THE CLINICAL NEUROPSYCHOLOGIST
https://doi.org/10.1080/13854046.2019.1699605

Routledge
Taylor & Francis Group

Cohort differences on the CVLT-II and CVLT3: evidence of a negative Flynn effect on the attention/working memory and learning trials

Lisa V. Graves[a,b], Lisa Drozdick[c], Troy Courville[d], Thomas J. Farrer[e], Paul E. Gilbert[a,b] and Dean C. Delis[a,f]

[a]San Diego State University/University of California San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA, USA; [b]Department of Psychology, San Diego State University, San Diego, CA, USA; [c]Pearson, San Antonio, TX, USA; [d]Professional Education, Georgia Institute of Technology, Atlanta, GA, USA; [e]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA; [f]Department of Psychiatry, University of California San Diego School of Medicine, La Jolla, CA, USA

ABSTRACT

Objective: Although cohort effects on IQ measures have been investigated extensively, studies exploring cohort differences on verbal memory tests, and the extent to which they are influenced by socioenvironmental changes across decades (e.g. educational attainment; ethnic makeup), have been limited.

Method: We examined differences in performance between the normative samples of the CVLT-II from 1999 and the CVLT3 from 2016 to 2017 on the immediate- and delayed-recall trials, and we explored the degree to which verbal learning and memory skills might be influenced by the cohort year in which norms were collected versus demographic factors (e.g. education level).

Results: Multivariate analysis of variance tests and follow-up univariate tests yielded evidence for a *negative* cohort effect (also referred to as negative Flynn effect) on performance, controlling for demographic factors ($p = .001$). In particular, findings revealed evidence of a negative Flynn effect on the attention/working memory and learning trials (Trial 1, Trial 2, Trial 3, Trials 1–5 Total, List B; $ps < .007$), with no significant cohort differences found on the delayed-recall trials. As expected, education level, age group, and ethnicity were significant predictors of CVLT performance ($ps < .01$). Importantly, however, there were no interactions between cohort year of norms collection and education level, age group, or ethnicity on performance.

Conclusions: The clinical implications of the present findings for using word list learning and memory tests like the CVLT, and the potential role of socioenvironmental factors on the observed negative Flynn effect on the attention/working memory and learning trials, are discussed.

ARTICLE HISTORY
Received 14 June 2019
Accepted 22 November 2019

KEYWORDS
Cohort differences; Flynn effect; verbal memory; California Verbal Learning Test

CONTACT Lisa V. Graves lvgraves@gmail.com VA San Diego Healthcare System, 3350 La Jolla Village Dr. (116B), San Diego, CA 92161, USA

Cohort effects reflect global societal changes that sweep across entire generations of individuals and result in subtle but statistically significant changes in performance on nationally normed psychometric instruments (Baltes, Lindenberger, & Staudinger, 2006; Brailean et al., 2018; Christensen et al., 2013; Dodge, Zhu, Lee, Chang, & Ganguli, 2014; Dodge et al., 2017; Flynn, 1987; Freedman, Aykan, & Martin, 2002; Karlsson, Thorvaldsson, Skoog, Gudmundsson, & Johansson, 2015; Langa et al., 2008; Llewellyn & Matthews, 2009; Pietschnig & Voracek, 2015; Schaie, Willis, & Pennak, 2005; Skirbekk, Stonawski, Bonsang, & Staudinger, 2013; Thorvaldsson, Karlsson, Skoog, Skoog, & Johansson, 2017). The Flynn effect is the most well-known and documented generational change (Flynn, 1987, 2012; Pietschnig & Voracek, 2015; Trahan, Stuebing, Hiscock, & Fletcher, 2014), and refers to the sustained increase in intelligence quotient (IQ) scores worldwide during the 20th century (approximately 3 IQ points per decade; Pietschnig & Voracek, 2015; Trahan et al., 2014). This subtle but steady rise in IQ levels per decade over the past century has been attributed to several positive socioenvironmental changes, including higher educational attainment and improvements in nutrition, standard of living, and healthcare services that reach broader segments of the general population (Bratsberg & Rogeberg, 2018; Dodge et al., 2014, 2017; Pietschnig & Voracek, 2015; Skirbekk et al., 2013). The term "Flynn effect," while originally used in reference to IQ changes, has since been expanded to refer to generational changes in other cognitive domains, including memory, inductive reasoning, visuospatial ability, and processing speed (Baxendale, 2010; Brailean et al., 2018; Rönnlund & Nilsson, 2008, 2009; Wongupparaj, Wongupparaj, Kumari, & Morris, 2017).

In recent years, however, a plateauing and even reversal of the Flynn effect in IQ scores has been observed in some countries, including Norway, Denmark, Australia, Britain, the Netherlands, Sweden, Finland, France, Germany, and Estonia, resulting in *lower* IQ and cognitive test scores in individuals born on the cusp between the 20th and 21st centuries (Bratsberg & Rogeberg, 2018; Dutton & Lynn, 2013, 2015; Dutton, van der Linden, & Lynn, 2016; Flynn & Shayer, 2018; Pietschnig & Gittler, 2015; Shayer & Ginsburg, 2007, 2009; Sundet, Barlaug, & Torjussen, 2004; Teasdale & Owen, 2005, 2008; Woodley & Meisenberg, 2013). This "negative Flynn effect" may reflect more recent socioenvironmental changes that might adversely affect cognitive test performance, including reduced quality of education due to varying standards and methods across educational settings (e.g. proliferation of online educational programs that differ in terms of their standards of practice; see Allen & Seaman, 2013; Jaggars & Bailey, 2010), and declines in healthcare and standard of living for certain sectors of the population (Dutton & Lynn, 2013; Rindermann, Becker, & Coyle, 2017).

Few studies have addressed cohort effects on tests of auditory attention/working memory or learning and memory of word lists. Two studies focused on cohort effects on list learning and memory tests and yielded mixed findings (Baxendale, 2010; Dodge et al., 2017). Specifically, Baxendale (2010) explored differences on a verbal list learning and memory test from the Adult Memory and Information Processing Battery ($n = 184$; ages 16–75) published in 1985 versus its successor, the Brain Injury Rehabilitation Trust Memory and Information Processing Battery ($n = 300$; ages 16–89 years) published in 2007. Baxendale (2010) reported no differences in performance across the two cohorts; however, potential limitations in this study were the relatively small sample size, especially in the first testing, and the use of two different word lists in the two batteries, thereby potentially confounding time of testing with changes in the target words. Another study by Dodge et al. (2017) reported a positive cohort effect on the Consortium to Establish a Registry for Alzheimer Disease (CERAD) Word List in a pooled sample of U.S. adults who were born between the years of 1902–1943 and examined between the years of 1987–2015 across two population studies. However, limitations in this study included (a) investigating only older adults (65 years or older); (b) notable differences in the proportions of adults in each age cohort within the pooled sample who were tested in the first versus second data collection period; (c) not addressing the fact that age cohort and period of testing were potentially confounded; and (d) using data harmonization techniques to convert Logical Memory scores from a subset of its data pool into CERAD Word List scores to facilitate analyses of cohort differences on verbal memory performance. In a third study, Wongupparaj et al. (2017) conducted a meta-analysis on cohort effects on attention and working memory measures, and they concluded that, while more basic attentional skills (e.g. digit span forward) have gradually improved over the past four decades (between 1973 and 2016), more complex attention/working memory abilities (e.g. digit span backward) have shown a gradual decline over the same time period. This conclusion is intriguing, because it suggests that different domains of cognition may be differentially affected by cohort changes.

