

## Remember this: Age moderation of genetic and environmental contributions to verbal episodic memory from midlife through late adulthood

Susan E. Luczak<sup>a,\*</sup>, Christopher R. Beam<sup>a</sup>, Shandell Pahlen<sup>b</sup>, Morgan Lynch<sup>a</sup>, Matthew Pilgrim<sup>a</sup>, Chandra A. Reynolds<sup>b</sup>, Matthew S. Panizzon<sup>c,d</sup>, Vibeke S. Catts<sup>e</sup>, Kaare Christensen<sup>f,g,h</sup>, Deborah Finkel<sup>i,j</sup>, Carol E. Franz<sup>c,d</sup>, William S. Kremen<sup>c,d</sup>, Teresa Lee<sup>e</sup>, Matt McGue<sup>k,l</sup>, Marianne Nygaard<sup>f</sup>, Brenda L. Plassman<sup>m</sup>, Keith E. Whitfield<sup>n</sup>, Nancy L. Pedersen<sup>o</sup>, Margaret Gatz<sup>i</sup>, for the IGEMS Consortium

<sup>a</sup> Department of Psychology, University of Southern California, Los Angeles, CA, USA

<sup>b</sup> Department of Psychology, University of California, Riverside, Riverside, CA, USA

<sup>c</sup> Department of Psychiatry, University of California, San Diego, San Diego, CA, USA

<sup>d</sup> Center for Behavior Genetics of Aging, University of California, San Diego, San Diego, CA, USA

<sup>e</sup> Centre for Healthy Brain Ageing (CHEBA), Discipline of Psychiatry and Mental Health, Faculty of Medicine and Health, University of New South Wales, Sydney, Australia

<sup>f</sup> The Danish Twin Registry, Department of Public Health, University of Southern Denmark, Denmark

<sup>g</sup> Department of Clinical Genetics, Odense University Hospital, Odense, Denmark

<sup>h</sup> Department of Clinical Biochemistry, Odense University Hospital, Odense, Denmark

<sup>i</sup> Center for Economic and Social Research, University of Southern California, Los Angeles, CA, USA

<sup>j</sup> Institute of Gerontology, School of Health and Welfare, Jönköping University, Sweden

<sup>k</sup> Department of Psychology, University of Minnesota, Minneapolis, MN, USA

<sup>l</sup> Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark

<sup>m</sup> Departments of Psychiatry and Neurology, Duke University, Durham, NC, USA

<sup>n</sup> Department of Psychology and Brain and Health, University of Nevada Las Vegas, Las Vegas, NV, USA

<sup>o</sup> Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

### ARTICLE INFO

#### Keywords:

Verbal episodic memory

Twin studies

Aging

### ABSTRACT

It is well documented that memory is heritable and that older adults tend to have poorer memory performance than younger adults. However, whether the magnitudes of genetic and environmental contributions to late-life verbal episodic memory ability differ from those at earlier ages remains unresolved. Twins from 12 studies participating in the Interplay of Genes and Environment in Multiple Studies (IGEMS) consortium constituted the analytic sample. Verbal episodic memory was assessed with immediate word list recall ( $N = 35,204$  individuals; 21,792 twin pairs) and prose recall ( $N = 3805$  individuals; 2028 twin pairs), with scores harmonized across studies. Average test performance was lower in successively older age groups for both measures. Twin models found significant age moderation for both measures, with total inter-individual variance increasing significantly with age, although it was not possible definitively to attribute the increase specifically to either genetic or environmental sources. Pooled results across all 12 studies were compared to results where we successively dropped each study (leave-one-out) to assure results were not due to an outlier. We conclude the models indicated an overall increase in variance for verbal episodic memory that was driven by a combination of increases in the genetic and nonshared environmental parameters that were not independently statistically significant. In contrast to reported results for other cognitive domains, differences in environmental exposures are comparatively important for verbal episodic memory, especially word list learning.

\* Corresponding author at: Department of Psychology, University of Southern California, 3620 S. McClintock Avenue, SGM 501, Los Angeles, CA 90089-1061, USA.  
E-mail address: [luczak@usc.edu](mailto:luczak@usc.edu) (S.E. Luczak).

## 1. Introduction

Age-related changes in episodic memory are a key aspect of older adult cognition (e.g., Nyberg, 2017; Zaheed et al., 2021). Verbal episodic memory is commonly assessed by asking the individual to remember either a list of words (Morris et al., 1989) or a prose story (Storandt, Botwinick, Danziger, Berg, & Hughes, 1984). Younger adults reliably perform better than older adults on immediately recalling the presented material, recalling the material after a short delay, and recognition of which material had or had not been presented (Murphy, West, Armilio, Craik, & Stuss, 2007). A deficit in episodic memory performance is a predominant early sign of Alzheimer's disease (AD; Sperling et al., 2011). In addition, in factor analyses with samples of middle-aged and older adults, episodic memory tends to have lower factor loadings on general intelligence (g) than other cognitive domains, further indicating its unique role in cognitive processes (Beam, Luczak, Reynolds, Panizzon, & Gatz, 2022; Royall & Palmer, 2012).

Twin studies of memory permit addressing the extent to which individual differences in verbal episodic memory performance derive from genetic or environmental sources, and the extent to which the balance of these influences differs with age, with implications for the etiology of age differences in verbal episodic memory. There are several possible patterns of correlations between members of twin pairs that would lead to differences in estimated heritability of a trait. If monozygotic (MZ) twins increase in similarity more than dizygotic (DZ) twins increase in similarity over age, the result would be an increased heritability with age. For instance, if expression of deleterious genes associated with cognitive impairment increases with age, then MZ twins would show greater increase in similarity than DZ twins, reflected in greater genetic influences and greater heritability at older ages (assuming constant total variance in verbal episodic memory). Age also, however, may lead to greater accumulation of unique environmental insults, resulting in increased non-shared environmental influences, and thereby lowered heritability at older ages (see Baltes, Lindenberger, & Staudinger, 2006; McClearn et al., 1997; Papenberg, Lindenberger, & Bäckman, 2015; Turkheimer, Beam, Sundet, & Tamba, 2017). An examination of the literature suggests that heritability of word list learning in the oldest cohorts (e.g., Lindgren, Kaprio, Rinne, & Vuoksima, 2019; McGue & Christensen, 2001) tends to be greater than in midlife samples (e.g., Panizzon et al., 2011). Heritability of prose recall shows a similar but less pronounced trend of higher heritability in older adults (Finkel & McGue, 1993; Giubilei et al., 2008) than in midlife samples (Finkel & McGue, 1993; Kremen et al., 2014). Comparing the two measures of verbal episodic memory, there was a tendency for heritability for recalling a prose passage to be greater than heritability for recalling a list of words (Giubilei et al., 2008; Swan et al., 1999; Wilson, Barral, et al., 2011). In addition, standardized estimates of non-shared environmental influences were correspondingly higher in midlife than in old age for both verbal episodic memory measures.

A meta-analysis of cross-sectional studies of cognitive aging concluded that heritability of episodic memory could not reliably be said either to increase with age or to remain statistically stable across the age spectrum (Reynolds & Finkel, 2015). However, their memory domain encompassed a more heterogeneous collection of verbal and visual tasks than the two considered here. A more recent meta-analysis of specific cognitive abilities (Procopio et al., 2022) did not report results for older adults for their storage and retrieval category. Here we add to the literature by including a large number of studies, especially of older adults, separating results by the specific measure of verbal episodic memory, and pooling the data into a single analytic sample instead of pooling the results from each study.

Some prior studies of heritability of episodic memory considered the confounding influence of education. Greater attained education is known to be related to better cognitive performance, including verbal episodic memory, in middle aged and older adults (cf., Lachman, Agrigoroaei, Murphy, & Tun, 2010). Furthermore, education is often shared

by twins within a family and can inflate the estimate of shared environmental influences if not statistically adjusted for in twin models (Johnson, Deary, & Iacono, 2009). In addition, earlier born cohorts tend to have fewer opportunities for advanced education (Wilson, Zozula, & Gove, 2011); thus, education addresses possible cohort effects underlying age differences.

