Published in final edited form as:

Psychol Sci. 2015 April ; 26(4): 433–443. doi:10.1177/0956797614567339.

When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the lifespan

Joshua K. Hartshorne and

Department of Psychology, Harvard University

Laura T. Germine

Department of Psychology, Harvard University & Center for Human Genetic Research,
Massachusetts General Hospital

Joshua K. Hartshorne: jharts@mit.edu; Laura T. Germine: lgermine@fas.harvard.edu

Abstract

Understanding how and when cognitive change occurs over the lifespan is a prerequisite for understanding normal and abnormal development and aging. Most studies of cognitive change are constrained, however, in their ability to detect subtle, but theoretically informative lifespan changes, as they rely on either comparing broad age groups or sparse sampling across the age range. Here, we present convergent evidence from 48,537 Web participants and a comprehensive analysis of normative data from standardized IQ and memory tests. Our results reveal considerable heterogeneity in when cognitive abilities peak: some abilities peak and begin to decline around high school graduation; some abilities plateau in early adulthood, beginning to decline in the 30s; still others do not peak until the 40s or later. These findings motivate a nuanced theory of maturation and age-related decline, where multiple, dissociable factors differentially affect different domains of cognition.

Early IQ tests lumped together all persons 16+ years-old into the homogenous category “adult” (Matarazzo, 1972). While it is now recognized that changes in cognition occur late in life, many researchers and laypeople alike share the intuition that there is some broad age range, post-development but pre-senescence, at which individuals’ cognitive ability is stable (neither improving nor declining) – an intuition that is reflected in studies of cognitive function relying on “typical adults” (usually 18–35 years old).

Nonetheless, it has long been known that this intuition cannot be quite right: Measures of fluid intelligence (e.g., short-term memory) peak early in adulthood, whereas measures of crystallized intelligence (e.g., vocabulary) peak in middle age (Bayley, 1970; Doppelt & Wallace, 1955; Fox & Birren, 1949; Sorenson, 1933; Shakow & Goldman, 1938). Even this may be too simple: Recent evidence shows that whereas short-term memory for names and inverted faces peaks around 22, neither short-term memory for faces nor quantity

Correspondence and requests for materials should be addressed to JKH (jkhartshorne@gmail.com).

Author contributions. Experiments were designed by JKH and LTG. Data for Exps. 2 & 3 were collected by LTG, with remaining data acquired by JKH. Analyses were discussed by both authors and performed by JKH. JKH took lead in preparing the manuscript.

Author Information. Raw data are available from the authors. The authors have no competing financial interests.

discrimination peaks until around 30, a fact difficult to assimilate into the fluid/crystallized intelligence dichotomy (Germine, Duchaine, & Nakayama, 2011; Halberda, Ly, Wilmer, Naiman, & Germine, 2012). Whether face memory and quantity discrimination are exceptions to the fluid/crystallized rule or represent more systematic and previously unrecognized patterns of age-related difference is an open question.

Comparing age-of-peak-performance across cognitive domains has several uses. If age-of-peak-performance is indeed far more variable than the fluid/crystallized intelligence distinction implies, that suggests that the fluid/crystallized intelligence construct needs revision (cf. Hampshire, Highfield, Parkin & Owen, 2012). More generally, distinct ages-of-peak-performance for two tasks suggests distinct underlying mechanisms. Delineating age-of-peak-performance also informs research methodology: the widespread use of college students as controls for development or aging studies may not be appropriate for functions that are still maturing at 18 or are already showing evidence of age-related decline. Finally, understanding the dynamics of age-related cognitive change can lead to more optimized educational interventions and methods of identifying and addressing age-related cognitive decline and quality-of-life among elder members of the population.

Despite increased interest in identifying and understanding differences in age-of-peak-performance, there has been little progress in determining which ages-of-peak-performance are reliably different from one another (cf. Ardila, 2007; Doppelt & Wallace, 1955; Halberda et al., 2012; Kaufman, 2001; Lee, Gorsuch, Saklofske & Patterson, 2008; Murre, Janssen, Rouw, & Meeter, 2013; Salthouse, 2003; Wisdom, Mignogna & Collins, 2012; but see Germine et al., 2011). Two main difficulties include lack of access to datasets of sufficient scale and lack of statistically sound methods for quantitatively comparing ages-of-peak-performance.