The California Verbal Learning Test (CVLT) is a widely used standardized measure of verbal learning and memory (Rabin, Barr, & Burton, 2005; Rabin, Paolillo, & Barr, 2016). The CVLT-II and CVLT3 normative samples were carefully matched to the U.S. Census data that were available at the time that each test was normed (i.e. the 1999 U.S. Census data for the CVLT-II and the 2015 U.S. Census data for the CVLT3). According to the U.S. Census Bureau, the time period spanning the development of the second and third editions of the CVLT coincided with an increase in the proportion of U.S. adults who completed post-secondary education. Some research suggests that cohort effects on other measures of word list learning and memory are not attenuated by education (Dodge et al., 2017). However, while significant (albeit relatively small) correlations have been found between education and verbal learning and memory performance on the various editions of the CVLT (e.g. education level explained only 4.5% of the variance on CVLT-II total immediate recall after accounting for other important moderating variables; see Delis, Kramer, Kaplan, & Ober, 2017), cohort differences on CVLT performance and the degree to which generational changes in educational attainment may influence or explain cohort effects on CVLT performance, have not been examined.

In the present study, we examined differences between the CVLT-II and CVLT3 normative samples in performance on the immediate- and delayed-recall trials of the measure. Specifically, we were interested in investigating the extent to which verbal learning and memory skills might be influenced by the cohort year of norms collection (i.e. 1999 for the CVLT-II versus 2016–2017 for the CVLT3) versus differences in education level, while accounting for other demographic variables (i.e. age; ethnicity). The CVLT-II and CVLT3 lend themselves well to the study of cohort effects because the instructions and target word lists have remained the same across versions, with the only substantive change between the two versions occurring on the Forced Choice Recognition trial (where the eight “abstract” distracter items were replaced with “concrete” distractor items); however, the Forced Choice Recognition trial was not included in the present study due to ceiling effects in the normative population.

Method

In conducting the national normative studies for the CVLT-II and CVLT3, stratified sampling plans were implemented to ensure that the normative samples were representative of the current U.S. population based on selected demographic variables, including age, gender, education, race/ethnicity, and geographic region. Specifically, the CVLT-II and CVLT3 normative samples were carefully matched to the U.S. Census data from 1999 and 2015, respectively (Delis, Kramer, Kaplan, & Ober, 2000; Delis et al., 2017). Age was stratified by seven levels: 16–19, 20–29, 30–44, 45–59, 60–69, 70–79, and 80–90 years. Gender was stratified by two levels: male and female. Education was stratified by five levels: 0–8 years, 9–11 years (without high school diploma), high school diploma or equivalent, some college or technical school or Associate’s degree, and Bachelor’s degree or higher (note: for examinees ages 16–24, parental education was used, and parental education was defined as the average of the highest grade completed by each parent). Ethnicity was stratified by four levels on the CVLT-II (White, African-American, Hispanic, and Other, with Asian participants subsumed under the Other category) and by five levels on the CVLT3 (White, African-American, Hispanic, Asian, and Other). In the present study, Asian participants in the CVLT3 normative sample were subsumed under the Other category to facilitate 1) a direct, comparative analysis of the ethnic composition of the CVLT-II and CVLT3 normative samples and 2) inclusion of ethnicity as a predictor in primary analyses.

Participants

Participants in the present study included the 1087 examinees in the CVLT-II normative sample (Delis et al., 2000) and the 700 examinees in the CVLT3 normative sample (Delis et al., 2017). The CVLT-II and CVLT3 were administered using standard procedures outlined in the test manuals. The CVLT-II and CVLT3 contain identical target words on the immediate- and delayed-recall trials, with, as noted above, the only substantive change occurring on the Forced Choice Recognition trial (where the eight “abstract” distractor items were replaced with “concrete” distractor items on the CVLT3).

Variables of interest

Outcome variables of interest in the present study included raw scores on the immediate-recall trials (Trial 1, Trial 2, Trial 3, Trial 4, Trial 5, Trials 1–5 Total, List B trial) and delayed-recall trials (Short Delay Free Recall [SDFR], Short Delay Cued Recall [SDCR], Long Delay Free Recall [LDFR], Long Delay Cued Recall [LDCR]). The Yes/No and Forced Choice Recognition trials were not included in the present analysis given evidence for skewed distributions of raw scores resulting in mild ceiling effects in the former, and strong ceiling effects in the latter, in samples of typically developing and aging (nonclinical) examinees (Delis et al., 2017). Predictor variables of primary interest included cohort year of norms collection (2 levels: 1999 for the CVLT-II versus 2016–2017 for the CVLT3), education level (five levels: 0–8 years, 9–11 years [without high school diploma], high school diploma or equivalent, some college or technical school or Associate's degree, and Bachelor's degree or higher), age group (seven levels: 16–19, 20–29, 30–44, 45–59, 60–69, 70–79, and 80–90 years), and ethnicity (four levels: White, African-American, Hispanic, Other).

Statistical analyses

Analyses were conducted using the Statistical Package for the Social Sciences (SPSS) Version 25. Prior to conducting the primary analyses, preliminary chi-square tests were conducted to explore differences between the CVLT-II and CVLT3 normative samples on key demographic variables (e.g. education level, age group, gender, ethnicity). Cramer's *V* effect size estimates associated with cohort differences on demographic variables were calculated and reported (Cohen, 1988: 0.10 = small; 0.30 = medium; 0.50 = large).

To address the primary aims of the study, two sets of multivariate analysis of variance (MANOVA) tests were conducted to examine the effects of cohort year of norms collection (two levels), education level (five levels), and other demographic variables (age group, ethnicity), as well as interactions between cohort year and demographic variables, on (a) the immediate-recall trials (five learning trials, Trials 1–5 Total, List B trial) and (b) the delayed-recall trials (SDFR, SDCR, LDFR, LDCR). MANOVAs were conducted in two sets given patterns of correlations observed among (a) the immediate-recall trials and (b) the four delayed-recall trials. Follow-up univariate tests and post-hoc pairwise comparisons with Bonferroni corrections for multiple comparisons (.007 for immediate recall trials and .013 for delayed recall trials) were conducted. Partial eta squared ($\eta_p{}^2$) effect size estimates associated with univariate tests of effects of cohort year of norms collection, education level, and other demographic variables (i.e. age, ethnicity) on immediate- and delayed-recall trials were calculated and reported (Cohen, 1988: 0.01 = small; 0.59 = medium; 0.14 = large). For ease of reading, "cohort year" is used to indicate cohort year of norms collection and "education" is used to indicate education level in descriptions of results and tables below.

