738

2015

SOEPpapers

on Multidisciplinary Panel Data Research

738-2015

Secular Changes in Late-life Cognition and Well-being: Towards a Long Bright Future with a Short Brisk Ending?

Denis Gerstorf, Gizem Hülür, Johanna Drewelies, Peter Eibich, Sandra Duezel, Ilja Demuth, Paolo Ghisletta, Elisabeth Steinhagen-Thiessen, Gert G. Wagner, Ulman Lindenberger

DIW SOEP

SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
http://www.diw.de/soeppapers

Secular Changes in Late-life Cognition and Well-being:

Towards a Long Bright Future with a Short Brisk Ending?

Denis Gerstorf[1,2], Gizem Hülür[1], Johanna Drewelies[1], Peter Eibich[2], Sandra Duezel[3]

Ilja Demuth[4], Paolo Ghisletta[5,6],

Elisabeth Steinhagen-Thiessen[4], Gert G. Wagner[2,3,7], & Ulman Lindenberger[3]

[1]Humboldt University, Berlin, Germany

[2]German Institute for Economic Research (DIW Berlin), Berlin, Germany

[3]Max Planck Institute for Human Development, Berlin, Germany

[4]Charité – Universitätsmedizin Berlin, Germany

[5]University of Geneva, Switzerland

[6]Distance Learning University, Sierre, Switzerland

[7]Berlin University of Technology (TUB), Germany

Address correspondence regarding this manuscript to: Denis Gerstorf, Humboldt University Berlin, Institute of Psychology, Rudower Chaussee 18, 12489 Berlin, Germany; phone: +49-(0)30-2093-9422; fax: +49-(0)30-2093-9351; denis.gerstorf@hu-berlin.de.

Abstract

How socio-cultural contexts shape individual functioning is of prime interest for psychological inquiry. Secular increases favoring later-born cohorts in fluid intelligence measures are widely documented for young adults. In the current study, we quantify such trends in old age using data from highly comparable participants living in a narrowly defined geographical area and examine whether these trends generalize to quality of life indicators. To do so, we compared data obtained 20 years apart in the Berlin Aging Study (in 1990–93) and the Berlin Aging Study II (in 2013–14), applied a case-matched control design (per cohort, n = 161, M_{age} = 75), quantified sample selection using a nationally representative sample as the reference, and controlled for number of physical diseases. The later cohort performed better on the fluid intelligence measure (d = .85) and reported higher morale, less negative affect, and more positive affect (ds > .39) than the earlier cohort. We conclude that secular advances have resulted in better cognitive performance and perceived quality of life among older adults and discuss when and how advantages of later cohorts reach their limits.

<div align="right">Words: 185</div>

Keywords: Cohort; Cognitive Ability; Well-Being; Sociocultural Factors; Individual differences

Secular Changes in Late-life Cognition and Well-being:

Towards a Long Bright Future with a Short Brisk Ending?

Lifespan psychological and life course sociological research have long been interested in understanding how the historical times people live in shape individual development (Baltes, Cornelius, & Nesselroade, 1979; Bronfenbrenner, 1986; Elder, 1974; Riley, 1973; Rosow, 1978; Ryder, 1965; Schaie, 1965). Secular increases favoring later-born cohorts in performance of adolescents and young adults on fluid intelligence measures are widely documented (Flynn, 1999; Trahan, Stuebing, Fletcher, & Hiscock, 2014). Empirical studies have repeatedly shown that positive secular trends persist into late adulthood (Bowles, Grimm, & McArdle, 2005; Christensen et al., 2013; Finkel, Reynolds, McArdle, & Pedersen, 2007; Langa et al., 2008; Rönnlund & Nilsson, 2008; Schaie, Willis, & Pennak, 2005), but not up to the final years of old age (Gerstorf, Ram, Hoppmann, Willis, & Schaie, 2011; Hülür, Infurna, Ram, & Gerstorf, 2013). Adding to this body of research, the current study provides a comprehensive quantification of the nature and size of such trends in old age using data from participants residing in a narrowly defined geographical area. It is also an open question whether historical trends exist in quality of life indicators. Ratings of well-being condense self-evaluations of how people function across multiple domains, including health and cognition (Diener, 1984). If well-being ratings of people in their early 70s nowadays would indeed be higher than those of 70-year olds several decades ago, then this would indicate that (the perception of) people's quality of life has improved historically and that people are now living more joyful and satisfying lives. Some studies indicate that later-born cohorts report higher well-being (Sutin et al., 2013), whereas others indicate that later cohorts report lower well-being (Blanchflower & Oswald, 2004; Schilling, 2005). In this study, we compare data obtained about 20 years apart in the Berlin Aging Study (BASE, in 1990–93) and the Berlin Aging Study II (BASE-II, in 2013–14) to examine whether and how older adults' levels of cognitive performance and well-being have changed over recent decades.

Cohort Differences in Factors Shaping Cognition and Well-Being

Factors that shape cognitive performance and well-being are known to differ across cohorts. To begin with, years of schooling have increased significantly over the last century (e.g., by about 5.5 years for cohorts born from 1889 to 1973: Schaie et al., 2005). Because schooling is central for cognitive development (Ceci, 1991; Cahan & Cohen, 1989), increases in educational attainments are considered one major reason for historical increases in cognitive performance. Education also relates to well-being, with better educated individuals typically reporting higher well-being (Blanchflower & Oswald, 2004). More educated people often do better economically, and higher socio-economic status is often associated with better cognitive performance (Stern, 2002) and well-being (Pinquart & Sörensen, 2000). We thus expect that historical increases in education would lead to better cognition and higher well-being among later cohorts.

Health also shapes individual development, particularly in old age, when poor health and functional limitations become frequent and severe (Gerstorf & Ram, 2013). Health conditions such as cardiovascular diseases, heart disease, or stroke are known to undermine cognitive functioning (Spiro & Brady, 2008). Similarly, health–well-being associations have long been documented (DeNeve & Cooper, 1998). However, findings on historical trends in health are mixed (see Crimmins & Beltran-Sanchez, 2011; Christensen, Doblhammer, Rau, & Vaupel, 2009): Physical functioning of older adults has improved from the 1980s onwards, and common diseases such as arthritis have become less disabling. In contrast, prevalence rates of chronic health conditions (e.g., cardio-vascular diseases, cancer) and multimorbidity, both associated with compromised well-being (Charles, 2010), are higher among later-born cohorts. In light of these conflicting findings, indicators of health should be taken into account when examining secular trends in cognition and well-being.

