^2	.613	.614	.616
F stat.	205.45 ***	201.38 ***	67.34 ***

Table 7: 2SLS coefficients on $DistR$ and its interaction with $LowEd$ (the asymptotic null distribution of the Sargan-Hansen J statistic for testing the overidentifying restrictions is χ^2_1 for Models 2A and 2B and χ^2_2 for Model 2C; p-values are as in Table 4).

	Men			Women		
	2A	2B	2C	2A	2B	2C
Orientation						
$DistR$	-.032 ***	-.031 ***	-.028 ***	-.005	-.003	-.007
$DistR * LowEd$			-.005			.010
J stat.	.71	.69	1.17	.47	.46	.53
Recall imm.						
$DistR$	-.025 ***	-.018 **	-.022 **	-.055 ***	-.051 ***	-.018
$DistR * LowEd$			.012			-.075 ***
J stat.	1.38	1.57	2.12	4.26	1.41	.95
Recall del.						
$DistR$	.009	.015 *	.015	-.029 ***	-.025 ***	-.003
$DistR * LowEd$			-.003			-.050 ***
J stat.	0.61	.72	.81	4.26 **	5.26 **	4.51
Fluency						
$DistR$	-.011	-.006	.003	-.025 **	-.023 **	.003
$DistR * LowEd$			-.027			-.055 ***
J stat.	3.51 *	4.22 **	4.24	5.01 **	6.34 **	8.99 **
Numeracy						
$DistR$	-.038 ***	-.029 ***	-.030 ***	-.046 ***	-.041 ***	-.014
$DistR * LowEd$			.005			-.059 ***
J stat.	.05	.11	1.08	.50	.87	.50
N	7,874	7,874	7,874	5,879	5,879	5,879

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 (p-values are as in Table 4).

	Men			Women		
	2C	2D	2E	2C	2D	2E
Orientation						
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.						
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.						
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						
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						
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
N	10,328	10,328	10,328	7,870	7,870	7,870

Table 9: Intra-household learning effects (p-values are as in Table 4).

	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 ***

Table 10: 2SLS coefficients on *DistR*, *LowEd* and their interaction using SHARELIFE and raw rather than adjusted test scores (Models 2B and 2C are as in Table 7; Models 3B and 3C include controls for initial conditions: math and language performance at age 10, number of books at home, and occupation of the main breadwinner in the household; p-values are as in Table 4).

	Men				Women			
	2B	2C	3B	3C	2B	2C	3B	3C
Orientation								
$DistR$	.002	.007	.002	.008	-.001	-.004	-.001	-.004
$LowEd$	-.113 ***	-.051	-.084 ***	-.019	-.073 ***	-.117 **	-.060 ***	-.108 **
$DistR * LowEd$		-.018		-.019		.010		.011
Recall imm.								
$DistR$	-.015	-.010	-.015 *	-.010	-.025 ***	-.010	-.022 **	-.010
$LowEd$	-.397 ***	-.387 ***	-.290 ***	-.282 ***	-.433 ***	-.254 ***	-.322 ***	-.164 **
$DistR * LowEd$		-.006		-.005		-.044 **		-.038 **
Recall del.								
$DistR$	.015	.029 *	.015	.029 *	-.021 **	-.005	-.017 *	-.004
$LowEd$	-.356 ***	-.270 ***	-.255 ***	-.168 ***	-.341 ***	-.157 **	-.225 ***	-.061
$DistR * LowEd$		-.029		-.029 *		-.045 ***		-.040 **
Fluency								
$DistR$	-.003	-.001	-.004	-.002	-.020 **	-.008	-.016 *	-.005
$LowEd$	-.391 ***	-.335 ***	-.236 ***	-.181 ***	-.419 ***	-.284 ***	-.256 ***	-.142 **
$DistR * LowEd$		-.014		-.015		-.034 **		-.029 *
Numeracy								
$DistR$	-.024 ***	-.032 ***	-.025 ***	-.030 ***	-.034 ***	-.021 *	-.031 ***	-.020 *
$LowEd$	-.507 ***	-.575 ***	-.343 ***	-.397 ***	-.559 ***	-.448 ***	-.406 ***	-.323 ***
$DistR * LowEd$		.021		.016		-.030		-.023
Initial conditions	No	No	Yes	Yes	No	No	Yes	Yes
N	5,901	5,901	5,901	5,901	4,501	4,501	4,501	4,501

