National variation in cognitive life cycle development

Vegard Skirbekk*, Daniela Weber+, and Valeria Bordone^

* International Institute for Applied Systems Analysis (IIASA) - World Population Program, Laxenburg (AT).
vegard.skirbekk@iiasa.ac.at
+ International Institute for Applied Systems Analysis (IIASA) - World Population Program, Laxenburg (AT).
daniela.weber@iiasa.ac.at
^ Vienna University of Economics and Business - Wittgenstein Centre for Demography and Global Human Capital, Vienna (AT).
valeria.bordone@wu.ac.at

Paper prepared for the European Population Conference (EPC) 2012

ABSTRACT

We study life cycle variation in cognitive abilities for nationally representative cohorts for the fifty plus populations in eight countries born from 1944 to 1952 - we observe their cognitive abilities in their teenage years and again when they are past 50 years of age. We use standardized test scores from 1964 IEA data on school performances of pupils of two age groups (around 13 years of age and around 19 years of age, for males and females belonging to the 1949-1952 and 1944-1947 cohorts, respectively). We observe these cohorts again at older ages in the mid-2000s (data from ELSA, HRS, JSTAR, and SHARE surveys that include standardized tests of cognition). We consider math tests both at younger and older ages and also analyze other cognitive tests for the seniors. Sweden along with the US improves their country-ranking-ordering the most across the life cycle, while Belgium and France decline the most.

Identifying ways to improve cognitive functioning among seniors is one of the greatest challenges to improve independence and productive capacities among elderly individuals in aging societies. It is therefore necessary to first recognize how cognition is shaped over the life cycle.

Several studies point that some cognitive changes tend to occur over the life cycle, where fluid abilities (including perceptual speed and reasoning abilities) decline considerably, while crystallized abilities (including vocabulary size and semantic meaning) remain more stable to older ages (Lindenberger & Ghisletta, 2009; K. W. Schaie & Willis, 1986; Warr, 1994). Most research on the determinants of cognitive ageing has focused on individual level characteristics, including medical conditions, such as diabetes or the APOE-4 genotype or injuries (Williams, Plassman, Burke, Holsinger, & Benjamin, 2010), and behavioral determinants, such as degree of cognitive engagement, educational and nutritional intake (Engelhardt, Buber, Skirbekk, & Prskawetz, 2008; Williams, et al., 2010). A relatively high degree of temporal stability in the rank-ordering of cognition has been found in several longitudinal studies. In the USA, general mental ability tests from ages 5 and 16 (Weinert & Schneider, 1999; Yule, Gold, & Busch, 1982), and 11 subtests of fluid and crystallized intelligence over a 12-year mid-life period (Schneider & Bullock, 2008; Zimprich & Mascherek, 2010) showed such a stability. This result has been confirmed in Europe: on Scottish data from age 11 to young adulthood and later life (Deary, Whiteman, Starr, Whalley, & Fox, 2004), and on British data for ages 8 to 17 and even stronger between 14 and 17 years old for general mental ability tests (Hindley & Owen, 1978). Similarly, population subgroups may also maintain constant differences (e.g. women vs. men (Blum, Fosshage, & Jarvik, 1972), or subgroups of high and low performers (Facon, 2008).

Still, so far there has been relatively little emphasis on how well people age cognitively over the life cycle across countries. Yet, country specific influences which may affect cognitive maintenance may arise from social structures, lifestyles and cultures, opportunities for physical and mental activity levels, environmental influences at the country level, as well as educational and employment structures, including variation in the length of the working life (Aneshensel, Ko, Chodosh, & Wight, 2011; Booth, Owen, Bauman, Clavisi, & Leslie, 2000; Ceci, 1991; Chee, Zheng, Goh, Park, & Sutton, 2010; Glymour, Kawachi, Jencks, & Berkman, 2008; Voelcker-Rehage, Godde, & Staudinger, 2011). Most existing international comparisons of cognitive skills are focusing on human capital and cognition of children and youth (Gonzales et al., 2008; Programme for International Student Assessment, 2009; Rindermann, 2007). However, they are not informative about the change or maintenance of cognitive abilities over the life cycle, which also is of high importance.

In this paper, we follow two cohort-groups identified in terms of cognitive performance for a selection of European countries, Japan, and the US over a 40-year life span. We collect and analyze nationally representative cross-sectional surveys of comparable measures of cognitive skills at younger and older ages. We first observe data from a standardized test from the early 1960s. Then, we observe nationally representative samples of the same cohorts in 2004 (2007 for Japan), who participated into survey interviews with up to four widely recognized cognitive tests.

Other than overcoming the lack of longitudinal data covering a population from teenage years to the old age, repeated cross-sectional testing offers several benefits compared to longitudinal studies. "Assuming that some process is declining over time [i.e. cognition], nonrandom dropout and retest-practice effects will have an effect of decreasing observed longitudinal changes […]" ((K. Warner Schaie & Hofer, 2001): 60). The use of a sequential cohort based approach relying on a combined study of cross-country databases, allows us to overcome both these drawbacks. The successive testing of large scale cross-sectional nationally representative surveys avoids bias from retest effects and selective (through attrition and refreshment) follow-up populations (Thorvaldsson, Hofer, Berg, & Johansson, 2006).

