

Research Articles

Executive Function Flynn Effects Are Independent of Psychometric *g* in Aging Populations Across Europe (2005-2022)

Denise Andrzejewski^{1,2} , Jonathan Fries¹, Jakob Pietschnig¹

¹ Department of Developmental and Educational Psychology, University of Vienna, ² SoSS / Psychology, Heriot-Watt University Dubai

Keywords: Flynn effect, aging, positive manifold of intelligence, cognitive abilities

<https://doi.org/10.65550/001c.157760>

Intelligence & Cognitive Abilities

Vol. 2, Issue 1, 2026

Systematic generational test score increases, commonly known as the Flynn effect, have been extensively documented in younger populations. However, much less is known about whether and how this phenomenon manifests in later adulthood. This study provides the first systematic examination of the Flynn effect and changes in the positive manifold of intelligence among older adults (aged 50+) across 22 countries from 2005 to 2022, using data from the Survey of Health, Ageing, and Retirement in Europe (SHARE). Results revealed predominantly positive Flynn effects across five cognitive domains, with verbal fluency exhibiting the strongest gains ($d = 0.48$), followed by working memory ($d = 0.38$) and free recall ($d = 0.27$) whilst numeracy showed weaker effects (d range 0.07-0.08). When accounting for age and education, the observed effects were substantially attenuated, and were differentiated across domains and countries in terms of strength and direction. Interestingly, the positive manifold appeared to somewhat increase in strength, but showed considerable country-specific differences. In all, our findings demonstrate cross-temporal changes in executive functioning that are differentiated according to specific domains, but are unrelated to shifts in the positive manifold of intelligence.

Introduction

The Flynn effect describes systematic generational increases in intelligence test performance that were observed throughout much of the 20th century. These gains, typically averaging approximately three IQ points per decade, have been documented across numerous populations and test instruments (Pietschnig & Voracek, 2012). Increases were particularly pronounced in measures of fluid intelligence, while crystallized intelligence exhibited comparatively modest gains.

However, these trends followed a nonlinear trajectory, characterized by alternating periods of accelerated and attenuated growth. Although effect sizes varied across national contexts, the overall direction of changes was consistently positive. More recently, however, accumulating evidence suggests a departure from this trend. Several countries now exhibit deceleration (e.g., United States: Rindermann & Thompson, 2013), stagnation (e.g., Australia: Cotton et al., 2005) or a reversal of the Flynn effect (e.g., Denmark: Dutton et al., 2016; USA: Dworak et al., 2023). However, Dworak et al. (2023) observed substantial domain heterogeneity, with largely stagnating or negative trends in most abilities but a clear positive Flynn effect in their most fluid measure the three-dimensional mental rotation (for an overview see Rodgers, 2023). This suggests that the Flynn effect may no longer yield a universally positive trajectory.

These inconsistent trends in cognitive performance have prompted renewed attention to the theoretical models underlying intelligence research. Contemporary frameworks move beyond global IQ scores and emphasize the structural complexity of cognitive abilities. Central to this perspective is psychometric *g*, which is a consequence of the positive manifold of cognitive abilities (i.e., consistently positive intercorrelations between virtually all intelligence (sub-)tests). Earlier models, such as Cattell's theory of fluid and crystallized intelligence (Cattell, 1963, 1987; Horn, 1968), postulated a hierarchical structure, with *g* at the highest level of abstraction and subordinate correlated broad abilities, but did not differentiate more specific ability domains. However, Cattell's taxonomy represented over many decades at least implicitly the theoretical underpinning of many cognitive test instruments. Because Flynn effect research depends on archival, typically incidentally collected data, past research on long term cognitive trends was often limited to change assessments in terms of crystallized and fluid as well as fullscale IQ. This limited the capacity to detect differentiated ability changes. Should IQ changes be differentiated according to more specific abilities, as increasingly suggested by contemporary findings (e.g. Lazaridis, 2022), traditional assessments based solely on fluid and crystallized indices would be insufficient to capture these patterns. As a result, the inconsistent trajectories now observed may stem from the increasing availability of subtest-level scores from modern test instruments that

allow for the detection of domain-specific variation that may have previously been obscured by cruder metrics.

The Cattell-Horn-Carroll model (CHC) has become one of the most widely accepted intelligence models in the field to date (Cattell, 1963, 1987; Horn, 1968; McGrew, 2018; J. Schneider & McGrew, 2013). This model organizes abilities into a hierarchical structure comprising correlated broad stratum II and narrow stratum I domains that are subordinate to the stratum III *g* factor. Whilst stratum II represents essentially the same level of abstraction as the fluid and crystallized intelligence *sensu* Cattell, another fourteen broad domains are assumed in this stratum, each of which comprises multiple narrow abilities (W. Schneider & McGrew, 2022). It is therefore necessary to examine cross-temporal test score changes beyond the more general stratum III and II abilities to account for this more refined understanding of intelligence.

Because cognitive ability tests have become more refined as well, these changes in intelligence assessments may be responsible for the domain-specific trends that have recently begun to emerge in contemporary studies and that may have remained obscured in prior research. For example, recent findings indicate consistent gains in domain-specific knowledge, learning efficiency, and comprehension knowledge. By contrast, negative trends have been reported for spatial orientation and working memory capacity, while processing speed appears to remain largely stable (Lazaridis, 2022). Therefore, investigating domain-specific trajectories, rather than aggregated higher-order scores, is necessary to identify the mechanisms underlying these observed changes.

Similarly, changes in memory-related domains appear to be inconsistent. Some studies report improvements in declarative memory (Rönnlund & Nilsson, 2008), visual recall (Baxendale, 2010), and short-term memory (Wongupparaj et al., 2017), while others suggest stagnation (Gignac, 2015) or even declines in working memory (Lazaridis, 2022). Mixed findings have also been observed for executive function components: Whilst data from the U.S. indicate decreases in attention and working memory (Graves et al., 2021), recent Austrian results suggest moderate gains in selective attention among adults (Andrzejewski, Oberleiter, et al., 2024; Bugelnig et al., 2025). Memory performance has frequently been shown to covary with executive functioning (e.g. McCabe et al., 2010). These observations emphasize the importance of domain-specific analyses in the study of generational cognitive trends and executive functioning provides a theoretically coherent domain for this. Specifically, it is important to examine the Flynn effect on CHC stratum I level to clarify the nature, meaning, and causes of the Flynn effect.

In general, although mean test and subtest scores have generally increased over time, several studies reported a negative association between the Flynn effect and psychometric *g* (e.g., Must et al., 2003; Pietschnig & Voracek, 2012; Woodley et al., 2014; but see, Colom et al., 2001, for contrasting evidence). This pattern has been interpreted as a consequence of a weakening of the positive manifold of intelligence (Pietschnig et al., 2021). Two concurring hy-

potheses have been proposed to explain this pattern, both of which may be attributed to an increasing cognitive ability specialization of the general population, due to environmental reinforcement:

On the one hand, it is possible that all specific cognitive abilities have changed in strength at the population level while maintaining their relative importance (i.e., either increasing or decreasing in an identical manner), thereby producing increasingly asymmetric individual cognitive profiles over time (Pietschnig et al., 2021, 2023). This interpretation is consistent with work suggesting that observed declines in the positive manifold may emerge as a consequence of broad, domain-general shifts that affect all abilities simultaneously, albeit to differing degrees (Andrzejewski, Zeilinger, et al., 2024; Oberleiter et al., 2024). On the other hand, some abilities may have become more important than others due to environmental changes (e.g., increased demands on visuospatial processing, reasoning complexity, or digital literacy) resulting in diverging cross-temporal trajectories in which certain domains show gains while others stagnate or decline (Pietschnig, 2025; see also Lazaridis, 2022 for tentative evidence of domain-specific differentiation). abilities yielding increases, whilst others show stagnation or decreases; Pietschnig, 2025). offer plausible mechanisms for how changes in the positive manifold could emerge across cohorts: either because ability profiles become increasingly asymmetric despite all abilities yielding directionally consistent changes (i.e., everyone becomes more specialized, but not in the same way) or because different abilities follow increasingly distinct developmental and societal relevance pathways (i.e., some abilities become more important than others, thus yielding broadly similar specializations and profile asymmetries across individuals). In fact, both of these ideas are consistent with evidence for declines in the strength of the positive manifold that have been observed among child cohorts in France (Lynn & Cooper, 1993), educational datasets in Italy (Pietschnig et al., 2023), and intelligence test standardization samples of Austrian adults (Andrzejewski, Zeilinger, et al., 2024; Oberleiter et al., 2024). Accordingly, it is crucial to systematically investigate changes in the positive manifold cross-temporally to examine their role in these observed ability trends.

One central challenge in Flynn effect research concerns whether observed test score changes represent genuine population level ability differences (i.e., they are consequences of true population ability changes) or reflect a measurement artifact arising from changes in item properties across assessment periods (i.e., measurement invariance). Differential item functioning refers to systematic shifts in item difficulty or discrimination that may result from evolving cultural contexts or test taking behaviors (e.g., Beaujean & Osterlind, 2008; Gonthier & Grégoire, 2022; Wicherts et al., 2004). Without evidence for cross-temporal measurement invariance, such item level changes have been shown to lead to inaccurate change scores in terms of strength and even direction (Gonthier et al., 2021). Consequently, meaningful interpretations of generational score differences require measurement invariance to ensure

that observed effects reflect true ability changes rather than differences in item functioning between assessments (Lazaridis, 2022).