Results

Preliminary analyses

Education

Chi-square tests revealed a significant difference between the CVLT-II and CVLT3 normative samples in their distributions of education levels, χ^2 (4, $N = 1787$) = 28.32, $p <$ .001, Cramer's V = .126. Relative to the CVLT-II cohort, the CVLT3 cohort had 1) a significantly lower proportion of individuals who completed 9–11 years of education (without high school diploma) or a high school diploma or equivalent, and 2) a

Table 1. Frequency distributions for education level, age group, and ethnicity in the CVLT-II and CVLT3 normative samples.

	CVLT-II	CVLT3
Education level	*n*	*n*
0–8 years	71	33
9–11 years (no high school diploma)	122	56
High school diploma or equivalent	371	195
Some college, technical school, or associate's degree	302	208
Bachelor's degree or higher	221	208
Age group (years)	*n*	*n*
16–19	150	100
20–29	190	100
30–44	200	100
45–59	150	100
60–69	145	100
70–79	145	100
80–90	107	100
Ethnicity	*n*	*n*
White	844	466
African-American	119	86
Hispanic	98	94
Other (including Asian)	26	54

Note: CVLT = California Verbal Learning Test.

significantly higher proportion of individuals who completed a Bachelor's degree or higher (see Table 1 for frequency distributions of education levels in the CVLT-II and CVLT3 normative samples). Given these cohort differences on education level and that education was a demographic variable of interest in the present study, education level was included as a predictor in the primary analyses.

Age group

Chi-square tests also revealed a significant difference between the CVLT-II and CVLT3 normative samples in their distributions of age groups, χ^2 (6, $N = 1787$) $= 14.92$, $p = .021$, Cramer's $V = .091$. Relative to the CVLT-II cohort, the CVLT3 cohort had 1) a significantly lower proportion of individuals ages 30–44 years, and 2) a significantly higher proportion of individuals ages 80–90 years. Given these cohort differences in age and that age has been shown to be an important moderating variable on CVLT performance (Delis et al., 2017), age group was included as a predictor in the primary analyses.

Gender

Although gender was not a demographic variable of primary interest in the present study (gender explained only 5.1% of the variance on CVLT-II total immediate recall after accounting for other important moderating variables; see Delis et al., 2017), it is worth noting that chi-square tests revealed no significant difference between the CVLT-II and CVLT3 normative samples in their gender distributions, χ^2 (1, $N = 1787$) $= 0.18$, $p > .05$, Cramer's $V = .010$. Moreover, although exploratory MANOVA tests revealed significant gender differences on CVLT-II/CVLT3 performance (with women generally outperforming men on immediate- and delayed-recall trials, *ps* $< .05$, consistent with past research), there were no significant cohort year x gender interactions (*ps* $> .05$).

Ethnicity

Chi-square tests revealed a significant difference between the CVLT-II and CVLT3 normative samples in their ethnic composition, χ^2 (3, $N = 1787$) $= 42.45$, $p < .001$, Cramer's $V = .154$. Relative to the CVLT-II cohort, the CVLT3 cohort had 1) a significantly lower proportion of White individuals, and 2) a significantly higher proportion of Hispanic and Other individuals; there were no cohort differences in the proportion of African-American individuals ($p > .05$). Some studies suggest that ethnicity is a significant predictor of CVLT performance (Aboudarham & Zalewski, 1995; Fox, Brook, Heilbronner, Susmaras, & Hanlon, 2019; Norman, Evans, Miller, & Heaton, 2000), although other research suggests that it does not account for much variance in CVLT performance (e.g. it accounted for only 0.3% of the variance on CVLT-II total immediate recall; Delis et al., 2017). Nevertheless, given the noted differences in the ethnic composition of the CVLT-II and CVLT3 normative samples, ethnicity was included as a predictor in the primary analyses.

Primary analyses

Descriptive statistics associated with performance on the immediate- and delayed-recall trials are provided in Tables 2–5 (stratified by cohort year, education level, age group, and ethnicity, respectively). MANOVA test results revealed significant effects of cohort year on the immediate-recall trials, but not on the delayed-recall trials (see Table 6). In contrast, MANOVA tests revealed significant effects of education, age group, and ethnicity on the immediate- and delayed-recall trials. No cohort year x education, cohort year x age group, or cohort year x ethnicity interaction effects were observed on the immediate- or delayed-recall trials. Results from follow-up univariate tests and post-hoc pairwise comparisons (with Bonferroni corrections for multiple comparisons) are described below. Inferential statistics associated with follow-up univariate tests are also provided in Table 7.

Table 2. Descriptive statistics associated with performance on the immediate- and delayed-recall trials in the CVLT-II and CVLT3 normative samples.

	CVLT-II ($n = 1087$)		CVLT3 ($n = 700$)	
Trial	*M*	*SE*	*M*	*SE*
Trial 1	6.04	0.06	5.66	0.08
Trial 2	8.58	0.08	8.22	0.10
Trial 3	10.02	0.08	9.80	0.11
Trial 4	10.90	0.09	10.71	0.11
Trial 5	11.55	0.08	11.37	0.11
Trials 1–5 Total	47.11	0.34	45.75	0.44
List B	5.55	0.07	4.95	0.07
SDFR	9.54	0.10	9.55	0.13
SDCR	10.94	0.09	10.72	0.11
LDFR	10.19	0.11	10.15	0.13
LDCR	11.11	0.10	10.98	0.12

Note: CVLT = California Verbal Learning Test; SDFR = Short Delay Free Recall; SDCR = Short Delay Cued Recall; LDFR = Long Delay Free Recall; LDCR = Long Delay Cued Recall; *M* = mean; *SE* = standard error.

Table 3. Descriptive statistics associated with performance on the CVLT-II/CVLT3 immediate- and delayed-recall trials, stratified by education level.