In sum, there is evidence at the individual level (and sometimes at population subgroups level) for a relatively strong stability in the relative cognitive performance. This however does not necessary imply that differences are similar across countries, where very different mechanisms are at work.

DATA

We make use of five datasets in order to observe two cohort groups. These datasets allow us to cover a period of 40 years in eight countries. In the following, we present first the data used to describe the performance at younger ages and then the performance among elderly.

Observations at ages 13 to 19

Data about mental abilities at young ages are provided by the First International Mathematics Study (FIMS). This survey was designed to be representative of the particular age and school year groups in each country and was conducted by the International Association for the Evaluation of educational Achievement (IEA) (Husén, 1967a, 1967b; Wolf, 1967). The data were collected in 1964. We use information on test scores of pupils about 13 years of age that are at the same school level[1] and of pupils who are exposed to a qualifying examination for the university or similar institution (i.e. pre-university students about 19 years old). The first belong to the cohorts born in 1949-1952 and the latter to those born between 1944 and 1947. Of the 12 countries included in the FIMS dataset, we are able to follow cohorts in eight countries up to older ages: Belgium, England, France, Western Germany (Federal Republic of Germany), Japan, the Netherlands, Sweden, and the United States - as we also have data on cognition in later life from surveys with comparable cognition indices.

In general, the test was carried out by the teachers in the classrooms. An international committee proposed that the markers of the test attempt to develop items to cover several categories of intellectual process, that is knowledge and information, techniques and skills, translation of data into symbols or schema and vice versa, comprehension, and inventiveness (for further details, refer to Husén, 1967b).

A summary of the data used is presented in Table 1 by country. In each country, sampled pupils were relatively equally distributed by gender (with an exception of the French and Swedish pre-university samples containing a majority of females).

Table 1: Summary statistics of FIMS data.

1964	a. 13 years old (1949-1952 cohorts)				b. 17 and 19 year olds (1944-1947 cohorts)			
	Sample size	Males (%)	Mean age	Number of pupils tested per 100,000 of pop.	Sample size	Males (%)	Mean age	Number of pupils tested per 100,000 of pop.
Belgium	2,645	65	14.0	29	1,004	54	18.0	11
England	3,089	52	14.3	7	1,782	52	17.9	4
France	3,449	54	13.6	7	192	21	18.8	0.4

[1] 13 years old grade level students are not all necessary aged exactly 13, but they all belong to the same grade level.

Western Germany	4,475	48	13.7	8	643	55	19.8	1.2
Japan	2050	51	13.4	2	4372	49	17.7	5
Netherlands	1,443	56	13.1	12	50	49	18.6	0.4
Sweden	2,828	51	13.7	37	222	27	19.6	3
US	6,544	51	14.0	3	2,042	41	17.8	1.1

Sources: a) Tables 14.3B and 14.4 (Husén, Vol.I (Husén, 1967b)) and 1.2 (Husén, Vol.II (Husén, 1967a)); b) Tables 14.3D and 14.4 (Husén, Vol.I (Husén, 1967b))

Cohort observations at ages 52-60

We follow the cohorts observed in the FIMS data over time, until they reach the age of 52-55 and 57-60, respectively, in 2004 (Japanese cohorts were observed 2007, so at the age of 55-58 or 60-63). At this point in time, we derive information on cognitive abilities in the eight countries considered from data collected by the English Longitudinal Study of Ageing (ELSA, wave 2), the Health and Retirement Study (HRS, wave 7), the Japanese Study of Aging and Retirement (JSTAR), and the Survey of Health, Ageing and Retirement in Europe (SHARE, wave 1). All these surveys are nationally representative of the non-institutionalized population aged 50 and over. Individual response rates are shown in table A.1 in the appendix. A brief description of the surveys used follows.

ELSA investigates on the lives of people in England who are aged 50 and over (and their partners, who may also be under 50). The study covers a broad range of topics such as people's health, economic situation and quality of life (Marmot & Banks, 2003).

HRS is a large-scale longitudinal project that studies the labor force participation and health transitions that individuals undergo toward the end of their work lives and in the years that follow. With its nationally representative sample of adults over the age of 50, the HRS provides a body of multidisciplinary data that address the challenges and opportunities of aging (National Institute on Aging, 2007).

JSTAR is a panel survey of elderly people aged 50 or older, conducted in Japan for the first time in 2007 by the Research Institute of Economy, Trade and Industry. It provides information on the economic, social, and health conditions of elderly people (Ichimura, Shimizutani, & Hideki, 2009).

SHARE is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status as well as social and family networks of more than 45,000 individuals aged 50 or over. Fifteen European countries participate in this longitudinal survey (Börsch-Supan & al., 2005).

SHARE is harmonized with the ELSA and HRS, and ELSA is in turn modeled on the HRS. Moreover, JSTAR is designed to ensure comparability with ELSA, HRS and SHARE to the maximum extent possible[2]. The items on which we rely to construct the indicators of cognitive abilities are the following.