Historical Trends in Cognitive Performance and Well-Being in Old Age

Evidence is accumulating that historical increases in cognitive performance among adolescents and young adults (Flynn, 1999; Trahan et al., 2014) persist into old age (Dodge, Zhu, Lee, Chang, & Ganguli, 2014; Zelinski & Kennison, 2007). For example, Danish 90+

year olds born later (1915) reached higher cognitive test scores than approximately same-aged

individuals born earlier (1905; Christensen et al., 2013). Similarly, 75+ year olds in Great

Britain in 2008/11 had lower prevalence rates of dementia than 75+ year olds in 1989/94

(Matthews et al., 2013). However, initial evidence indicates that these historical

improvements do not persist into the last years of life (Hülür et al., 2013). For example,

Gerstorf and colleagues (2011) demonstrated that differences between ages 50 and 80 years

favoring later-born cohorts (1914–1948, $M = 1921$) over earlier-born cohorts (1883–1913, M

$= 1904$) in the Seattle Longitudinal Study were not present anymore in the very last years of

life, probably resulting from population selection (e.g., Lindenberger, Singer, & Baltes, 2002;

see also Christensen, McGue, Petersen, Jeune, & Vaupel, 2009): Because of increases in life

expectancy, a larger portion of lower-functioning population segments have reached old age

among later-born cohorts than among earlier-born cohorts.

Studies examining historical trends in well-being have revealed inconclusive results.

Sutin and colleagues (2013) indicate that, for example, 75-year olds born later (1925) reported

higher well-being compared with same-aged cohorts born earlier (1905). In contrast, other

studies did not find evidence for historical increases in well-being and even report lower well-

being among later cohorts (Blanchflower & Oswald, 2004; Schilling, 2005). It remains

unclear whether cohort differences reported in other domains generalize to well-being as a

central quality of life indicator in old age.

The Present Study

We examined cohort differences in two central psychological domains, cognition and

well-being. Cognition functions as a resource people draw on when facing obstacles, and

constitutes a key component of successful aging (Rowe & Kahn, 1987). We used test scores

on the Digit Symbol subtest of the Wechsler Adult Intelligence Test (WAIS; Wechsler, 1955)

to index cognitive performance (cf. Hoyer, Stawski, Wasylyshyn, & Verhaeghen, 2004). In

longitudinal work, tests of perceptual speed have been found to be valid and reliable

indicators of cognitive decline in old age (e.g., Ghisletta & Lindenberger, 2012; Lindenberger

& Ghisletta, 2009; Tucker-Drob et al., 2014). Well-being represents a summary measure of

how people perceive their functioning across multiple domains of life (Diener, 1984). To cover a broad construct space, we include measures of morale as well as two emotional facets that tap into the frequency with which people had experienced positive affect and negative affect. To control for differences in relevant socio-demographic characteristics and for sampling differences between studies, we used propensity score matching (Coffman, 2011; Jackson, Thoemmes, Jonkmann, Lüdtke, & Trautwein, 2012) and identified case-matched controls in either cohort based on age, gender, and education. In follow-up analyses, we also covaried for number of physical diseases.

Method

In this report, we used data from the BASE (obtained 1990–93) and the BASE-II (obtained 2013–14). Detailed descriptions of participants, variables, and procedures can be found in previous publications (BASE: Baltes & Mayer, 1999; BASE-II: Bertram et al., 2014). Select details relevant to this report are given below.

Participants and Procedure

Berlin Aging Study (BASE): The initial BASE longitudinal sample consisted of 516 residents of former West-Berlin districts (age: $M = 84.92$, $SD = 8.66$, range = 70–103; 50% women) identified based on the obligatory city registry, recruited and tested from 1990–93. Participants were stratified by age and gender into six age brackets (70-74 years, 75-79 years, 80-84 years, 85-89 years, 90-94 years, and 95+ years). Here, we included data from the 447 participants (age: $M = 83.91$, $SD = 8.44$, range = 70–102; 49% women; education: $M = 10.98$ years, $SD = 2.04$, range = 7–18) who had provided information on the speed tasks (which required functional visual acuity), the three well-being indicators, and the socio-demographic variables (age, gender, education). Testing took place at the participants' place of residence (i.e., private household or institution) and was carried out in individual face-to-face sessions by trained research assistants. Sessions required an average of 90 minutes and, when necessary, were split into shorter units of assessment. Participants were allowed corrective visual aids (most frequently glasses).

Berlin Aging Study II (BASE-II): The BASE-II sample included residents of the greater metropolitan area of Berlin. Potential participants were recruited via a participant pool at the Max Planck Institute for Human Development (Berlin) and via advertisements in local newspapers and the public transportation system. We used a case-matched control design (see below) to minimize the effects of differences in sampling strategy between BASE and BASE-II. In this study, we included data from the 708 participants in the older subgroup of BASE-II (age: $M = 70.39$, $SD = 3.73$, range = 61–88; 51% women; education: $M = 14.41$, $SD = 2.85$, range = 7–18) who had provided valid information on the outcome and the socio-demographic variables. The cognitive test used here was administered as part of a comprehensive cognitive test battery and was carried out by trained interviewers in group sessions of three to six participants. The well-being measures were obtained as part of a take-home questionnaire. The younger subgroup of BASE-II (aged 20–35) was not considered here.

Measures

Cognitive performance was in both studies operationally defined by performance on the same version of the Digit Symbol test (Wechsler, 1955). The test consists of a code box with nine digit symbol pairs where each digit is paired with a corresponding symbol, and rows of double boxes with a digit in the top box and an empty lower box. Participants are asked to fill in as many corresponding symbols as possible in 90 seconds. The score indicates the number of correctly filled boxes, with penalty for wrong answers (score = correct – wrong). The Digit Symbol test was the only cognitive test administered in both BASE studies in a comparable fashion.