	0–8 years ($n = 104$)		9–11 years ($n = 178$)		HS diploma or equivalent ($n = 566$)		Some college, technical school, or AA ($n = 510$)		Bachelor's degree or more ($n = 429$)	
Trial	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Trial 1	4.95	0.18	5.38	0.14	5.63	0.08	6.19	0.09	6.32	0.10
Trial 2	6.83	0.20	7.63	0.18	8.10	0.10	8.79	0.11	9.21	0.13
Trial 3	8.20	0.27	8.81	0.22	9.65	0.12	10.32	0.12	10.75	0.13
Trial 4	8.82	0.27	9.78	0.23	10.44	0.12	11.28	0.11	11.72	0.13
Trial 5	9.74	0.27	10.24	0.22	11.13	0.11	12.00	0.11	12.26	0.13
Trials 1–5 Total	38.54	1.01	41.83	0.87	44.96	0.46	48.58	0.46	50.25	0.54
List B	4.48	0.20	4.61	0.17	5.04	0.09	5.64	0.09	5.79	0.10
SDFR	7.53	0.31	8.50	0.25	9.09	0.14	9.99	0.15	10.54	0.16
SDCR	9.51	0.27	9.87	0.23	10.48	0.12	11.22	0.13	11.65	0.13
LDFR	8.38	0.31	8.84	0.28	9.71	0.14	10.67	0.15	11.18	0.16
LDCR	9.53	0.28	9.92	0.24	10.63	0.13	11.48	0.14	11.98	0.14

Note: CVLT = California Verbal Learning Test; HS = high school; AA = Associate's degree; SDFR = Short Delay Free Recall; SDCR = Short Delay Cued Recall; LDFR = Long Delay Free Recall; LDCR = Long Delay Cued Recall; *M* = mean; *SE* = standard error.

Cohort year

Follow-up univariate tests revealed a significant effect of cohort year on Trial 1, Trial 2, Trial 3, Trials 1–5 Total, and the List B trial. An examination of mean values indicated that the CVLT3 cohort performed significantly *worse* than the CVLT-II cohort on these trials, and these significant cohort effects were associated with small effect sizes ($\eta_p^2 < .010$; see Table 7).

Education

Follow-up univariate tests revealed a significant effect of education on all immediate- and delayed-recall trials. Consistent with past research (Delis et al., 2017), post-hoc comparisons generally revealed a significant positive, albeit relatively small, association between education and performance ($\eta_p^2 = .026$–$.060$ on immediate-recall trials; $\eta_p^2 = .030$–$.038$ on delayed-recall trials; see Table 7).

Age group

Follow-up univariate tests revealed a significant effect of age group on all immediate- and delayed-recall trials. Consistent with past research (Delis et al., 2017), post-hoc comparisons revealed a significant negative association between age group and performance ($\eta_p^2 = .142$–$.221$ on immediate-recall trials; $\eta_p^2 = .127$–$.183$ on delayed-recall trials; see Table 7).

Of note, an initial analysis (in which MANOVA tests examining cohort effects were stratified by age group) indicated that, for participants ages 70–79 years, there was a significant, *positive* cohort effect on delayed recall performance (i.e. the CVLT3 cohort performed significantly *better* than the CVLT-II cohort on delayed-recall trials); this cohort effect appeared to be driven more by changes in education level than by cohort year of norms collection (based on follow-up regression analyses indicating that more variance in performance was explained by education than by cohort year,

Table 4. Descriptive statistics associated with performance on the CVLT-II/CVLT3 immediate- and delayed-recall trials, stratified by age group.

	16–19 years (*n* = 250)		20–29 years (*n* = 290)		30–44 years (*n* = 300)		45–59 years (*n* = 250)		60–69 years (*n* = 245)		70–79 years (*n* = 245)		80–90 years (*n* = 207)	
Trial	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Trial 1	6.49	0.12	6.68	0.11	6.63	0.11	6.07	0.12	5.52	0.11	4.98	0.10	4.29	0.10
Trial 2	9.16	0.15	9.39	0.14	9.42	0.15	8.80	0.15	7.96	0.15	7.25	0.14	6.39	0.14
Trial 3	11.08	0.15	11.17	0.14	11.14	0.15	10.16	0.16	9.10	0.17	8.50	0.16	7.51	0.17
Trial 4	11.98	0.14	12.01	0.13	11.82	0.15	11.05	0.17	10.07	0.17	9.58	0.18	8.43	0.19
Trial 5	12.70	0.13	12.59	0.14	12.48	0.14	11.60	0.17	10.76	0.17	10.34	0.17	9.04	0.19
Trials 1–5 Total	51.41	0.57	51.84	0.54	51.50	0.58	47.67	0.66	43.41	0.67	40.65	0.64	35.65	0.67
List B	6.16	0.13	6.32	0.12	5.98	0.13	5.36	0.12	4.77	0.11	4.30	0.11	3.71	0.11
SDFR	11.15	0.17	10.89	0.17	10.87	0.17	9.76	0.21	8.45	0.20	8.09	0.21	6.57	0.20
SDCR	11.73	0.15	11.79	0.16	11.92	0.16	11.02	0.18	10.26	0.17	9.95	0.18	8.51	0.20
LDFR	11.52	0.17	11.56	0.17	11.50	0.17	10.31	0.22	9.22	0.21	8.71	0.22	7.38	0.23
LDCR	12.04	0.16	12.27	0.16	12.20	0.16	11.23	0.20	10.32	0.19	9.93	0.20	8.55	0.21

Note: CVLT = California Verbal Learning Test; SDFR = Short Delay Free Recall; SDCR = Short Delay Cued Recall; LDFR = Long Delay Free Recall; LDCR = Long Delay Cued Recall; *M* = mean; *SE* = standard error.

Table 5. Descriptive statistics associated with performance on the CVLT-II/CVLT3 immediate- and delayed-recall trials, stratified by ethnicity.

	White		African-American		Hispanic		Other	
Trial	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Trial 1	5.92	0.06	5.63	0.13	5.99	0.14	5.91	0.24
Trial 2	8.47	0.07	7.91	0.17	8.69	0.18	8.72	0.32
Trial 3	9.90	0.08	9.43	0.19	10.52	0.20	10.40	0.31
Trial 4	10.86	0.08	10.19	0.19	11.23	0.20	10.90	0.31
Trial 5	11.53	0.08	10.75	0.21	11.84	0.18	11.71	0.31
Trials 1–5 Total	46.68	0.31	43.92	0.78	48.27	0.78	47.65	1.33
List B	5.34	0.06	4.98	0.16	5.65	0.17	5.03	0.24
SDFR	9.54	0.09	8.66	0.24	10.50	0.22	9.70	0.39
SDCR	10.93	0.08	10.04	0.21	11.22	0.19	10.78	0.31
LDFR	10.19	0.10	9.13	0.26	10.97	0.23	10.63	0.38
LDCR	11.13	0.09	10.05	0.23	11.61	0.20	11.14	0.34

Note: CVLT = California Verbal Learning Test; SDFR = Short Delay Free Recall; SDCR = Short Delay Cued Recall; LDFR = Long Delay Free Recall; LDCR = Long Delay Cued Recall; *M* = mean; *SE* = standard error.