We evaluated age differences in genetic and environmental influences on verbal episodic memory, using data from 12 studies participating in the Interplay of Genes and Environment in Multiple Studies (IGEMS) consortium (Pedersen et al., 2013; Pedersen, Gatz, et al., 2019). To do so, we first created harmonized scores for word list immediate recall and prose recall immediate recall to enable pooled analyses. Second, using those harmonized scores, we examined phenotypic relationships of verbal episodic memory to age. Third, we estimated heritability and tested whether genetic and environmental variance differed with age. We focused on raw variance to capture differences in total variance by age in addition to the genetic and environmental proportional contributions to variance. Prior studies have found that sex does not affect heritability differently by age (e.g., Pahlen et al., 2018; Read et al., 2006), so we did not test for sex differences but did include sex in the models as a covariate. We examined models with and without attained education covaried, and covaried for study in pooled genetic analyses or examined results split by study in phenotypic analyses.

We expected to find age differences on both verbal episodic memory measures, as has been documented in the literature, with performance lower in older adults than in middle-aged adults. Based on the literature, we predicted a weak trend toward greater genetic influences in older than in younger adults. We separately considered word list and prose recall to determine whether they suggest similar conclusions, while predicting higher heritability for prose recall than word list. This study extends our existing knowledge by testing these relationships in a pooled twin sample over mid- to late-adulthood with adequate power to test for significant genetic and environmental effects.

## 2. Method

All studies included in the analyses are part of the IGEMS consortium (Pedersen et al., 2013; Pedersen, Gatz, et al., 2019). We included 12 studies from 4 countries (Sweden, Australia, Denmark, and the US). Each participating study had to have administered a test of verbal episodic memory, either word list or prose recall. Because of scarce representation at the ends of the age range, we selected participants aged 40 to <90 years at the time of the first administration of the test. Using the first administration eliminated any need to adjust for practice effects. The combined final analytic sample size was 35,204 individuals and 21,792 twin pairs for word list and 3805 individuals and 2028 twin pairs for prose recall (see Table 1).

All word list tests asked participants to listen to or read aloud 10 to 16 related or unrelated words, and then immediately to repeat back as many words as they could remember. Some studies administered one learning trial and some more than one; and some but not all studies included delayed recall with quite disparate durations of delay. In the interest of maximizing comparability of procedures, only the first immediate recall learning trial was used for harmonization. Prose recall is a verbal episodic memory test that requires participants to recall a prose story immediately after it is read aloud. As only some studies included delayed prose recall, harmonization used the immediate recall score. We refer to these two measures as verbal episodic memory measures, but recognize that the first immediate recall trial of word list includes attention components (Gavett & Horwitz, 2012), prose recall is often termed logical memory (Wechsler, 1945, 1987, 1997), and both are considered verbal declarative memory measures (Lezak, 2004).

**Table 1**  
Analytic sample characteristics by study.

Study	N	% female	MZ Families	DZ Families	Mean Age (SD, Range) at Testing	Measure
Sweden						
SATSA	603	60%	141	237	70.2 (9.4, 51–89)	WL
GENDER	485	50%	0	248	74.0 (2.7, 69–83)	WL
OCTO-Twin	620	67%	143	188	82.7 (2.7, 79–89)	PR
Australia						
OATS	565	65%	165	137	70.7 (5.5, 65–89)	WL
	591	65%	171	140	70.6 (5.3, 65–89)	PR
Denmark						
MADT	4268	49%	799	1618	56.4 (10.7, 45–69)	WL
MIDT	9728	53%	1241	5633	56.7 (9.4, 40–80)	WL
LSADT	4195	58%	1024	2140	76.5 (4.9, 70–89)	WL
US						
CAATSA	489	61%	104	192	55.5 (11.1, 40–89)	WL
	480	61%	102	190	55.4 (11.1, 40–89)	PR
VETSA	1482	0%	435	312	56.4 (3.3, 51–66)	WL
	1480	0%	434	312	56.3 (3.3, 51–66)	PR
MTSADA	634	60%	204	144	59.3 (9.0, 40–86)	PR
MIDUS	1150	56%	272	449	60.4 (12.0, 40–86)	WL
NAS-NRC	12,239	0%	3076	3569	66.6 (3.0, 62–81)	WL

Notes. N indicates the number of individual twins in the analytic sample. MZ families = number of monozygotic families (whether one or both twins participated); DZ families = number of dizygotic families (whether one or both twins participated); WL = word list; PR = prose recall.

## 2.1. Participating studies

Brief descriptions of each of the participating studies, by country of origin, are given below. Additional details concerning the methodology for each study can be found in the citations provided. Note that three of the studies were also included in the meta-analysis by Reynolds and Finkel (2015).

### 2.1.1. Sweden

Three Swedish studies were drawn from Swedish Twin Register, a population-based register of twins born in Sweden since 1886 (Lichtenstein et al., 2006).

Swedish Adoption/Twin Study of Aging (SATSA; N. L. Pedersen et al., 1991) is a longitudinal study that began in 1984. It includes same-sex twins who indicated they had been reared apart, and a matched sample of twins reared together. All SATSA twins who reached age 50 years were invited for in-person testing, with word list included from 1999 through 2012.

Ageing in Women and Men: A Longitudinal Study of Gender Differences in Health Behavior and Health among Elderly (GENDER; Gold, Malmberg, McClearn, Pedersen, & Berg, 2002) is a longitudinal study of opposite-sex twin pairs born between 1916 and 1925, and followed from 1995 through 2005. The first in-person testing at age 70–79 years included word list. Word list as administered in SATSA and GENDER consisted of 10 unrelated words (Consortium to Establish a Registry for Alzheimer's Disease, CERAD; Morris et al., 1989).

Origins of Variance in the Oldest-Old (OCTO-Twin; McClearn et al., 1997) includes same-sex twin pairs over the age of 80 at baseline in 1991, with assessments conducted through 2002. Prose recall in OCTO-Twin was the Johansson story (Berg, 1980) with a maximum score of 16.

### 2.1.2. Australia

The Older Australian Twins Study (OATS; Sachdev et al., 2009, 2013) began in 2006 with the recruitment of twins aged 65+ years, including twins from the Australian Twin Registry as well as new volunteers from the community. Word list was an expanded version of the original Rey Auditory Verbal Learning Test (RAVLT) consisting of 15 unrelated words (Giles & English, 2002; Rey, 1964) and prose recall was the Logical Memory subtest from Wechsler Memory Scale-III (WMS-III; Wechsler, 1997) with a maximum score of 25.

### 2.1.3. Denmark

The three Danish twin studies were drawn from the Danish Twin Registry and include the Middle Age Danish Twins Study (MADT; Pedersen, Larson, et al., 2019), Middle age Danish Twin Study (MIDT; Skytthe et al., 2013), and Longitudinal Study of Aging Danish Twins (LSADT; McGue & Christensen, 2007). MADT is a longitudinal study of same and opposite-sex twin pairs born between 1931 and 1952 first assessed in 1998, resulting in an age range from 45 to 68 years. MIDT is a study of twins born between 1931 and 1969 who were not already included in MADT and were assessed between 2008 and 2011, resulting in an age range from 40 to 80 years. LSADT is a cohort-sequential study of same sex twin pairs that began in 1995 and includes age 70 to 96 years. In all studies, word list was 12 unrelated words from the RAVLT.