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

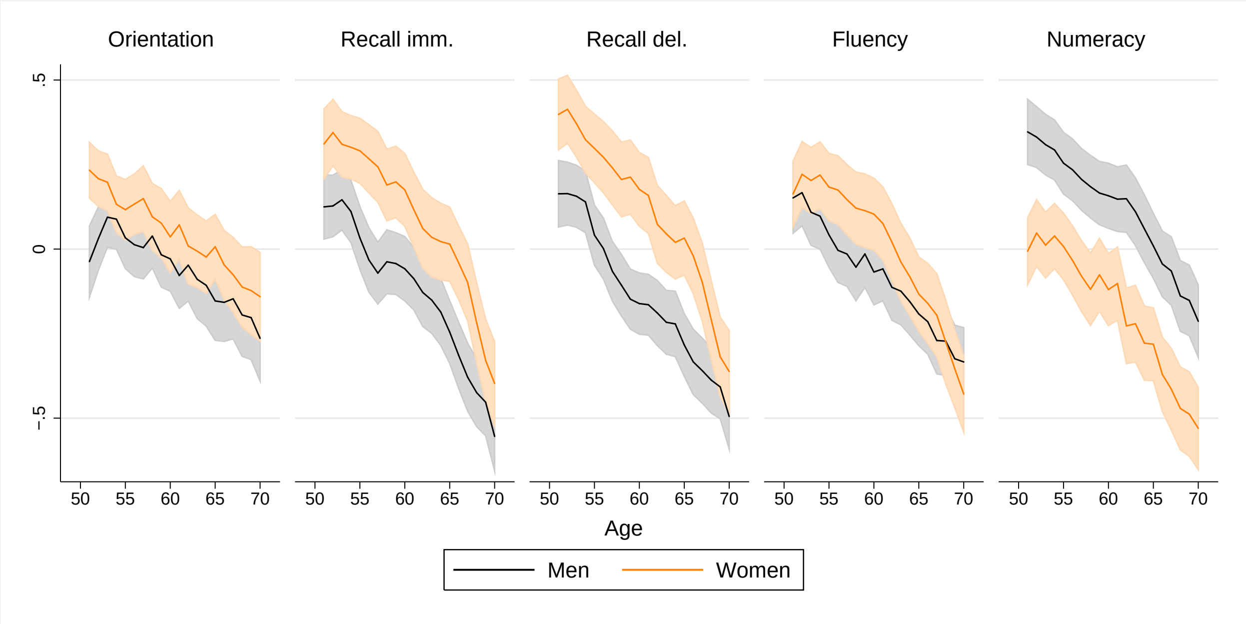


Figure 2: Age-profiles of average test scores by education level.