Finally, most of the available evidence about test score changes to date is based on rather young samples (e.g., military conscript-based large-scale assessments; Pietschnig & Voracek, 2012), thus leaving the Flynn effect in older adults underexplored. However, Flynn effect research into aging populations is increasing, with accumulating evidence indicating that secular cognitive gains extend into later adulthood and may be strongest in fluid or mixed-process abilities (e.g., Clouston et al., 2021; Dickinson & Hiscock, 2010; O’Keefe et al., 2023; Skirbekk et al., 2013; Zhang et al., 2024). Moreover, systematic investigations of cross-temporal changes of the positive manifold of intelligence remain scarce and are largely limited to specific countries or subpopulations. The present study addresses this gap by means of a cross-temporal analysis of stratum I cognitive ability scores across nine cohorts of adults in retirement age from 22 countries spanning the years 2005 to 2022. In addition, we examine changes in the positive manifold to assess potential changes in population-level cognitive differentiation. To this end, we use archival executive functioning data from the Survey of Health, Ageing and Retirement in Europe (SHARE), a longitudinal dataset that captures individual level data on health, well being, and retirement among individuals aged 50 years and older. The SHARE dataset provides an opportunity to examine whether Flynn effect patterns extend into late adulthood, where gains in fluency and memory would align with established FE trajectories.

Methods

The present study was preregistered on January 22, 2025, before accessing or analyzing any data. The preregistration protocol, supplementary materials, and corresponding analytic code are available on the Open Science Framework (OSF; <https://osf.io/5qdrp/>).

The Survey of Health, Ageing and Retirement in Europe (SHARE)

We used data from the Survey of Health, Ageing and Retirement in Europe (SHARE; <https://share-eric.eu>), a large-scale longitudinal survey that collects individual level data on the health, wellbeing, and further characteristics of individuals aged 50 years and older. The survey commenced in 2004, with data collection conducted across multiple waves in an approximately biannual rhythm. Wave 1 data were collected in 2004 and have continued until Wave 9 so far which has been completed in 2022 (Wave 9 release date was March 28, 2024). The SHARE dataset includes representative samples of non-institutionalized individuals aged 50 years and older from 29 countries.

Given the multinational scope of SHARE, all survey instruments were translated into the respective languages of participating countries. A standardized translation procedure was implemented, involving multiple independent translations followed by country-level validation to ensure

both linguistic and cultural comparability (Harkness, 2005). The SHARE questionnaire covers a broad range of domains, including demographics, physical and mental health, behavioral risk factors, cognitive functioning, healthcare utilization, financial situation, and social engagement. Data collection involved a combination of face-to-face computer-assisted personal interviews and self-report questionnaires (Börsch-Supan et al., 2013).

For the purposes of this study, we used cognitive functioning data from Wave 1 through Wave 9. The cognitive function battery in SHARE comprises indicators of working memory, free recall, verbal fluency, and basic mathematical reasoning. These measures are generally considered to reflect aspects of executive functioning (e.g. Amunts et al. 2018, McCabe et al., 2010).

Participants

Because we were interested in between-cohort changes, we included only unique participants across all waves. In other words, we omitted all follow-up interview responses of participants who participated in earlier waves but only included newly recruited participants in later waves. At least 100 participants within a country had to participate within a wave to be included in our analyses.

The final sample in the current study consists of 128,610 participants that averaged 64.29 years ($SD = 10.11$; 52.9% women) from 22 countries. For participant education level we used the International Standard Classification of Education (ISCED; UNESCO, 2011) provided within the SHARE dataset. Within-wave participant characteristics are provided in [Table 1](#) and country-specific characteristics are available from the online supplemental materials S1.

Measures

Memory

In the memory task, participants listen to a list of 10 words once and are then immediately asked to recall as many of them as they can in any order. The number of words recalled equals the resulting score, ranging from 0 to 10. Moreover, delayed recall is assessed by asking participants to recall any of these words after having completed distractor tasks (i.e., the numeracy I and II and verbal fluency tasks). Although SHARE provides an aggregated memory score, we examined immediate and delayed recall separately because they map on distinct Stratum I domains within the Cattell-Horn-Carroll (CHC) framework (J. Schneider & McGrew, 2013). Specifically, immediate recall corresponds to working memory capacity (Gwm-Wc), while delayed recall corresponds to free recall memory (Gl-M6).

Verbal Fluency

Verbal fluency was assessed using a semantic category fluency test, in which participants have to name as many animals as possible within 60 seconds. The total score is the number of correctly generated words. Verbal fluency has been well-established as a measure of executive function-

Table 1. Sample characteristics for all data across cohorts

	2005	2007	2011	2013	2015	2017	2020	2022
N	29,619	14,415	35,497	20,733	10,477	29,475	6,940	10,929
Age M (SD)	63.7 (10.5)	62.8 (10.5)	63.9 (10.4)	64.0 (10.6)	62.0 (10.6)	62.0 (10.2)	62.0 (10.8)	62.7 (9.6)
Sex								
Men	13,149 (44.4 %)	6,345 (44 %)	15,152 (42.7 %)	9,575 (46.2 %)	4,638 (44.3 %)	13,352 (45.3%)	3,156 (45.5 %)	4,895 (44.8 %)
Women	16,470 (55.6 %)	8,070 (56 %)	20,345 (57.3 %)	11,158 (53.8 %)	5,839 (55.7 %)	16,123 (54.7%)	3,784 (54.5 %)	6,034 (55.2 %)
Education								
ISCED Code 1	8,463 (19.25%)	3,502 (14.17%)	6,034 (11.19%)	2,943 (8.96%)	1,762 (8.73%)	1,999 (6.78%)	264 (1.92%)	729 (3.75%)
ISCED Code 2	5,289 (12.03%)	2,406 (9.74%)	7,341 (13.61%)	3,740 (11.39%)	1,824 (9.04%)	2,802 (9.51%)	939 (6.84%)	1,587 (8.17%)
ISCED Code 3	8,193 (18.64%)	4,550 (18.41%)	13,054 (24.20%)	6,864 (20.90%)	4,205 (20.83%)	8,219 (27.88%)	3,351 (24.41%)	5,338 (27.46%)
ISCED Code 4	766 (1.74%)	769 (3.11%)	2,021 (3.75%)	828 (2.52%)	315 (1.56%)	1,134 (3.85%)	368 (2.68%)	465 (2.39%)
ISCED Code 5	5,504 (12.52%)	2,704 (10.94%)	6,818 (12.64%)	5,233 (15.93%)	2,286 (11.33%)	3,350 (11.37%)	1,976 (14.39%)	2,778 (14.29%)
ISCED Code 6	143 (0.33%)	48 (0.19%)	365 (0.68)	214 (0.65%)	54 (0.27%)	69 (0.23%)	81 (0.59%)	110 (0.57%)

ing (Aita et al., 2019), involving cognitive flexibility, strategy formation, and rapid retrieval of semantic knowledge (Barbey et al., 2012). Within the CHC model, this measure as corresponds to Naming Facility (Gr-NA), reflecting rapid lexical retrieval and semantic fluency.

Numeracy

In the numeracy I task, the ability to perform basic numerical operations is assessed by means of four items. Participants are asked to (i) calculate 10% of a certain number; (ii) find half of a number; (iii) compute a number based on two thirds of its value; and (iv) successively add 10% of its own value twice to a certain number. In the numeracy II task, participants' ability to perform serial arithmetic subtraction is assessed by instructing them to sequentially subtract seven from an initial value of 100, repeating this operation five consecutive times. Both numeracy tasks correspond to Quantitative Reasoning (Gf-RQ) within the CHC framework. Due to the differences in item composition and availability across waves, we analyzed these variables separately, examining the numeracy I variable from Waves 1 through 9, and the numeracy II variable from Waves 4 through 9.

Statistical Analyses

For the purposes of this study, we used data from Wave 1 through Wave 9 wherever available. Although some waves covered two calendar years, the year the survey concluded was used as the survey year. Accordingly, the survey years used for Waves 1 through 9 were 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2020, and 2022, respectively. Values marked as "Don't know" or "Not answered" were treated

as missing for demographic variables and as incorrect responses across all analytical procedures

Given the large sample size of the dataset, conventional null-hypothesis significance testing is of limited interpretive value. Accordingly, here we emphasize the interpretation of effect sizes rather than *p*-values and adopt the thresholds proposed by Funder and Ozer (2019) to assess the relevance of effects. Effect sizes were interpreted with Cohen *d* lower effect thresholds as very small (0.10), small (0.20), medium (0.41), large (0.63), and very large (0.87) as well as for η^2 as very small (.0025), small (.01), medium (.04), large (.09), and very large (.16), respectively.

We calculated all pairwise standardized mean differences (Cohen *d*) between the cognitive function scores between the earliest and latest available measurements available within each country. For ease of interpretation, Cohen *d* values of raw scores were transformed to changes per decade in the IQ metric (DIQ) via the following formula: $DIQ(\text{interval}) = [(d \times 15) / \text{interval}] \times 10$ (see Lazaridis, 2022). Consequently, positive values of *d* indicate cross-temporally increasing test performance, thus representing positive Flynn effects, while negative *d* values represent negative ones.

