Descriptive Finding

A cross-country comparison of math achievement at teen age and cognitive performance 40 years later

Vegard Skirbekk

Valeria Bordone

Daniela Weber

Table of Contents

A cross-country comparison of math achievement at teen age and cognitive performance 40 years later

Vegard Skirbekk[1]

Valeria Bordone[2]

Daniela Weber[3]

Abstract

BACKGROUND
Maintaining cognitive functioning through mid- to late-life is relevant for the individual and societal aim of active ageing. Evidence shows considerable stability in individual-level rank-ordering of cognitive functioning, but little attention has been given to cohort performance over the life cycle and macro-level factors that could affect it.

OBJECTIVE
The main goal of this paper is to address cross-national variation in mental performance from younger to older ages.

METHODS
Using a quasi-longitudinal approach, we compare the relative country ranking in standardised mathematical test scores of young teenagers in 1964 from the First International Mathematics Study (FIMS) and cognitive test performance at mid-life in 2004, based on the Survey of Health, Ageing and Retirement in Europe (SHARE) for the cohort born between 1949 and 1952.

RESULTS
Our results show that those countries which had the highest scores in math tests taken by 13-year-old-grade-level students are not the same countries that, 40 years later, have the top performing scores in cognitive tests among mid-age adults.

[1] Wittgenstein Centre (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Austria. E-Mail: skirbekk@iiasa.ac.at.
[2] Wittgenstein Centre (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Austria. E-Mail: bordone@iiasa.ac.at.
[3] Wittgenstein Centre (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Austria. E-Mail: weberd@iiasa.ac.at.

CONCLUSIONS

This article highlights the importance of considering country-level influences on cognitive change over the life cycle, in addition to individual characteristics, and provides some descriptive findings that could be incorporated with further research on the link between specific contextual factors and cognitive functioning.

1. Introduction

Maintaining cognitive functioning through mid- to late-life is relevant for both the individual and the societal aim of active and productive ageing.

Evidence has shown considerable stability in individual-level rank-ordering of cognitive functioning; i.e., a person with high cognitive performance at younger ages tends also to score high in cognitive tests later in life (e.g., Deary et al. 2000; Zimprich and Mascherek 2010). Health status, genetic traits, behaviour, socio-economic status as well as the degree of cognitive engagement and education all contribute to the path of individual ageing (Comijs et al. 2011; Engelhardt et al. 2010; Glymour et al. 2008). Life-cycle stability in cognition at the individual-level does not necessarily imply that a cohort's performance over the life cycle will also be stable in a cross-country comparative framework. Macro-level conditions in early life (such as the environment where children are born or grow up), human capital investment at younger ages, and also life course events (such as employment and retirement patterns) could significantly affect how well a certain cohort in a specific country maintains its cognitive skills over time.

The absence of international comparable data which track student achievement from kindergarten or elementary school through high school age – not to mention mid-life or old age – resulted in an empirical gap within studies addressing cross-national variation in mental performance over different stages of life. With a quasi-longitudinal approach, this paper sets out to explore the life cycle variation of cognitive functioning at the country-level: we compare math achievement in the early teenage years in 1964 and cognitive performance at mid-life in 2004 across five European countries for the birth cohort 1949–1952. The results of this study can inform the development of policies to promote cognitive health of older people.

2. Data and methods

This study uses data from two datasets that allow for a cross-country comparison of a cohort's cognitive performance at two points in time. Data at younger ages are derived from the 1964 First International Mathematics Study (FIMS) on 13-year-old-grade-level pupils (i.e., birth cohort 1949–1952) conducted by the International Association for the Evaluation of Educational Achievement (IEA) (Husén 1967a, 1967b; Wolf 1967). The FIMS was designed to be representative of particular age- and school year-groups in twelve countries (Australia, Belgium, England, Finland, France, Germany (Federal Republic of Germany before reunification), Israel, Japan, the Netherlands, Scotland, Sweden, and the United States) and was generally carried out by teachers within the classroom. An international committee developed items to cover several categories of intellectual process. In Table 1, sample descriptive statistics as well as the percentage of the school timetable devoted to the study of mathematics in the specific grade level are provided for each country considered in this study. For further survey and questionnaire details, as well as regulations specifying when children should begin and cease mandatory schooling, we refer to Table 13.1 in Husén (1967a) and to Appendix II in Husén (1967b).