Below, we address both issues. We use modern statistical analysis techniques to compare age-of-peak-performance across thirty different cognitive tasks. To achieve sufficient sample size, we present both novel re-analyses of standardized test normative data as well as new, massive Web-based samples. We find similar results across datasets, strengthening confidence in the validity and reliability of the findings.

Note that in contrast to many lifespan studies that employ factor analysis in order to control random noise and other nuisance factors, we take the approach more commonly used in developmental psychology and cognitive neuroscience: employ well-understood tasks that are purposefully chosen because they are expected to dissociate. This allows us to treat unshared variance between tasks as potential sources of signal, rather than noise, with differences due to random noise addressed by our very large samples. This was the preferred approach given our specific hypotheses about potential domain-specificity in ages-of-peak performance (cf. Wilmer et al., 2012).

Exp. 1: Re-Analysis of Standardized Tests

To examine the degree of heterogeneity in age-of-peak-performance, we analyzed published demographically-stratified normative data from two standardized test batteries: WAIS-III, a widely-used intelligence test consisting of 15 subtests tapping a range of mental abilities,

and WMS-III, consisting of 16 subtests tapping different aspects of short-term and long-term memory (Wechsler, 1997).

Method

The tests are described in Table 1. Note that we replaced the unintuitive task names “symbol search”, “logical memory”, “verbal paired associates”, and “mental control” with the more descriptive terms “visual search”, “stories”, “word pairs”, and “reversed lists.”¹ The WAIS-III sample consisted of 2,450 healthy, cognitively-unimpaired Americans recruited in geographically-diverse locations, aged 16–89 (200 participants in each of the following age bins: 16–17, 18–19, 20–24, 25–29, 30–34, 35–44, 45–54, 55–64, 65–69, 70–74, 75–79; 150 80–84 year-olds; 100 85–89 year-olds). The WMS-III sample consisted of exactly half as many at each age.

Results

We generated bootstrapped estimates for age-of-peak performance (cf. Germine et al., 2011) based on norms reported by the WAIS-III and WMS-III. The WAIS III and WMS III manuals provide a fine-grained approximation of the normal distribution of scores for each age group (the “scaled” scores). We then used these distributions to draw N_g samples from each age group, where N_g is the number of participants used to generate norms for that age group. Resampled scaled scores were then converted back to raw scores using age-specific normative data, and age group with the highest score was identified. This procedure was repeated 2,500 times for each task in order to provide the distribution on age-of-peak-performance used for analysis and for Figure 1.

Intuitively, the width of the distribution returned reflects the range of ages at which participants are near peak performance. If the lifespan curve is sharply peaked, most of the bootstrapped ages-of-peak-performance will fall in a narrow window (assuming sufficient statistical power). If individuals remain at peak over a broad range of ages, the bootstrapped ages-of-peak-performance will fall across a similarly wide window.

We compared ages-of-peak-performance by conducting t-tests using the means and standard errors generated by our bootstrapping method (all pairwise comparisons are shown in Tables S1–S3 in the Supplemental Materials available online). A significant result indicates that the two distributions are substantially non-overlapping.²

We observed the previously-reported pattern of earlier peaks for fluid intelligence than crystallized intelligence (Baltes, 1987; Cattell, 1971). The pattern of age-related differences

¹The WMS-III includes several scoring methods. We used the “primary” method, with the following exceptions. For “word pairs,” we considered only the first presentation of the first word list. Scoring the long-term memory tests in terms of the proportion of all items recalled yielded results identical to the short-term memory task, so we adopted the alternate method of scoring the proportion of items recalled out of just those items correctly recalled during the short-term test. Since no norms are provided for this latter scoring method for the “stories” test, the long-term version of the “stories” test was excluded. We analyze forward and backward spatial span separately (unfortunately, digit span was not similarly decomposed into forwards and backwards scores). Finally, we include the “secondary” recognition test for “visual reproduction”. There was a similar, “secondary” recognition test for the word pairs, but since all age groups were at ceiling, analysis is not informative.

²Note that a more sophisticated comparison method would be needed to distinguish a sharply peaked distribution from a broad distribution with the same mean. Because we saw little clear evidence for such situations in our data, we used the more familiar t-test. Future researchers should keep this possibility in mind and use alternative methods of comparing distributions if needed.

for representative early- and late-peaking tasks are shown in Figure 2. In particular, the five tasks invoking learned knowledge (vocabulary, information, comprehension, arithmetic, similarities) peaked significantly later than nearly every other task ($ps < .05$; see Supplementary Data).