Immediate recall "Now, I am going to read a list of words from my computer screen. We have purposely made the list long so it will be difficult for anyone to recall all the words. Most people recall just a few. Please listen carefully, as the set of words cannot be repeated. When I have finished, I will ask you to recall aloud as many of the words as you can, in any order. Is this clear?" The interviewer reads 10 words and the interviewee has one minute to recall as many as possible[3].

Delayed recall Approximately 5 minutes after the previous memory item, the interviewer asks "A little while ago, I read you a list of words and you repeated the ones you could remember. Please tell me any of the words that you can remember now?'' Again, the respondent has 1 minute for recalling[4].

Fluency "Now I would like you to name as many different animals as you can think of. You have one minute to do this. Ready, go.'' Valid answers are any members of the animal kingdom, real

---

[2] While having the same standardized items for memory as in SHARE and ELSA, HRS does not include fluency and numeracy tests. ELSA has both the memory and the fluency tests in common with SHARE, although it also does not include the numeracy item. JSTAR does not include the fluency tests.

[3] ELSA and HRS respondents have up to two minutes to recall the words. Based on the results from HRS 2006, we however know that the two-minute answer time does not alter our results, as 94.8% of respondents answered immediate recall within 60 seconds, and 94.3% answered the delayed recall question within 60 seconds. We acknowledge that respondents who took between 1 and 2 minutes to respond answered slightly more words, on average, than respondents who completed their answer within 60 seconds.

[4] As for immediate recall, ELSA and HRS respondents have two minutes of time to answer.

or mythical, specifically species name any accompanying breeds within the species as well as male, female and infant names within the species; not valid are repetitions and proper nouns.

Numeracy A first mathematical task is given to the interviewee.

I) "If the chance of getting a disease is 10 per cent, how many people out of one thousand would be expected to get the disease?"

If the first item was answered wrongly, the interviewee got the following question, after which the numeracy test was stopped, independently of whether the answer was correct. "In a sale, a shop is selling all items at half price. Before the sale, a sofa costs 300 (local currency). How much will it cost in the sale?" If the first item was answered correctly, a second question was posed.

II) "A second hand car dealer is selling a car for 6,000 (local currency). This is two-thirds of what it costs new. How much did the car cost new?"

Only if both the first and second numeracy items were answered correctly, the interviewee is asked to answer a last question:

III) "Let's say you have 2,000 (local currency) in a savings account. The account earns ten percent interest each year. How much would you have in the account at the end of two years?"

We calculate a numeracy score that ranges from 1 (poor numerical ability, if the interviewee has not answered correctly any mathematical task) to 5 (excellent numerical ability, if the interviewee has answered correctly all the mathematical tasks)[5].

Table 2.a provides an overview of the data about seniors belonging to the cohorts 1949-1952 in 2004 (i.e. the "about 13 years old" pupils in the FIMS data). Similarly, the following columns (Table 2.b) refer to the seniors belonging to the at least secondary educated 1944-1947 cohorts in 2004 (i.e. the "about 18 years old" pupils in the FIMS data).

Table 2: a.) Summary statistics on 1949-1952 cohorts; b.) Summary statistics on 1944-1947 cohorts.

[5] The indicator has value 2 if the interviewee does not answer correctly the first question, but gives the proper answer to the successive mathematical task. Value 3 corresponds to the case of a correct answer only at the first question. Value 4 corresponds to two correct answers (and both first and second task).

2004	a. Cohorts 1949-1952 (52-55 years old)			b. Cohorts 1944-1947 (57-60 years old)		
	Sample size	Males (%)	Mean age	Sample size	Males (%)	Mean age
Belgium	484	52.4	54.44	335	52.51	59.41
England	1207	49.56	53.89	944	53.41	58.32
France	279	55.12	53.47	140	43.55	58.06
Western Germany	329	51.62	53.46	230	51.85	58.68
Japan	530	48.87	56.60	384	50.00	61.40
Netherlands	355	48.65	53.41	327	47.84	58.27
Sweden	331	50.07	53.44	316	49.18	58.36
US	1774	53.02	53.04	1349	45.84	57.85

Sources: ELSA2 for England, HRS7 for the USA, JSTAR for Japan and SHARE1 for the European countries. Authors' calculation. Note: Japanese data are from 2007 and, contrary to those for the other countries, do not have weights available.

Comparability of tests

Earlier national and individual comparisons of different facets of cognitive skills reveal strong correlations between different types of ability measures that are believed to study the same or a similar underlying measure. General mental ability highly correlates with subsequent math performance in school (Furnham & Monsen, 2009; Hemmings, 1996; Reynolds, 1991; Yates, 2000). Koenig et al. (Koenig, Frey, & Detterman, 2008) find high associations between various general mental ability tests, where for instance g levels from the National Longitudinal Survey of Youth 1979 were correlated with American College Test (ACT) in the US, with a correlation of r: 0.77. Evidence from Europe shows still stronger correlations. Correlations within recent international school achievement tests (PISA, TIMSS and PIRLS) were very high, r: 0.60–0.98, between different student assessment studies (such as between PISA and TIMSS), r: 0.82–0.83; and between student assessment tests and general mental ability (g) tests were equally high, r: 0.85–0.86 (Rindermann, 2007).