Table 6. Inferential statistics associated with multivariate analysis of variance (MANOVA) tests examining main and interaction effects of cohort year, education, age group, and ethnicity on performance on the CVLT-II/CVLT3 immediate- and delayed-recall trials.

	Immediate recall				Delayed recall			
Predictor	*F*	df	*p*	Wilk's Λ	*F*	df	*p*	Wilk's Λ
Cohort year	3.75	6, 1754.00	.001*	.987	1.01	4, 1756.00	.400	.998
Education	5.77	24, 6120.19	<.001*	.925	5.03	16, 5365.31	<.001*	.956
Age group	15.55	36, 7705.11	<.001*	.735	17.00	24, 6127.16	<.001*	.799
Ethnicity	3.38	18, 4961.55	<.001*	.966	5.51	12, 4646.23	<.001*	.963
Cohort year x education	0.69	24, 6120.19	.869	.991	1.11	16, 5365.31	.339	.990
Cohort year x age group	1.03	36, 7705.11	.425	.979	1.34	24, 6127.16	.124	.982
Cohort year x ethnicity	1.28	18, 4961.55	.191	.987	1.52	12, 4646.23	.108	.990

Note: CVLT = California Verbal Learning Test.
*Significant at level of .05.

coupled with chi-square tests showing that there was a significant increase (from 14% to 28%) in the proportion of individuals ages 70–79 years who completed a Bachelor's degree or more across the CVLT-II and CVLT3 normative samples). However, these differences failed to reach statistical significance in our final analyses; that is, no significant cohort year × age group interaction effects were found on CVLT-II/CVLT3 performance, precluding the need to follow up with univariate tests exploring potential cohort differences across different age groups (which, as discussed above, would have revealed the positive cohort effect in delayed-recall performance among individuals ages 70–79 noted above).

Ethnicity

Follow-up univariate tests revealed a significant effect of ethnicity on all immediate- and delayed-recall trials. Post-hoc comparisons revealed a pattern such that, consistent with some past research (e.g. Norman et al., 2000, 2011), performance on multiple immediate- and delayed-recall trials was significantly higher among White and Hispanic individuals relative to African-American individuals ($\eta_p{}^2$ = .008–.022 on immediate-recall trials; $\eta_p{}^2$ = .023–.028 on delayed-recall trials; see Table 7).

Table 7. Inferential statistics associated with follow-up univariate tests demonstrating significant effects of cohort year, education, age group, and ethnicity on performance on the CVLT-II/CVLT3 immediate- and delayed-recall trials.

Predictor	Trial	*F*	df	*p*	η_p^2
Immediate recall					
Cohort year	Trial 1	9.81	1, 1759	.002*	.006
	Trial 2	12.10	1, 1759	.001*	.007
	Trial 3	7.52	1, 1759	.006*	.004
	Trial 4	5.26	1, 1759	.022	.003
	Trial 5	3.30	1, 1759	.069	.002
	Trial 1–5 Total	10.31	1, 1759	.001*	.006
	List B	16.39	1, 1759	<.001*	.009
Education	Trial 1	11.57	4, 1759	<.001*	.026
	Trial 2	19.76	4, 1759	<.001*	.043
	Trial 3	19.45	4, 1759	<.001*	.042
	Trial 4	23.52	4, 1759	<.001*	.051
	Trial 5	22.04	4, 1759	<.001*	.048
	Trial 1–5 Total	28.13	4, 1759	<.001*	.060
	List B	13.47	4, 1759	<.001*	.030
Age group	Trial 1	51.58	6, 1759	<.001*	.150
	Trial 2	48.59	6, 1759	<.001*	.142
	Trial 3	66.72	6, 1759	<.001*	.185
	Trial 4	57.97	6, 1759	<.001*	.165
	Trial 5	60.77	6, 1759	<.001*	.172
	Trial 1–5 Total	83.41	6, 1759	<.001*	.221
	List B	52.84	6, 1759	<.001*	.153
Ethnicity	Trial 1	4.97	3, 1759	.002*	.008
	Trial 2	7.93	3, 1759	<.001*	.013
	Trial 3	7.88	3, 1759	<.001*	.013
	Trial 4	10.96	3, 1759	<.001*	.018
	Trial 5	13.29	3, 1759	<.001*	.022
	Trial 1–5 Total	12.77	3, 1759	<.001*	.021
	List B	8.07	3, 1759	<.001*	.014
Delayed recall					
Cohort year	SDFR	0.55	1, 1759	.457	<.001
	SDCR	2.49	1, 1759	.115	.001
	LDFR	2.17	1, 1759	.141	.001
	LDCR	2.50	1, 1759	.114	.001
Education	SDFR	16.02	4, 1759	<.001*	.035
	SDCR	13.68	4, 1759	<.001*	.030
	LDFR	17.43	4, 1759	<.001*	.038
	LDCR	16.63	4, 1759	<.001*	.036
Age group	SDFR	65.77	6, 1759	<.001*	.183
	SDCR	42.48	6, 1759	<.001*	.127
	LDFR	54.16	6, 1759	<.001*	.156
	LDCR	47.58	6, 1759	<.001*	.140
Ethnicity	SDFR	14.72	3, 1759	<.001*	.024
	SDCR	13.81	3, 1759	<.001*	.023
	LDFR	15.99	3, 1759	<.001*	.027
	LDCR	16.83	3, 1759	<.001*	.028

Note: CVLT = California Verbal Learning Test; SDFR = Short Delay Free Recall; SDCR = Short Delay Cued Recall; LDFR = Long Delay Free Recall; LDCR = Long Delay Cued Recall; *M* = mean; *SE* = standard error.

*Significant at level of .007 for immediate-recall trials and .013 for delayed-recall trials, after Bonferroni corrections for multiple comparisons.

Discussion

In the present study, we examined differences in performance on the immediate- and delayed-recall trials between the CVLT-II and CVLT3 normative samples. Specifically, we explored the extent to which verbal learning and memory skills were influenced by the cohort year in which norms were collected (i.e. 1999 for the CVLT-II versus

2016–2017 for the CVLT3) versus differences in education level. Of note, differences in education level between the CVLT-II and CVLT3 normative samples mirrored an increase in the proportion of U.S. adults who completed post-secondary education during the time period spanning the development of the CVLT-II and CVLT3.