Table 1: **Summary statistics of data on 13-year-old-grade-level students in 1964**

Country	Sample size[1]	Males (%)[1]	Mean age[1]	Location in the school system[2]	Percent Mathematics[3]
Belgium	2,645	65	14.0	5 e. (2 e A 3 in Enseignement Technique)	11 (Academic)- 14 (Vocational)- 12 (Others)
France	3,449	54	13.6	5 e. (CFE in Ecole Primaire)	13 (Lycée sauf classes term.)- 15 (CEG)- 18 (Ecole Primaire)
Germany	4,475	48	13.7	7. Klasse (Schulleistungsjahr)	12
Netherlands	1,443	56	13.1	6 e. in primary schools	11
Sweden	2,828	51	13.7	Arskurs 7	11

Sources: FIMS, [1]Table 14.3B; [2]Table 13.4; [3]Table 13.7. All Tables from Husén (1967a).

FIMS was the first of several subsequent internationally comparable school tests (e.g., Mullis et al. 2009; OECD 2011), and the only one so far whose teenage interviewees have already reached mid-age.

We are able to hypothetically follow the 1949–1952 cohort, observing a representative sample again at mid-age in 2004 (i.e., aged 52–55), in Belgium, France, Germany, the Netherlands, and Sweden using the Survey of Health, Ageing and Retirement in Europe (SHARE) (Börsch-Supan et al. 2005). SHARE is a multidisciplinary and cross-national panel database of micro data on health, socioeconomic status as well as social and family networks of more than 85,000 individuals in nineteen European countries. Table 2 provides summary statistics related to the 1949–1952 cohort in the five countries included in the comparative analysis.

Table 2:   Summary statistics of data on 52–55 year olds in 2004

Country	Sample size	Males (%)	Mean age
Belgium	484	52.40	54.44
France	279	55.12	53.47
Germany	329	51.62	53.46
Netherlands	355	48.65	53.41
Sweden	331	50.07	53.44

Sources: SHARE, wave 1. Authors' calculation.

Although SHARE does not track the same individuals tested by the FIMS, we observe nationally representative samples of these populations over a 40-year life span. Of particular interest is the (re-)examination at an age when age-related cognitive decline is already visible (i.e., between ages 52 and 55).

Different domains of cognitive functioning in adulthood may vary differently over time (Richards et al. 2004). As indicators of these domains, we rely on four items: *immediate recall* tests the ability to recall as many words as possible within one minute, out of ten common words read by the interviewer; the *delayed recall* test consists in recalling, with a few minutes of delay, as many words as possible out of the ten words read before the immediate recall task; *fluency* refers to the performance of naming as many animals as possible within one minute; *numeracy* tests the ability of performing some simple calculations based on real life situations. While recall measures range from 0 to 10, fluency has no fixed ceiling and numeracy scores from 1 to 5, indicating at which step of the numerical battery the interviewee stopped answering correctly. For the exact wording of the tests, we

refer the reader to the SHARE questionnaire, available on the website http://www.share-project.org/. This study follows a descriptive approach in order to investigate whether the cognitive stability usually found at the individual level (i.e., performing well at younger ages is correlated with higher cognitive performance at older ages) is confirmed at the country level. We first create a relative rank-ordering of countries for the 1949–1952 cohort, based on the results from the mathematical test at teen age in 1964. Second, we create four relative rank-ordering of countries for the same cohort group, based on the performances in the respective cognitive tests at ages 52–55 in 2004. Third, we compare the relative orderings at the two points in time to identify whether nations with good students are also top performers in terms of cognitive functioning when the cohort analysed is forty years older.