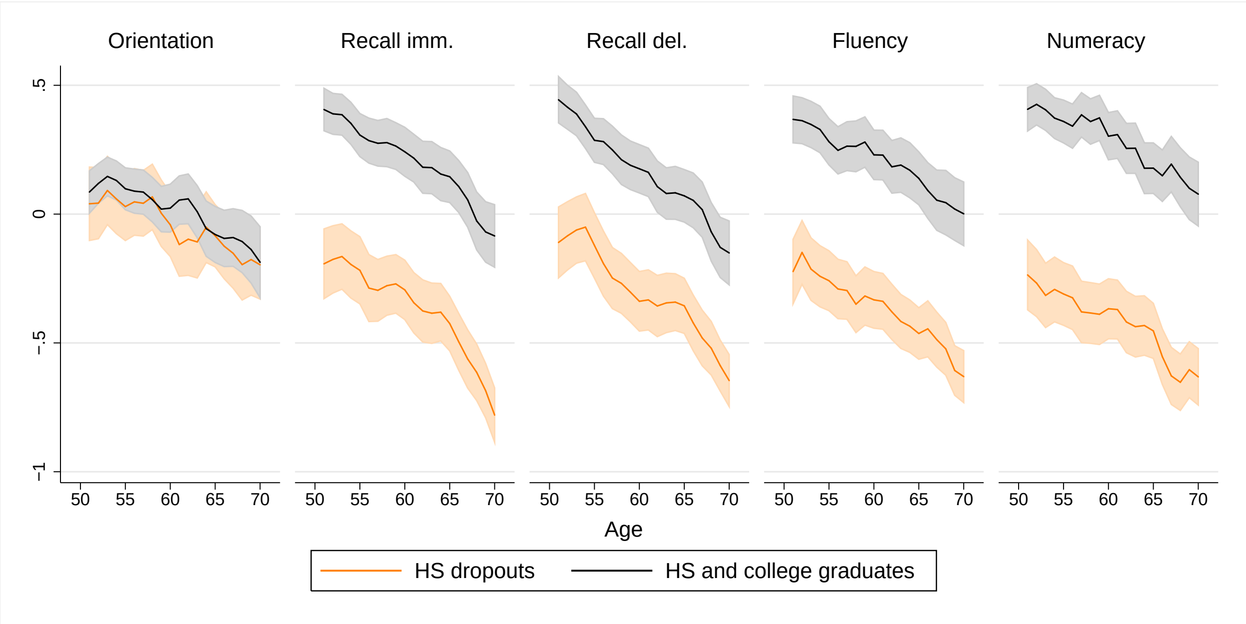


Figure 3: Age-profiles of average test scores by macro-region.

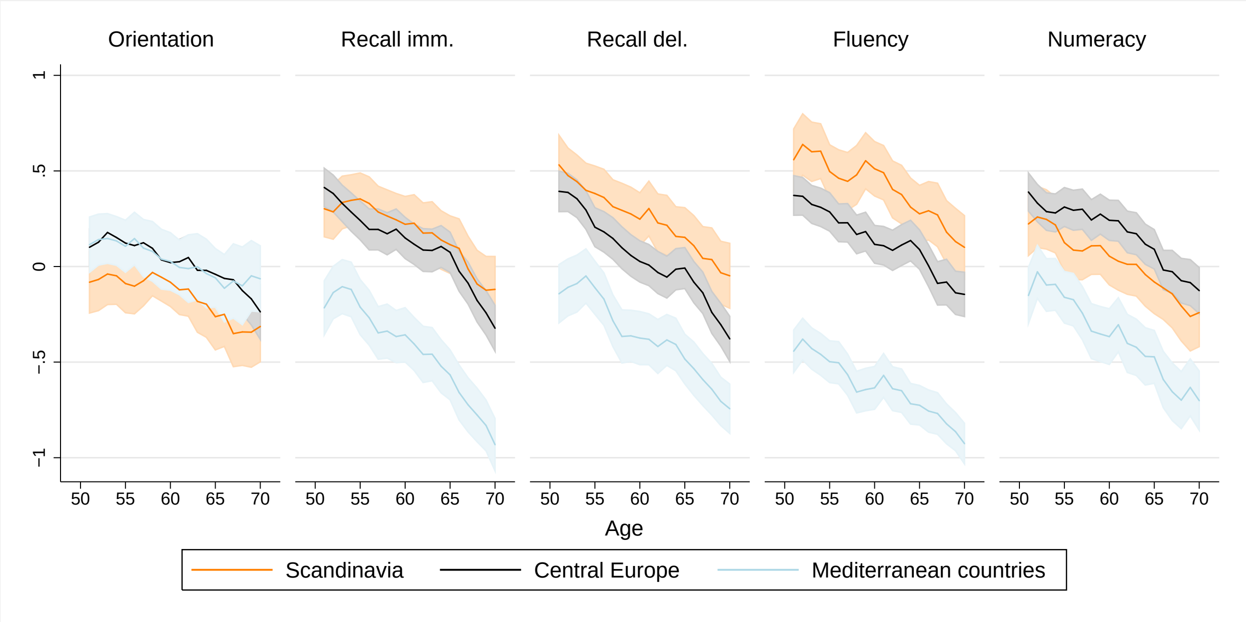


Figure 4: Age-profiles of average test scores by employment status.