The present study revealed evidence of a *negative* Flynn effect on the attention/working memory and learning trials of the CVLT-II/CVLT3, with the CVLT3 cohort performing significantly *worse* than the CVLT-II cohort on Trial 1, Trial 2, Trial 3, Trials 1–5 Total, and List B). In contrast, no significant cohort differences were found on the delayed-recall trials. Consistent with past research, education level, age group, and ethnicity were shown to be significant predictors of overall CVLT performance. Education level and age group were positively and negatively associated with CVLT-II/CVLT3 performance, respectively. With regard to ethnicity, performance on multiple immediate- and delayed-recall trials was significantly higher among White and Hispanic individuals relative to African-American individuals. *Nevertheless, none of these demographic variables were shown to have an interactive effect with cohort year of norms collection on performance.*

The present study overcomes some of the limitations of previous studies that examined Flynn or cohort effects on learning and memory of word lists (e.g. use of relatively small sample sizes; limited age ranges; confounding time of testing with changes in the target words; using data harmonization techniques to convert Logical Memory scores to CERAD Word List scores). Of note, the present study offers the advantage of using the same word lists administered to large normative samples that represent a wide age range and that were matched to the demographic makeup of the U.S. census at the time that the testing occurred in order to explore potential cohort effects on a standardized measure of verbal learning and memory. Further, the present findings are in line with recent research suggesting that a negative Flynn effect may be occurring not only on IQ tests, but also on measures of auditory attention/working memory and learning of word lists. That is, given that negative cohort effects were observed only on immediate-recall trials (and appeared to be driven by cohort differences on the first three learning trials in particular), the present findings provide further evidence that the attention/working memory aspects of verbal memory may be particularly vulnerable to negative cohort effects (Wongupparaj et al., 2017).

As discussed above, the present study indicated that the CVLT3 normative sample was more highly educated, on average, than the CVLT-II normative sample, and this difference mirrored the increase over the past two decades in the proportion of U.S. adults who completed post-secondary education. However, while the present study yielded evidence for a significant *positive* (albeit relatively small) association between education level and CVLT-II/CVLT3 performance, the evidence for a *negative* cohort effect on performance persisted even after accounting for differences in education level and cohort year × education interactions (which were nonsignificant). Furthermore, the observed negative cohort effect on immediate recall was present across all age and ethnic groups, with cohort year x age group and cohort year x ethnicity interactions also being nonsignificant.

The present results are consistent with the findings from a meta-analysis of cohort effects on attention and working memory measures conducted by Wongupparaj et al.

(2017), who found a gradual decline in more complex auditory attention/working memory skills (e.g. digit span backward) over the past four decades. Although the current findings differ from those of Dodge et al. (2017), in which a positive cohort effect was reported for word list learning and memory performance (including immediate and delayed recall), there were a number of limitations in that study that make it difficult to directly compare results from the two investigations (e.g. investigating only older adults [65 years or older]; notable differences in the proportions of adults in each age cohort within the pooled sample who were tested in the first versus second data collection period; not addressing the fact that age cohort and period of testing were potentially confounded; and using data harmonization techniques to convert Logical Memory scores from a subset of its data pool into CERAD Word List scores to facilitate analyses of cohort differences on verbal memory performance).

The results of the present study raise intriguing questions about the effects of socioenvironmental changes that have unfolded during the time period spanning the development of the second and third editions of the CVLT. In particular, the present findings suggest that socioenvironmental changes may have occurred since 2000 that (a) might be negatively impacting working memory and verbal learning skills, (b) are not disproportionately affecting certain age or ethnic groups, and (c) are occurring independent of generational changes in educational attainment. While education level was examined in the present study, a number of researchers have highlighted distinctions between educational *attainment* and education *quality*, and have suggested that "educational attainment" as a homogeneous variable may have become diluted in recent years due to varying standards and quality required for degrees across educational settings (Allen & Seaman, 2013; Bratsberg & Rogeberg, 2018; Hamad et al., 2019; Jaggars & Bailey, 2010; Nguyen et al., 2016; Rindermann et al., 2017). The lack of an observed cohort year x education level interaction effect found in the present study may reflect, in part, these recent concerns about the homogeneity of educational attainment. This is important to consider for the present findings given that the negative Flynn effect that has recently been found on IQ measures has been partly attributed to reduced quality of education in those studies (Allen & Seaman, 2013; Jaggars & Bailey, 2010).

It is difficult to escape the observation that the time period spanning the development of the second and third editions of the CVLT (1999/2000 versus 2016/2017) also coincided with a profound societal change: the digital revolution. As noted in recent reviews (Rindermann et al., 2017; Wilmer, Sherman, & Chein, 2017), the use of digital technology, while offering multiple advantages, may have subtle but significant adverse effects on working memory and rote memorization skills. While relationships between the use of digital technology and verbal learning and memory performance were not formally investigated in the present study, the current findings invite the intriguing hypothesis that increased use of digital tools may inadvertently have an adverse effect on working memory and learning abilities. Unfortunately, there has been a paucity of studies investigating associations of self-reported and performance-based internet use with cognition. While there is evidence that the ability to perform different tasks on the internet is significantly correlated with performance on cognitive tests (Woods et al., 2019), no studies have directly investigated whether varying degrees of internet, mobile phone, or other digital technology usage may positively or negatively affect the development and maintenance of different domains of cognition. Future research should explore potential differences between high and low internet users on neuropsychological test performance. In addition, the present study was also limited in that we were unable to assess relationships between other socioenvironmental changes that may have occurred in the years spanning the development of the CVLT-II and CVLT3 (e.g. generational changes in healthcare or standard of living; see Dutton & Lynn, 2013; Rindermann et al., 2017).

The present findings were likely related to true cohort differences in verbal learning and memory skills, and not to differences between the makeup of the CVLT-II versus the CVLT3, given that (a) the lists of target words are identical across the two versions of the test; (b) the negative cohort effect was only observed on select trials, thereby indicating that one version is not uniformly harder or easier than the other; and (c) other recent studies have also found evidence for a negative Flynn effect on attention/working memory components of verbal memory (Wongupparaj et al., 2017). One question that arises is whether the observed negative cohort effect found on the CVLT in the present study was due to a negative Flynn effect specifically on attention/working memory and learning skills versus a broader effect on IQ in general, which has also been reported in recent years (Bratsberg & Rogeberg, 2018; Dutton & Lynn, 2013, 2015; Dutton et al., 2016; Flynn & Shayer, 2018; Pietschnig & Gittler, 2015; Shayer & Ginsburg, 2007, 2009; Sundet et al., 2004; Teasdale & Owen, 2005, 2008; Woodley & Meisenberg, 2013). Given that the CVLT-II and CVLT3 were not co-normed with IQ tests, we cannot directly investigate this relationship. However, IQ has been shown to correlate robustly with education level, and education was not shown to drive or moderate any of the observed cohort effects in the present study. These findings suggest that the present findings were related to true cohort differences in attention/working memory and learning skills independent of any cohort changes that might also be occurring for IQ functions in general.