We therefore rely on the assumption that the cognitive tests at younger and older ages are largely indicative of the same underlying latent construct, allowing a meaningful comparison of the rank-ordering of cognition at different ages.

METHOD

Our multiple-survey dataset enables us to run comparisons across life cycle variation in cognition that was not possible before. The FIMS data at young ages and the ELSA, HRS, JSTAR, and SHARE data above the age of 50, assessing the cognitive skills of the same cohorts at different ages, make it feasible to study the national level of cognitive ageing. As outlined in Figure 1, we observe a certain cohort in 1964, say born in 1945, which is aged 19 years old. At a later stage in life, in 2004, we re-observe the same cohort at 59 years old. By proceeding in the same way for all the countries considered in both the points in time, we rank the countries and compare them. In addition to fill in a gap in comparative studies on skills variation, the cohort approach we adopt explores whether the stability found at individual level (i.e. if one is performing well at younger ages, one is also good at older ages) is confirmed at country-level.

We summarize each ability individually and survey by survey with weighted means and standard deviations. Weighted means are our measure of the country-specific level of ability of a certain age group. Standard deviations provide information about the size of the age-specific gap within each country. The countries are ordered according to their weighted mean[6].

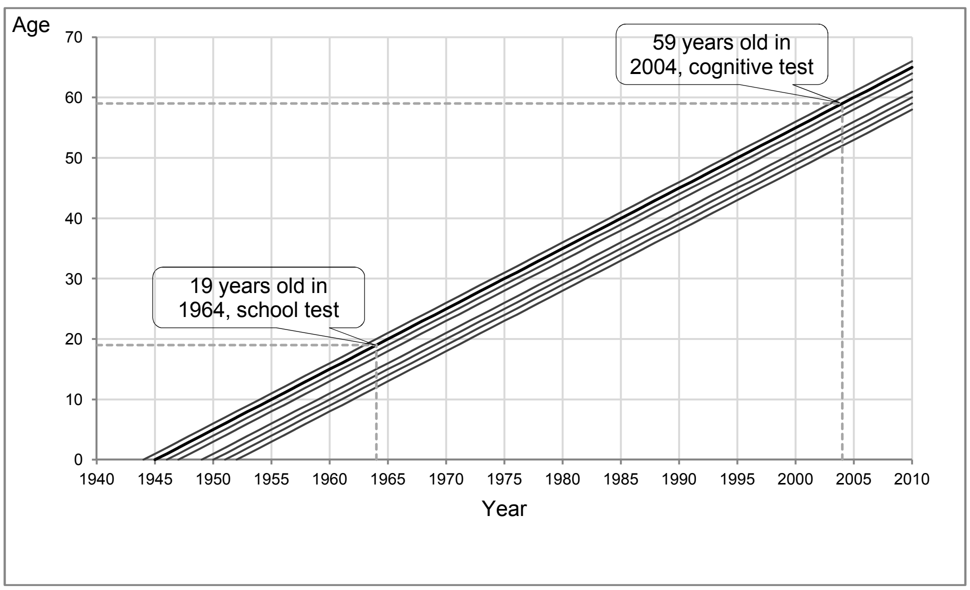

Figure 1: Lexis diagram showing our data collection approach.

As the findings are based on different scales, it is not possible to directly compare a country across the various measures of cognition. In other words, we cannot say whether the average cognitive skill level for a cohort in country *A* between 1964 and 2004 decreased, was preserved or rose. However, we are able to compare the country ranking, both over time and across

[6] For numeracy, we use the relative frequencies of individuals within each numerical ability group.

different cognitive tests. That is, once we observe that country *A* has a higher level of immediate memory than country *B*, we compare the relative position of country *A* to country *B* in another measure of cognition at the same age or in the school test at their youth 40 years before. The results for the four cognitive items at older ages are therefore reported separately. In this way, we can induce which countries age better.

RESULTS

The results described in the following report countries' scores using the "raw" data for each selected indicator, drawing attention to the top performers in terms of immediate recall, delayed recall, fluency, and numeracy unadjusted for context. Such ranking does not take into account the differences in the challenges faced and variation in available structures and resources or welfare.

Observations at ages 13 to 19

Differences in the performance in the math test at young ages are highlighted in Figure 2 by the mean math score (vertical axis). This ranges from 15.3 for Sweden and 17.8 for the USA up to 31.2 for Japan at age 13 in 1964. At pre-university age in the same year, USA and Sweden register as well the lowest mean math scores (8.3 and 12.6, respectively), while Western Germany, with 27.7, heads the list, followed by France and Japan.

The within country standard deviation, contained in Tables 3a and 3b alongside of the test scores, shows the level of inequalities between students in a certain country. Japan and England result to be the countries, among those considered, with a wider gap between their own pupils' math scores in 1964.