However, the pattern of results was more complicated than this dichotomy would suggest. Among the earlier-peaking tasks, reversed lists and backward spatial span peaked significantly later than word pairs and stories ($ps < .05$) but earlier than vocabulary, information and comprehension. Backward spatial span additionally peaked earlier than arithmetic and similarities ($ps < .05$). No other differences were significant, though some of the qualitative patterns matched those observed in previous work (Germine et al., 2011; Logie & Maylor, 2009).

Experiments 2 & 3

Experiment 1 suggested some heterogeneity in age-of-peak performance across fluid intelligence tasks, but the broad, coarse-grained age bins in the normative data limited the ability to identify subtle differences between tasks. In Experiments 2–3, we used Web-based methods to collect very large samples across five specific cognitive tasks, allowing a more fine-grained analysis. We focused on digit symbol coding, digit span, and vocabulary, which in Experiment 1 peaked (respectively) in the late teens, early twenties, and around 50 years-old. The comparison of these three tasks is of particular interest in light of the long-standing debate about how central a role working memory plays in fluid intelligence (cf. Nisbett, Aronson, Blair, Dickens, Flynn, Halpern, & Turkheimer, 2012).

The amount of heterogeneity in age-of-peak-performance might be even greater if we looked beyond the relatively narrow range of intelligence and memory tasks used so far to other areas of behavior such as social cognition, perception, and linguistic processing. As a first step in this direction, in Experiment 3 we investigated a widely-used test of emotion perception (Baron-Cohen, Wheelwright, & Hill, 2001).

Method

Participants—Participants in Experiment 2 ($N=10,394$, 10–69 years-old) and Experiment 3 ($N=11,532$; 10–71 years-old) were visitors to TestMyBrain.org, who took part in experiments in order to contribute to scientific research and in exchange for performance-related feedback.³ We continued data-collection for each experiment for approximately one year, sufficient to obtain around 10,000 participants, allowing fine-grained age-of-peak-performance analysis.

Web-based methods enable the rapid recruitment and testing of very large samples. Systematic comparisons between data collected from lab vs. Web-based samples have demonstrated that Web data can be as reliable as data collected in the lab and/or using

³Following common practice, we excluded participants who used a device other than a laptop or desktop, repeated the experiment, reported visual or psychiatric problems, or indicated that they had technical difficulties. These exclusions were decided prior to data-collection. The age ranges were chosen such that there were at least 30 subjects at each age. As a result, in Experiment 2, we excluded 16 subjects under 10 y.o. and 74 over 69 y.o. In Experiment 3, we excluded 138 subjects under 11 y.o. or over 67 y.o. We found that this cutoff struck an acceptable balance between minimizing noise and providing as broad an age range as possible.

traditional methods (Germine, Nakayama, Duchaine, Chabris, Chatterjee & Wilmer, 2012; Meyerson & Tryon, 2003).

Materials and Procedure—Experiment 2 consisted of tests of digit symbol coding, visual working memory, verbal working memory (forward digit span), and vocabulary. Digit symbol coding (also known as digit-symbol substitution) and forward digit span were adapted from the WAIS-III (see Table 1). The visual working memory task was adapted from a standard change-detection paradigm for testing visual working memory (Phillips, 1974): On each of 42 trials, participants viewed an array of 4 non-nameable novel shapes. After a brief retention period, they determined whether a single probe shape was a member of the memory set. The 20-question multiple-choice vocabulary test was modeled on the General Social Surveys' WORDSUM test (Smith, Marsden, & Hout, 2013). Experiment 3 consisted of a series of pictures of faces cropped such that only the eye region is visible; the participant selects the most appropriate emotion word from a list (for full method, see Baron-Cohen et al., 2001).

Analysis—Estimates and standard errors for age of peak performance were calculated using a bootstrap resampling procedure identical to the one described in Experiment 1, but applied to raw performance data. In order to dampen noise, means for each age were smoothed using a moving three-year window prior to identifying age-of-peak performance in each sample. Other methods of dampening noise provide similar results. In Experiment 2, age-of-peak-performance was compared across tasks with paired t-tests. Within-participant data was not available in Experiment 3.