It is also worth noting that the negative cohort effects observed in the present study were associated with relatively small effect size estimates (i.e. $\eta_p^2 < .010$ on immediate-recall trials). However, the cohort effects are unlikely due to random chance given the robust statistical power rendered by our large sample size. Moreover, from a clinical perspective, even a small difference in raw scores can have a notable impact on the conversion to standardized scores, which in turn can impact decisions about an examinee's level of cognitive functioning. For example, for an individual within the age range of 45–54 years, a raw score of 4 on Trial 1 yields a *z*-score of −1.5 based on CVLT-II norms versus a scaled score of 7 based on CVLT3 norms (note that the CVLT3 now uses scaled scores rather than *z*-scores); thus, this individual's Trial 1 performance could be interpreted as mildly impaired using CVLT-II norms and low average using CVLT3 norms.

The present results have other important implications for clinical practice. In a recent position paper, Bush et al. (2018) discussed the advantages and disadvantages of using newer versus older versions of neuropsychological tests. The authors note that an advantage of an older version of a neuropsychological test is that it may be grounded more in empirical data supporting its validity, whereas a newer version may lack such empirical support. Additionally, older versions of tests offer the advantage of increased familiarity and ease of interpretation for clinicians. However, Bush et al. (2018) also note that if cohort differences are found in the normative data between the older and new versions of a test, then the use of the older version may provide inaccurate standardized scores in a present-day evaluation (see also Alenius et al., 2019). Given the present findings, the continued use of the CVLT-II's 1999 norms in today's assessments may provide artificially lower standardized scores on indices of attention/working memory and learning across the immediate-recall trials (e.g. Trial 1, Trial 2, Trial 3, Trials 1–5 Total, List B). Further, given that the target lists and Yes/No Recognition trial are the same on the CVLT3 as those used on the CVLT-II, 1) the validity studies that have been conducted to date for the CVLT-II (over 1,000 published studies; Delis et al., 2017) likely still have relevance for the CVLT3, and 2) familiarity and ease of interpretation should be relatively equivalent across the two test versions. Finally, the present results also suggest that the normative data that are currently being used for other verbal learning and memory tests (e.g. California Verbal Learning Test – Children's Version; Rey Auditory Verbal Learning Test; Hopkins Verbal Learning Test), which were initially collected before 2000 and have not undergone any major revisions since the early 2000s, may also have become outdated and are in need of re-norming in the near future.

In summary, the current study found evidence of a *negative* Flynn effect on the attention/working memory and learning trials of the CVLT-II/CVLT3. The findings have clinical implications for the use of word list learning and memory tests like the CVLT, and raise intriguing questions about the possible adverse effects of recent socioenvironmental changes on attention, working memory, and learning skills.

Disclosure statement

Dean C. Delis, Ph.D. is a co-author of the CVLT-II and CVLT3 and receives royalties for the tests. Lisa Drozdick, Ph.D., and Troy Courville, Ph.D., are or were employees of Pearson and contributed to the development of the CVLT.

References

Aboudarham, J.-F., & Zalewski, C. (1995). The relationship between ethnicity and performance on the California Verbal Learning Test. *Archives of Clinical Neuropsychology*, *10*(4), 285–286. doi:10.1093/arclin/10.4.285

Alenius, M., Koskinen, S., Hallikainen, I., Ngandu, T., Lipsanen, J., Sainio, P., … Hänninen, T. (2019). Cognitive performance among cognitively healthy adults aged 30-100 years. *Dementia and Geriatric Cognitive Disorders Extra*, *9*(1), 11–23. doi:10.1159/000495657

Allen, I. E., & Seaman, J. (2013). *Changing course: Ten years of tracking online education in the United States*. Babson Park, MA: Babson Survey Research Group and Quahog Research Group, LLC.

Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (2006). Lifespan theory in developmental psychology. In R. M. Lerner & W. Damon (Eds.), *Theoretical models of human development. Handbook of child psychology* (Vol. 1, 6th ed., pp. 569–664). Hoboken, NJ: Wiley.

Baxendale, S. (2010). The Flynn effect and memory function. *Journal of Clinical and Experimental Neuropsychology*, *32*(7), 699–703. doi:10.1080/13803390903493515

Brailean, A., Huisman, M., Prince, M., Prina, A. M., Deeg, D. J. H., & Comijs, H. (2018). Cohort differences in cognitive aging in the Longitudinal Aging Study Amsterdam. *Journals of Gerontology: Psychological Sciences*, *73*(7), 1214–1223. doi:10.1093/geronb/gbw129

Bratsberg, B., & Rogeberg, O. (2018). Flynn effect and its reversal are both environmentally caused. *Proceedings of the National Academy of Sciences*, *115*(26), 6674–6678. doi:10.1073/pnas.1718793115

Bush, S. S., Sweet, J. J., Bianchini, K. J., Johnson-Greene, D., Dean, P. M., & Schoenberg, M. R. (2018). Deciding to adopt revised and new psychological and neuropsychological tests: An inter-organizational position paper. *The Clinical Neuropsychologist*, *32*(3), 319–325. doi:10.1080/13854046.2017.1422277

Christensen, K., Thinggaard, M., Oksuzyan, A., Steenstrup, T., Andersen-Ranberg, K., Jeune, B., … 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*(9903), 1507–1513. doi:10.1016/S0140-6736(13)60777-1

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test–II* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2017). *California Verbal Learning Test–3* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Dodge, H. H., Zhu, J., Hughes, T. F., Snitz, B. E., Chang, C.-C. H., Jacobsen, E. P., & Ganguli, M. (2017). Cohort effects in verbal memory function and practice effects: A population-based study. *International Psychogeriatrics*, *29*(1), 137–148. doi:10.1017/S1041610216001551

Dodge, H. H., Zhu, J., Lee, C.-W., Chang, C.-C H., & Ganguli, M. (2014). Cohort effects in age-associated cognitive trajectories. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *69*(6), 687–694. doi:10.1093/gerona/glt181

Dutton, E., & Lynn, R. (2013). A negative Flynn effect in Finland, 1997-2009. *Intelligence*, *41*(6), 817–820. doi:10.1016/j.intell.2013.05.008

Dutton, E., & Lynn, R. (2015). A negative Flynn Effect in France, 1999 to 2008-9. *Intelligence*, *51*, 67–70. doi:10.1016/j.intell.2015.05.005

Dutton, E., van der Linden, D., & Lynn, R. (2016). The negative Flynn effect: A systematic literature review. *Intelligence*, *59*, 163–169. doi:10.1016/j.intell.2016.10.002

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*(2), 171–191. doi:10.1037/0033-2909.101.2.171

Flynn, J. R. (2012). *Are we getting smarter?: Rising IQ in the twenty-first century*. Cambridge University Press: Cambridge, UK.