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

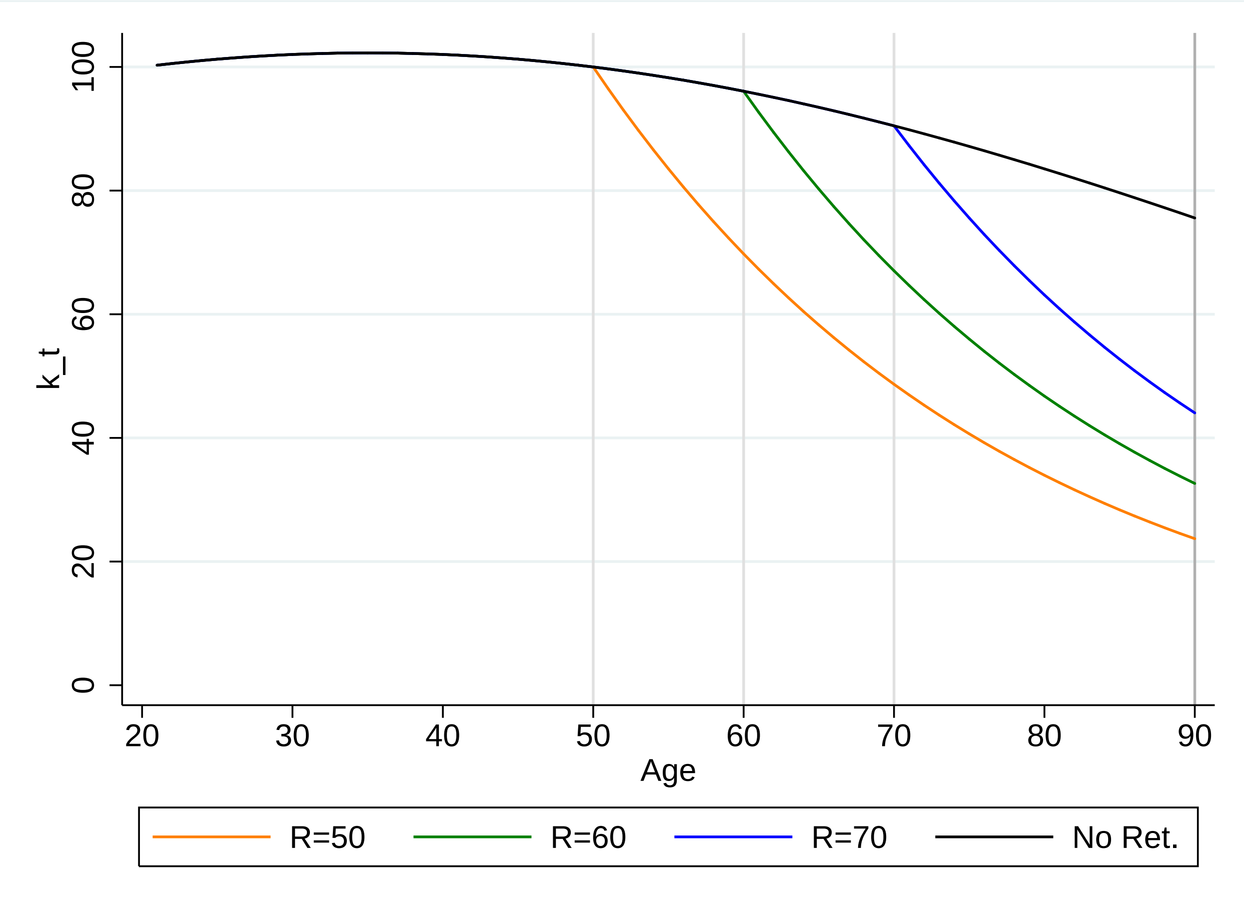


Figure 6: Predicted test scores at baseline by age, education and retirement at the age of 60.