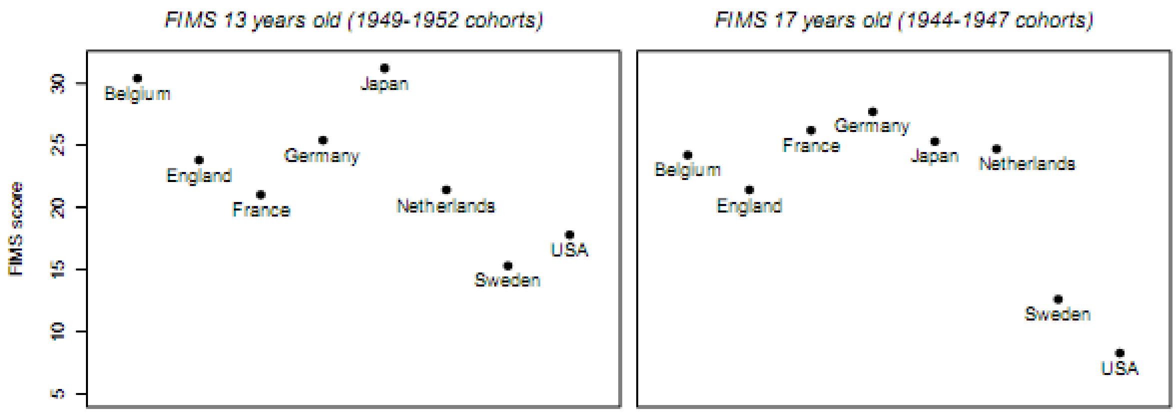

Figure 2: Mean FIMS test score in 1964 by cohort group and country. Source: 1.2, Tables 1.2 and 1.4 (Husén, Vol. II (Husén, 1967a)).

Observations at ages 52-60

In the 2000s, the same cohorts are exposed to four tests of cognitive abilities. We distinguish the two cohort groups for older ages as we did for the younger ages, focusing first on immediate recall, delayed recall, and fluency. Looking at those aged 52 to 55 years old in 2004, belonging to the cohorts 1949-1952, we notice some changes in the ranking as compared to the 1964 picture for students (Table 3a). While Japan is among the top-three only in the ranking for delayed recall, England has the top score in both immediate and delayed recall, sharing the leadership with Germany and Sweden in the fluency task. The most striking result is the top position of England and the USA in the recall tests. This could imply either relative "improvement" in these countries or a poor ability in the others to preserve cognitive abilities over the life cycle. As for Japan, which appears to be among the lowest cognitive skilled countries when considering immediate recall in older ages, we should highlight the remark that the available data refer to a sample population 3 years older than in the other countries. At the bottom of the ranking, France finds its position in the immediate recall, delayed recall, and fluency tests.

As regarding seniors aged 57-60 in 2004 (around 18 years old in 1964, belonging to the cohorts 1944-1947), we observe a similar change over time that supports our previous findings (Table 3b). France, together with Belgium, confirms its bottom ranking in the recall tests; however, this country is second only to Sweden in the fluency test in older ages.

Table 3a: Rank ordering (and mean scores) of math test at 13 years old grade level in 1964 and of immediate recall, delayed recall and fluency at older ages in 2004. 1949-1952 cohorts.

	Cohorts 1949-1952							
	FIMS ~13 years old	S.D.	immediate recall	S.D.	delayed recall	S.D.	fluency	S.D.
Belgium	2 (30.40)	13.7	5 (0.55)	0.16	6 (0.40)	0.18	4 (21.51)	6.17
England	4 (23.80)	18.5	1 (0.63)	0.16	1 (0.52)	0.18	3 (22.32)	6.38
France	6 (21.00)	13.2	7 (0.50)	0.20	7 (0.36)	0.20	5 (21.30)	7.59
Western Germany	3 (25.40)	11.7	2 (0.60)	0.18	4 (0.45)	0.18	2 (22.80)	7.20
Japan	1 (31.20)	16.9	6 (0.54)	0.16	2 (0.49)	0.20	-	-
Netherlands	5 (21.40)	12.1	5 (0.55)	0.17	5 (0.43)	0.18	6 (20.83)	5.94
Sweden	8 (15.30)	10.8	4 (0.57)	0.14	3 (0.47)	0.16	1 (25.53)	7.15
USA	7 (17.80)	13.3	3 (0.59)	0.15	2 (0.49)	0.18	-	-

Sources: 1.2, Tables 1.2 and 1.4 (Husén, Vol. II (Husén, 1967a)), ELSA2 for England, HRS7 for the USA, JSTAR for Japan and SHARE1 for the European countries. Authors' calculation. Note: Japanese data are from to 2007 and refer to the age group 55-58.

Table 3b: Rank ordering (and mean scores) of math test at pre-university level in 1964 and of immediate recall, delayed recall and fluency at older ages in 2004. 1944-1947 cohorts.