To evaluate the meaningfulness of observed cross-temporal changes, we assessed measurement invariance using multi group confirmatory factor analysis for scales with appropriate response formats. Measurement invariance indicates whether scales retain the same measurement properties across different groups or time points, and it represents a necessary condition for interpreting observed differences as reflecting true population ability changes rather than measurement artifacts (Meade & Lautenschlager, 2004; Putnick & Bornstein, 2016). Given item level data availability in the SHARE dataset, measurement invariance testing

was only feasible for the numeracy II measure. Response formats of the other cognitive tasks precluded formal measurement invariance assessments. Therefore, we interpret raw score-based changes for the remaining measures. This is reasonable because measurement non-invariance is most likely to affect verbal ability domains, but has comparatively little effects on numeracy, fluency, or recall tasks (e.g. Horn & Mcardle, 1992; Wicherts et al., 2004).

To examine changes in the intercorrelation of intelligence subdomains (i.e., the positive manifold of intelligence), we performed a factor analysis (FA), using the subscale scores of the SHARE data separately for each sample. In this vein, we determined the explained variance (R^2) of psychometric g in each cohort using a forced single-factor analysis, with subtest factor loadings representing the psychometric g loadings of the positive manifold. Changes in the positive manifold strength were assessed by comparing explained variances (i.e., R^2) of the single-factor analysis between the two cohorts. Two separate analyses were performed: In the first analysis, working memory, free recall memory, verbal fluency, and numeracy I were included. In the second one working memory, free recall memory, verbal fluency, and numeracy I and II were included. This was done because the numeracy II variable was introduced to the SHARE dataset only in 2011, limiting its availability to earlier waves. The proportion of explained variance (R^2) from each factor solution was used as an indicator of the positive manifold's strength, with larger R^2 values suggesting a stronger general factor characterized by higher intercorrelations among the cognitive measures. Between-cohort changes in R^2 may thus be interpreted as changes in the extent to which cognitive domains are interrelated at the population level (i.e., the positive manifold of intelligence). For cohorts 2011 to 2022, we compared McDonald's ω_h between samples (McDonald, 1999) to examine changes in the g saturation over time (see Revelle & Wilt, 2013).

We examined associations between survey year and cognitive function using weighted linear regression analyses, with weights reflecting cohort specific sample sizes to account for differences in estimate precision. Separate models were conducted for each cognitive domain. Missing data on either independent or dependent variables resulted in casewise exclusion from the respective analyses. To evaluate the extent to which age and education account for observed time trends, we employed hierarchical regression models, entering survey year in the first step and adding participant age and education as predictors in the second and third step, respectively.

All analyses were first conducted across all and subsequently within each country to assess the generality of observed patterns. We used R 4.3.2 for all our analyses (R Core Team, 2023) and psych (Revelle, 2024), ggplot2 (Wickham, 2016) for creating the Figures, and maturalearth (South, 2017) for creating heat maps.

Results

Pairwise mean differences

Descriptive statistics for all cognitive function domains across cohorts are presented in Table 2 and displayed in Figure 1. Standardized mean differences based on raw scores from the total sample yielded consistent positive Flynn effects across all assessed domains when comparing the earliest and latest available measurements within each country (Table 3, top row; Figure 2). In the full analytic sample, effect sizes ranged from very small to medium (Cohen's $d = 0.08$ to 0.48), corresponding to estimated gains of approximately 0.9 to 4.0 IQ points per decade.

Given that the number and timing of assessment waves varied across countries, all reported effects refer to the interval between the first and last available measurement occasion within each country. Detailed country-specific intervals are reported in Table 3.

When examining domain specific trends, verbal fluency exhibited the most consistent and largest positive effects across countries (d range: 0.15 to 0.90 ; Table 3). Only three countries deviated from this pattern, with small and large negative effects observed for Israel ($d = -0.24$) and Greece ($d = -0.48$), respectively and Hungary showing virtually no change ($d = 0.04$). Working memory also yielded highly consistent positive effects, with very small to very large effect sizes observed across nearly all countries (d range: 0.13 to 1.17). The only exception was Croatia, where a small negative effect was identified ($d = -0.18$).

For free recall memory, the results were less consistent. While small to very large positive effects were observed in most countries (d range: 0.19 to 1.13), several countries showed trivial or no change, including Israel, Latvia, Poland, and Portugal (d range: -0.05 to 0.07). Additionally, Croatia again showed a small negative effect ($d = -0.15$).

Numeracy I yielded small to very large positive effects in the majority of countries (d range: 0.10 to 0.80). However, trivial changes were observed for Germany, Greece, Israel, and Portugal (d range: 0.01 to 0.07). Small negative effects were also identified for Croatia ($d = -0.18$) and the Czech Republic ($d = -0.15$).

The least consistent pattern emerged for numeracy II. Some countries exhibited very small to medium positive effects (d range: 0.10 to 0.55), while others, including Croatia, Israel, and Portugal, showed small to medium negative effects (d range: -0.25 to -0.46). Additionally, virtual stagnation was observed for Denmark, Germany, Latvia, Luxembourg, the Netherlands, and Switzerland (d range: -0.03 to 0.05). Latent mean analyses for numeracy II yielded broadly comparable patterns, providing additional support for the observed trends based on raw scores.

Regression Analyses

We conducted precision-weighted hierarchical regression analyses to examine associations between survey year and cognitive functioning across all domains. These analyses were performed both for the full analytic sample and separately for each country. In all models, survey year was

Table 2. Descriptive statistics for cognitive function domains across cohorts

	Working Memory Capacity			Free-recall Memory			Verbal Fluency			Numeracy I			Numeracy II		
	M	SD	M _{change}	M	SD	M _{change}	M	SD	M _{change}	M	SD	M _{change}	M	SD	M _{change}
2005	4.80	1.86	0.00	3.34	2.02	0.00	18.61	7.46	0.00	3.29	1.16	0.00	-	-	-
2007	5.00	1.82	0.20	3.54	2.05	0.20	18.36	7.63	-0.25	3.35	1.14	0.05	-	-	-
2011	5.21	1.87	0.41	3.76	2.19	0.42	19.86	8.00	1.25	3.31	1.14	0.01	3.96	1.59	0.00
2013	5.24	1.86	0.44	3.89	2.18	0.56	20.59	7.80	1.97	3.40	1.11	0.11	4.06	1.53	0.10
2015	5.39	1.76	0.59	3.85	2.10	0.51	18.51	7.90	-0.10	3.34	1.00	0.04	4.17	1.37	0.21
2017	4.83	1.86	0.03	3.45	2.09	0.11	-	-	-	-	-	-	-	-	-
2020	5.51	1.76	0.71	4.01	2.14	0.67	21.77	8.23	3.16	3.47	1.06	0.17	4.16	1.39	0.20
2022	5.51	1.76	0.71	3.90	2.10	0.56	22.31	8.33	3.70	3.41	1.06	0.12	4.08	1.50	0.12

Note: Reference mean is set to 2005, hence M_{change} values are relative to the year 2005; M denotes mean and SD standard deviation.

Table 3. Test score changes of pairwise cohort comparisons across cognitive function domains and countries reported in Cohen's *d* and DIQ

		Working Memory (Cf008)		Free Recall (Cf016)		Verbal Fluency (Cf010)		Numeracy I (numeracy)		Numeracy II (numeracy2)	
		<i>d</i>	DIQ	<i>d</i>	DIQ	<i>d</i>	DIQ	<i>d</i>	DIQ	<i>d</i>	DIQ
All Countries	2005 - 2022	0.38***	3.2	0.27***	2.3	0.48***	4.0	0.1***	0.9		
	2011 - 2022									0.08***	1.0
Austria	2005 - 2022	0.58**	4.8	0.50**	4.1	0.47**	4.0	0.31**	2.6		
	2011 - 2022									0.08***	0.96
Belgium	2005 - 2022	0.64***	5.3	0.70***	5.8	0.64***	5.3	0.35***	2.9		
	2011 - 2022									0.09*	1.1
Croatia	2015 - 2022	-0.18***	-3.4	-0.15***	-2.9	0.15***	2.8	-0.18***	-3.4	-0.30***	-5.6
Czech Republic	2007 - 2022	0.35***	3.3	0.30***	2.8	0.9***	8.4	-0.15***	-1.4		
	2011 - 2022									0.16***	2.0
Denmark	2005 - 2022	0.57***	4.8	0.52***	4.3	0.70***	5.8	0.40***	3.4		
	2011 - 2022									0.04	0.4
Estonia	2011 - 2022	0.71***	8.9	0.59***	7.3	0.73***	9.2	0.26***	3.2	0.32***	4.0
Finland	2017 - 2022										
France	2005 - 2022	0.59***	4.9	0.52***	4.3	0.41***	3.4	0.19***	1.6		
	2011 - 2022									0.10*	1.3
Germany	2005 - 2022	0.28***	2.3	0.40***	3.4	0.52***	4.3	0.02	0.2		
	2011 - 2022									0.01	0.1
Greece	2007 - 2015	0.17***	2.8	0.19***	3.2	-0.48***	-8.1	0.01	0.2		
Hungary	2011 - 2022	0.43***	5.4	0.36***	4.6	0.04	0.5	0.32***	4.0	0.43***	5.4
Israel	2005 - 2020	0.13*	1.2	0.04	0.4	-0.24***	-2.3	0.03	0.2		
	2013 - 2020									-0.46***	-8.7
Italy	2005 - 2015	0.55***	7.5	0.5***	6.9	0.47***	6.5	0.25***	3.4		
	2011 - 2015									0.18***	5.4
Latvia	2017 - 2022	0.18***	4.4	-0.05	-1.2						
	2020 - 2022									-0.03	-1.3
Luxembourg	2013 - 2015	0.15**	7.7	0.10	4.8	0.29***	14.2	0.16**	7.9	0.02	0.9
Netherlands	2005 - 2013	0.16***	2.7	0.24***	4.0	0.31***	5.1	0.15***	2.5		
	2011 - 2013									-0.02	-1.0