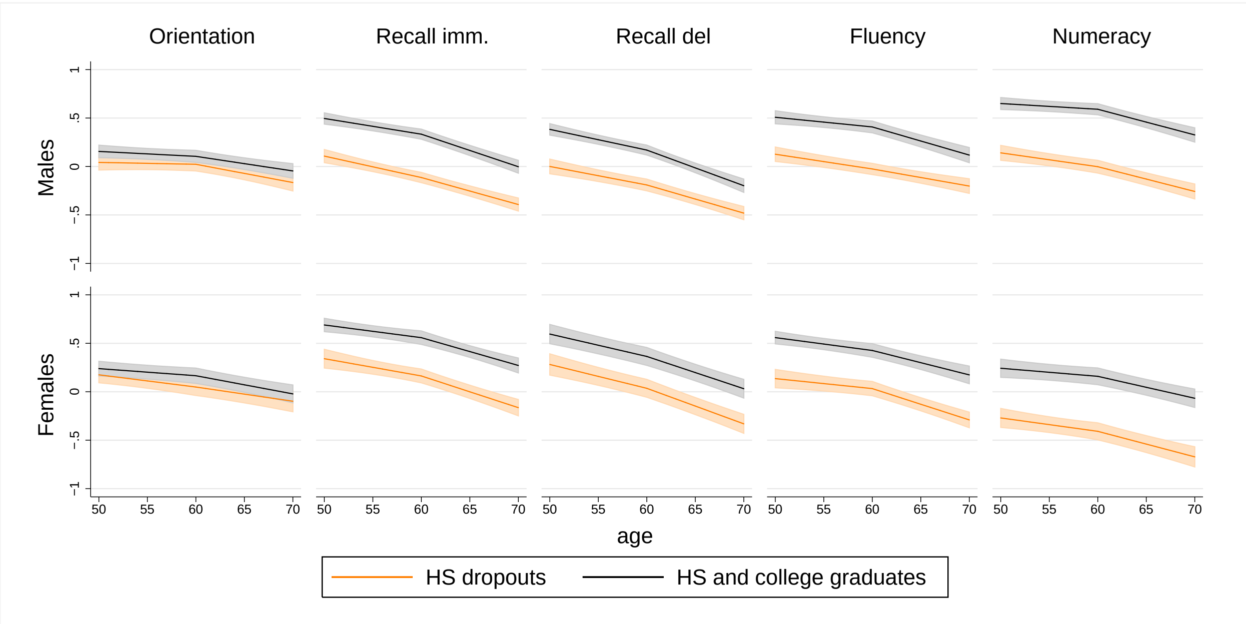


Figure 7: Age-profiles of predicted test scores by education level and macro-region and retirement at the age of 60.