Well-being was indicated by three measures, morale, positive affect, and negative affect. *Morale* was assessed with three items selected from the Philadelphia Geriatric Center Morale Scale (PGCMS; Lawton, 1975) rated on a scale from 1 ("strongly disagree") to 5 ("strongly agree"). The responses to the items ("I sometimes feel that life isn't worth living", "I have a lot to be sad about", "I take things hard") were reverse-coded so that higher scores indicated higher morale (and higher well-being). *Positive affect and negative affect* were

assessed with four items each from the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). Participants were asked to rate how often they experienced a given affect (positive affect: attentive, inspired, active, and determined; negative affect: nervous, upset, jittery, and distressed) over the last year on a scale from 1 ("not at all") to 5 ("very often"). We acknowledge that these brief scales are not optimal for adequately measuring the constructs represented.

Socio-demographic variables. Three socio-demographic variables known to be associated with cognitive performance and well-being were used to pair BASE and BASE-II participants using propensity score matching procedures. *Age* was calculated as the difference between the date the Digit Symbol test was administered and a participant's date of birth and scaled in years. *Gender* was indicated by a binary variable (0 = men; 1 = women). *Education* was measured as the number of years the individual had spent in formal schooling. Because educational levels have been increasing across recent decades (Schaie et al., 2005), we note that the same level of formal education could be associated with a different standing in the socio-economic distribution across cohorts. For example, a person with a high school degree (which was not common in the earlier-born cohort) might have been more successful in finding employment in the earlier-born cohort as compared to the later-born cohort for whom many jobs have required a college degree. To control for such confounds, we standardized the years of education variable using population means and standard deviations separately for each cohort using reference data drawn from the German Socio-Economic Panel study (SOEP; Headey, Muffels, & Wagner, 2010). The reference population was 70+ year olds in 1990 for the BASE sample (*mean* = 10.69 years of education, *SD* = 2.06) and 60+ year olds in 2010 for the BASE-II sample (*mean* = 11.83 years of education, *SD* = 2.72).

In follow-up analyses, we also covaried for number of physical diseases. Based on participant reported and physician-observed medical diagnoses of moderate to severe mostly chronic illnesses, we computed a morbidity index largely based on the categories of the Charlson index (Charlson, Pompei, Ales, & MacKenzie, 1987) that represents a weighted classification scheme of common conditions, including cancer (e.g., leukemia), cardiovascular

(e.g., congestive heart failure) and metabolic diseases (e.g., diabetes mellitus). In both BASE studies, the diagnoses were determined via participant reports and clinical examinations and were supported by additional blood laboratory assessments (for details, see BASE: Steinhagen-Thiessen & Borchelt, 1999; BASE-II: Bertram et al., 2014).

Data Preparation

To minimize possible confounds and equate the cohort samples as closely as possible on age, gender, and education, we used propensity score matching procedures (Coffman, 2011; Foster, 2010; McCaffrey, Ridgeway, & Morral, 2004; Thoemmes & Kim, 2011). Calculating a logistic regression, we used 1:1 matching methods to select for each participant from the BASE cohort ($n = 447$) a 'twin' participant from the BASE-II cohort ($n = 708$) who had the same (or as similar as possible) age at baseline, gender, and cohort-normed education. To calculate a between-group distance matrix, the propensity score was logit-transformed as recommended in the propensity score matching literature (e.g., Rosenbaum & Rubin, 1985). We matched nearest neighbors with a caliper matching algorithm. The caliper (maximum allowable distance between matched participants) was continuously increased by steps of 0.01 until cohort differences in the matching variables were no longer reliably different from 0 at $p < .05$. Each participant in BASE was allocated the nearest neighbor from BASE-II only if the neighbor fell within the caliper distance. With a caliper of $c = 0.18$ SD, the matched BASE and BASE-II cohorts did no longer differ in age, gender, and cohort-normed years of formal education. A suitable neighbor in BASE-II could be identified for 161 BASE participants. Figure 1 shows cohort differences in the matching variables before and after the propensity score matching procedures. Descriptive statistics for study measures are given in Table 1 separately for the matched cohorts. The propensity score procedure was particularly successful for the covariates. For example, while the original age ranges of the BASE and BASE II studies vary considerably (70–103 vs. 61–88 years), the selected subsamples of either study resulted in a final, common age range of 65–89 years. We note that the large majority of participants in the matched samples were between ages 70 and 80 years. None of the BASE participants and only very few BASE-II participants were younger than age 70

years. We also note that the East vs. West-Berlin and the physical disease covariates that we examined in follow-up analyses could have alternatively been included as part of the propensity score-matching procedure. However, we faced a considerable trade-off between the number of variables included in the matching and the resulting sample size. Balancing the need for statistically controlling covariates with statistical power considerations based on sample size, we selected in an iterative fashion those variables and criteria for matching that simultaneously reduced group differences and secured moderate group sizes.

Results

Cohort Differences in Cognitive Performance and Well-being

Table 1 additionally presents the standardized mean difference between cohorts in the Digit Symbol and the three indicators of well-being. One-factorial ANOVAs with cohort membership (BASE vs. BASE-II) as the independent variable and the Digit Symbol and each of the three indicators of well-being, respectively, as the dependent variable showed that these cohort differences were reliably different from zero (F [1, 320] = 58.25, $p < 0.001$ for the Digit Symbol test; F [1, 320] = 12.49, $p < 0.001$ for the PGCMS; F [1, 320] = 32.68, $p < 0.001$ for positive affect; and F [1, 320] = 12.50, $p < 0.001$ for negative affect). The standardized mean difference between BASE and BASE-II cohorts in Digit Symbol test performance amounted to more than 80% of a *SD* unit ($d = 0.85$). On average, BASE-II participants correctly scored 9.02 items more than BASE participants. A meta-analysis by Hoyer et al. (2004) had estimated the average decline of the Digit Symbol score at −0.46 items per year of chronological age. Based on this estimate, the cognitive performance of BASE-II participants was on average 19.61 years "younger" relative to the BASE cohort. Moderately sized effects were found for indicators of well-being, with the BASE-II cohort reporting higher morale as measured with the PGCMS items ($d = 0.39$), more positive affect ($d = 0.64$), and fewer negative affect ($d = −0.39$) relative to the BASE cohort. Figure 2 illustrates both average cohort differences and the considerable individual difference in the four major outcomes under study. Taken together, these findings support notions of substantial historical improvements in cognitive performance and well-being in old age.