Results

Results for Experiment 2 (Figures 3A and 3C) show the same ordering in age-of-peak-performance as in the standardized test results: the two working memory tasks peaked at around 30 years, significantly later than processing speed ($ps < .01$) and significantly earlier than vocabulary ($ps < .0001$) (for additional detail, see online supplementary materials). These results are consistent with models on which working memory is distinguishable from other tasks that load on fluid intelligence (cf. Nisbett et al., 2012). While verbal working memory's age-of-peak-performance was later than that of visual working memory, the difference was not significant ($t < 1$).

Results for the emotion perception task (Experiment 3; Figures 3B and 3C) reveal a peak significantly later than the peak for either working memory task ($ps < .05$), and a trend towards peaking earlier than vocabulary performance ($t(41595.1) = 1.8$, $p = .07$). The peak in emotion recognition ability was also much broader than the peaks for any of the other tasks, reflecting a long period of relative stability in performance between the ages of 40 and 60.⁴

Given the recent concern about the replicability of findings in psychological research (Hartshorne & Schachner, 2012; Open Science Collaboration, 2012), we attempted to confirm a subset of the above findings with separate datasets. 12,073 participants aged 10–

⁴Note: degrees of freedom were estimated using the Welch-Satterthwaite equation to correct for unequal variances

66 years old completed a separate digit span task (identical to Exp. 2), and 8,300 participants 15–73 years-old completed a slight variation on Exp. 2's visual working memory task, on a different site (GamesWithWords.org). Resulting peak age estimates (Figure 3C) were not significantly different from those of Experiment 2 (digit span: $t(54006.2) < 1$; visual working memory: $t(55612.0) < 1$).

Experiment 4: Cohort effects

Experiments 2–3 reveal the same general pattern as the demographically-stratified Wechsler norming samples, with digit-symbol coding performance peaking first, followed by working memory, and then finally by vocabulary. Thus, the different results for these tasks cannot be explained by differences between Web-based and in-person testing nor by cohort effects (see additional discussion in online supplementary materials). This provides additional evidence that Web-based testing methods and traditional testing procedures provide similar results (cf. Germine et al., 2012; Meyerson & Tryon, 2003).

Interestingly, however, vocabulary age-of-peak-performance was later for the Web-based data (~65) than in WAIS-III (~50). This could suggest confounds in one or both datasets. Alternatively, this may reflect cohort differences: The Wechsler data was collected two decades ago. With the increase in the proportion of adults engaged in cognitively-demanding careers, it may be that ages-of-peak-performance are later in the more recent Web sample, particularly for vocabulary. This could be related to the Flynn Effect: IQ has increased steadily in modern times, possibly due to increasing amounts of time devoted to mental activity (Flynn, 1997). We test this hypothesis in Experiment 4.

Method

We reanalyzed published results for 26,850 participants tested 1974–2012 on a 10-question vocabulary included as part of the General Social Surveys (Smith et al., 2013). In order to track changes over time, we divided the dataset by year of testing into three epochs with roughly equivalent numbers of participants: 1974–1987 ($N=9,155$, 5,200 female), 1988–1997 ($N=8,440$, 4,811 female), 1998–2012 ($N=9,255$, 5,191 female).

Results

We first confirmed that the dataset was sufficient in size and sensitivity to detect the cohort differences of interest. In particular, we found that the dataset replicated the standard Flynn Effect, with vocabulary scores increasing significantly across epochs ($t(26,848) < .001$, Figure 4E).

Consistent with our observation above of a later peak in the more recent dataset, analysis of age-related differences in performance for the three epochs showed visibly later peaks with each epoch (Figure 4 A–C). We followed this qualitative observation with quantitative age-of-peak-performance estimates, following the method outlined for Experiments 2 & 3. These analyses similarly showed later peaks for more recent samples (Figure 4 D). Linear regression shows that this represents an average annual increase in the age-of-peak-performance of .90 years, a result which trended towards significance ($p=.078$).⁵ Combining this dataset with the vocabulary data from WAIS-III (collected 1995) and Exp. 2 (collected

2010) resulted in an estimated annual increase in age-of-peak-performance of 0.96 years, a result which reached significance ($p=.0003$).⁶

Thus, it is likely that the later ages-of-peak-performance in our data relative to the Wechsler data are at least partly due to generational differences, with later peaks seen in more recent generations, at least for vocabulary.