		Working Memory (Cf008)		Free Recall (Cf016)		Verbal Fluency (Cf010)		Numeracy I (numeracy)		Numeracy II (numeracy2)	
Poland	2007 - 2022	0.21 ^{***}	2.0	-0.04	-0.4	0.71 ^{***}	6.7	0.35 ^{***}	3.3		
	2015 - 2022									-0.2 ^{***}	-3.8
Portugal	2011 - 2022	0.20 ^{***}	2.5	0.07	0.9	0.29 ^{***}	3.6	0.07	0.9	-0.25 ^{***}	-3.1
Slovenia	2011 - 2022	0.56 ^{***}	7.0	0.36 ^{***}	4.5	0.57 ^{***}	7.1	0.38 ^{***}	4.8	0.21 ^{***}	2.6
Spain	2005 - 2022	1.17 ^{***}	9.8	1.13 ^{***}	9.4	0.64 ^{***}	5.3	0.80 ^{***}	6.7		
	2011 - 2022									0.55 ^{***}	6.8
Sweden	2005 - 2020	0.51 ^{***}	4.8	0.62 ^{***}	5.8	0.45 ^{***}	4.3	0.29 ^{***}	2.7		
	2013 - 2020									0.25 ^{***}	4.7
Switzerland	2005 - 2020	0.51 ^{***}	4.7	0.63 ^{***}	5.9	0.75 ^{***}	7.0	0.21 ^{**}	2.0		
	2011 - 2020									0.05	0.7

Note: changes in IQ per decade (DIQ) can be calculated with the following formula: $DIQ(\text{interval}) = ((d * 15) / \text{interval})$

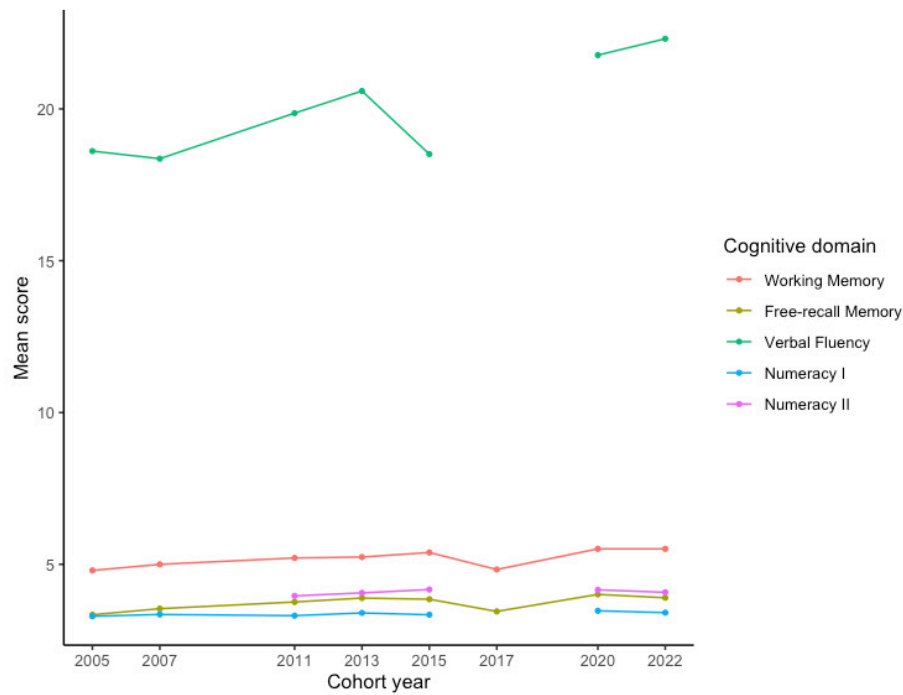


Figure 1. Cognitive function domains across cohorts

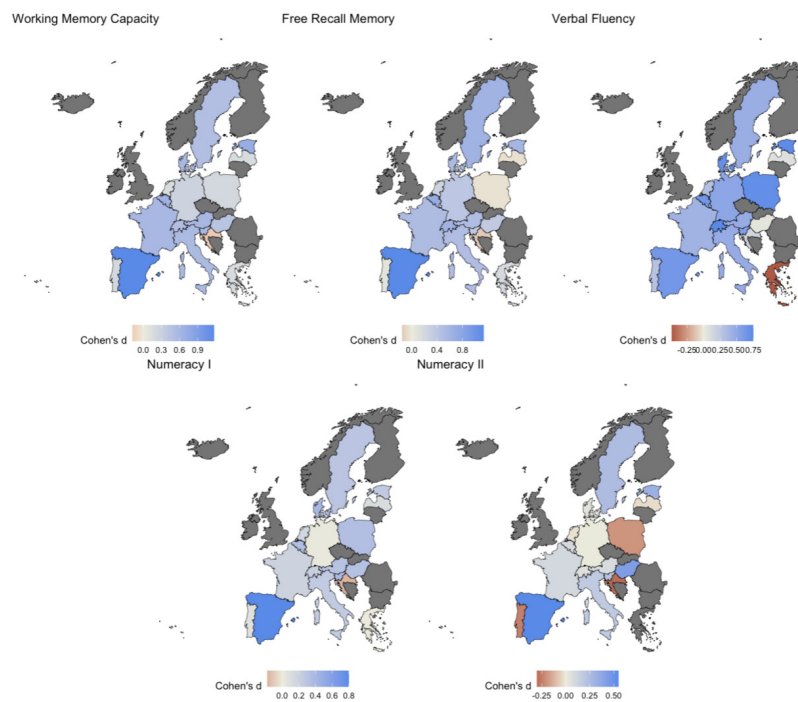


Figure 2. Heat map of cohort comparisons across domains

entered as the primary predictor in the first step, followed by participant age and education as additional predictors in subsequent steps (Table 6; detailed country level results are presented in the Supplementary Materials S1).

For the full analytical sample, survey year yielded statistically significant but trivial positive effects across all cognitive domains (partial eta squared range: .001 to .018; Table 6). In contrast, country level models revealed differentiated patterns depending on both cognitive domain and

country (see Supplementary Materials S1). Working memory capacity showed meaningful time trends in seven countries, with small to medium effect sizes (partial eta squared range: .011 to .044). Free recall memory exhibited small positive effects in ten countries (partial eta squared range: .011 to .038) and a large effect in one country (Poland: partial eta squared = .113). For verbal fluency, small positive effects were observed in seven countries (partial eta squared range: .010 to .030), with medium effects in two

Table 4. Correlation matrix for full analytic sample

	Working Memory	Free Recall	Verbal Fluency	Numeracy I	Numeracy II
Working Memory	-	.729	.489	.431	.416
Free Recall	.729	-	.446	.396	.376
Verbal Fluency	.489	.446	-	.404	.374
Numeracy I	.431	.396	.404	-	.542
Numeracy II	.416	.376	.374	.542	-

Note. Correlations are computed by pooling all eligible observations across countries and cohorts using all five subscales. Country- and cohort-specific correlations and matrices can be found in the supplement (<https://osf.io/5qdrp/files/osfstorage>).

Table 5. Hierarchical Omega (ω^2) Across Countries and Cohorts

	2011	2013	2015	2020	2022
Austria	0.708	-	-	0.641	0.498
Belgium	0.642	0.584	0.626	0.607	0.640
Croatia	-	-	0.683	0.589	0.716
Denmark	0.332	0.539	0.303	0.380	-
Estonia	0.683	-	0.539	0.634	0.510
France	0.607	-	0.469	0.505	0.571
Germany	-	0.593	-	0.531	0.407
Hungary	0.596	-	-	0.499	0.338
Israel	-	0.441	-	0.651	-
Italy	0.588	0.634	0.463	-	-
Latvia	-	-	-	0.587	0.508
Luxembourg	-	0.602	0.685	-	-
Netherlands	0.635	0.577	-	-	-
Poland	-	-	0.509	0.650	0.570
Slovenia	0.535	0.574	0.624	0.595	0.548
Spain	0.704	0.741	-	-	0.408
Sweden	-	0.467	-	-	0.433
Switzerland	0.517	-	-	0.377	-

Note. Values represent hierarchical omega coefficients (ω^2), indexing the proportion of reliable variance attributable to a general factor. Estimates are reported only for cohorts with at least 100 complete cases and a minimum of five indicators. Dashes indicate insufficient data or unstable factor solutions.

countries (Czech Republic: .075; Poland: .041). Non-trivial positive time trends for numeracy I were identified in three countries (partial eta squared range: .011 to .019), and for numeracy II in one country (Hungary: .011).