	Cohorts 1944-1947							
	FIMS pre-university	S.D.	immediate recall	S.D.	delayed recall	S.D.	fluency	S.D.
Belgium	5 (24.20)	9.50	7 (0.54)	0.16	6 (0.39)	0.17	5 (21.41)	5.75
England	6 (21.40)	10.00	1 (0.64)	0.16	2 (0.53)	0.18	3 (23.02)	6.16
France	2 (26.20)	9.50	8 (0.51)	0.17	7 (0.37)	0.17	2 (23.20)	8.28
Western Germany	1 (27.70)	7.60	3 (0.60)	0.15	5 (0.42)	0.19	4 (21.73)	6.96
Japan	3 (25.30)	14.30	6 (0.55)	0.16	3 (0.49)	0.20	-	-
Netherlands	4 (24.70)	9.80	5 (0.57)	0.18	5 (0.42)	0.21	6 (20.85)	5.88
Sweden	7 (12.60)	6.20	4 (0.59)	0.15	4 (0.45)	0.16	1 (25.04)	7.26
USA	8 (8.30)	9.00	2 (0.63)	0.15	1 (0.54)	0.18	-	-

Sources: 1.2, Tables 1.2 and 1.4 (Husén, Vol. II (Husén, 1967a)), ELSA2 for England, HRS7 for the USA, JSTAR for Japan and SHARE1 for the European countries. Authors' calculation. Note: Japanese data are from 2007 and refer to the age group 60-63.

For what it concerns numeracy ability, we rely on the relative frequencies of individuals within each numerical ability level of performance in order to draw the ranking. The results in Table 4 show a rather different ranking, as compared to that based on the math test scores on pupils in 1964. While Japan maintains a leader position, scoring within the top-three countries at both younger and older ages and across both the cohort groups under observation, Germany keeps heading the ranking only for the 1949-1952 cohorts. Once we look at the 1944-1947 cohorts ranking, we find Germany among the bottom-three countries. Even more striking is the totally opposite position of France and Belgium, which passed from heading the math test ranking for

1964 pupils to being at the top of the poor performers in 2004. On the contrary, Sweden has the highest percentage of excellent performances for the 57-60 years old.

Table 4: a.) Relative frequencies of numeracy performances at seniors' age. 1949-1952 cohorts; b.) Relative frequencies of numeracy performances at seniors' age. 1944-1947 cohorts.

	a. Cohorts 1949-1952 (52-55 years old)					b. Cohorts 1944-1947 (57-60 years old)				
	poor	fair	good	very good	excellent	poor	fair	good	very good	excellent
Belgium	3.64	11.63	29.62	37.93	17.18	2.21	7.79	35.69	37.23	17.07
England	-	-	-	-	-	-	-	-	-	-
France	7.79	12.77	25.45	32.87	21.13	4.01	11.05	26.1	41.37	17.47
Western Germany	1.44	7.33	18.61	38.62	34.0	0	15.08	31.22	28.04	25.66
Japan	0.96	13.63	26.1	29.75	29.56	2.11	8.58	21.14	38.33	29.84
Netherlands	2.02	9.53	28.02	27.93	32.5	1.67	6.6	24.14	33.81	33.78
Sweden	0.42	9.37	25.67	35.17	29.37	0.86	6.48	26.28	31.98	34.4
USA	-	-	-	-	-	-	-	-	-	-

Sources: ELSA2 for England, HRS7 for the USA, JSTAR for Japan and SHARE1 for the European countries. Authors' calculation. Note: "poor" = failing in answering all questions; "excellent" = being able to answer all questions. Japanese data are from 2007 and refer to the age group 60-63.

DISCUSSION

We constructed a large international dataset from a selection of different surveys to study nationally representative cohorts from their teens to the post 50s in several countries. We used standardized test scores from 1964 IEA data on school performances of 13 years old grade level pupils and pre-university students (1944-1947 and 1949-1952 cohorts). We have re-observed these cohorts at older ages (57-60 and 52-55 years old, respectively) in SHARE, ELSA and HRS surveys in 2004. Using JSTAR data, we re-observed the Japanese cohorts in 2007. All these surveys include standardized tests of cognition. In this way, we are not only able to describe the level of cognitive ageing, but we can also deal with life-cycle changes in cognition across countries. To our knowledge, no past study on cognitive abilities follows representative cohorts over long age intervals across countries.

Our results show that forty years after the first tests, to which 13 and 19 years old students were exposed in 1964, changes in the ranking position of the eight countries considered can be observed. While some of the countries that performed well at younger ages are poorly performing four decades later, other countries are improving their relative performance. We find differences in the relative performance of life cycle development of skills between the two cohort groups under study and also observe variation in the relative performance of the four different cognitive tests at senior ages (immediate recall, delayed recall, fluency, and numeracy tests).

Although the tests we consider are not identical at younger and older ages, some of the underlying concepts that are tested (e.g., mathematical skills) are common. Furthermore, as noted above, a number of earlier studies find a high degree of correlation between several different types of cognitive tests (even between achievement and aptitude tests), which would suggest that our test measures are, to a large extent, comparable.

We acknowledge that our approach may also have some limitations. Standard survey techniques suggest that the same questions should be used at different waves in order to avoid confounding measurement change with substantive change (Cook & Campbell, 1979). This is however a problem that measures of life cycle variation often face: the retest effect (repeated tests can imply an improvement over time due to learning of test procedures, familiarity with the testing situation, and a shorter "warm-up" phase) (Ferrer, Salthouse, Stewart, & Schwartz, 2004). The data used for this study (partly) allow overcoming differential retest effects due to test exposure, anxiety, procedures or items familiarity, which could otherwise inflate the results. The pupils tested in 1964 are not necessarily the same adults interviewed in 2004. Moreover, we are using the first wave of SHARE and the first wave of JSTAR, excluding therefore any possible retest effect for six out of eight countries in the study. For the younger American respondents (born in 1949-1952), we also do not have any retest effect: this cohort-group was a part of the sample of HRS only from 2004 (the year we observe them). The concern might regard the 1944-1947 cohorts from the US which could be interviewed already up to three times before the 2004 wave of HRS, and the English sample (at its second interview in 2004).