### 2.1.4. United States

The Carolina African-American Twin Study of Aging (CAATSA) is a cross-sectional population-representative sample of African American twins ranging in age from 20s through 80s living in North Carolina recruited in 1999 through 2003 (Whitfield, 2013; Whitfield, Brandon, Wiggins, Vogler, & McClearn, 2003). Word list was 10 unrelated words from the Telephone Interview for Cognitive Status (TICS; Brandt, Spencer, & Folstein, 1988). Prose recall was the Logical Memory subtest from WMS-R (Wechsler, 1987) with a maximum score of 25.

The Vietnam Era Twin Study of Aging (VETSA; Kremen, Franz, & Lyons, 2013) is a longitudinal study of a national sample of male twins who served in the military at some time during the Vietnam era (1965–1975). Twins were 51 to 61 years of age at initial assessment wave in 2003 through 2008. Word list was 16 words in four categories from the Second edition of the California Verbal Learning Test (CVLT-2; Delis, Kramer, Kaplan, & Ober, 2000), and prose recall was the Logical Memory subtest from the WMS-III.

The Minnesota Twin Study of Adult Development and Aging (MTSADA; Finkel & McGue, 1993; McGue, Hirsch, & Lykken, 1993) is a population-based sample of same-sex pairs. Assessments were conducted between 1984 and 1994 with age at intake ranging from 26 to 87 years. Prose recall was the Logical Memory subtest from the WMS (Wechsler, 1945) with a maximum score of 24.

The nationally-representative Midlife Development in the United States (MIDUS; Kendler, Thornton, Gilman, & Kessler, 2000) study includes a twin subsample supplemented through snowball recruitment (Brim, Ryff, & Kessler, 2004). Cognitive tests were administered by telephone in 2004 to 2006 and in 2013 to 2014 for individuals with ages

ranging from 34 to 82 years. Word list was the 15-item expanded RAVLT.

The National Academy of Sciences-National Research Council Twin Registry (NAS-NRC; Gatz et al., 2019) is a comprehensive registry of approximately 16,000 White male twin pairs born between 1917 and 1927 with both twins having served in the military. Cognitive testing was conducted by telephone between 1990 and 2002 (Plassman et al., 2006). Word list was the immediate recall of 10 words from the modified TICS (TICSm; Welsh, Breitner, & Magruder-Habib, 1993).

## 2.2. Harmonization approach

To maximize comparability across all IGEMS studies, we created harmonized T-scores for the first administration of word list and prose recall. First, consistent with prior work, we converted scores into POMP (percentage of maximum possible) values to retain raw score information and provide consistent scaling and adequate range of variances across measures (Cohen, Cohen, Aiken, & West, 1999; Reynolds, Gatz, & Pedersen, 2002). To create the standardized T-score variable within each study, we identified a robust normative sample. Here, we took the first trial of the first administration of the test for individuals who were aged 65 to <70 years at that time. If a study had a measure of cognitive impairment (e.g., dementia diagnosis, cognitive screening), individuals considered cognitively impaired were excluded from the standardization sample. To meet the 65 to <70 age range criterion, the Swedish studies were harmonized together using SATSA as the standardization sample, and the Danish studies were harmonized together using MADT as the standardization sample. Otherwise, each study was harmonized individually. For VETSA, we included everyone new to the study in wave 2 who met the age range. Within the standardization sample for each study, mean T-score = 50 and standard deviation = 10, thus equating for difficulty of the different memory tests used in different studies. To assign scores to the full sample, we applied the T-score values to scores on word list and/or prose recall within each respective study.

Attained education was harmonized across IGEMS using the International Standard Classification of Education (ISCED; United Nations Educational, Scientific, and Cultural Organization, 2011).

## 2.3. Data analysis

Mixed effects regression analysis was used to investigate whether predicted verbal episodic memory scores change linearly or nonlinearly as a function of age at testing with sex covaried and with and without adjusting for attained education. Age in pooled analyses was centered at 60, and age in the individual study analyses included as supplementary material was centered on each study's sample mean age. Model estimation was performed in the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in R 3.6.1 (R Core Team, 2019).

Estimation of the genetic and environmental variances underlying both measures of verbal episodic memory was a two-step process. First, we estimated MZ and DZ twin correlations of word list and prose recall within each study and in the pooled sample. We examined correlations by age group and then across the entire age range and then used Falconer's formulae as an initial pass to assess the contributions of genetic and environmental influences underlying the twin correlations (Falconer & Mackay, 1998).

$$h^2 = 2*(r_{MZ} - r_{DZ})$$

$$c^2 = 2*r_{DZ} - r_{MZ}$$

$$e^2 = 1 - r_{MZ}$$

where  $h^2$  represents heritability,  $r_{MZ}$  and  $r_{DZ}$  represent twin correlations for MZ and DZ twins respectively,  $c^2$  represents environmental variance in common to both twins in a pair, and  $e^2$  represents environmental

variance unique to each twin in a pair.

Second, to test for whether age moderates heritability, we then estimated univariate gene-by-environment interaction models (Van der Sluis, Posthuma, & Dolan, 2012) in which twins' age at testing was treated as a within-family moderator in *Mplus* 8.8 (Muthén & Muthén, 1998-2017). Fig. 1 provides a graphical display of the model used. Age in these pooled analyses was centered at 60. Phenotypic variance in each verbal episodic memory variable is decomposed into additive genetic variance, shared environmental variance, and nonshared environmental variance. Additive genetic variance, denoted A, comprises the total genetic effect across the genome that makes a randomly selected pair of twins similar in their verbal episodic memory performance. Shared (or common) environmental variance, denoted C, comprises the total nongenetic effect that makes a randomly selected pair of twins similar in their verbal episodic memory performance. Nonshared environmental variance, denoted E, comprises any factor, including measurement error, that makes twins different from one another in their verbal episodic memory performance. In conventional twin (ACE) models that include MZ and DZ twins, additive genetic variance is identified because genetic influences on MZ twins are perfectly correlated whereas genetic influences on DZ twins correlate 0.50, on average. Shared environmental variance is identified under the assumption that some nongenetic factors, like neighborhood or work experience, similarly influence twins regardless of their zygosity. Nonshared environmental variance constitutes unique variance in the model and thus it is uncorrelated across twins. The fixed correlations between the genetic and shared environmental variance components between twin 1 and twin 2 in Fig. 1 meet these modeling assumptions.

The age variable in Fig. 1 is defined as the moderator and consists of a main effect on the verbal episodic memory measure and a moderator of the ACE variances. The ACE variances were moderated by age as follows:

$$Var(A) = (A_0 + A_1Age_i + A_2Age_i^2)^2$$

$$Var(C) = (C_0 + C_1Age_i + C_2Age_i^2)^2$$

$$Var(E) = (E_0 + E_1Age_i + E_2Age_i^2)^2$$

$$Var(T) = Var(A) + Var(C) + Var(E)$$

Proportions of variance across age are constructed by dividing each vector of age-conditioned ACE variances by the vector of age-conditioned total variance estimates. We present the raw ACE variance estimates. As implied by the figure, the ACE components correlate with age of measurement, which are highly correlated within twin pairs. In order to keep the ACE variance estimates as unbiased as possible, effects of twin 1's verbal episodic memory scores are regressed on twin 2's age and age<sup>2</sup> terms (and vice versa for twin 2) (Van der Sluis, Posthuma, & Dolan, 2012). Sex was included as a main effect in the model. Additionally, as IGEMS consists of multiple studies and therefore cohorts, we adjusted for cohort heterogeneity (e.g., country, birth years, specific memory measure, language, and other sample differences) in memory performance due to study membership by including dummy-coded variables to account for variance in memory due to study. Nine dummy coded variables were included for word list whereas four were included for prose recall. Finally, we ran models with and without attained education as a main effect to examine whether attained education influenced age-related findings and present these findings as supplemental material.