General Discussion

The present study demonstrates that age-related changes in cognitive ability are considerably more heterogeneous and complex than the fluid/crystallized intelligence distinction suggests. Above, we show evidence for at least 3–4 distinct patterns.

These results were reliable across samples: we directly replicated the visual and verbal working memory findings of Exp. 2, and we obtained converging results for several tasks in both Web-based and traditional samples. Moreover, this convergence rules out a significant role for several possible confounds in the Web-based data, such as older adults having less experience with computers or differential representativeness at different ages; such confounds would have resulted in differences between the Web-based data and the demographically-stratified, paper-and-pencil data.⁷ This convergence adds to the growing body of work indicating that Web-based data is highly reliable (e.g., Germine et al., 2012).

One potential concern with cross-sectional data is that it may be subject to cohort effects. Our findings in Experiment 2 are consistent with the possibility that people born in 1945 have unusually large vocabularies, people born in 1980 have unusually good working memory, and people born in 1990 have unusually fast processing speed. Such concerns can be mitigated by converging results from cross-sectional datasets collected at different times (Schaie, 2005). Here, we compared results derived from Web cross-sectional data with results derived from WAIS-III and WMS-III cross-sectional data collected 20 years earlier. Thus, if the results in Exp. 2 and its replications were driven by cohort effects, all the peaks in these earlier cross-sectional studies should have occurred 20 years younger. Instead, digit span and digit-symbol coding showed similar ages-of-peak-performance in all datasets.

One difference was observed between Web-based and traditional samples: earlier age-of-peak-performance for vocabulary in the latter. This is unlikely to be related to testing method, since it also appears in a long-term paper-and-pencil study (Exp. 4). This novel finding may also explain a current puzzle in the literature: while average vocabulary of both adults and children has increased in recent years, the increase has been much larger for

⁵Given the small number of data-points (3 epochs), we assessed significance using a permutation test. The distribution of regression coefficients under the null hypothesis (no effect of epoch on age-of-peak-performance) was assessed through 2,500 permutation samples. On each sample, participants were randomly re-assigned to epoch, with the constraint that the number of participants of each age in each epoch remain consistent. Age-of-peak-performance was assessed for each epoch, and the regression coefficient was measured. The p-value is the number of such coefficients at least as large as the actual coefficient (.897), including the actual coefficient itself. We converted this one-tailed p-value to a two-tailed p-value by doubling the one-tailed p-value.

⁶Because it was not possible to do use bootstrapping – this analysis involved three different vocabulary tests – we used a standard linear regression instead.

⁷Note here and below that although we found converging results for digit span, digit-symbol coding, and vocabulary, such convergence may not generalize to visual working memory and emotion perception. Conclusions about the latter tasks must be more tentative.

adults than for children, a fact only partly explained by the increase in tertiary education (Flynn, 2010). Our data offer an explanation: vocabulary learning is continuing later into adulthood, possibly due to environmental factors (e.g., continued exposure to new words). Some purchase may be gained by exploring whether other tasks show similar generational changes.

Our findings have practical and theoretical implications. On the practical side, not only is there no age at which humans are performing at peak at all cognitive tasks, there may not be an age at which humans are at peak on *most* cognitive tasks.

Studies that compare the young or elderly to “normal adults” must carefully select the “normal” population. For instance, comparing college freshman and 65 year-olds on emotion recognition would result in no difference, leading to the erroneous conclusion of no age-related change (Figure 3B). This may explain why studies differ in whether or not they show age-related decline in aspects of social perception (Moran, 2013). Critically, these studies compare different age groups. Similarly, clinicians attempting to determine whether an individual exhibits early signs of abnormal decline must consider both the type of cognitive task and the individual’s age.

On the theoretical side, the complexities described above provide a rich, challenging set of phenomena for theories of development, maturation, and aging. While heterogeneity in lifespan curves results from differences in biological maturation and aging of the underlying neural substrates (Greenwood, 2007; Paus, 2005), this cannot easily account for tasks that show continued improvement past early adulthood. Salthouse (2003, 2004) suggests that these are precisely those tasks that depend on experience, which necessarily increases with age. However, this alone does not explain why visual working memory, which shows minimal effects of practice and experience (Eng, Chen & Jiang, 2005), peaks later than digit-symbol coding, nor why emotion recognition peaks before vocabulary. Some purchase on this problem may be gained by better understanding differences in the learning problems presented by experience-dependent tasks. For instance, while vocabulary size depends heavily on encountering the words in question, digit span depends heavily on explicit strategies that must be learned (Gathercole, Adams, & Hitch, 1994). Another important factor determining when performance begins to decline due to aging is the degree to which different tasks allow for compensatory strategies (Greenwood, 2007).