In addition to survey year effects, participant age consistently exhibited negative associations with cognitive function scores across all domains. The strongest age-related declines were observed for working memory capacity (partial eta squared = 0.145) and free recall memory (partial eta squared = 0.142; Table 6). Country-level analyses yielded comparable patterns. Model fit consistently improved with the stepwise inclusion of predictors, with Step 2 explaining a significantly larger proportion of variance than step 1, and step 3 yielding further significant incremental gains over step 2.

Highest educational qualification showed consistently positive associations with cognitive functions, with effect sizes in the medium to large range (partial eta squared = .073 to .124), thus reflecting better cognitive outcomes for individuals with higher ISCED education levels. Full coun-

try specific regression results for all predictors and domains are provided in Supplementary S1.

Positive manifold changes

Our examination of positive manifold changes revealed inconsistent patterns across the full analytical sample, with R^2 values reflecting both periods of increasing and decreasing strength. However, between 2005 and 2022 positive manifold strength as measured via R^2 significantly increased from .841 (95% CI [0.838; 0.844]) to .850 (95% CI [0.845; 0.855]) for the full analytical sample with substantial differentiation emerging according to country.

Notably, significant increases in the strength of the positive manifold were observed in Hungary and Israel, indicating somewhat progressively stronger intercorrelations between cognitive domains over time. Non-significant increases were observed in Austria, Belgium, Croatia, Czech Republic, France, Germany, Netherlands, Portugal, and Switzerland.

Table 6. Regressions for All data across domains

Model	Predictors	Working Memory Capacity			Free Recall Memory			Verbal Fluency			Numeracy I			Numeracy II		
		<i>adj. R</i> ²	<i>b</i>	η_p^2	<i>adj. R</i> ²	<i>b</i>	η_p^2	<i>adj. R</i> ²	<i>b</i>	η_p^2	<i>adj. R</i> ²	<i>b</i>	η_p^2	<i>adj. R</i> ²	<i>b</i>	η_p^2
1	Year	.007***	0.031***	.007	.005***	0.028***	.005	.018***	0.205***	.018	.001***	0.008***	.001	.001***	0.009***	.001
2	Year	.151***	0.029***	.008	.146***	0.026***	.005	.106***	0.189***	.017	.060***	0.006***	.001	.047***	0.009***	.001
	Age		-0.068***	.145		-0.077***	0.142		-0.226***	.090		-0.026***	.058		-0.031***	.046
3	Year	.212***	0.011***	.001	.198***	0.008***	<.001	.182***	0.107***	.006	.166***	-0.009***	.002	0.117***	-0.006***	<.001
	Age		-0.053***	.097		-0.063***	.100		-0.164***	.051		-0.015***	.021		-0.016***	.014
	Education			.086			.073			.094			.124			.086
	ISCED Code 2		0.483***			0.379***			2.385***			0.378***			0.641***	
	ISCED Code 3		0.946***			0.869***			4.551***			0.769***			1.099***	
	ISCED Code 4		1.094***			1.087***			5.603***			0.892***			1.124***	
	ISCED Code 5		1.533***			1.611***			6.959***			1.133***			1.320***	
	ISCED Code 6		1.730***			1.882***			7.955***			1.265***			1.374***	

In contrast, other countries, such as Greece and Latvia, exhibited significant decreases in R^2 values, suggesting a weakening of the positive manifold in those regions. Non-significant decreases were observed in Denmark, Estonia, Italy, Luxembourg, Poland, Slovenia, Spain, and Sweden. All R^2 trajectories are displayed in 3 and country specific analyses and bivariate intercorrelation matrices can be found in Supplementary Materials. Hierarchical omega ($\omega_\diamond$) estimates were broadly consistent with the R^2 -based trends, with most, but not all, countries displaying identical directional changes across both indices (Table 5)

Discussion

Here, we examined longitudinal trends in cognitive function and the strength of the positive manifold among older adults across 22 countries in the SHARE data between 2005 and 2022. Our findings indicate consistent positive Flynn effects across all five cognitive domains, while there were no consistent changes in cross-temporal positive manifold strength. These findings present several points of interest, as we discuss below.

Consistent with prior research on cognitive gains in adulthood (Pietschnig & Voracek, 2012) and older age (e.g., Clouston et al., 2021; Dickinson & Hiscock, 2010; Hessel et al., 2018; Zhang et al., 2024), we observed predominantly positive Flynn effects across all domains. Among these, verbal fluency exhibited the strongest and most consistent gains, whereas both numeracy domains showed comparatively smaller improvements. This pattern partially contrasts with previous literature that generally reported larger Flynn effects for fluid relative to crystallized abilities (Pietschnig & Voracek, 2012). In contrast, the verbal fluency task used here is not treated as a crystallized measure but as an indicator of executive functioning, consistent with contemporary frameworks that emphasize cognitive control and retrieval fluency (e.g. Amunts et al., 2018). However, recent work by Massa et al. (2024) also identified verbal fluency as a particularly robust and informative cognitive marker across the lifespan. It is possible that these seeming contradictions are a consequence of jingle-jangle fallacies in terms of effect classifications (Kelley, 1927) rather than representing genuine inconsistencies. Another plausible explanation for these contrasting results may be rooted in the distinct age-related trajectories of cognitive domains.

Abilities such as verbal fluency, which integrate crystallized lexical knowledge with fluid executive and speed components (retrieval, switching), may be considered to be comparatively stable in older adulthood due to preserved knowledge, whereas more abilities that are more closely related to fluid and reasoning abilities typically decline with aging-related neurobiological changes, particularly in frontal brain regions (Dickinson & Hiscock, 2010; Harada et al., 2013; Salthouse, 2015). In addition to these age-related mechanisms, cohort-linked environmental influences may have also contributed to domain-specific trends. Cohort differences observed here likely also reflect cumulative environmental advantages experienced by later-born cohorts echoing evidence that early-life conditions and soci-

etal improvements can produce lifelong resilience in cognitive functioning. For example, increasing societal and educational emphasis on communication and language-based skills may have reinforced verbal fluency performance relative to numeracy domains, which is consistent with evidence of environmentally driven gains in verbal abilities (Flynn, 2012; Greenfield, 2009; Rönnlund & Nilsson, 2008).

Our analyses also revealed small to moderate gains in memory-related domains, specifically working memory and free recall which involve core executive function processes. These findings support prior observations of increases in verbal and visiospatial working memory (Baxendale, 2010; Wongupparaj et al., 2017) but contrast indications of stagnation or decline (Gignac, 2015; Graves et al., 2021; Lazaridis, 2022). Such inconsistencies across studies may be conceivably attributable to cohort-specific historical and environmental exposures that shape cognitive reserve and health differently across generations. For example, older cohorts often had limited access to formal education, physically demanding occupations, and less comprehensive healthcare systems, all of which are associated with lower baseline cognitive performance and accelerated decline in later life (e.g. Sauerberg, 2021; Van Der Noordt et al., 2019). In contrast, more recent cohorts have typically benefited from extended schooling, cognitively stimulating work environments, and improved preventive and medical care, which are known to enhance cognitive reserve and mitigate age-related decline (e.g. Montemurro et al., 2025). These structural differences likely contribute to differentiated cognitive aging trajectories observed across studies. Given the established vulnerability of memory functions to age-related decline (Murman, 2015), the positive trends observed here suggest that protective influences have operated within more recent cohorts. Conceivably, these gains may reflect the emergence of cognitively healthier cohorts, consistent with evidence of long-run improvements in health and functional capacity in older European populations (Börsch-Supan et al., 2021). Cohort-related improvements in morbidity profiles, cardiovascular health, and everyday functioning have been identified as major contributors to cognitive advantages in later-born aging cohorts (e.g., Clouston et al., 2021; Zheng, 2021; Zhang et al., 2024). These mechanisms are characteristic for aging samples and differ fundamentally from conventional Flynn effect research.

In line with our broader findings, most countries exhibited small-to-moderate gains in both numeracy I and II. These domains primarily reflect fluid cognitive processes, such as quantitative reasoning and applied problem solving, which are generally considered vulnerable to neurocognitive aging and rely substantially on executive functioning components. The presence of positive Flynn effects in these fluid domains is therefore notable, particularly among older adults. Previous meta-analytic research has consistently demonstrated that gains in fluid abilities, including numerical reasoning, have occurred across various age groups (Pietschnig & Voracek, 2012).

Several factors may contribute to the observed trends across all domains. First, rising levels of formal education and greater participation in lifelong learning activities may have differentially preserved cognitive abilities. According to the cognitive reserve hypothesis (Stern, 2009; Whalley et al., 2004), sustained intellectual engagement fosters more efficient neural networks, particularly benefiting higher order cognitive functions such as working memory and verbal abilities. Higher levels of formal education are also associated with increased participation in lifelong learning activities (Nermond, 2024), which may further reinforce domain-specific gains, particularly in verbal abilities. However, while the comparatively larger improvements in verbal fluency may plausibly reflect the cumulative effects of educational and post-educational engagement, it is notable that these gains remain robust even when controlling for education, but are attenuated when adjusting for age (see [Table 6](#)). This pattern implies that factors beyond formal education (e.g., lifelong cognitive stimulation) contribute to cohort differences. Nonetheless, these verbal domain changes should be interpreted cautiously, given the potential influences of measurement non-invariance that could not be assessed in the present study due to the scale response format. However, these potential biases may be assumed to be limited due to the task's simplicity (i.e., mere production of words) that can be considered to be largely impervious to non-invariance. Of note, while verbal fluency is often considered a higher-order ability, its operationalization in the SHARE study may overlap with elementary retrieval processes, thereby limiting the specificity of inferences about educational effects.