Model estimation proceeded as follows. The baseline model was unrestricted and fit all ACE parameters, age moderation effects, and effects of sex and study. In the next step, we tested whether C effects could be removed from the model as shared environmental variance was expected to be negligible given the age range of the sample. Next, we tested whether quadratic effects of age moderation of the ACE

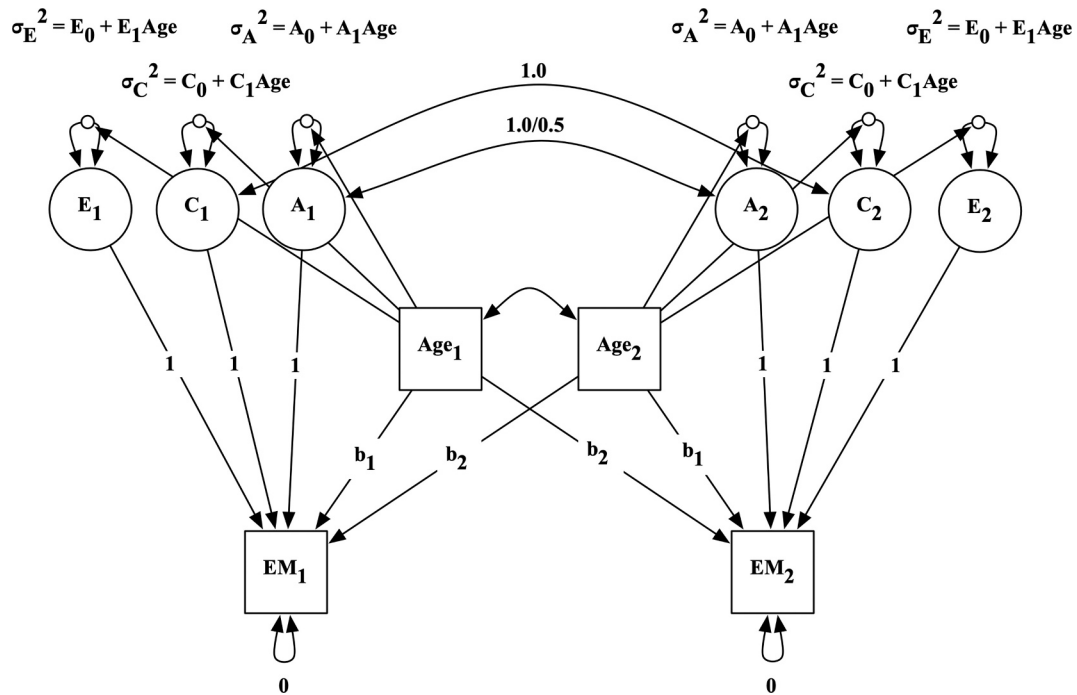


Fig. 1. Gene-by-environment interaction model with age moderation.

Notes. EM = verbal episodic memory. A = additive genetic variance. C = common environmental variance. E = environmental variance that is not shared in common between twins in a pair. Subscripts denote twin1 and twin2 within a pair. In all models, we statistically adjusted for main effects of sex and study.

components could be removed from the model without statistically significant loss of model fit. We then tested whether linear age moderation of the ACE components could be removed from the model. Plots of the raw genetic and environmental variances over the age range from the best fitting model are presented. Lastly, we conducted a series of leave-one-out tests in which we re-estimated the baseline model less one study (e.g., LSADT) to evaluate whether the baseline model estimates were the same when that study was included in the model.

Models were fit using maximum likelihood estimation in *Mplus* 8.8 (Muthén & Muthén, 1998-2017). Competing models are nested and were compared using Likelihood Ratio Tests. Additionally, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) statistics were used to compare models. Models with lower AIC and BIC imply better overall model fit (West, Taylor, & Wu, 2012). Improvement in model fit was accepted if the probability value was below an alpha cut-off value of 0.05.

### 3. Results

#### 3.1. Descriptive results

Mean scores on word list were similar for MZ and DZ twins (49.3 vs. 49.5, respectively) as were mean scores on prose recall (51.2. vs., 50.1, respectively). Means and standard deviations (accounting for twin pair dependence) for each study and the pooled sample are shown in

Table 2  
Mixed effects regression model estimates for word list and prose recall.

	Word List Est (SE), p	Prose Recall Est (SE), p
Total Sample		
Intercept	49.87 (0.083), <0.001	51.15 (0.319), <0.001
Age	-0.53 (0.006), <0.001	-0.20 (0.020), <0.001
Age <sup>2</sup>	-0.03 (0.001), <0.001	-0.01 (0.003), <0.001

Notes. Est = unstandardized regression estimates. SE = standard error. Sex covaried.

Table S1.

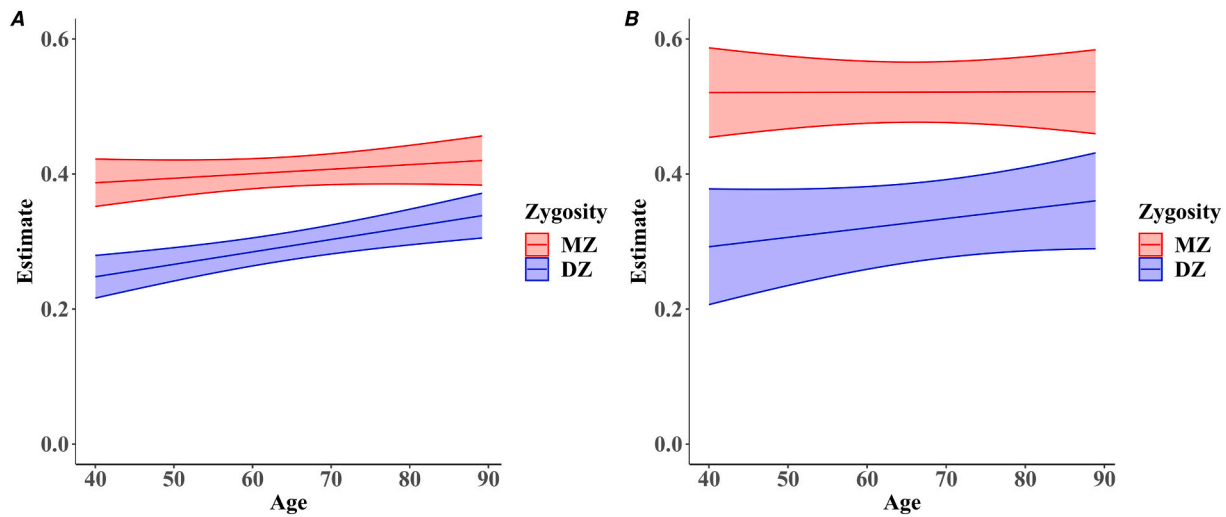
Table 2 shows results of the test for linear and nonlinear age differences with sex covaried, and Table S2 shows results covaried for sex and for sex and attained education in the total sample and by study. When pooled across the IGEMS studies, both word list and prose recall scores declined by age, as indicated in Table 2 by the negative linear and quadratic effects of age. For example, word list was 5.8 T-score units lower at age 75 than age 65. Not all individual studies showed significant negative effects of age (see Table S2), with age effects generally significant for studies with wider age ranges but not for those with narrower age ranges. Age effects were similar regardless of whether or not education was included in the model (see Table S2).

Fig. 2 presents unadjusted univariate twin correlations for word list and prose recall by age for the pooled sample (correlations for the pooled sample and individual studies by age decade are in Table S3). The relative stability of the MZ twin correlations across age suggests a constant effect of the nonshared environment whereas the tendency for the DZ twin correlations to converge toward the MZ correlation suggests that the relative proportion of variance attributed to additive genetic effects declines with increasing age. For word list, across ages, the twin correlation is 0.39 for MZ and 0.28 for DZ twins. Using Falconer's formulae,  $h^2$  is 0.22,  $c^2$  is 0.17, and  $e^2$  is 0.61. For prose recall, the twin correlation is 0.51 for MZ and 0.34 for DZ twins. Using Falconer's formulae,  $h^2$  is 0.34,  $c^2$  is 0.17, and  $e^2$  is 0.49.