Importantly, the present data and method provide a powerful new constraints on theories of cognition. Researchers in the aging and intelligence literatures have more typically employed factor analysis. Factor analysis has analytic and conceptual advantages in that it attempts to directly model underlying factors shared across tasks, abstracting away from nuisance factors such as task-specific strategies. While influential and informative, factor analysis studies have left numerous questions unresolved, in part because these studies do not agree on the number or nature of dissociable factors relevant to aging (Ghisletta, Rabbit, Lunn & Lindenberger, 2012; Goh, An, & Resnick, 2012; Tucker-Drob, 2011). Power can be an issue: Each participant must complete a large battery of task, making collecting large samples difficult (though see Hampshire et al., 2012; Johnson, Logie, & Brockmole, 2010).

Moreover, by focusing on broad pools of shared variance across tasks, factor analysis may miss smaller but theoretically relevant differences (cf. see Wilmer et al., 2012).

As such, the above method – which builds on methodologies more common in developmental psychology and cognitive neuroscience – provides a valuable new tool, where carefully-selected tasks are directly compared. Noise is controlled through sample size, and potential nuisance factors such as scientifically uninteresting task strategies can be tested for experimentally by comparing variants of the same task. Age-of-peak-performance analyses make it possible to directly compare the results of different tasks measured on different scales and performed by different participants. Such datasets are now increasingly easy to obtain through Web-based testing (e.g., Germine et al. 2010; Germine et al., 2011; Halberda et al., 2012; Hartshorne, 2008; Hampshire et al., 2012; Johnson et al., 2010; Logie & Maylor, 2009; Maylor & Logie, 2010).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Thanks are due to Timothy O'Donnell and Tim Brady for help with analyses, Jeremy Wilmer, Manizeh Khan, Ken Nakayama, Lucia Garrido, Erich Eich, and three anonymous reviewers for comments, and to NDSEG, NIH NRSA (JKH) and NSF GRFP, NIH NRSA (LTG).

References

- Ardila A. Normal aging increases cognitive heterogeneity: Analysis of dispersion in WAIS-III scores across age. *Archives of Clinical Neuropsychology*. 2007; 22:1003–1011. [PubMed: 17904332]
- Baltes PB. Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*. 1987; 23:611–626.
- Baron-Cohen S, Wheelwright S, Hill J. The 'Reading the Mind in the Eyes' test revised version: A study with normal adults and adults with Asperger Syndrome or High-Functioning autism. *Journal of Child Psychology and Psychiatry*. 2001; 42:241–252. [PubMed: 11280420]
- Bayley, N. Development of mental abilities. In: Mussen, PH., editor. *Carmichael's Manual of Child Psychology*. 3. Vol. 1. New York: Wiley; 1970. p. 1163–1209.
- Cattell, RB. *Abilities: Their Structure, Growth & Action*. Boston: Houghton Mifflin; 1971.
- Doppelt JE, Wallace WL. Standardization of the Wechsler adult intelligence scale for older persons. *Journal of abnormal psychology*. 1955; 51:312–330. [PubMed: 13263050]
- Eng HY, Chen D, Jian Y. Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*. 2005; 12:1127–1133. [PubMed: 16615339]
- Flynn, J. *What is intelligence?*. Cambridge University Press; Cambridge, UK: 2007.
- Flynn J. Problems with IQ gains: The huge vocabulary gap. *Journal of Psychoeducational Assessment*. 2010; 28(5):412–433.10.1177/0734282910373342
- Fox C, Birren JE. Some factors affecting vocabulary size in later maturity: Age, education, and length of institutionalization. *Journal of Gerontology*. 1949; 4(1):19–26. [PubMed: 18125187]
- Gathercole SE, Adams AM, Hitch GJ. Do young children rehearse? An individual-differences analysis. *Memory and Cognition*. 1994; 22(2):201–207. [PubMed: 8035696]
- Germine L, Nakayama K, Duchaine BC, Chabris CF, Chatterjee G, Wilmer JB. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review*. 201210.3758/s13423-012-0296-9