Furthermore, selective attrition, selective remainders, and selective refreshment groups could also imply that test score performance is artificially inflated. That is, as Cooney et al. (Cooney, Schaie, & Willis, 1988) noted, people with prior experience are not necessarily comparable with those without prior experience because the people who return for a second assessment often have a higher level of initial functioning than those who do not return. Bias can occur through selective attrition, selective remainders and selective refreshment samples. As a consequence, at least a portion of the difference in performance between those tested at a second occasion and those tested at the same time but for their first occasion may be attributable to a higher initial level of functioning among people who return for a second assessment (Ferrer, et al., 2004). Still, the quite similar ranking for memory tests on cohorts 1944-1947, where some concerns about retest effects are present, and 1949-1952, where retest effects are excluded, supports the notion that retest effects are not so important in our case.

We additionally acknowledge that response rates in the various countries might be different. However, the surveys we are using are all aimed to be representative (at country level), both among the young and among the (non-institutionalized) population aged 50 and older. A deeper look into the breakdown of all samples by country can indeed highlight a rather homogeneous response rate which satisfies our need in order to draw valid conclusions on the comparability of the data. Hence, notwithstanding these concerns, we view these efforts as an important first step in meaningful comparisons.

REFERENCES

Aneshensel, C. S., Ko, M. J., Chodosh, J., & Wight, R. G. (2011). The Urban Neighborhood and Cognitive Functioning in Late Middle Age. *Journal of Health and Social Behavior, 52*(2), 163-179. doi: 10.1177/0022146510393974

Blum, J. E., Fosshage, J. L., & Jarvik, L. F. (1972). Intellectual changes and sex differences in octogenarians: A twenty-year longitudinal study of aging. [doi:10.1037/h0033001]. *Developmental Psychology, 7*(2), 178-187. doi: 10.1037/h0033001

Booth, M. L., Owen, N., Bauman, A., Clavisi, O., & Leslie, E. (2000). Social–Cognitive and Perceived Environment Influences Associated with Physical Activity in Older Australians. [doi: 10.1006/pmed.2000.0661]. *Preventive Medicine, 31*(1), 15-22.

Börsch-Supan, A., & al., e. (2005). Ageing and Retirement in Europe - First Results from the Survey of Health, Ageing and Retirement in Europe. Mannheim: MEA.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. [doi:10.1037/0012-1649.27.5.703]. *Developmental Psychology, 27*(5), 703-722. doi: 10.1037/0012-1649.27.5.703

Chee, M. W. L., Zheng, H., Goh, J. O. S., Park, D., & Sutton, B. P. (2010). Brain Structure in Young and Old East Asians and Westerners: Comparisons of Structural Volume and Cortical Thickness. [doi: 10.1162/jocn.2010.21513]. *Journal of Cognitive Neuroscience, 23*(5), 1065-1079. doi: 10.1162/jocn.2010.21513

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation : design & analysis issues for field settings*. Chicago: Rand McNally College Pub. Co.

Cooney, T. M., Schaie, K. W., & Willis, S. L. (1988). The Relationship Between Prior Functioning on Cognitive and Personality Dimensions and Subject Attrition in Longitudinal Research. *Journal of Gerontology, 43*(1), P12-P17. doi: 10.1093/geronj/43.1.P12

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The Impact of Childhood Intelligence on Later Life: Following Up the Scottish Mental Surveys of 1932 and 1947. [doi:]. *Journal of Personality and Social Psychology, 86*(1), 130-147.

Engelhardt, H., Buber, I., Skirbekk, V., & Prskawetz, A. (2008). Social Engagement, Behavioural Risks and Cognitive Functioning Among the Aged *European Demographic Research Papers*. Vienna, Austria: Vienna Institute of Demography of the Austrian Academy of Sciences.

Facon, B. (2008). How does the strength of the relationships between cognitive abilities evolve over the life span for low-IQ vs high-IQ adults? [doi: 10.1016/j.intell.2007.11.004]. *Intelligence, 36*(4), 339-349.

Ferrer, E., Salthouse, T. A., Stewart, W. F., & Schwartz, B. S. (2004). Modeling Age and Retest Processes in Longitudinal Studies of Cognitive Abilities. [doi:10.1037/0882-7974.19.2.243]. *Psychology and Aging, 19*(2), 243-259. doi: 10.1037/0882-7974.19.2.243

Furnham, A., & Monsen, J. (2009). Personality traits and intelligence predict academic school grades. [doi: 10.1016/j.lindif.2008.02.001]. *Learning and Individual Differences, 19*(1), 28-33.