Follow-up Analyses

To substantiate our results, we conducted four sets of follow-up analyses. First, BASE participants had resided in districts of former West Berlin in 1990–93, whereas BASE-II participants were residents of the entire Berlin metropolitan area in 2013–14, including districts of East Berlin located in the former German Democratic Republic (GDR) prior to German reunification in 1989. To control for this potential confound, we excluded BASE-II participants from our sample who had resided at least for one year in the former GDR prior to 1989. We applied the above propensity score matching procedures to this smaller subsample ($n = 120$ in each cohort) and corroborated the earlier reported cohort differences with similar effect sizes: Relative to the earlier-born BASE sample, later-born BASE-II participants performed better on the Digit Symbol ($d = 1.02$), and reported higher morale ($d = 0.57$), more positive affect ($d = 0.86$), and fewer negative affect ($d = -0.55$).

Second, the BASE-II was conducted at multiple institutions and another Digit Symbol test performance had already been obtained for 609 out of 708 BASE-II participants an average of two years before. Thus, those BASE-II respondents had already been "trained" when the cognitive test under consideration was performed. In follow-up analyses, we used this earlier Digit Symbol score when available and redefined age accordingly. In these analyses, participants were on average 1.92 years younger, thereby reducing the number of BASE-II participants who could be matched with the initially older BASE participants ($n = 133$ in each cohort). Despite reduction in sample size, we replicated the cohort difference in Digit Symbol performance with this subsample and found an even larger effect size ($d = 1.21$) in favor of the BASE-II cohort, thereby ruling out the confounding effect of retest.

Third, attempting to account for individual and cohort differences in health, we calculated follow-up analyses that covaried for physical diseases. For participants with valid medical information (missing data: $n = 15$ in BASE, $n = 14$ in BASE-II), the physical disease status did not reliably differ across cohorts ($M = 0.78$, $SD = 0.85$ in BASE, $M = 0.74$, $SD = 1.25$ in BASE-II, $p = .75$). However, as one would expect, the earlier-born BASE cohort was more often suffering from, for example, myocardial infarction (13% in BASE, 3% in BASE-

II, χ^2 (1, N = 315) = 11.7, p < .001) and diabetes (29% in BASE, 14% in BASE-II, χ^2 (1, N = 311) = 9.6, p < .001) than the later-born BASE-II cohort. We applied the above propensity score matching procedures to this smaller subsample (n = 147 in each cohort) and again corroborated the earlier reported cohort differences with similar effect sizes: Relative to the earlier-born BASE sample, later-born BASE-II participants performed better on the Digit Symbol (d = 0.94), and reported higher morale (d = 0.47), more positive affect (d = 0.62), and fewer negative affect (d = −0.44). Based on these findings, we conclude that cohort differences in physical health cannot explain the observed differences in indicators of cognitive performance and well-being. At the same time, we do not exclude the possible role of other health factors that were not included in our analyses.

Finally, we examined the extent of selectivity of each BASE sample relative to the larger populations from which they were drawn, in a first step with the larger BASE and BASE-II samples and in a second step with the propensity-matched samples. Again, data obtained from participants aged 70+ in the nationally representative SOEP study in 1990 served as the reference for the BASE sample and SOEP data obtained from participants aged 60+ in 2010 served as the reference for the BASE-II sample. In particular, we first generated a variable indicating sample of origin (e.g., BASE vs. SOEP) and then estimated (separate) logistic regression analyses of education and household income on sample of origin (see Sassenroth, Kroh, & Wagner, 2013). Results for the larger BASE and BASE-II samples revealed that higher education (BASE: OR = 1.07, 95% Confidence Interval =1.01–1.13, p < .05; BASE-II: OR = 1.24, CI = 1.21–1.27, p < .001) and higher income (BASE: OR = 3.15, CI = 2.28–4.33, p < .001; BASE-II: OR = 1.37, CI = 1.15–1.64, p < .001) were each associated with participation in the BASE studies. This suggests that both BASE samples represent positive selections of the larger populations they were each drawn from (for an in-depth analyses of selection effects in BASE, see Lindenberger et al., 1999). Most important for our research question is that the amount of selection on education was relatively comparable across the BASE studies though the non-overlapping confidence intervals indicate that the later-born BASE-II cohort sample was slightly more select on education than

the earlier-born BASE cohort. In contrast, the earlier-born BASE cohort sample was considerably more select on income than the later-born BASE-II cohort sample. With our design that compares case-matched controls for education (as well as age and gender), we controlled for the slightly more positive selection of the BASE-II sample on education. Accordingly, the propensity-matched BASE sample remained positively select on education (OR = 1.11, CI = 1.02–1.21, p = .013), whereas the matched BASE-II sample was highly comparable with the national reference (OR = 1.05, CI = 0.99–1.11, p = .096). In a similar vein, the matched BASE sample remained positively select on income (OR = 3.24, CI = 2.11–4.98, p < .001), whereas the matched BASE-II sample was highly comparable with the national reference (OR = 1.08, CI = 0.75–1.55, p < .6). These selectivity analyses suggest that our case-matched control design represents a fair, if not conservative test of cohort differences.

Discussion

Our major objective in this study was to examine cohort differences in cognitive performance and well-being in old age. To do so, we compared case-matched control samples of cohorts of participants who were assessed either in 1990–93 or in 2013–14. Results revealed that the later-born BASE-II cohort showed higher levels of cognitive performance and reported higher well-being. To put our findings in perspective, we will discuss possible factors underlying these results and suggest that culture-based efforts (beyond more formal education) have been successful in improving cognitive performance and well-being in old age. Following Baltes and Smith (2003), we define culture very broadly as encompassing material and economic environments, medical practice, educational and media systems as well as psychological resources such as reading, writing, and computer literacy. We will also consider potential confounds and debate whether advantages for later-born cohorts may decrease in the future.

Our findings of cohort differences in cognitive performance add to previous reports by showing substantial historical increases (d = 0.85) in an established cognitive measure among people residing in a narrowly defined geographical area. Average increases in well-

being (*d*s from 0.39 to 0.64) were smaller, presumably reflecting the subjective nature of well-being ratings. Cohorts may not only differ in their living conditions, but also in the standards used to evaluate these conditions (Idler, 1993). For example, later cohorts may have higher standards for evaluating their current situation because they spent their earlier life under more favorable living conditions than later cohorts. Following this reasoning, the moderate increase in well-being within only 20 years of historical time is striking.