#### 3.2. Results from model estimation for age moderation

##### 3.2.1. Word list

Model fit results for the pooled sample on word list are presented in Table 3. Model 2 indicated that all three of the C variance parameters could be removed. Model 3 showed that the quadratic effects of age on additive genetic and on nonshared environmental variance could be removed from the model. Model 4 shows that the linear effects of age on additive genetic and on nonshared environment cannot be removed, but further modeling could not distinguish whether A is moderated by age or E is moderated by age. Thus, model 3 with positive linear effects of age



**Fig. 2.** Twin correlations and 95% confidence intervals by age for word list (panel A) and prose recall (panel B).  
*Notes.* MZ = monozygotic; DZ = dizygotic twins; WL = word list; PR = prose recall. Age based on Twin 1. Unadjusted for other covariates.

**Table 3**  
 Model fit results for word list.

Model	Description	-2LL	Parameters	$\Delta$ -2LL	$\Delta$ df	<i>p</i>	AIC	BIC
1	ACE	-33,947.66	39	-	-	-	67,973.33	68,265.98
2	AE	-33,949.67	36	2.01	3	0.570	67,971.35	68,241.49
3	AE (Quadratic effects removed)	-33,953.79	34	4.11	2	0.128	67,975.57	68,230.70
4	AE (Linear effects removed)	-39,331.87	32	5378.08	2	0.000	78,727.74	78,967.86

*Notes.* A = additive genetic variance; C = shared environmental variance; E = nonshared environmental variance; -2LL = -2 log likelihood;  $\Delta$ -2LL = difference in -2LL between models being compared;  $\Delta$ df = difference in degrees of freedom between models being compared; *p* = probability; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

on both variances in verbal episodic memory was accepted as the best fitting model.

The best fitting model parameter estimates are given in Table 4. Fig. 3A shows raw variances for word list. There is 12% increase in total

variance over age 40 to 89, reflecting the combined effects of linear increases with age in shared genetic and nonshared environmental variance. The contribution of non-shared environmental influences is notably greater than the contribution of additive genetic influences.

**Table 4**  
 Parameter estimates for genetic and environmental components of variance, for word list (left) and prose recall (right).

Word List				Prose Recall			
Parameter	Estimate	SE	0.95 CI	Parameter	Estimate	SE	0.95 CI
A <sub>0</sub>	0.31	0.01	0.28, 0.33	A <sub>0</sub>	0.51	0.03	0.45, 0.57
A <sub>1</sub>	0.02	0.01	0.00, 0.03	A <sub>1</sub>	0.02	0.01	-0.01, 0.04
A <sub>2</sub>	-	-	-	A <sub>2</sub>	-	-	-
C <sub>0</sub>	-	-	-	C <sub>0</sub>	-	-	-
C <sub>1</sub>	-	-	-	C <sub>1</sub>	-	-	-
C <sub>2</sub>	-	-	-	C <sub>2</sub>	-	-	-
E <sub>0</sub>	0.80	0.03	0.78, 0.83	E <sub>0</sub>	0.59	0.03	0.54, 0.63
E <sub>1</sub>	0.01	0.01	-0.00, 0.02	E <sub>1</sub>	0.00	0.01	-0.02, 0.02
E <sub>2</sub>	-	-	-	E <sub>2</sub>	-	-	-
MZ parameters				MZ parameters			
Age <sub>1</sub>	-0.33	0.11	-0.55, -0.11	Age <sub>1</sub>	-0.40	0.44	-1.26, 0.32
Age <sub>1</sub> <sup>2</sup>	-1.67	0.64	-2.92, -0.43	Age <sub>1</sub> <sup>2</sup>	-1.58	1.36	-4.25, 0.67
Age <sub>2</sub>	-0.06	0.11	-0.28, 0.19	Age <sub>2</sub>	0.10	0.44	-0.76, 0.82
Age <sub>2</sub> <sup>2</sup>	0.81	0.64	-0.44, 2.05	Age <sub>2</sub> <sup>2</sup>	0.92	1.36	-1.75, 3.17
Sex <sub>1</sub>	0.32	0.04	0.25, 0.39	Sex <sub>1</sub>	0.18	0.07	0.05, 0.28
Sex <sub>2</sub>	-	-	-	Sex <sub>2</sub>	-	-	-
DZ parameters				DZ parameters			
Age <sub>1</sub>	-0.39	0.10	-0.59, -0.19	Age <sub>1</sub>	-0.12	0.71	-1.51, 1.05
Age <sub>1</sub> <sup>2</sup>	-1.81	0.55	-2.90, -0.73	Age <sub>1</sub> <sup>2</sup>	-4.59	1.89	-8.29, -1.49
Age <sub>2</sub>	-0.00	0.10	-0.20, 0.20	Age <sub>2</sub>	-0.13	0.71	-1.52, 1.04
Age <sub>2</sub> <sup>2</sup>	0.81	0.55	-0.27, 1.89	Age <sub>2</sub> <sup>2</sup>	3.69	1.89	0.17, 7.41
Sex <sub>1</sub>	0.34	0.02	0.29, 0.38	Sex <sub>1</sub>	0.15	0.08	0.00, 0.28
Sex <sub>2</sub>	-0.04	0.02	-0.09, -0.00	Sex <sub>2</sub>	0.02	0.08	-0.13, 0.15
WL <sub>int</sub>	5.44	0.08	5.29, 5.59	PR <sub>int</sub>	5.30	0.11	5.08, 5.48

*Notes.* All estimates adjusted for study effects. For A, C, and E, 0 subscripts indicate intercept, 1 subscripts indicate linear slope, 2 subscripts indicate quadratic trend. For age and sex, 1 subscripts indicate Twin 1, 2 subscripts indicate Twin 2. The intercept represents the parameter estimate at age 60.

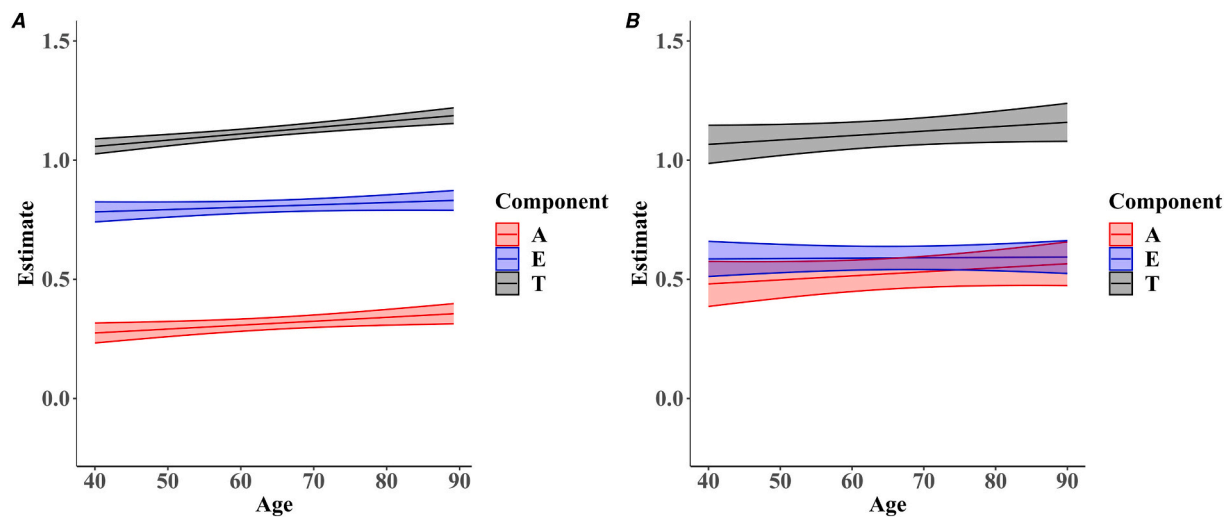


Fig. 3. Raw genetic and environmental variance estimates for word list (panel A) and prose recall (panel B).