Second, the increasingly complex cognitive demands of everyday life particularly through widespread engagement with digital technology may have contributed to cognitive maintenance and gains in older adulthood. The technological reserve hypothesis (Stern et al., 2023) posits that frequent interaction with digital devices, online platforms, and cognitively stimulating applications may enhance cognitive functioning through ongoing mental engagement and compensatory mechanisms (Benge & Scullin, 2025; Bordone et al., 2015; Charness & Schaie, 2010). However, substantial variation exists in the degree of technological engagement among older adults across European countries. Differences in economic resources, digital literacy, and cohort-specific attitudes toward technology adoption may moderate these effects and partly explain the observed cross-country variability in cognitive trends. Empirical evidence supports this view: older adults' internet use and digital competence are differentiated according to education, income, and national digital infrastructure (Cruz-Jesus et al., 2012; Rasi-Heikkinen & Doh, 2023; Seifert et al., 2021).

Positive manifold changes

Our analyses of positive manifold trends did not yield any clearly discernible pattern across cohorts, thus contrasting the idea of changes in the positive manifold strength as a potential cause for Flynn effect reversals that have been observed in other studies. While the full analytic

sample indicated a statistically significant overall increase in positive manifold strength between the earliest and latest assessment periods, our results suggest that these changes were neither gradual nor uniform. Instead, shifts in g -saturation appeared erratic between waves, both across and within countries (see [Figure 3](#)), complicating attempts to infer meaningful temporal structure. While previous research has reported temporally non-linear cognitive trends shaped by contextual factors such as education, economic conditions, or population health (Dutton et al., 2016; Flynn, 2012; Pietschnig & Voracek, 2015; Sundet et al., 2004), the present results do not point to a consistent pattern in the strength of intercorrelations. We note that in contrast to typical Flynn effect studies, aging populations are simultaneously affected by age-related decline, survivorship and attrition dynamics, and cohort-linked health differences (e.g. Salthouse, 2016; Zhang et al., 2021), thus possibly obscuring positive manifold change patterns.

Positive manifold trajectories also varied substantially across countries, suggesting the influence of country-specific contextual factors. For example, Italy showed a decline in positive manifold strength during the observed period, a finding broadly consistent with earlier reports of downward trends in Italian highschool student samples (Pietschnig et al., 2023). In contrast, Austria exhibited an increase in positive manifold strength across the observation period in our present data, despite earlier reports indicating declines based on cognitive test standardization data from younger Austrian cohorts (Andrzejewski, Zeilinger, et al., 2024; Oberleiter et al., 2024). The inconsistent patterns observed across countries and time points provide preliminary indications that Flynn effects related to executive functions may operate for the most part independently from trends in psychometric g .

It is important to note that most prior cross-temporal investigations of the positive manifold have focused on assessing standard cognitive ability domains. By contrast, the cognitive assessments used in the SHARE dataset primarily target executive functions rather than the broader range of cognitive abilities typically measured in psychometric ability tests. This distinction may limit the comparability with other examinations of positive manifold changes, because the presently used proxy for g is based on a comparatively narrow number of executive functioning domains.

Executive functions refer to higher-order cognitive processes involved in regulating goal-directed behavior, including working memory, inhibitory control, cognitive flexibility, and planning (Diamond, 2013; Miyake & Friedman, 2012). In contrast, traditional cognitive abilities as measured in psychometric intelligence tests encompass standard cognitive ability domains such as reasoning, spatial visualization, verbal comprehension, or abstract problem solving (Carroll, 1993; Jensen, 1998). Whilst cognitive functioning and other more standard cognitive ability measures are correlated, the intercorrelations between the presently rather narrow set of variables may not necessarily be expected to yield identical patterns compared to more salient operationalizations of the positive manifold of intelligence.

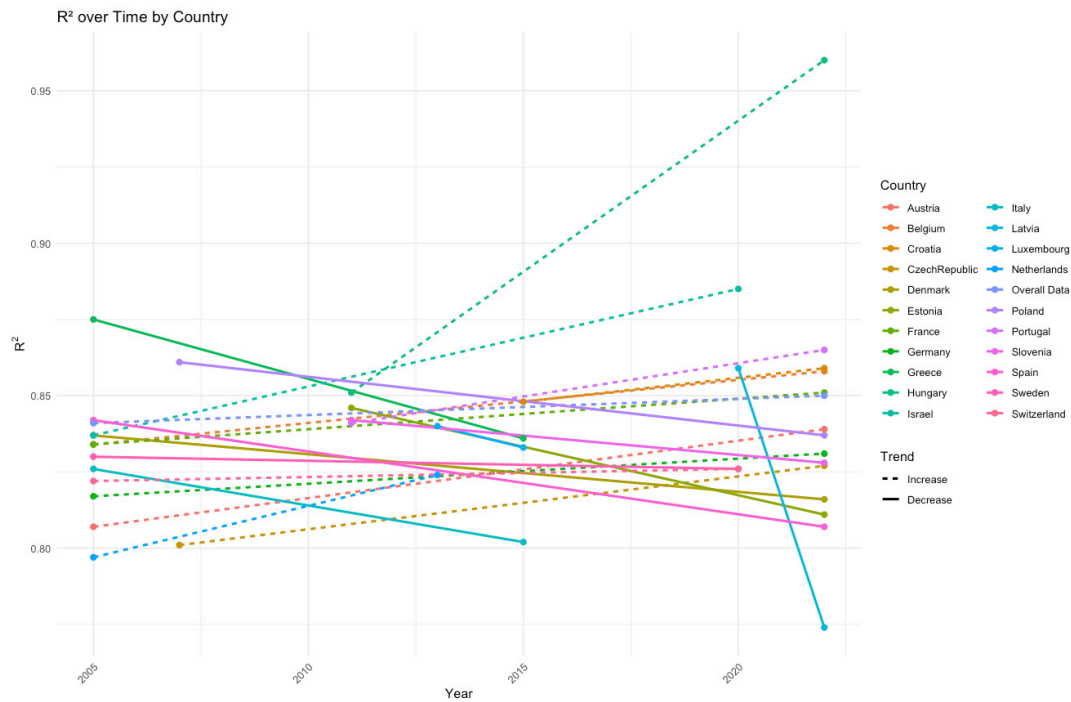


Figure 3. R^2 values 2005 – 2022 for All Countries

Table 7. Fit statistics for Numeracy II and measurement invariance levels for 2011 to 2022 (All Countries)

Model	χ^2	df	$p(\chi^2)$	CFI	Model Comp.	$\Delta\chi^2 (\Delta df)$	ΔCFI	Decision
M1: Overall	1275.834	5	< .001	1.000	-			
M2: Configural	1303.774	37	< .001	1.000	-			
M3: Strict	1434.240	57	< .001	1.000	M2	102.71 (20)	.000	

Limitations

In the present analyses, we applied a forced single factor extraction to estimate the positive manifold, modeling shared variance across the available cognitive tasks. However, the interpretation of this extracted general factor critically depends on the breadth and content of the cognitive indicators included (Johnson et al., 2004, 2008). Psychometric g , as traditionally conceptualized, is derived from a broad range of higher order cognitive abilities, whereas the present g factor likely reflects an executive function-specific construct given the domain composition of the SHARE battery (Johnson et al., 2008). Consequently, our results reflect longitudinal trends in executive functions rather than general intelligence per se.

Second, measurement invariance analyses could only be performed for one of the five cognitive tasks (numeracy II, see Table 7), due to subscale measurement properties. This raises concerns about the meaningfulness of the observed test score changes in the remaining four domains and in particular when it comes to verbal fluency. However, the non-verbal subscales of our cognitive functioning assessments should be largely unaffected from measurement non-invariance-related problems.

Third, some of the decadal IQ (DIQ) estimates derived in this study appear implausibly large and warrant careful interpretation. Because DIQs are standardized based on the length of time intervals between samples, even modest performance differences may yield disproportionately large DIQ values when intervals are brief. Thus, high DIQ estimates may exaggerate the magnitude of true cognitive changes, reflecting short-term interval sensitivity rather than substantial cognitive gains or losses. However, changes were for the most part remarkably consistent in terms of sign across countries, thus indicating meaningful evidence for predominantly positive cognitive functioning Flynn effects in our data.

Finally, due to our analytical approach, our results rely on refreshment samples added across SHARE waves. However, refreshment sampling was not consistently implemented across all countries or waves, introducing temporal gaps in the available data. For example, Greece, Italy, and Luxembourg provided no new samples after 2015, restricting temporal continuity, while Finland and Latvia joined SHARE only in 2017, thereby limiting the observation period for these countries. Moreover, in the interpretation of cohort differences in older-adult samples methodological challenges specific to aging research have to be accounted for. In contrast to studies of younger populations,

cognitive data from older adults are shaped by demographic processes such as age-related decline and selective survival. These factors affect who remains available for testing in later life. Although the current design compares cohorts at virtually identical ages and does not estimate within-person change, these background processes may conceivably influence the sampled populations. As a result, they may subtly affect observed differences. Previous studies have emphasized the importance of these demographic influences when analyzing secular trends in aging populations (e.g., Clouston et al., 2021; Rodgers, 2023). The present findings should be interpreted with these considerations in mind.