For math performance at younger ages and both fluency and recall tests at mid-life, the mean is our measure of the country-specific level of ability. Consequently, countries are ordered according to their mean achievement. Standard deviations (S.D.) provide information about the size of the age-specific cognitive variation within each country. For numeracy at mid-life, country scores are calculated as the proportion of the population with the top two performance levels on the five-point scale. Individual sampling weights are applied in all our calculations.

3. Results

3.1 Math achievement at teen age and cognitive performance at mid-age

As shown in Table 3, the FIMS scores range from 15.3 points for Sweden up to 30.4 points for Belgium. Due to the small sample size – five countries – we use effect size measures instead of significance tests to compare country means. Here, the mean difference between two countries in their math performance is represented by Cohen's d (Cohen 1988), which on average is 0.56, indicating moderate differences in country means.

Table 3: **Mean performance in math test scores, rank-ordering of countries and standard deviations (S.D.) for 13-year-old-grade-level students in 1964**

Country	Mean math test scores	S.D.
Belgium	30.4 (1)	13.7
France	21.0 (4)	13.2
Germany	25.4 (2)	11.7
Netherlands	21.4 (3)	12.1
Sweden	15.3 (5)	10.8

Sources: FIMS, Table 1.2 (Husén 1967b).

The standard deviations give us additional information: Belgium registers the greatest within-country variation at teen age in 1964 (S.D. = 13.7), while Sweden is the country with the lowest variation in pupils' performance (S.D. = 10.8) among those considered in this study.

The scores of the four tests of cognitive functioning undertaken in 2004 at age 52–55 are reported in Table 4. Among the continental European countries considered, in terms of immediate recall, delayed recall, fluency, and numeracy, Germany and Sweden score the highest: on average, Germans aged 52–55 recall six words out of ten immediately and 4.5 with a few minutes of delay, and they name almost 23 animals within one minute. Their Swedish counterparts recall 5.7 and 4.7 words respectively, and name about 26 animals. France is the country with the largest within-country variation in memory and fluency tasks among the 52–55 year olds, with an average of only five words recalled immediately, 3.6 words in the delayed recall, and 21 animals named on average in 1 minute (S.D. = 2 in recall tests and = 7.6 in fluency). The smallest standard deviation in recall measures is registered in Sweden (S.D. immediate recall = 1.4; S.D. delayed recall = 1.6). Among Germans aged 52–55, about 73% reached the last question of the numerical battery. In France, only 54% reached the final stage of the numerical test.

Table 4:   **Performance in cognitive tests and rank-ordering of countries for 52–55 year olds in 2004: Mean scores and standard deviations (S.D.) of immediate recall, delayed recall, and fluency tests; relative frequencies reaching the final question of the numerical test battery**

Country	Immediate recall	S.D.	Delayed recall	S.D.	Fluency	S.D.	Numeracy
Belgium	5.5 (3)	1.6	4.0 (4)	1.8	21.5 (3)	6.2	55.1% (4)
France	5.0 (5)	2.0	3.6 (5)	2.0	21.3 (4)	7.6	54.0% (5)
Germany	6.0 (1)	1.8	4.5 (2)	1.8	22.7 (2)	7.2	72.6% (1)
Netherlands	5.5 (3)	1.7	4.3 (3)	1.8	20.8 (5)	5.9	60.4% (3)
Sweden	5.7 (2)	1.4	4.7 (1)	1.6	25.5 (1)	7.2	64.5% (2)

Sources: SHARE, wave 1. Authors' calculation.

The between-country difference in the mean cognitive performances is rather small among the 52–55 year olds. In particular, the smallest effect size and hence also the smallest Cohen's d is for immediate recall, where d is on average 0.25 (calculated as the average of ds for each couple of countries). The difference in mean delayed recall is d = 0.3, whereas the maximal difference is identified for fluency (Cohen's d = 0.32).