- Germine LT, Duchaine B, Nakayama K. Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition*. 2011; 118:201–210.10.1016/j.cognition.2010.11.002 [PubMed: 21130422]
- Ghisletta P, Rabbit P, Lunn M, Lindenberger U. Two thirds of the age-based changes in fluid and crystallized intelligence, perceptual speed, and memory in adulthood are shared. *Intelligence*. 2012; 40:260–68.
- Goh JO, An Y, Resnick SM. Differential trajectories of age-related changes in components of executive and memory processes. *Psychology and Aging*. 2012; 27(3):707–719. [PubMed: 22201331]
- Greenwood PM. Functional plasticity in cognitive aging: Review and hypothesis. *Neuropsychology*. 2007; 21(6):657–673. [PubMed: 17983277]
- Halberda J, Ly R, Wilmer JB, Naiman Daniel Q, Germine L. Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:11116–11120.10.1073/pnas.1200196109 [PubMed: 22733748]
- Hampshire A, Highfield RR, Parkin BL, Owen AM. Fractionating Human Intelligence. *Neuron*. 2012; 76:1225–1237.10.1016/j.neuron.2012.06.022 [PubMed: 23259956]
- Hartshorne JK. Visual working memory capacity and proactive interference. *PLoS ONE*. 2008; 3(7)
- Hartshorne JK, Schachner A. Tracking replicability as a method of postpublication open evaluation. *Frontiers in Computational Neuroscience*. 2012; 6:8.10.3389/fncom.2012.00008 [PubMed: 22403538]
- Johnson W, Logie RH, Brockmole JR. Working memory tasks differ in factor structure across age cohorts: Implications for dedifferentiation. *Intelligence*. 2010; 38:513–528.10.1016/j.intell.2010.06.005
- Kaufman AS. WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. *Intelligence*. 2001; 29:131–167.10.1016/S0160-2896(00)00046-5
- Lee HF, Gorsuch RL, Saklofske DH, Patterson DA. Cognitive differences for ages 16 to 89 years (Canadian WAIS-III). *Journal of Psychoeducational Assessment*. 2008; 26:382–394.
- Logie RH, Maylor EA. An Internet study of prospective memory across adulthood. *Psychology and Aging*. 2009; 24:767–774.10.1037/a0015479 [PubMed: 19739935]
- Matarazzo, JD. Wechsler's measurement and appraisal of adult intelligence. Baltimore, MD: The Williams & Wilkins Company; 1972.
- Meyerson P, Tryon WW. Validating Internet research: A test of the psychometric equivalence of Internet and in-person samples. *Behavior Research Methods, Instruments, & Computers*. 2003; 35:614–620.
- Moran JM. Lifespan development: The effects of typical aging on theory of mind. *Behavioral Brain Research*. 2013; 237:32–40.
- Murre JMJ, Janssen SMJ, Rouw R, Meeter M. The rise and fall of immediate and delayed memory for verbal and visuospatial information from late childhood to late adulthood. *Acta Psychologica*. 2013; 142:96–107.10.1016/j.actpsy.2012.10.005 [PubMed: 23261419]
- Nisbett RE, Aronson J, Blair C, Dickens W, Flynn J, Halpern DF, Turkheimer E. Intelligence: New findings and theoretical developments. *American Psychologist*. 2012; 67(2):130–159. [PubMed: 22233090]
- Open Science Collaboration. An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*. 2012; 7(6):657–660.
- Paus T. Mapping brain maturation and cognitive development during adolescence. *TRENDS in Cognitive Sciences*. 2005; 9(2):60–68. [PubMed: 15668098]
- Salthouse TA. Memory and aging from 18 to 80. *Alzheimer Disease and Associated Disorders*. 2003; 17(3):162–167. [PubMed: 14512830]
- Salthouse TA. What and when of cognitive aging. *Current Directions in Psychological Science*. 2004; 13:140–144.
- Schaie, KW. Developmental influences on adult intelligence: The Seattle Longitudinal Study. New York: Oxford University Press; 2005.