Final words

Taken together, our findings provide robust evidence for positive Flynn effects in executive functioning in older European adults that are differentiated in strength according to country and domain. While verbal and working memory

functions exhibited the most consistent improvements, effects for numeracy were less unequivocal. In all, no consistent systematic changes in the strength of the positive manifold emerged, thus indicating that the presently observed Flynn effect operates independently of psychometric *g*.

Credit Statement

Conceptualization: D.A. and J.P.; methodology: D.A. and J.P.; software: D.A.; validation: J.F. and J.P.; formal analysis: D.A.; investigation: D.A.; data curation: D.A. and J.F.; writing original draft preparation: D.A.; writing—review and editing: J.F. and J.P.; visualization: J.F. and D.A.; supervision: J.P.; project administration: D.A."

Submitted: October 22, 2025 CDT. Accepted: February 11, 2026 CDT. Published: March 16, 2026 CDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Aita, S. L., Beach, J. D., Taylor, S. E., Borgogna, N. C., Harrell, M. N., & Hill, B. D. (2019). Executive, language, or both? An examination of the construct validity of verbal fluency measures. *Applied Neuropsychology: Adult*, 26(5), 441–451. <https://doi.org/10.1080/23279095.2018.1439830>
- Andrzejewski, D., Oberleiter, S., Vetter, M., & Pietschnig, J. (2024). Increasing IQ Test Scores and Decreasing g: The Flynn Effect and Decreasing Positive Manifold Strengths in Austria (2005–2018). *Journal of Intelligence*, 12(12), 130. <https://doi.org/10.3390/jintelligence12120130>
- Andrzejewski, D., Zeilinger, E. L., & Pietschnig, J. (2024). Is there a Flynn effect for attention? Cross-temporal meta-analytical evidence for better test performance (1990–2021). *Personality and Individual Differences*, 216, 112417. <https://doi.org/10.1016/j.paid.2023.112417>
- Barbey, A. K., Colom, R., Solomon, J., Krueger, F., Forbes, C., & Grafman, J. (2012). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain*, 135(4), 1154–1164. <https://doi.org/10.1093/brain/aws021>
- Baxendale, S. (2010). The Flynn effect and memory function. *Journal of Clinical and Experimental Neuropsychology*, 32(7), 699–703. <https://doi.org/10.1080/13803390903493515>
- Beaujean, A. A., & Osterlind, S. J. (2008). Using Item Response Theory to assess the Flynn Effect in the National Longitudinal Study of Youth 79 Children and Young Adults data. *Intelligence*, 36(5), 455–463. <https://doi.org/10.1016/j.intell.2007.10.004>
- Benge, J. F., & Scullin, M. K. (2025). A meta-analysis of technology use and cognitive aging. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-025-02159-9>
- Bordone, V., Scherbov, S., & Steiber, N. (2015). Smarter every day: The deceleration of population ageing in terms of cognition. *Intelligence*, 52, 90–96. <https://doi.org/10.1016/j.intell.2015.07.005>
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., Zuber, S., & on behalf of the SHARE Central Coordination Team. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, 42(4), 992–1001. <https://doi.org/10.1093/ije/dyt088>
- Börsch-Supan, A., Ferrari, I., & Salerno, L. (2021). Long-run health trends in Europe. *The Journal of the Economics of Ageing*, 18, 100303. <https://doi.org/10.1016/j.jeoa.2020.100303>
- Bugelnig, A., Gruber, M., Birner, A., Langer, C., Voracek, M., & Pietschnig, J. (2025). Assessing the Flynn effect on attention test performance in Austrian conscripts (2010–2022). *ISSID 2025*.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1–22. <https://doi.org/10.1037/h0046743>
- Cattell, R. B. (Ed.). (1987). *Intelligence: Its structure, growth, and action*. Elsevier Science Publishers.
- Charness, N., & Schaie, K. W. (Eds.). (2010). *Impact of technology on successful aging*. Springer Pub.
- Clouston, S. A. P., Terrera, G. M., Rodgers, J. L., O’Keefe, P., Mann, F. D., Lewis, N. A., Wänström, L., Kaye, J., & Hofer, S. M. (2021). Cohort and Period Effects as Explanations for Declining Dementia Trends and Cognitive Aging. *Population and Development Review*, 47(3), 611–637. <https://doi.org/10.1111/padr.12409>
- Colom, R., Juanespinosa, M., & Garcia, L. (2001). The secular increase in test scores is a “Jensen effect.” *Personality and Individual Differences*, 30(4), 553–559. [https://doi.org/10.1016/S0191-8869\(00\)00054-4](https://doi.org/10.1016/S0191-8869(00)00054-4)
- Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R., & Crewther, S. G. (2005). A normative and reliability study for the Raven’s Coloured Progressive Matrices for primary school aged children from Victoria, Australia. *Personality and Individual Differences*, 39(3), 647–659. <https://doi.org/10.1016/j.paid.2005.02.015>
- Cruz-Jesus, F., Oliveira, T., & Bacao, F. (2012). Digital divide across the European Union. *Information & Management*, 49(6), 278–291. <https://doi.org/10.1016/j.im.2012.09.003>
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dickinson, M. D., & Hiscock, M. (2010). Age-related IQ decline is reduced markedly after adjustment for the Flynn effect. *Journal of Clinical and Experimental Neuropsychology*, 32(8), 865–870. <https://doi.org/10.1080/13803391003596413>
- Dutton, E., Van Der Linden, D., & Lynn, R. (2016). The negative Flynn Effect: A systematic literature review. *Intelligence*, 59, 163–169. <https://doi.org/10.1016/j.intell.2016.10.002>
- Dworak, E. M., Revelle, W., & Condon, D. M. (2023). Looking for Flynn effects in a recent online U.S. adult sample: Examining shifts within the SAPA Project. *Intelligence*, 98, 101734. <https://doi.org/10.1016/j.intell.2023.101734>
- Flynn, J. R. (2012). *Are we getting smarter? Rising IQ in the twenty-first century*. Cambridge University Press.
- Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Gignac, G. E. (2015). The magical numbers 7 and 4 are resistant to the Flynn effect: No evidence for increases in forward or backward recall across 85 years of data. *Intelligence*, 48, 85–95. <https://doi.org/10.1016/j.intell.2014.11.001>