Notes. A = additive genetic variance, E = nonshared environmental variance, T = total variance. Sex and study covaried.

Results were similar with educational attainment included in the model (see Tables S4 and S5).

### 3.2.2. Prose recall

Model fit results from the pooled sample for prose recall are presented in Table 5. Model 2 indicated that all three of the C variance parameters could be removed from the model. Model 3 shows that the quadratic effect of age on additive genetic variance components and non-shared environmental components could be removed, while model 4 shows that the linear effects of age on the additive genetic and non-shared environmental variance components could not be removed from the model. Further modeling again could not distinguish whether A is moderated by age or E is moderated by age. Thus, model 3 with positive linear effects of age on both variances in verbal episodic memory is the best fitting model.

The best fitting model parameter estimates are given in Table 5 and plotted in Fig. 3B. There is a 9% increase in total variance over the age range, reflecting the combined effects of linear increases with age in shared genetic and nonshared environmental variance. At the oldest ages, additive genetic and non-shared environmental sources contribute similarly to total variance. Results were similar with educational attainment included in the model (see Tables S5 and S6).

### 3.3. Leave-one-out analysis

Finally, we examined whether the baseline model fitting results were influenced by any one study. For word list, the leave-one-out analysis (see Table S7) suggested similar baseline results when each study was removed from the overall analysis. Additionally, with educational attainment covaried (see Table S8), there was stronger evidence of an interaction between the nonshared environment and age, although results were inconsistent across studies. When educational attainment was

not covaried, there was some evidence of linear shared environmental effects, which suggests that twins' similarity in education accounted for nongenetic variance in word list that twins share. Similarly, for prose recall, the leave-one-out analysis (see Table S9) also suggested that baseline estimates were similar regardless of which study was removed from the analysis. With educational attainment covaried (see Table S10), results were virtually the same except that linear effects of the shared environment were present when OCTO-Twin was removed. Again, the difference suggests that twins' similarity in educational attainment accounts for nongenetic similarity in their prose recall performance.

## 4. Discussion

Current findings provided modest evidence of age moderation of two measures of verbal episodic memory across middle and older adulthood, reflected in an increase in variance underlying immediate recall of both word list and story prose learning measures. The overall increase in variance was driven by a combination of increases in additive genetic and unique environmental parameters that were not independently statistically significant. Thus, we could not parse whether the increase in variance underlying verbal episodic memory was attributed to additive genetic variance, nonshared environmental variance, or both.

Nevertheless, across ages, estimated heritability for word list was 0.22 whereas for prose recall it was 0.40. These heritability estimates are consistent with those published for word list and prose recall (Giubilei et al., 2008; Kremen et al., 2014; Lindgren et al., 2019; McGue & Christensen, 2001; Swan et al., 1999; Wilson, Barral, et al., 2011). Prior studies show greater heritability of prose recall than word list tasks administered to the same participants within the same study (e.g., Kremen et al., 2014; Swan et al., 1999). Although all neuropsychological tests are multidetermined, the verbal episodic memory measures in the

Table 5  
Model fit results for prose recall.

Model	Description	-2LL	Parameters	$\Delta$ -2LL	$\Delta$ df	p	AIC	BIC
1	ACE	-4619	29	-	-		9296.00	9455.00
2	AE	-4621.31	26	2.31	3	0.510	9294.63	9437.18
3	AE (Quadratic effects removed)	-4622.48	24	1.17	2	0.558	9292.96	9424.55
4	AE (Linear effects removed)	-5095.82	22	473.34	2	0.000	10,235.64	10,356.26

Notes. A = additive genetic variance; C = shared environmental variance; E = nonshared environmental variance; -2LL = log likelihood;  $\Delta$ -2LL = difference in -2LL between models being compared;  $\Delta$ df = difference in degrees of freedom between models being compared; p = probability; AIC = Akaike Information Criterion; BIC=Bayesian Information Criterion.

current study are not identical and may assess independent facets of memory functioning. For example, there are differing task demands in processing a story versus a list of words. Moreover, the word list learning measure included in the current study is the first trial and may have a learning component that prose recall does not. How those different task demands would be related to differences in the extent of genetic influences on individual differences remains an interesting, but unresolved, question.

Twin correlations for word list show that similarity between twins in both MZ and DZ twin pairs increased over the age range studied, more evidently for DZ than for MZ pairs. For prose recall, similarity between twins in MZ pairs did not show a strong age trend while similarity between DZ pairs increased over age. Yet, after taking into account effects of sex and study (as well as educational attainment), heritability only changed by about 0.05 units across the entire age range suggesting relative stability from age 40 to 90. Thus, although we found that both genetic and environmental variance components contributed to greater total variability in verbal episodic memory at older ages, additive genetic variance may contribute to slight increases in heritability at older ages.

Heritability of general cognitive ability has been found to be higher in older age compared to earlier in adulthood (e.g., Pedersen, Plomin, Nesselrode, & McClearn, 1992). However, different patterns of aging have been reported across different cognitive domains (e.g., McGue & Christensen, 2013; Pahlen et al., 2018; Reynolds & Finkel, 2015), which would be obscured if only measuring general cognitive ability. Generally, heritability for specific domains, particularly non-verbal memory, has been lower than for general cognitive ability (Pedersen et al., 1992). Moreover, age differences in additive genetic variance differ across cognitive domains. For example, a previous study using a subsample of the analytic sample in the current study showed that genetic variance for all but verbal and memory measures declined with age (Finkel, Pedersen, Plomin, & McClearn, 1998).

As we fit models to raw variances, we can compare to previous analyses of IGEMS cognitive data that have considered similar pooled models for other cognitive domains (Gustavson et al., 2021; Pahlen et al., 2018). For word list, we found greater environmental variance than genetic variance. Pahlen et al. (2018) similarly found greater environmental than genetic influences for digit span, which is generally considered to be a measure of attention and concentration. This similarity in findings fits with the observation of Gavett and Horwitz (2012) who suggested that the immediate recall of a word list, in particular, reflects both short term memory and attention processes, rather than purely verbal episodic memory. For prose recall, we found more equal contributions from genetic and environmental sources. In comparison, Gustavson et al. (2021) reported equal magnitudes of genetic and environmental influences for verbal fluency, whereas Pahlen et al. (2018) reported greater genetic than environmental influences for verbal ability and processing speed.

Differences in age moderation have also been reported across cognitive domains in IGEMS. Pahlen et al. (2018) found significant age moderation of genetic effects for digits forward, digits backwards, block design, and symbol digit, but found no significant age moderation for vocabulary, a measure considered indicative of crystallized intelligence (Horn & Cattell, 1966, 1967). Gustavson et al. (2021) also found no significant age moderation for verbal fluency. We note that memory abilities, where we report age moderation of overall variance, are often found to be distinct from other cognitive abilities. For example, in prior exploratory factor analyses of cognitive abilities in IGEMS studies, memory did not significantly load on a measure of general cognitive ability (Beam et al., 2022; see also Horn & Cattell, 1966).

The current findings have implications for late life cognitive development. First, phenotypically, for both measures, verbal episodic memory scores were lower at higher ages, as would be expected (e.g., Friedman et al., 2007). Age differences in verbal episodic memory were not explained by attained education. Second, for both measures of

verbal episodic memory, total variance increased across age, which is the summation of additive genetic variance and non-shared environmental variance. The best fitting models, however, did not clarify whether age moderated additive genetic, nonshared environmental, or both variance components. As the twin correlations suggest that DZ correlations tend to converge toward MZ correlations with increasing age, it is likely that there is a gene-by-environment transaction process that unfolds with age.