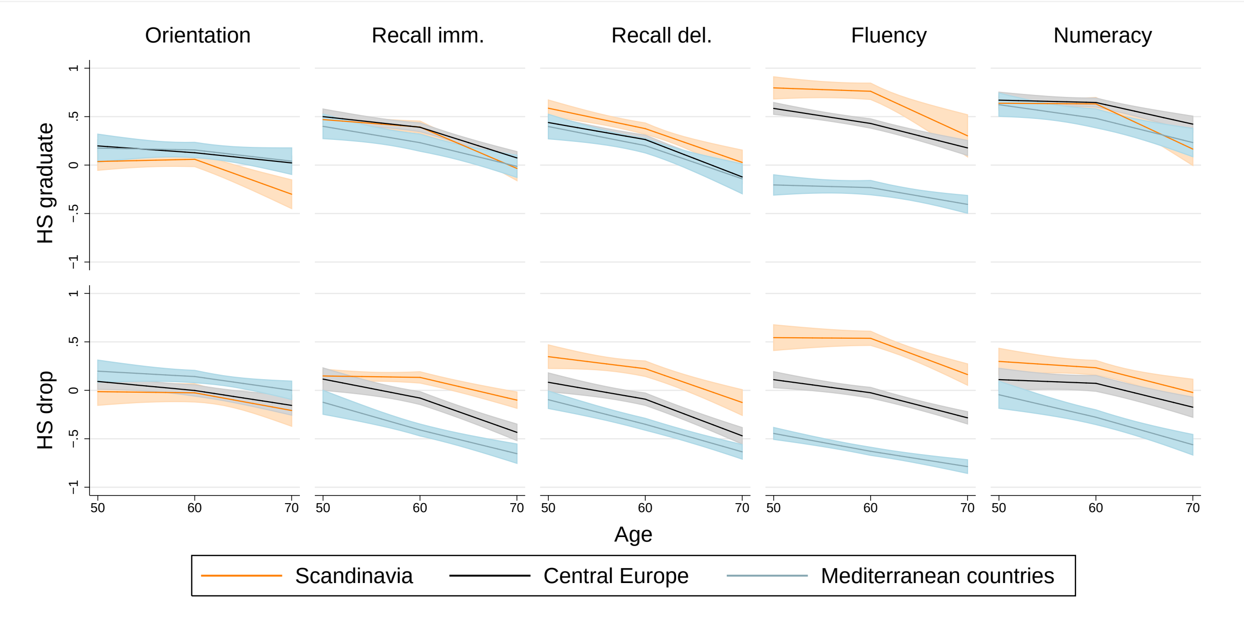


Figure 8: Early and normal eligibility ages for pension benefits, by country (men).