3.2 Cross-country comparison over time

Of interest for this study is the relative performance of the 1949–1952 cohort in the five countries considered at the two points in time. Our results show that the countries which had the top performing scores in math tests undertaken by 13-year-old-grade-level students in 1964 are not the same countries that, 40 years later, have the top performing mid-age adults in cognitive tests.

In particular, a positive shift in the relative ordering of Sweden is observed, in spite of its weak performance at younger ages. Comparing Sweden when it is the top performer (i.e., in delayed recall and fluency) with the bottom performer, we can see not only that Swedes at age 52–55 perform better at every level of education, but also that Swedish low educated perform better than the high educated in France and the Netherlands, respectively. Although the older Swedes are recalling only 0.7 words more than the bottom performer in immediate recall, for example, we should notice that the difference, on average, between the low and the

high educated in immediate recall among the 52–55 year olds considered is 0.6 words. In contrast, Belgium shows a decline in its relative position. In delayed recall, for example, Belgium differs from the top scorer (Sweden) of 0.7 words, thus roughly the same magnitude as the difference between low and high educated among the 52–55 year olds. Germany confirms its position among the top performers from teen age (in the mathematical school test) to mid-life (in the recall, fluency, and numeracy cognitive tests). In fluency, the range between top scorer and bottom scorer is 4.7 words. This range is almost as big as the difference between low and high educated (4.9).

Comparing rankings of only five countries is a challenge for any statistical test. Nevertheless, we calculate the Kendall's coefficient of concordance (Kendall's W) which ranges from 0 (no agreement) to 1 (complete agreement), to assess the agreement between raters (Kendall and Smith 1939). Here, showing the tendency for discordance between the ranking of FIMS scores and those of the cognitive performances at higher age, Kendall's Ws range between 0.3 (for math and fluency scores) and 0.58 (for math and immediate recall scores), while Kendall's W for math and delayed recall tests scores = 0.35.

4. Summary and discussion

Understanding how cognitive functioning is maintained from the teens to the fifties is of great importance, not only for the individual, but also for national prosperity in ageing economies (OECD 2006, Engelhardt et al. 2010; 2011; Skirbekk et al. 2013). This study identifies reordering over a significant time period in the national-level ranking of cognitive performance for the cohort group born in 1949-1952, suggesting that more weight should be given to national-level influences accompanying individual-level analysis to understand life-cycle variation in cognitive functioning. The average country performance may represent more than the simple sum of individual capabilities.

The quasi-longitudinal approach used in this study has the important advantage of avoiding problems originating from non-random dropout and retest-practice effects associated with longitudinal surveys (Thorvaldsson et al. 2006). Nevertheless, we acknowledge that this approach cannot compensate for potential distortions arising from differences at the country level in survey procedures, target populations, response rates and comparability of different tests (Goldstein 2004; Jerrim 2011). Tesch-Römer and Kondratowitz (2006) for example referred to the translation problem in cross-cultural analyses (e.g., the ten items to be repeated in the recall tests are common place, yet they might be perceived differently in the

various countries). With the available data, we are unfortunately not able to test whether such differences in survey procedures could affect results and, if so, in which direction. However, all the samples we use were meant to be nationally representative.

Furthermore, the cognitive tests in SHARE allow us to examine cognitive abilities, as measured in the psychological and epidemiological literature (Christelis, Jappelli, and Padula 2010). In addition, as discussed in Mazzonna and Peracchi (2012), the test format adopted by SHARE assesses cognitive functions in a way that is highly correlated with the Mini-Mental State Examination (MMSE) (Folstein, Folstein, and Mc Hugh 1975), a screening tool frequently used by health-care providers around the world to assess overall brain function.