- Shakow D, Goldman R. The effect of age on the Stanford-Binet vocabulary score of adults. *Journal of Educational Psychology*. 1938; 29:241–256.
- Smith, TW.; Marsden, PV.; Hout, M. *General Social Surveys, 1972–2012: Cumulative Codebook*. National Opinion Research Center; Chicago, IL: 2013.
- Sorenson H. Mental ability over a wide range of adult ages. *Journal of Applied Psychology*. 1933; 17:729–741.
- Tucker-Drob EM. Global and domain-specific changes in cognition throughout adulthood. *Developmental Psychology*. 2011; 47:331–343. [PubMed: 21244145]
- Wechsler, D. *WAIS-III, WMS-III Technical Manual*. San Antonio, TX: The Psychological Corporation; 1997.
- Wilmer J, Germine L, Chabris C, Chatterjee G, Gerbasi M, Nakayama K. Capturing specific abilities as a window into human individuality: the example of face recognition. *Cognitive Neuropsychology*. 2012; 29(5–6):360–392. [PubMed: 23428079]
- Wisdom NM, Mignogna J, Collins RL. Variability in Wechsler Adult Intelligence Scale-IV subtest performance across age. *Archives of clinical neuropsychology: the official journal of the National Academy of Neuropsychologists*. 2012; 27:389–397.10.1093/arclin/acs041 [PubMed: 22512934]

Figure 1. Bootstrapped age of peak performance for WAIS-III and WMS-III tasks (Exp. 1)
 Represented are the medians (central line), inter-quartile ranges (boxes) and 95% confidence intervals (whiskers). WM = Working Memory (immediate test after each trial); STM = Short-Term Memory (test soon after stimulus presentation); LTM = Long-Term Memory (test 20–30 min. after stimulus presentation).

Figure 2. Lifespan curves for representative early- and late-peaking WAIS-III and WMS-III tasks (Exp. 1). Shaded areas represent standard errors.

Figure 3. Results of Exps. 2 & 3

Panel A: Mean z-scored performance for each task in Exp. 2. Shaded regions represent SEs.

Panel B: Mean z-scored performance in Exp. 3. **Panel C:** Bootstrapped age of peak performance for Exps. 1–2, plus replications.

Figure 4.

Lifespan curves for vocabulary for participants tested in 1974–1987 (Panel A), 1988–1997 (Panel B), and 1998–2012 (Panel C). Estimated age-of-peak-performance is shown in Panel D, and mean score for each cohort is shown in Panel E.

Table 1

Names and descriptions of WAIS-III and WMS-III tasks.

Task Name	Description
Vocabulary	Provide definitions for words
Information	Answer general knowledge questions
Comprehension system?)	Explain why things happen (ex: why do we have parole
Arithmetic	Answer arithmetic problems
Similarities	In what way are pairs of items alike (ex: fork, spoon)?
Reversed lists	Produce memorized lists (ex: alphabet) in forwards then backwards as quickly as possible
Backward spatial span	Tap a set of cubes in reverse order from experimenter
Digit span	Repeat lists of digits, either in same order or reversed order. This task appears in both the WAIS and WMS
Picture Completion	Find missing part in each picture.
Picture Arrangement	Arrange pictures in sequence to tell a coherent story
Object assembly	Assemble puzzles
Block design	Recreate visually-depicted geometric patterns using blocks.
Forward spatial span	Tap a set of cubes in same order as experimenter.
Digit symbol coding	Digits 1–3 are each paired with a symbol. Given list of symbols, write down corresponding digit as fast as possible
Visual search	A speeded visual search task
Letter/Number sequencing	Given list of interspersed numbers and letters. From memory repeat numbers in ascending order then letters in alphabetical order. This task was included in both WAIS and WMS
Matrix Reasoning	A variant of Raven's Progressive Matrices
Faces *	Exposed to faces for 2 sec. each. Then discriminate these from novel faces
Stories *	Two stories, read by examiner. Then retell the stories
Word pairs *	Learn lists of word pairs. During test, one word is provided and other must be recalled.
Family pictures *	Exposed to scenes of family activities. Asked to recall which characters were in scene, where they were positioned, and what they were doing.
Word lists *	A list of 12 words is presented. Asked to recall in any order.
Visual reproduction *	Exposed to geometric design for 10 sec. Asked to reproduce.

* These tasks come in short-term memory (STM) and long-term memory (LTM) variants. Participants were tested immediately after exposure to the stimulus set (STM) and then again later in the session (LTM).