A lifespan developmental perspective might lead to the expectation that environmental influences unique to people are greater at older ages than younger ages (e.g., Baltes et al., 2006). Indeed, we saw greater total variance at older ages. These increased differences would include not only unique environmental influences but also influences resulting from gene by environment interplay, which would be encompassed in estimates of genetic influences. Alternatively, given the tendency for DZ twins correlations to converge toward MZ correlations, unmodeled gene-shared environment correlation could also account for the increase in total variance.

There are important study limitations. First, pooling cognitive measures across studies may obscure important between-study differences. However, looking at each study individually confounds differing age ranges with substantively relevant between-study differences, and mainly results in loss of age differences. Our solution was to harmonize by creating T-scores within each study prior to pooling, to include study as a covariate, and also to conduct leave-one-out analyses to assure that one study was not unduly influencing the pooled sample results. Second, our harmonized measures, while providing a range of scores for examining individual differences, nonetheless suffer from a need to rely on available measures from each verbal episodic memory instrument used across participating studies. Thus, there was no measure of delayed recall, and word list learning used only the first presentation of the list, so our conclusions about episodic memory are limited to those based on results from two specific measures. Third, few studies had both measures, and different groups of studies chose to administer word list versus prose recall. Thus, the comparison of the two measures is confounded with which studies chose which measure. Fourth, since nonshared environmental effects are confounded with measurement error in classic ACE models, differences in the reliability of the measures could contribute to the observed difference in nonshared environmental variance and estimated heritability for the two memory tasks. Moreover, potential age effects on the reliability of these measures may also be confounded with age effects on the performance in these tasks. The literature suggests test-retest reliability is higher for prose recall (Lo, Humphreys, Byrne, & Pachana, 2012) than for word list (Alioto et al., 2017; Vlahou et al., 2013; see Calamia, Markon, & Tranel, 2013). Fifth, IGEMS studies, although from different parts of the world, are all from relatively high-income countries, with limited ethnic diversity, which may limit the generalizability of the findings. Sixth, data were cross-sectional. To help address this limitation, we used both the pooled sample covarying for study and analyses of individual studies to examine consistency of results across studies that consist of individuals from different countries with varying age ranges and representing different birth cohorts. We also controlled for a key cohort difference by including attained education as a covariate in supplemental analyses. Results were largely consistent in the pooled sample, with some indication that education may account for some of the shared environment effects.

In summary, we conclude that, using a cross-sectional design pooling across studies to achieve a large number of participants, individual variation in verbal episodic memory performance is greater at older ages. We conclude that genetic influences are of continuing or possibly slightly increasing relevance to memory performance in older adults. At the same time, accumulated environmental exposures unique to the individual remain influential. In contrast to reported results for other cognitive domains, differences in environmental exposures are comparatively important for verbal episodic memory, especially word



list learning.

## Funding

This work was supported by the National Institutes of Health, National Institute on Aging (NIH Grant Nos. R01 AG060470 to MG and NLP, and RF1 AG058068 to Pike, LaDu, and MG) and the Alzheimer's Association (AARF-17-505302 to CRB).

## Declaration of Competing Interest

None.

## Data availability

Each study of the consortium has different sharing rules, no data were collected specifically for this study.

## Acknowledgements

We thank the participating studies. SATSA was supported by grants R01 AG04563, R01 AG10175, the John D. and Catherine T. MacArthur Foundation Research Network on Successful Aging, the Swedish Council For Working Life and Social Research (FAS) (97:0147:1B, 2009-0795) and Swedish Research Council (825-2007-7460, 825-2009-6141), to NLP. GENDER was supported by the MacArthur Foundation Research Network on Successful Aging (to McClearn), The Axel and Margaret Ax:son Johnson's Foundation, The Swedish Council for Social Research, and the Swedish Foundation for Health Care Sciences and Allergy Research, to Malmberg. OCTO-Twin was supported by grant R01 AG08861. Gender was supported by the MacArthur Foundation Research Network on Successful Aging, The Axel and Margaret Ax:son Johnson's Foundation, The Swedish Council for Social Research, and the Swedish Foundation for Health Care Sciences and Allergy Research. The Danish Twin Registry was supported by grants from The National Program for Research Infrastructure 2007 from the Danish Agency for Science and Innovation, the Velux Foundation and the US National Institute of Health (P01 AG08761), to Vaupel and KC. The Minnesota Twin Study of Adult Development and Aging (MTSADA) was supported by NIA grant R01 AG06886. VETSA was supported by National Institute of Health grants NIA R01 AG050595, R01 AG022381, and R01 AG076838 to WSK and CEF. The Cooperative Studies Program of the Office of Research & Development of the United States Department of Veterans Affairs has provided financial support for the development and maintenance of the Vietnam Era Twin (VET) Registry. The MIDUS study was supported by the John D. and Catherine T. MacArthur Foundation Research Network on Successful Midlife Development and by National Institute on Aging Grant AG20166 (Ryff). The NAS-NRC Twin Registry received funding from NIA grants R01 AG08549 and R21 AG039572. We acknowledge the contribution of the OATS research team (<https://cheba.unsw.edu.au/project/older-australian-twins-study>) to this study. The OATS study has been funded by a National Health & Medical Research Council (NHMRC) and Australian Research Council (ARC) Strategic Award Grant of the Ageing Well, Ageing Productively Program (ID No. 401162); NHMRC Project (seed) Grants (ID No. 1024224 and 1025243); NHMRC Project Grants (ID No. 1045325 and 1085606); and NHMRC Program Grants (ID No. 568969 and 1093083) to PS. OATS participant recruitment was facilitated through Twins Research Australia, a national resource in part supported by a Centre for Research Excellence Grant (ID: 1079102), from the National Health and Medical Research Council. The Carolina African American Twin Study of Aging (CAATSA) was funded by NIA grant R01 AG13662 to KEW. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIA/NIH, or the VA.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.intell.2023.101759>.