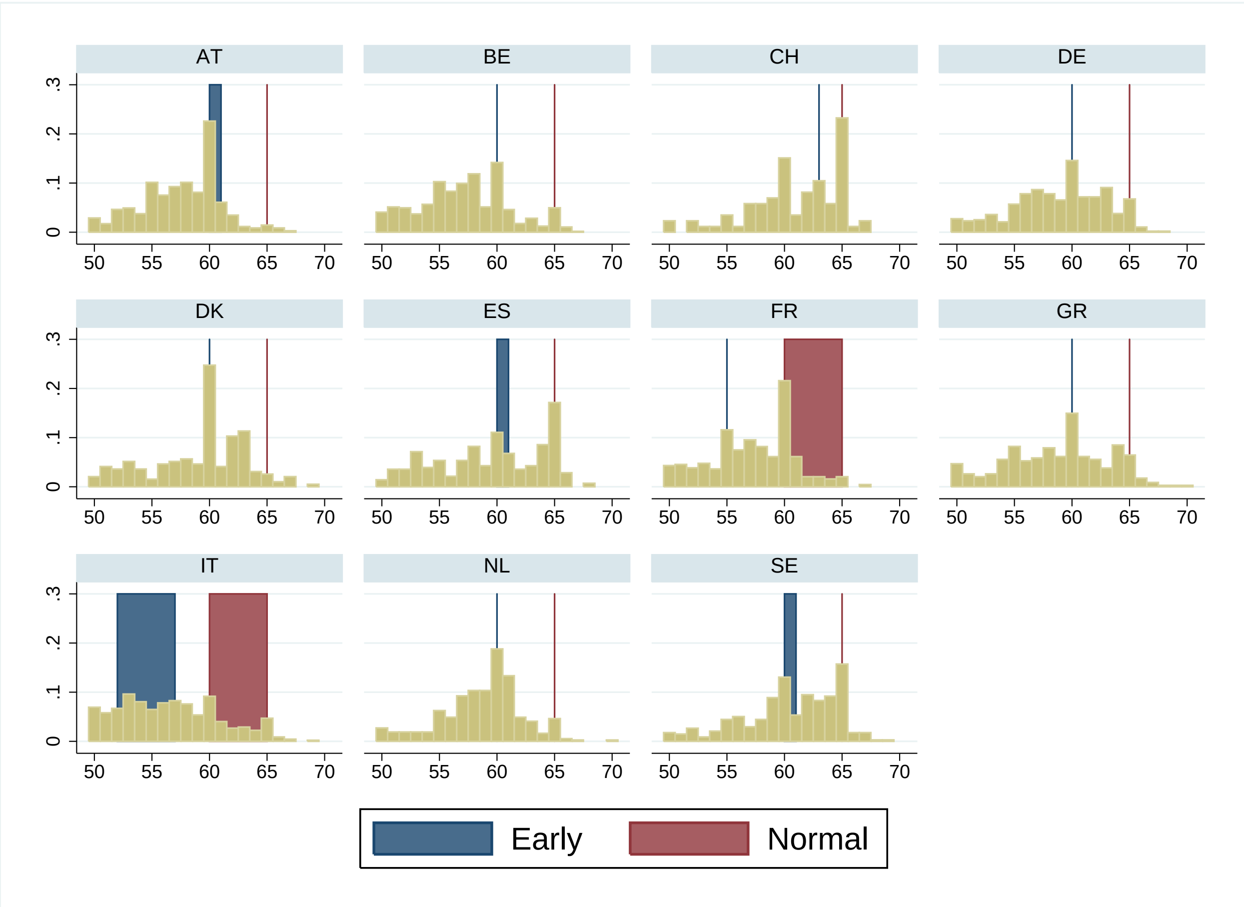
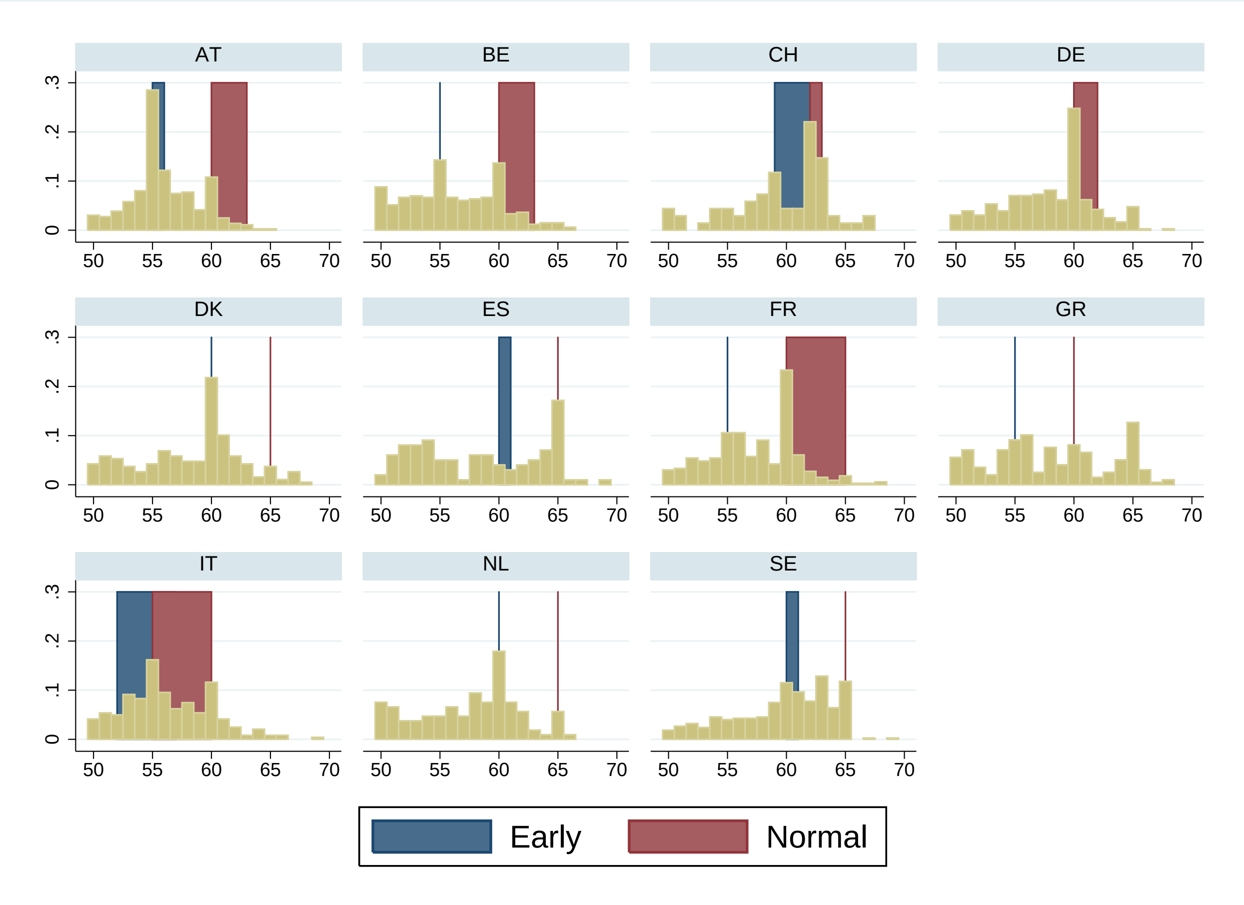


Figure 9: Early and normal eligibility ages for pension benefits, by country (women).



A 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.

B 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) and Angelini et al. (2009).

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 (in 1992, 1995 and 1997) changed repeatedly the criteria for eligibility.