Before discussing potential mechanisms, we note that causal conclusions can of course not be inferred from our correlational, albeit quasi-experiment with respect to age, gender, and education, study. Although historical increases in education are one central reason for improved cognitive performance over the last century (Schaie et al., 2005), we found differences in cognition and well-being also after matching cohorts on years spent in formal education. However, we could only control for the quantity of education relative to one's cohort, not for its quality. The later-born BASE-II cohort probably benefitted from a higher-quality education by, for example, elementary mathematics curricula being shifted from less flexible forms of learning such as rote memorization to cognitively demanding mathematical tasks emphasizing fluid skills (Blair, Gamson, Thorne, & Baker, 2005). It would also be instrumental if future research would explore how further mechanisms, including self-regulation or working memory (see Jaeggi, Buschkuehl, Jonides, & Perrig, 2008), may have changed across cohorts, particularly in old age. Follow-up analyses also showed that health differences in number of medically diagnosed physical illnesses (primarily those that predict mortality directly) did not explain our findings. However, we could not consider cohort differences in other aspects of compromised health such as the severity of illnesses, diagnostic criteria, treatment options, or (chronic) limitations in physical functioning. For example, common ailments such as arthritis have become less disabling over recent decades and fewer individuals are nowadays transitioning from less to more severe forms of disability (Crimmins & Beltran-Sanchez, 2011).

To reconcile our findings with reports that secular advantages for later cohorts do not carry into late life (Gerstorf et al., 2011; Hülür et al., 2013), we heuristically sketch three

scenarios for how cohort differences manifest in old age (Fries, 1980; Olshansky, Hayflick, & Carnes, 2002; Vaupel, 2010). The scenarios conjointly assume that (a) later-born cohorts live longer than earlier-born cohorts, (b) death late in life is preceded by a period of steep deterioration, and (c) average functioning at the time of death has not changed historically (i.e., people die at older ages, but are not healthier or sicker at death than before). The scenarios differ in the supposed historical changes in the onset of decline in a given domain and the rate of decline. In scenario 1, the age at onset of decline remained stable historically, but later cohorts experience shallower rates of decline. Thus, people are aging longer and more slowly nowadays. In scenario 2, the onset of decline is postponed to later ages in later cohorts, but the rate of decline remains invariant historically. Thus, aging processes set-in later nowadays, but once started, these proceed at a similar rate. In scenario 3, the onset of decline is postponed even further into late life, but late-life decline is steeper among later cohorts. Thus, late-life deteriorations are "compressed" into a relatively short time window nowadays, but decrements are exacerbated (see Cheng, 2014). Of course, combinations of scenarios are possible and our cross-sectional data do not allow drawing firm conclusions. However, we would argue that the moderate to large effect sizes observed in the present study suggest that scenario 1 is by far the least likely of the three. Cohort differences in performance levels notwithstanding (Flynn, 1999), it is difficult to imagine how a gain of 19 years of age in average cognitive performance per 20 years of historical time can be attained by decline that is merely slowed rather than postponed to later ages. Instead, and in line with other reports and conceptual considerations (Lindenberger, 2014; Small, Dixon, & McArdle, 2011; Vaupel, 2010; Wilson et al., 2011), it appears that the onset of decline has shifted to older ages. To spawn scientific inquiry and societal debate, we borrowed from the title of Laura Carstensen's 2011 book on the future of aging, A Long Bright Future, and added for consideration that the end of life following this future may actually turn out to be short and brisk. At the same time, we note that independent replications and longitudinal extensions of our initial findings are needed to empirically test the three scenarios against one another.

Limitations and Outlook

We note several limitations of our study. First, our case-matched control design, which was necessary to ensure samples' comparability on three crucial demographic variables, reduces but does not eliminate the effects of sampling differences between studies (registry-based vs. convenience sample). Acknowledging that differences in study design hamper direct comparability of results obtained in the two studies, we attempted to quantify selectivity of both BASE samples using data that were obtained in the same time window in the nationally representative SOEP study (i.e., 1990 and 2010). These selectivity analyses revealed that although BASE started with a random draw from the city registry ($N = 1908$), participants in the final sample ($N = 516$) were about as positively selected as is known from other studies of older adults, including BASE-II. By applying the propensity score procedure to education (as well as to age and gender), we controlled for the slightly more positive selection of the larger BASE-II study on education. In the matched sample, however, BASE participants were, relative to the SOEP sample, more positively selected than BASE-II participants were on both education and income. Despite the fact that we were comparing more (educated and) affluent (relative to their same aged-peers) BASE participants with less (educated and) affluent (relative to their same aged-peers) BASE-II participants, considerable cohort differences emerged in favor of the latter. Although less positively select, the later-born BASE-II cohort performed better on the fluid intelligence measure and also reported higher well-being than the more positively select BASE cohort. As a consequence, the effect sizes reported here can be interpreted to represent a lower bound estimate of the true cohort difference. Relatedly, the limited age range of our matched sample (65–89 years; average of 75 years, standard deviation of less than 4 years) and the sparse data below age 70 years and above age 80 years limit the generalizability of our findings. It was thus also not possible to examine whether historical increases in old age reported here generalize to very old age and late life when broad-based dysfunctionalities prevail (Gerstorf & Ram, 2013). Based on our preceding considerations, we would expect smaller effects in very old age. It would also be

instructive to investigate in detail how secular trends were modulated by whether and at what age members of each of the two cohorts lived through one or both World Wars.