Unfortunately, data on the cognitive abilities of nationally representative samples from the English Longitudinal Study of Ageing (ELSA) and the Health and Retirement Study (HRS) in the USA differ from SHARE in the inclusion and/or the formulation of some cognitive tests. For this reason, we have not considered them in this study (such analysis is available on request).

Nevertheless, the results of this study offer a motivation for more data collection and extensive analyses that could allow comparing different cohorts. Identifying the specific mechanisms at work and, furthermore, identifying which subgroups of the population perform poorly in terms of life cycle development of skills can have important policy implications, but goes beyond the aim of this paper. We hypothesise that these may be related to both obligatory schooling and basic training, but also targeted through intermediary or advanced formal education programs (Schneeweis, Skirbekk, and Winter-Ebmer 2014).

Moreover, we believe that variations in economic activity (i.e., employment and retirement patterns) could potentially play an important role. Past studies have found that working until higher ages results in greater investment in cognitive skills, higher participation in on-the-job-training, healthier lifestyles, and greater mental activity levels (Bonsang, Adam, and Perelman 2012). The large differences in sex-specific employment, in particular the entry of the female population into the labour market that has taken place at different times across the countries considered (Boeri, Del Boca, and Pissarides 2005), may additionally account for cross-country variation in cognition. Among a number of other national-level factors, social structures as well as culture and lifestyle could significantly affect the development of cognition over time. Diet, for example, can negatively affect mental performance through excessive undernutrition, but also excessive intake of oils, salt, sugar, and saturated fats, or if the diet lacks specific nutrients and vitamins (Leyse-Wallace 2013). It has also been shown that a greater level of atmospheric pollution negatively affects cognitive development and functioning at the national level

(WHO 2013). Opportunities for physical and mental activities, but also access to local facilities are positively associated with being active and this can in turn inform the development of policies to promote cognitive health of the population (Booth et al. 2000). Furthermore, attitudes towards ageing are likely to help explain the between-country variation in cognitive performances over time (for a discussion, see e.g., Gutchess and Indeck 2009). Additional cross-country comparisons of average cognitive functioning at different ages, on different data for the same cohort of individuals, or on different cohorts can be performed once younger cohorts for which school tests are available reach older ages and are observed in ageing surveys. Further results may provide support for our findings and help overcome possible national differences in schooling testing and assessment traditions.

5. Acknowledgments

References

Boeri, T., Del Boca, D., and Pissarides, C.A. (2005). *Women at work: An economic perspective*. OUP Catalogue. Oxford: Oxford University Press.

Börsch-Supan, A., Brugiavini, A., Jürges, H., Mackenbach, J., Siegrist, J., and Weber, G. (Eds.). (2005). *Health, ageing and retirement in Europe - First results from the Survey of Health, Ageing and Retirement in Europe*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).

Bonsang, E., Adam, S., and Perelman, S. (2012). Does retirement affect cognitive functioning? *Journal of Health Economics* 31(3): 490–501. doi:10.1016/j.jhealeco.2012.03.005.

Booth, M.L., Owen, N., Bauman, A., Clavisi, O., and Leslie, E. (2000). Social-cognitive and perceived environment influences associated with physical activity in older Australians. *Preventive Medicine* 31(1): 15–22. doi:10.1006/pmed.2000.0661.

Christelis, D., Jappelli, T., and Padula, M. (2010). Cognitive abilities and portfolio choice. *European Economic Review* 54(1): 18–38. doi:10.1016/j.euroecorev.2009.04.001.

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. New York: Lawrence Erlbaum Associates.

Comijs, H.C., van den Kommer, T.N., Minnaar, R.W.M., Penninx, B.W.J.H., and Deeg, D.J.H. (2011). Accumulated and differential effects of life events on cognitive decline in older persons: Depending on depression, baseline cognition, or ApoE ε4 status? *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 66B(1): 111–120. doi:10.1093/geronb/gbr019.