Glymour, M. M., Kawachi, I., Jencks, C. S., & Berkman, L. F. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology and Community Health, 62*(6), 532-537. doi: 10.1136/jech.2006.059469

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2008). Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context: National Center for Education Statistics.

Hemmings, B. C. (1996). A longitudinal study of Australian senior secondary school achievement *Issues in Educational Research, 6*(1), 13-37.

Hindley, C. B., & Owen, C. F. (1978). The extent of individual changes in I.Q. for ages between 6 months and 17 years, in a British longitudinal sample. *Journal of Child Psychology and Psychiatry, 19*(4), 329-350. doi: 10.1111/j.1469-7610.1978.tb00480.x

Husén, T. (Ed.). (1967a). *A Comparison of Twelve Countries: International Study of Achievement in Mathematics* (Vol. 2). Uppsala: Almqvist & Wiksells Boktryckeri AB.

Husén, T. (Ed.). (1967b). *A Comparison of Twelve Countries: International Study of Achievement in Mathematics* (Vol. 1). Uppsala: Almqvist & Wiksells Boktryckeri AB.

Ichimura, H., Shimizutani, S., & Hideki, H. (2009). JSTAR First Results 2009 Report: Research Institute of Economy, Trade and Industry (RIETI).

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. [doi: 10.1016/j.intell.2007.03.005]. *Intelligence, 36*(2), 153-160.

Lindenberger, U., & Ghisletta, P. (2009). Cognitive and sensory declines in old age: gauging the evidence for a common cause. *Psychol Aging, 24*(1), 1-16. doi: 2009-03151-001 [pii] 10.1037/a0014986

Marmot, M., & Banks, J. (2003). *Health, wealth and lifestyles of the older population in England : the 2002 English longitudinal study of ageing*. London: Institute for Fiscal Studies.

National Institute on Aging. (2007). *Growing Older in America: The Health and Retirement Study*. Washington, DC: National Institutes of Health.

Programme for International Student Assessment. (2009). *PISA 2009 Assessment Framework: Key competencies in reading, mathematics and science*: OECD.

Reynolds, A. J. (1991). The middle schooling process: influences on science and mathematics achievement from the longitudinal study of American youth. *Adolescence, 26*(101), 133-158.

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality, 21*(5), 667-706. doi: 10.1002/per.634

Schaie, K. W., & Hofer, S. M. (2001). longitudinal studies in aging research. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 53 - 77). San Diego, CA: Academic Press. (Reprinted from: 5th edition).

Schaie, K. W., & Willis, S. L. (1986). Can decline in intellectual functioning be reversed? *Developmental Psychology, 22*(2), 223-232.

Schneider, W., & Bullock, M. (2008). *Human Development from Early Childhood to Early Adulthood: Findings from a 20 Year Longitudinal Study* (first ed.): Psychology Press.

Thorvaldsson, V., Hofer, S. M., Berg, S., & Johansson, B. (2006). Effects of Repeated Testing in a Longitudinal Age-Homogeneous Study of Cognitive Aging. *J Gerontol B Psychol Sci Soc Sci 61*(6), 348-354.

Voelcker-Rehage, C., Godde, B., & Staudinger, U. M. (2011). Cardiovascular and coordination training differentially improve cognitive performance and neural processing in older adults. [Original Research]. *Frontiers in Human Neuroscience, 5*. doi: 10.3389/fnhum.2011.00026

Warr, P. (1994). Age and employment. In H. Triandis, M. Dunnette & L. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (pp. 485-550). Thousand Oaks, California: Consulting Psychologist Press, 2nd ed.

Weinert, F. E., & Schneider, W. (1999). *Individual Development from 3 to 12: Findings From the Munich Longitudinal Study*: Cambridge University Press.

Williams, J. W., Plassman, B. L., Burke, J., Holsinger, T., & Benjamin, S. (2010). Preventing Alzheimer's Disease and Cognitive Decline *Evidence Report/Technology Assessment*: AHRQ Publication.

Wolf, R. M. (1967). *Data Bank Manual. Phase I: International Study of Achievement in Mathematics: A Comparison of Twelve Countries*. Uppsala: Almqvist & Wiksells Boktryckeri AB.

Yates, S. M. (2000). Task involvement and ego orientation in mathematics achievement: A three year follow-up. *Issues In Educational Research, 10*(1), 77-91.

Yule, W., Gold, R. D., & Busch, C. (1982). Long-term predictive validity of the WPPSI: An 11-year follow-up study. [doi: 10.1016/0191-8869(82)90075-7]. *Personality and Individual Differences, 3*(1), 65-71.

Zimprich, D., & Mascherek, A. (2010). Five views of a secret: does cognition change during middle adulthood? *European Journal of Ageing, 7*(3), 135-146. doi: 10.1007/s10433-010-0161-5

APPENDIX

Table A.1. Individual response rates

Country	Individual Response Rate
Belgium	90.5%
England	67.0%
France	93.3%
Germany	86.2%
Japan	60.0%
Netherlands	87.8%
Sweden	84.6%
USA	88.6%

Sources: http://www.share-project.org; http://hrsonline.isr.umich.edu/sitedocs/sampleresponse.pdf; http://www.ifs.org.uk/elsa/report03/ch1.pdf.