Second, the studies also differed in the cognitive testing situation (individual vs. group tests). We are not aware of reports documenting sizeable advantages of group testing, but if factors that undermine performances of older adults (e.g., test anxiety) were reduced in a group setting, the test situation would have favored the later cohort. We also acknowledge that our "solution" to the identifiability problem of age-period-cohort effects (Schaie, 1965) was to largely ignore differences attributable to the period of testing (e.g., the 1990's vs. 2010's). Independent replication of our findings is thus needed before firm conclusions are warranted. Finally, we note that previous studies have criticized the Flynn effect (Rodgers, 1999; Russell, 2007; Sundet, Barlaug, & Torjussen, 2004; Teasdale & Owen, 2008). For example, Wicherts et al. (2004) have argued that metric invariance does not hold across cohorts. Although we acknowledge the points raised, in our perspective the argument does not fully apply to our study because we have been working with a single and reliable test. Unfortunately, the Digit Symbol subtest was the only cognitive test in the two BASE studies that was administered in exactly the same way in the two cohorts. As a consequence, we cannot draw strong inferences about the extent to which the observed advantage for later-born cohorts on one cognitive test generalizes to the level of underlying cognitive abilities, particularly to those that are less susceptible to individual and cohort differences in physical health and education. Thus, the present study does not provide a definite answer to the question whether the Flynn effect operates at the level of abilities or at the level of individual tests. However, given that the Digit Symbol subtest of the WAIS loads highly on a factor of general intelligence and is highly sensitive to cognitive decline (e.g., Tucker-Drob et al., 2014), it is at least conceivable that some portion of the cohort difference observed in the present study generalizes to the level of cognitive abilities.

Conclusions

We examined cohort differences in cognitive performance and well-being. Relative to the earlier-born BASE cohort (assessed 1990–93), the later-born BASE-II cohort (assessed

2013–14) showed better cognitive performance and reported higher well-being, presumably due to culture-based advances in the course of the past century. Our results suggest that historical trends favoring later-born cohorts in cognitive performance (as widely documented for early life) carry into old age, constitute strong effects at age 75 years, and generalize to multiple key indicators of perceived quality of life (well-being). Future research should address the underlying mechanisms and explore in detail when and how advantages of later cohorts reach their limits.

Words: 5,531

References

Baltes, P. B., & Mayer, K. U. (Eds.). (1999). *The Berlin Aging Study: Aging from 70 to 100*. New York, NY: Cambridge University Press.

Baltes, P. B., & Smith, J. (2003). New frontiers in the future of aging: From successful aging of the young old to the dilemmas of the fourth age. *Gerontology*, *49*, 123–135. doi:10.1159/000067946

Baltes, P. B., Cornelius, S. W., & Nesselroade, J. R. (1979). Cohort effects in developmental psychology. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 61–87). New York, NY: Academic Press.

Bertram, L., Böckenhoff, A., Demuth, I., Düzel, S., Eckardt, R., Li, S.-C., Lindenberger, U., Pawelec, G., Siedler, T., Wagner, G. G., & Steinhagen-Thiessen, E. (2014). Cohort profile: The Berlin Aging Study II (BASE-II). *International Journal of Epidemiology*, *43*, 703-712. doi:10.1093/ije/dyt018

Blair, C., Gamson, D. A., Thorne, S., & Baker, D. P. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex. *Intelligence*, *33*, 93–106. doi:10.1016/j.intell.2004.07.008

Blanchflower, D. G., & Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, *88*, 1359-1386. doi:10.1016/S0047-2727(02)00168-8

Bowles, R. P., Grimm, K. J., & McArdle, J. J. (2005). A structural factor analysis of vocabulary knowledge and relations to age. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *60*, P234–P241.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development research perspectives. *Developmental Psychology, 22,* 723–742. doi:10.1037/0012-1649.22.6.723

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, *60*, 1239–1249. doi: 10.1111/1467-8624.ep9676597

Carstensen, L. (2011). *A long bright future*. New York, NY: Public Affairs.

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

Charles, S. T. (2010). Strength and vulnerability integration: A model of emotional well-being across adulthood. *Psychological Bulletin*, *136*, 1068–1091. doi:10.1037/a0021232.

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation.

Journal of Chronic Disease, *40*, 373-383.

Cheng, S. T. (2014). Double compression: A vision for compressing morbidity and caregiving in dementia. *The Gerontologist*, 54, 901–908. doi: 10.1093/geront/gnu015.

Christensen, K., Thinggaard, M., Oksuzyan, A., Steenstrup, T., Andersen-Ranberg, K., Jeune, B., McGue, M., & Vaupel, J. W. (2013). Physical and cognitive functioning of people older than 90 years: a comparison of two Danish cohorts born 10 years apart. *The Lancet*, *382*, 1507-1513. doi: 10.1016/S0140-6736(13)60777-1

Christensen, K., Doblhammer, G., Rau, R., & Vaupel, J. W. (2009). Ageing populations: The challenges ahead. *Lancet*, *374*, 1196-1208. doi:10.1016/S0140-6736(09)61460-4

Christensen, K., McGue, M., Petersen, I., Jeune, B., & Vaupel, J. W. (2009). Exceptional longevity does not result in excessive levels of disability. *Proceedings of the National Academy of Sciences*, *105*, 13274-13279. doi: 10.1073/pnas.0804931105

Coffman, D. L. (2011). Estimating causal effects in mediation analysis using propensity scores. *Structural Equation Modeling*, *18*, 357-369. doi: 10.1080/10705511.2011.582001

Crimmins, E. M., & Beltrán-Sánchez, H. (2011). Mortality and morbidity trends: is there compression of morbidity? *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *66*, P75–P86. doi:10.1093/geronb/gbq088

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A metaanalysis of 137 personality traits and subjective well-being. *Psychological Bulletin*, *124*, 197–229. doi:10.1037/0033-2909.124.2.197

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*, 542-575. doi: 10.1037/0033-2909.95.3.542.

Dodge, H. H., Zhu, J., Lee, C-W., Chang, C.-C. H., & Ganguli, M. (2014). Cohort effects in age-associated cognitive trajectories. *Journals of Gerontology: Series A: Biological Sciences and Medical Sciences*, *69*, M687–M694. doi: 10.1093/gerona/glt181

Elder, G. H., Jr. (1974). *Children of the Great Depression: Social change in life experience.* Chicago, IL: University of Chicago Press.

Finkel, D., Reynolds, C. A., McArdle, J. J., & Pedersen, N. L. (2007). Cohort differences in trajectories of cognitive aging. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *62*, P286–P294

Flynn, J. R. (1999). Searching for justice: the discovery of IQ gains over time. *American Psychologist*, *54*, 5-20. 10.1037/0003-066X.54.1.5

Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology, 46,* 1454-1480. doi: 10.1037/a0020204.