- Gonthier, C., & Grégoire, J. (2022). Flynn effects are biased by differential item functioning over time: A test using overlapping items in Wechsler scales. *Intelligence*, 95, 101688. <https://doi.org/10.1016/j.intell.2022.101688>
- Gonthier, C., Grégoire, J., & Besançon, M. (2021). No negative Flynn effect in France: Why variations of intelligence should not be assessed using tests based on cultural knowledge. *Intelligence*, 84, 101512. <https://doi.org/10.1016/j.intell.2020.101512>
- Graves, L. V., Drozdick, L., Courville, T., Farrer, T. J., Gilbert, P. E., & Delis, D. C. (2021). Cohort differences on the CVLT-II and CVLT3: Evidence of a negative Flynn effect on the attention/working memory and learning trials. *The Clinical Neuropsychologist*, 35(3), 615–632. <https://doi.org/10.1080/13854046.2019.1699605>
- Greenfield, P. M. (2009). Linking social change and developmental change: Shifting pathways of human development. *Developmental Psychology*, 45(2), 401–418. <https://doi.org/10.1037/a0014726>
- Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal Cognitive Aging. *Clinics in Geriatric Medicine*, 29(4), 737–752. <https://doi.org/10.1016/j.cger.2013.07.002>
- Harkness, J. (2005). SHARE Translation Procedures and Translation Assessment. In A. Börsch-Supan & H. Jürges (Eds.), *The Survey of Health, Aging, and Retirement in Europe – Methodology*. MEA. https://share-eric.eu/fileadmin/user_upload/Methodology_Volumes/Methodology_2005.pdf
- Hessel, P., Kinge, J. M., Skirbekk, V., & Staudinger, U. M. (2018). Trends and determinants of the Flynn effect in cognitive functioning among older individuals in 10 European countries. *Journal of Epidemiology and Community Health*, 72(5), 383–389. <https://doi.org/10.1136/jech-2017-209979>
- Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 75(3), 242–259. <https://doi.org/10.1037/h0025662>
- Horn, J. L., & Mcardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117–144. <https://doi.org/10.1080/03610739208253916>
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.
- Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, 32(1), 95–107. [https://doi.org/10.1016/S0160-2896\(03\)00062-X](https://doi.org/10.1016/S0160-2896(03)00062-X)
- Johnson, W., Nijenhuis, J. T., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, 36(1), 81–95. <https://doi.org/10.1016/j.intell.2007.06.001>
- Kelley, T. L. (1927). *Interpretation of educational measurements*.
- Lazaridis, A. (2022). Domain-specificity of Flynn effects in the CHC-model: Stratum II test score changes in Germanophone samples (1996–2018). <https://doi.org/10.1016/j.intell.2022.101707>
- Lynn, R., & Cooper, C. (1993). A secular decline in Spearman's g in France. *Learning and Individual Differences*, 5(1), 43–48. [https://doi.org/10.1016/1041-6080\(93\)90025-N](https://doi.org/10.1016/1041-6080(93)90025-N)
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24(2), 222–243. <https://doi.org/10.1037/a0017619>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Psychology Press. <https://doi.org/10.4324/9781410601087>
- McGrew, K. S. (2018). *The new Cattell-Horn-Carroll (CHC) periodic table of cognitive elements: Build your own. IQ's Corner*. <http://www.iqscorner.com/2018/08/the-new-cattell-horn-carroll-chc.html>
- Meade, A. W., & Lautenschlager, G. J. (2004). A Comparison of Item Response Theory and Confirmatory Factor Analytic Methodologies for Establishing Measurement Equivalence/Invariance. *Organizational Research Methods*, 7(4), 361–388. <https://doi.org/10.1177/1094428104268027>
- Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, 21(1), 8–14. <https://doi.org/10.1177/0963721411429458>
- Montemurro, S., Rumiat, R. I., Pucci, V., Nucci, M., & Mondini, S. (2025). Cognitive reserve can impact trajectories in ageing: A longitudinal study. *Aging Clinical and Experimental Research*, 37(1), 93. <https://doi.org/10.1007/s40520-025-03000-z>
- Murman, D. (2015). The Impact of Age on Cognition. *Seminars in Hearing*, 36(03), 111–121. <https://doi.org/10.1055/s-0035-1555115>
- Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect. *Intelligence*, 31(5), 461–471. [https://doi.org/10.1016/S0160-2896\(03\)00013-8](https://doi.org/10.1016/S0160-2896(03)00013-8)
- Nermond, O. (2024). *Persistent inequalities in lifelong learning: What can we learn from the Adult Participation in Learning Survey 2024?* Learning & Working Institute. https://learningandwork.org.uk/news-and-policy/persistent-inequalities-in-lifelong-learning-what-can-we-learn-from-the-adult-participation-in-learning-survey-2024/?utm_source=chatgpt.com
- Oberleiter, S., Fries, J., Dejardin, F., Heller, J., Schaible, C., Vetter, M., Voracek, M., & Pietschnig, J. (2024). Inconsistent Flynn effect patterns may be due to a decreasing positive manifold: Cohort-based measurement-invariant IQ test score changes from 2005 to 2024. *Intelligence*, 107, 101867. <https://doi.org/10.1016/j.intell.2024.101867>
- O'Keefe, P., Muniz-Terrera, G., Voll, S., Clouston, S., Wanström, L., Mann, F. D., Lee Rodgers, J., & Hofer, S. M. (2023). Cohort Changes and Sex Differences After Age 50 in Cognitive Variables in the English Longitudinal Study of Ageing. *The Journals of Gerontology: Series B*, 78(10), 1636–1641. <https://doi.org/10.1093/geronb/gbad089>

- Pietschnig, J. (2025). Symposium: Effect differentiation in cognitive ability changes: The role of domain specificity in the Flynn effect. *ISSID 2025*.
- Pietschnig, J., Deimann, P., Hirschmann, N., & Kastner-Koller, U. (2021). The Flynn effect in Germanophone preschoolers (1996–2018): Small effects, erratic directions, and questionable interpretations. *Intelligence*, 86, 101544. <https://doi.org/10.1016/j.intell.2021.101544>
- Pietschnig, J., Oberleiter, S., Toffalini, E., & Giofrè, D. (2023). Reliability of the g factor over time in Italian INVALSI data (2010–2022): What can achievement-g tell us about the Flynn effect? *Personality and Individual Differences*, 214, 112345. <https://doi.org/10.1016/j.paid.2023.112345>
- Pietschnig, J., & Voracek, M. (2012). 100 years of IQ gains: A meta-analysis. *International Journal of Psychology*, 47, 120–120.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.3.2) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rasi-Heikkinen, P., & Doh, M. (2023). Older adults and digital inclusion. *Educational Gerontology*, 49(5), 345–347. <https://doi.org/10.1080/03601277.2023.2205743>
- Revelle, W. (2024). *psych: Procedures for psychological Psychometric, and personality research*. R package version 2.4.3. <https://CRAN.R-project.org/package=psych>
- Revelle, W., & Wilt, J. (2013). The general factor of personality: A general critique. *Journal of Research in Personality*, 47(5), 493–504. <https://doi.org/10.1016/j.jrp.2013.04.012>
- Rindermann, H., & Thompson, J. (2013). Ability rise in NAEP and narrowing ethnic gaps? *Intelligence*, 41(6), 821–831. <https://doi.org/10.1016/j.intell.2013.06.016>
- Rodgers, J. L. (2023). Eleven articles and 27 authors pay tribute to James Flynn: A summary and critique of special issue articles on the Flynn effect. *Intelligence*, 101, 101794. <https://doi.org/10.1016/j.intell.2023.101794>
- Rönnlund, M., & Nilsson, L.-G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analyses of data from a Swedish cohort study. *Intelligence*, 36(3), 192–209. <https://doi.org/10.1016/j.intell.2007.05.002>
- Salthouse, T. A. (2015). Implications of the Flynn effect for age–cognition relations. *Intelligence*, 48, 51–57. <https://doi.org/10.1016/j.intell.2014.10.007>
- Sauerberg, M. (2021). The impact of population's educational composition on Healthy Life Years: An empirical illustration of 16 European countries. *SSM - Population Health*, 15, 100857. <https://doi.org/10.1016/j.ssmph.2021.100857>
- Schneider, J., & McGrew, K. (2013). *The Cattell-Horn-Carroll (CHC) Model of Intelligence v2.2: A visual tour and summary*. <http://www.iapsych.com/chcv2.pdf>
- Schneider, W., & McGrew, K. S. (2022). The Cattell–Horn–Carroll Theory of Cognitive Abilities. In *Contemporary Intellectual Assessment (Fourth Edition)*. Guilford Press.
- Seifert, A., Cotten, S. R., & Xie, B. (2021). A Double Burden of Exclusion? Digital and Social Exclusion of Older Adults in Times of COVID-19. *The Journals of Gerontology: Series B*, 76(3), e99–e103. <https://doi.org/10.1093/geronb/gbaa098>
- Skirbekk, V., Stonawski, M., Bonsang, E., & Staudinger, U. M. (2013). The Flynn effect and population aging. *Intelligence*, 41(3), 169–177. <https://doi.org/10.1016/j.intell.2013.02.001>
- South, A. (2017). *Rnaturalearth: World Map Data from Natural Earth* [Computer software]. <https://github.com/ropenscilabs/rnaturalearth>
- Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, 47(10), 2015–2028. <https://doi.org/10.1016/j.neuropsychologia.2009.03.004>
- Stern, Y., Albert, M., Barnes, C. A., Cabeza, R., Pascual-Leone, A., & Rapp, P. R. (2023). A framework for concepts of reserve and resilience in aging. *Neurobiology of Aging*, 124, 100–103. <https://doi.org/10.1016/j.neurobiolaging.2022.10.015>
- Sundet, J., Barlaug, D., & Torjussen, T. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32(4), 349–362. [https://doi.org/10.1016/S0160-2896\(04\)00052-2](https://doi.org/10.1016/S0160-2896(04)00052-2)
- UNESCO. (2011). *International Standard Classification of Education (ISCED)*. UNESCO. <https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>
- Van Der Noordt, M., Hordijk, H. J., IJzelenberg, W., Van Tilburg, T. G., Van Der Pas, S., & Deeg, D. J. H. (2019). Trends in working conditions and health across three cohorts of older workers in 1993, 2003 and 2013: A cross-sequential study. *BMC Public Health*, 19(1), 1376. <https://doi.org/10.1186/s12889-019-7736-7>
- Whalley, L. J., Deary, I. J., Appleton, C. L., & Starr, J. M. (2004). Cognitive reserve and the neurobiology of cognitive aging. *Ageing Research Reviews*, 3(4), 369–382. <https://doi.org/10.1016/j.arr.2004.05.001>
- Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., Van Baal, G. C. M., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, 32(5), 509–537. <https://doi.org/10.1016/j.intell.2004.07.002>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed.). Springer International Publishing: Imprint: Springer. <https://doi.org/10.1007/978-3-319-24277-4>
- Wongupparaj, P., Wongupparaj, R., Kumari, V., & Morris, R. G. (2017). The Flynn effect for verbal and visuospatial short-term and working memory: A cross-temporal meta-analysis. *Intelligence*, 64, 71–80. <https://doi.org/10.1016/j.intell.2017.07.006>

Woodley, M. A., te Nijenhuis, J., Must, O., & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from g: The return of the Brand effect. *Intelligence*, 43, 27–34. <https://doi.org/10.1016/j.intell.2013.12.004>

Zhang, Y., Rodgers, J. L., O’Keefe, P., Hou, W., Voll, S., Muniz-Terrera, G., Wänström, L., Mann, F., Hofer, S. M., & Clouston, S. A. P. (2024). The Flynn effect and cognitive decline among americans aged 65 years and older. *Psychology and Aging*, 39(5), 457–466. <https://doi.org/10.1037/pag0000833>