Flynn, J. R., & Shayer, M. (2018). IQ decline and Piaget: Does the rot start at the top? Intelligence, *66*, 112–121. doi:10.1016/j.intell.2017.11.010

Fox, J. M., Brook, M., Heilbronner, R. L., Susmaras, T., & Hanlon, R. E. (2019). Neuropsychological and criminological features of female homicide offenders. *Journal of Forensic Sciences*, *64*(2), 460–467. doi:10.1111/1556-4029.13911

Freedman, V. A., Aykan, H., & Martin, L. G. (2002). Another look at aggregate changes in severe cognitive impairment: Further investigation into the cumulative effects of three survey design issues. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(2), S126–S131. doi:10.1093/geronb/57.2.S126

Hamad, R., Nguyen, T. T., Glymour, M. M., Vable, A., Manly, J. J., & Rehkopf, D. H. (2019). Quality and quantity: The association of state-level educational policies with later life cardiovascular disease. *Preventative Medicine*, *126*. doi:10.1016/j.ypmed.2019.06.008.

Jaggars, S. S., & Bailey, T. (2010). Effectiveness of fully online courses for college students: Response to a Department of Education meta-analysis. New York, NY: Community College Research Center, Teachers College, Columbia University. Retrieved from https://ccrc.tc.columbia.edu/publications/effectiveness-fully-online-courses.html

Karlsson, P., Thorvaldsson, V., Skoog, I., Gudmundsson, P., & Johansson, B. (2015). Birth cohort differences in fluid cognition in old age: Comparisons of trends in levels and change trajectories over 30 years in three population-based samples. *Psychology and Aging*, *30*(1), 83–94. doi:10.1037/a0038643

Langa, K. M., Larson, E. B., Karlawish, J. H., Cutler, D. M., Kabeto, M. U., Kim, S. Y., … Rosen, A. B. (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 & Dementia*, *4*(2), 134–144. doi:10.1016/j.alz.2008.01.001

Llewellyn, D. J., & Matthews, F. E. (2009). Increasing levels of semantic verbal fluency in elderly English adults. *Neuropsychology, Development, and Cognition, Section B: Aging. Neuropsychology and Cognition*, *16*(4), 433–445. doi:10.1080/13825580902773867

Nguyen, T. T., Tchetgen, E. J. T., Kawachi, I., Gilman, S. E., Walter, S., Liu, S. Y., … Glymour, M. M. (2016). Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk. *Annals of Epidemiology*, *26*(1), 71–76. doi:10.1016/j.annepidem.2015.10.006

Norman, M. A., Evans, J. D., Miller, W. S., & Heaton, R. K. (2000). Demographically corrected norms for the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*, *22*(1), 80–94. doi:10.1076/1380-3395(200002)22:1;1-8;FT080

Norman, M. A., Moore, D. J., Taylor, M., Franklin, D., Jr., Cysique, L., Ake, C., … the HNRC Group. (2011). Demographically corrected norms for African Americans and Caucasians on the Hopkins Verbal Learning Test-Revised, Brief Visuospatial Memory Test-Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-Card Version. *Journal of Clinical and Experimental Neuropsychology*, *33*(7), 793–804. doi:10.1080/13803395.2011.559157

Pietschnig, J., & Gittler, G. (2015). A reversal of the Flynn effect for spatial perception in German-speaking countries: Evidence from a cross-temporal IRT-based meta-analysis (1977-2014). Intelligence, *53*, 145–153. doi:10.1016/j.intell.2015.10.004

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn Effect (1909-2013). *Perspectives on Psychological Science*, *10*(3), 282–306. doi:10.1177/1745691615577701

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*(1), 33–65. doi:10.1016/j.acn.2004.02.005

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology*, *31*(3), 206–230. doi:10.1093/arclin/acw007

Rindermann, H., Becker, D., & Coyle, T. R. (2017). Survey of expert opinion on intelligence: The Flynn effect and the future of intelligence. *Personality and Individual Differences*, *106*, 242–247. doi:10.1016/j.paid.2016.10.061

Rönnlund, M., & Nilsson, L.-G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analyses of data from a Swedish cohort study. Intelligence, *36*(3), 192–209. doi:10.1016/j.intell.2007.05.002

Rönnlund, M., & Nilsson, L.-G. (2009). Flynn effects on sub-factors of episodic and semantic memory: Parallel gains over time and the same set of determining factors. Neuropsychologia, *47*(11), 2174–2180. doi:10.1016/j.neuropsychologia.2008.11.007

Schaie, K. W., Willis, S. L., & Pennak, S. (2005). An historical framework for cohort differences in intelligence. *Research in Human Development*, *2*(1), 43–67. doi:10.1207/s15427617rhd0201&2_3

Shayer, M., & Ginsburg, D. (2007). Thirty years on – A large anti-Flynn effect? The Piagetian test volume & heaviness norms 1975-2003. British Journal of Educational Psychology, *77*(1), 25–41. doi:10.1348/000709906X96987

Shayer, M., & Ginsburg, D. (2009). Thirty years on – A large anti-Flynn effect? (II): 13- and 14-year-olds. Piagetian tests of formal operations norms 1976-2006/7. *British Journal of Educational Psychology*, *79*(3), 409–418. doi:10.1348/978185408X383123

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

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, *32*(4), 349–362. doi:10.1016/j.intell.2004.06.004

Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn effect in reverse. *Personality and Individual Differences*, *39*(4), 837–843. doi:10.1016/j.paid.2005.01.029

Teasdale, T. W., & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn effect. Intelligence, *36*(2), 121–126. doi:10.1016/j.intell.2007.01.007

Thorvaldsson, V., Karlsson, P., Skoog, J., Skoog, I., & Johansson, B. (2017). Better cognition in new birth cohorts of 70 y ear olds, but greater decline thereafter. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, *72*(1), 16–24. doi:10.1093/geronb/gbw125

Trahan, L., Stuebing, K. K., Hiscock, M. K., & Fletcher, J. M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin*, *140*(5), 1332–1360. doi:10.1037/a0037173

Wilmer, H. H., Sherman, L. E., & Chein, J. M. (2017). Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. *Frontiers in Psychology*, *8*(605), 1–16. doi:10.3389/fpsyg.2017.00605

Wongupparaj, P., Wongupparaj, R., Kumari, V., & Morris, R. G. (2017). The Flynn effect for verbal and visuospatial short-term and working memory: A cross-temporal meta-analysis. Intelligence, *64*, 71–80. doi:10.1016/j.intell.2017.07.006

Woods, S. P., Kordovski, V. M., Tierney, S. M., & Babicz, M. A. (2019). The neuropsychological aspects of performance-based Internet navigation skills: A brief review of an emerging literature. *The Clinical Neuropsychologist, 33*(2), 305–326. doi:10.1080/13854046.2018.1503332.

Woodley, M., & Meisenberg, G. (2013). In the Netherlands the anti-Flynn effect is a Jensen effect. *Personality and Individual Differences*, *54*(8), 871–876. doi:10.1016/j.paid.2012.12.022