Gerstorf, D., Ram, N., Hoppmann, C. A., Willis, S. L., & Schaie, K. W. (2011). Cohort differences in cognitive aging and terminal decline in the Seattle Longitudinal Study. *Developmental Psychology*, *47*, 1026-1041. doi: 10.1037/a0023426

Gerstorf, D. & Ram, N. (2013). Inquiry into terminal decline: Five objectives for future study. *The Gerontologist*, *53,* 727-737. doi: 10.1093/geront/gnt046.

Ghisletta, P., Rabbitt, P., Lunn, M., & Lindenberger, U. (2012). Two thirds of the age-based changes in fluid and crystallized intelligence, perceptual speed, and memory in adulthood are shared. *Intelligence*, *40*, 260–268. doi: 10.1016/j.intell.2012.02.008

Headey, B., Muffels, R., & Wagner, G. G. (2010). Long-running German panel survey shows that personal and economic choices, not just genes, matter for happiness. *PNAS Proceedings of the Academy of Sciences of the United States of America*, *107*, 17922–17926. doi:10.1073/pnas.1008612107

Hoyer, W. J., Stawski, R. S., Wasylyshyn, C., & Verhaeghen, P., (2004) Adult age and digit symbol substitution performance: A meta-analysis. *Psychology and Aging, 19,* 211-214. doi: 10.1037/0882-7974.19.1.211

Hülür, G., Infurna, F. J., Ram, N., & Gerstorf, D. (2013). Cohorts based on decade of death: No evidence for secular trends favoring later cohorts in cognitive aging and terminal decline in the AHEAD study. *Psychology and Aging*, *28*, 115-127. doi: 10.1037/a0029965

Idler, E. L. (1993). Age differences in self-assessments of health: Age changes, cohort differences, or survivorship? *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *48*, P289–P300. doi: 10.1093/geronj/48.6.S289

Jackson, J. J., Thoemmes, F., Jonkmann, K., Lüdtke, O., & Trautwein, U. (2012). Military training and personality trait development: Does the military make the man or does the man make the military? *Psychological Science*, *23*, 270-277. doi: 10.1177/0956797611423545

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*, 6829–6833. doi: 10.1073/pnas.0801268105

Langa, K. M., Larson, E., Karlawish, J., Cutler, D., Kabeto, M., Kim, S., Rosen, A. (2008). Trends in the prevalence and mortality of cognitive impairment in the United States: Is there evidence of a compression of cognitive morbidity? *Alzheimer's and Dementia*, *4*, 134-144. doi: 10.1016/j.jalz.2008.01.001

Lawton, M. P. (1975). The Philadelphia Geriatric Center Morale Scale: A revision. *Journal of Gerontology*, *30*, 85-89. doi: 10.1093/geronj/30.1.85

Lindenberger, U. (2012). Human cognitive aging: Corriger la fortune? *Science*, *346*, 572-578. doi: 10.1126/science.1254403

Lindenberger, U., & Ghisletta, P. (2009). Cognitive and sensory decline in old age: Gauging the evidence for a common cause. *Psychology and Aging*, *24*, 1–16. doi:10.1037/a0014986

Lindenberger, U., Singer, T., & Baltes, P. B. (2002). Longitudinal selectivity in aging populations: Separating mortality-associated versus experimental components in the Berlin Aging Study (BASE). *Journals of Gerontology Series B: Psychological Sciences*, *57B*, P474-P482. doi: 10.1093/geronb/57.6.P474

Lindenberger, U., Gilberg, R., Litte, T. D., Nuthmann, R., Pötter, U., & Baltes, P. B. (1999). Sample selectivity and generalizability of the results of the Berlin Aging Study. In P. B. Baltes & P. B. Baltes (Eds.), *The Berlin Aging Study: Aging from 70 to 100* (pp. 56-82). New York: Cambridge University Press.

Matthews, F. E, Arthur, A., Barnes, L. E, Bond, J., Jagger, C., Robinson, L., & Brayne, C. (2013). A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. *The Lancet*, *382*, 1405-1412. doi: 10.1016/S0140-6736(13)61570-6

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods, 9,* 403-425. doi: 10.1037/1082-989X.9.4.403

Olshansky, S. J., Hayflick, L., & Carnes, B. A. (2002). Position statement on human aging. *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences, 57,* B292–B297. doi:10.1093/gerona/57.8.B292

Pinquart, M., & Sörensen, S. (2000). Influences of socioeconomic status, social network, and competence on subjective well-being in later life: A meta-analysis. *Psychology and Aging*, *15*, 187-224. doi: 10.1037/0882-7974.15.2.187

Riley, M. W. (1973). Aging and cohort succession: Interpretations and misinterpretations. *Public Opinion Quarterly, 37,* 35–49. doi:10.1086/268058

Rodgers, J. L. (1999). A critique of the Flynn effect: Massive IQ gains, methodological artifacts, or both? *Intelligence*, *26*, 337–356.

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*, 192-199. doi: 10.1016/j.intell.2007.05.002

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*, 33–38. doi: 10.2307/2683903

Rosow, I. (1978). What is a cohort and why. *Human Development, 21,* 65–75. doi:10.1159/000271575

Rowe, J. W., & Kahn, R. L. (1997). Successful aging. *The Gerontologist*, *37*, 433-440. doi: 10.1093/geront/37.4.433

Russell, E. W. (2007). Commentary: The Flynn effect revisited. *Applied Neuropsychology*, *14*, 262–266.

Ryder, N. B. (1965). The cohort as a concept in the study of social changes. *American Sociological Review, 30,* 843–861. doi:10.2307/2090964

Sassenroth, D., Kroh, M. & Wagner, G. G. (2013). Selectivity Processes in and Weights for the Berlin Aging Study II (BASE-II). *SOEPpapers on Multidisciplinary and Panel Data Research*, *608*.

Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64,* 92–107. doi: 10.1037/h0022371

Schaie, K. W., Willis, S. L., & Pennak, S. (2005). An historical framework for cohort differences in intelligence. *Research in Human Development*, *2*, 43–67. doi: 10.1080/15427609.2005.9683344

Schilling, O. K. (2005). Cohort- and age-related decline in elder's life satisfaction: is there really a paradox? *European Journal of Ageing*, *2*, 254-263. doi: 10.1007/s10433-005-0016-7

Small, B. J., Dixon, R. A., & McArdle, J. J. (2011). Tracking cognition–health changes from 55 to 95 years of age. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 66B*, i153–i161, doi:10.1093/geronb/gbq093

Spiro, A., III, & Brady, C. B. (2008). Integrating health into cognitive aging research and theory: Quo vadis? In S. M. Hofer & D. F. Alwin (Eds.), Handbook of cognitive aging: Interdisciplinary perspectives (pp. 260 –283). Thousand Oaks, CA: Sage. doi:10.4135/9781412976589.n16

Steinhagen-Thiessen, E., & Borchelt, M. (1999). Morbidity, medication, and functional limitations in very old age. In P. B. Baltes & K. U. Mayer (Eds.), The Berlin Aging Study: Aging from 70 to 100 (pp. 131–166). New York, NY: Cambridge University Press.

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

10.1017/S1355617702813248

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The End of the Flynn Effect? *Intelligence*, *32*, 349–362.

Sutin, A. R., Terracciano, A., Milaneschi, Y., An, Y., Ferrucci, L., & Zonderman, A. B. (2013). Cohort effect on well-being: The legacy of economic hard times. *Psychological Science*, *24*, 379-385. doi: 10.1177/0956797612459658

Teasdale, T. W., & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence*, *36*, 121–126. doi:10.1016/j.intell.2007.01.007

Thoemmes, F., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, *46*, 90–118. doi:10.1080/00273171.2011.540475

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn Effect: A Meta-Analysis. *Psychological Bulletin*, *140*, 1332–1360. doi: 10.1037/a0037173

Tucker-Drob, E. M., Reynolds, C. A., Finkel, D., & Pedersen, N. L. (2014). Shared and unique genetic and environmental influences on aging-related changes in multiple cognitive abilities. *Developmental Psychology*, *50,* 152–166. doi: 10.1037/a0032468.

Vaupel, J. W. (2010). Biodemography of human ageing. *Nature*, *464*, 536-542. doi: 10.1038/nature08984

Watson, D., Clark, L.A., & Tellegen, A. (1988). A development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Wechsler D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York, NY: Psychological Corporation. doi: 10.1177/073428298300100310

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*, 509–537.

Wilson, R. S., Leurgans, S. E., Boyle, P. A., & Bennett, D. A. (2011). Cognitive decline in Prodromal Alzheimer Disease and Mild Cognitive Impairment. *Archives of Neurology, 68,* 351-356. doi:10.1001/archneurol.2011.31

Zelinski, E. M., & Kennison, R. F. (2007). Not your parents' test scores: Cohort reduces psychometric aging effects. *Psychology and Aging*, *22*, 546-557. doi: 10.1037/0882-7974.22.3.546

Author Notes

Denis Gerstorf, Humboldt University, Berlin, Germany and German Institute for Economic Research (DIW Berlin), Berlin, Germany; Gizem Hülür, Humboldt University, Berlin, Germany; Johanna Drewelies, Humboldt University, Berlin, Germany; Peter Eibich, German Institute for Economic Research (DIW), Berlin, Germany; Sandra Duezel, Max Planck Institute for Human Development, Berlin, Germany; Ilja Demuth, Charite – Universitätsmedizin Berlin, Germany; Paolo Ghisletta, University of Geneva, Geneva, Switzerland, and Distance Learning University, Sierre, Switzerland; Elisabeth Steinhagen-Thiessen, Charite – Universitätsmedizin Berlin, Germany; Gert G. Wagner, Max Planck Institute for Human Development, Berlin, German Institute for Economic Research, Berlin, and Berlin University of Technology, Germany; Ulman Lindenberger, Max Planck Institute for Human Development, Berlin, Germany.

e-mail: denis.gerstorf@hu-berlin.de; gizem.hueluer@hu-berlin.de; johanna.drewelies@hu-berlin.de; PEibich@diw.de; duezel@mpib-berlin.mpg.de; ilja.demuth@charite.de; Paolo.Ghisletta@unige.ch; elisabeth.steinhagen-thiessen@charite.de; gwagner@mpib-berlin.mpg.de; lindenberger@mpib-berlin.mpg.de.

Table 1. Descriptive statistics and cohort differences in study measures

	BASE (n = 161)		BASE-II (n = 161)		
	M	*SD*	*M*	*SD*	*Cohen's d*
Age (65 – 89)[a]	75.12	3.43	74.43	3.19	−0.21
Gender (0=men; 1=women)	0.49	0.50	0.46	0.50	−0.06
Cohort-normed education (−1.78 – 3.55)	0.42	1.12	0.36	0.95	−0.06
Digit Symbol (0 – 66)	31.51	9.39	40.53	11.69	0.85*
Morale (1 – 5)[b]	3.69	0.80	4.02	0.87	0.39*
Positive affect (1 – 5)	3.51	0.62	3.89	0.57	0.64*
Negative affect (1 – 5)	2.78	0.68	2.50	0.75	−0.39*

Note. Positive values of Cohen's *d* indicate differences in favor of the BASE-II cohort. Participants in the matched earlier-born BASE cohort were born 1901 through 1922 (*M* = 1916; *SD* = 3.53 years) and those in the matched later-born BASE-II cohort 1925 through 1948 (*M* = 1939; *SD* = 3.22 years). [a] Please note that the large majority of participants in the matched samples were between ages 70 and 80 years. None of the BASE participants and only very few BASE-II participants were younger than age 70 years. [b] Higher scores indicate higher morale (and higher well-being).

<center>**Figure Captions**</center>

Figure 1. Illustrating standardized mean differences between the BASE and BASE-II cohorts in socio-demographic variables. Negative (positive) numbers signify greater scores for BASE (BASE II) participants. After the matching, the cohort differences were small and not reliably different from zero at $p < 0.05$.

Figure 2. Illustrating average cohort differences and individual differences in cognitive performance (Panel A: Digit Symbol) and indicators of well-being (Panel B: morale; Panel C: positive affect; Panel D: negative affect). The dots are raw data from participants in the matched BASE ($n = 161$; open circles) and BASE-II ($n = 161$, closed grey circles) samples. Sample means and standard errors for each cohort are displayed separately. Higher scores indicate higher morale (and higher well-being). Participants in the BASE-II cohort (data obtained in 2013–14) showed higher levels of cognitive performance and well-being compared to the BASE cohort (data obtained in 1990–93).