Deary, I.J., Whalley, L.J., Lemmon, H., Crawford, J.R., and Starr, J.M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence* 28(1): 49–55. doi:10.1016/S0160-2896(99)00031-8.

Engelhardt, H., Buber, I., Skirbekk, V., and Prskawetz, A. (2010). Social involvement, behavioural risks and cognitive functioning among older people. *Ageing and Society* 30(5): 779–809. doi:10.1017/S0144686X09990626.

Folstein, M.F., Folstein, S.E., and Mc Hugh, P.R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 12(3): 189–198. doi:10.1016/0022-3956(75)90026-6.

Glymour, M.M., Kawachi, I., Jencks, C.S., and 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.

Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education: Principles, Policy & Practice* 11(3): 319–330. doi:10.1080/0969594042000304618.

Gutchess, A.H. and Indeck, A. (2009). Cultural Influences on Memory. In Chiao, J.Y. (ed.). *Progress in Brain Research (*Volume 178). München: Elsevier: 137–150. doi:10.1016/S0079-6123(09)17809-3.

Husén, T. (Ed.). (1967a). *A comparison of twelve countries: International study of achievement in mathematics* (Vol. 1). Uppsala: Almqvist & Wiksells.

Husén, T. (Ed.). (1967b). *A comparison of twelve countries: International study of achievement in mathematics* (Vol. 2). Uppsala: Almqvist & Wiksells.

Jerrim, J. (2011). England's "plummeting" PISA test scores between 2000 and 2009: Is the performance of our secondary school pupils really in relative decline?. London: University of London (DoQSS Working Paper; 11–09).

Kendall, M.G. and Smith, B.B. (1939). The problem of m rankings. *Annals of Mathematical Statistics* 10(3): 275–287. doi:10.1214/aoms/1177732186.

Leyse-Wallace, R. (2013). *Nutrition and mental health*. CRC Press. doi:10.1201/b13745.

Mazzonna, F. and Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review* 56(4): 691–710. doi:10.1016/j.euroecorev.2012.03.004.

Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., and Preuschoff, C. (2009). *TIMSS 2011 Assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

OECD (2006). *Live longer, work longer. Ageing and employment policies*. Paris: OECD Publishing. doi:10.1787/9789264035881-en.

OECD (2011). *PISA 2009 results: Students on line*. Paris: OECD Publishing.

Richards, M., Shipley, B., Fuhrer, R., and Wadsworth, M.E.J. (2004). Cognitive ability in childhood and cognitive decline in mid-life: Longitudinal birth cohort study. *British Medical Journal* 328(7439): 552–554. doi:10.1136/bmj.37972.513819.EE.

Schneeweis, N., Skirbekk, V., and Winter-Ebmer, R., (2014). Does education improve cognitive performance four decades after school completion? *Demography* 51(2) 619–643. doi:10.1007/s13524-014-0281-1.

Skirbekk, V., Stonawski, M., Bonsang, E., and Staudinger, U.M. (2013). The Flynn effect and population aging. *Intelligence* 41(3): 169–177. doi:10.1016/j.intell.2013.02.001.

Tesch-Römer, C. and von Kondratowitz, H.-J. (2006). Comparative ageing research: A flourishing field in need of theoretical cultivation. *European Journal of Ageing* 3(3): 155–167. doi:10.1007/s10433-006-0034-0.

Thorvaldsson, V., Hofer, S.M., Berg, S., and Johansson, B. (2006). Effects of repeated testing in a longitudinal age-homogeneous study of cognitive aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 61(6): 348–354. doi:10.1093/geronb/61.6.P348.

WHO (2013). *Review of evidence on health aspects of air pollution – REVIHAAP Project*. Copenhagen, Denmark: World Health Organization.

Wolf, R.M. (1967). *Data bank manual. Phase I: International study of achievement in mathematics: A comparison of twelve countries*. Uppsala: Almqvist & Wiksells.

Zimprich, D. and 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.