## References

- Alioto, A. G., Kramer, J. H., Borish, S., Neuhaus, J., Saloner, R., Wynn, M., & Foley, J. M. (2017). Long-term test-retest reliability of the California Verbal Learning Test - second edition. *Clinical Neuropsychology*, 31(8), 1449–1458. <https://doi.org/10.1080/13854046.2017.1310300>
- Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (2006). Life span theory in developmental psychology. In R. M. Lerner, & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 569–664). John Wiley & Sons Inc.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beam, C. R., Luczak, S. E., Reynolds, C. A., Panizzon, M. S., ... Gatz, M., & the IGEMS consortium. (2022). Estimation of a latent dementia index in 10 genetically informed studies. *Journal of Alzheimer's Disease*, 90, 1187–1201. <https://doi.org/10.3233/JAD-220472>.
- Berg, S. (1980). Psychological functioning in 70- and 75-year-old people: A study in an industrialized city. *Acta Psychiatrica Scandinavica*, 62(Suppl), 288.
- Brandt, J., Spencer, M., & Folstein, M. (1988). The Telephone Interview for Cognitive Status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 1(2), 111–117.
- Brim, O. G., Ryff, C. D., & Kessler, R. C. (2004). The MIDUS national survey: An overview. In O. G. Brim, C. D. Ryff, & R. C. Kessler (Eds.), *How healthy are we? A national study of well-being at midlife* (pp. 1–36). University of Chicago Press.
- Calamia, M., Markon, K., & Tranel, D. (2013). The robust reliability of neuropsychological measures: Meta-analyses of test-retest correlations. *The Clinical Neuropsychologist*, 27(7), 1077–1105. <https://doi.org/10.1080/13854046.2013.809795>
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, 34, 315–346. [https://doi.org/10.1207/S15327906MBR3403\\_2](https://doi.org/10.1207/S15327906MBR3403_2)
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test* (2nd ed.). The Psychological Corporation.
- Falconer, D. S., & Mackay, T. F. (1998). *Introduction to quantitative genetics* (4th ed.). Essex, England: Longman Group, Ltd.
- Finkel, D., & McGue, M. (1993). The origins of individual differences in memory among the elderly: A behavior genetic analysis. *Psychology and Aging*, 8(4), 527–537. <https://doi.org/10.1037/0882-7974.8.4.527>
- Finkel, D., Pedersen, N. L., Plomin, R., & McClearn, G. E. (1998). Longitudinal and cross-sectional twin data on cognitive abilities in adulthood: The Swedish Adoption/Twin Study of Aging. *Developmental Psychology*, 34(6), 1400. <https://doi.org/10.1037/0012-1649.34.6.1400>
- Friedman, D., Nessler, C., & Johnson, R. (2007). Memory encoding and retrieval in the aging brain. *Clinical EEG and Neuroscience*, 38(1), 2–7. <https://doi.org/10.1177/155005940703800105>
- Gatz, M., Plassman, B. L., Tanner, C. M., Goldman, S. M., Swan, G. E., Chanti-Ketterl, M., ... Butler, D. A. (2019). The NAS-NRC Twin Registry and Duke Twins Study of Memory in Aging: An update. *Twin Research and Human Genetics*, 22(6), 757–760. <https://doi.org/10.1017/thg.2019.45>
- Gavett, B. E., & Horwitz, J. E. (2012). Immediate list recall as a measure of short-term episodic memory: Insights from the serial position effect and item response theory. *Archives of Clinical Neuropsychology*, 27(2), 125–135. <https://doi.org/10.1093/arclin/acr104>
- Giles, G. G., & English, D. R. (2002). The Melbourne collaborative cohort study. *IARC Scientific Publications*, 156, 69–70.
- Giubilei, F., Medda, E., Fagnani, C., Bianchi, V., De Carolis, A., Salvetti, M., ... Stazi, M. A. (2008). Heritability of neurocognitive functioning in the elderly: Evidence from an Italian twin study. *Age and Ageing*, 37(6), 640–646. <https://doi.org/10.1093/ageing/afn132>
- Gold, C. H., Malmberg, B., McClearn, G. E., Pedersen, N. L., & Berg, S. (2002). Gender and health: A study of older unlike-sex twins. *The Journals of Gerontology: Series B*, 57(3), 168–176. <https://doi.org/10.1093/geronb/57.3.S168>
- Gustavson, D. E., Panizzon, M. S., Kremen, W. S., Reynolds, C. A., Pahlen, S., Nygaard, M., ... IGEMS Consortium. (2021). Genetic and environmental influences on semantic verbal fluency across midlife and later life. *Behavior Genetics*, 51(2), 99–109. <https://doi.org/10.1007/s10519-021-10048-w>
- Horn, J. L., & Cattell, R. B. (1966). Age differences in primary mental ability factors. *Journal of Gerontology*, 21(2), 210–220. <https://doi.org/10.1093/geronj/21.2.210>
- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26(2), 107–129. [https://doi.org/10.1016/0001-6918\(67\)90011-X](https://doi.org/10.1016/0001-6918(67)90011-X)
- Johnson, W., Deary, I. J., & Iacono, W. G. (2009). Genetic and environmental transactions underlying educational attainment. *Intelligence*, 37(5), 466–478. <https://doi.org/10.1016/j.intell.2009.05.006>
- Kendler, K. S., Thornton, L. M., Gilman, S. E., & Kessler, R. C. (2000). Sexual orientation in a U.S. national sample of twin and nontwin sibling pairs. *American Journal of Psychiatry*, 157(11), 1843–1846. <https://doi.org/10.1176/appi.ajp.157.11.1843>
- Kremen, W. S., Franz, C. E., & Lyons, M. J. (2013). VETSA: The Vietnam Era Twin Study of Aging. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, 16(1), 399–402. <https://doi.org/10.1017/thg.2012.86>

- Kremen, W. S., Panizzon, M. S., Franz, C. E., Spoon, K. M., Vuoksimaa, E., Jacobson, K. C., ... Lyons, M. J. (2014). Genetic complexity of episodic memory: A twin approach to studies of aging. *Psychology and Aging*, 29(2), 404–417. <https://doi.org/10.1037/a0035962>
- Lachman, M. E., Agrigoroaei, S., Murphy, C., & Tun, P. A. (2010). Frequent cognitive activity compensates for education differences in episodic memory. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, 18(1), 4–10. <https://doi.org/10.1097/JGP.0b013e3181ab8b62>
- Lezak, M. D. (2004). *Neuropsychological assessment* (4th ed.). London: Oxford University Press.
- Lichtenstein, P., Sullivan, P. F., Cnattingius, S., Gatz, M., Johansson, S., Carlström, E., ... Pedersen, N. L. (2006). The Swedish Twin Registry in the third millennium: An update. *Twin Research and Human Genetics*, 9(6), 875–882. <https://doi.org/10.1375/twin.9.6.875>
- Lindgren, N., Kaprio, J., Rinne, J. O., & Vuoksimaa, E. (2019). Immediate verbal recall and familial dementia risk: Population-based study of over 4000 twins. *Journal of Neurology, Neurosurgery, and Psychiatry*, 90(1), 90–97. <https://doi.org/10.1136/jnnp-2018-319122>
- Lo, A. H., Humphreys, M., Byrne, G. J., & Pachana, N. A. (2012). Test-retest reliability and practice effects of the Wechsler Memory Scale-III. *Journal of Neuropsychology*, 6(2), 212–231. <https://doi.org/10.1111/j.1748-6653.2011.02023.x>
- McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Pettrill, S. A., & Plomin, R. (1997). Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science (New York, N.Y.)* (Vol. 276,(5318)), 1560–1563. <https://doi.org/10.1126/science.276.5318.1560>
- McGue, M., & Christensen, K. (2001). The heritability of cognitive functioning in very old adults: Evidence from Danish twins aged 75 years and older. *Psychology and Aging*, 16(2), 272–280. <https://doi.org/10.1037/0882-7974.16.2.272>
- McGue, M., & Christensen, K. (2007). Social activity and healthy aging: A study of aging Danish twins. *Twin Research and Human Genetics*, 10(2), 255–265. <https://doi.org/10.1375/twin.10.2.255>
- McGue, M., & Christensen, K. (2013). Growing old but not growing apart: Twin similarity in the latter half of the lifespan. *Behavior Genetics*, 43(1), 1–12. <https://doi.org/10.1007/s10519-012-9559-5>
- McGue, M., Hirsch, B., & Lykken, D. T. (1993). Age and the self-perception of ability: A twin study analysis. *Psychology and Aging*, 8(1), 72–80. <https://doi.org/10.1037/0882-7974.8.1.72>
- Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., ... Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD): I. clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39, 1159–1165. <https://doi.org/10.1212/WNL.39.9.1159>
- Murphy, K. J., West, R., Armilio, M. L., Craik, F. I., & Stuss, D. T. (2007). Word-list-learning performance in younger and older adults: Intra-individual performance variability and false memory. *Neuropsychology, Development, and Cognition. Section B Aging, Neuropsychology and Cognition*, 14(1), 70–94. <https://doi.org/10.1080/138255890969726>
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide*, v. 8.8.
- Nyberg, L. (2017). Functional brain imaging of episodic memory decline in ageing. *Journal of Internal Medicine*, 281(1), 65–74. <https://doi.org/10.1111/joim.12533>
- Pahlen, S., Hamdi, N. R., Dahl Aslan, A. K., Horwitz, B. N., Panizzon, M. S., Petersen, I., Zavala, C., Christensen, K., Finkel, D., Franz, C. E., Gatz, M., Johnson, W., Kremen, W. S., Krueger, R. F., Neiderhiser, J. M., Reynolds, C. A., Pedersen, N. L., & McGue, M. (2018). Age-moderation of genetic and environmental contributions to cognitive functioning in mid- and late-life for specific cognitive abilities. *Intelligence*, 68, 70–81. <https://doi.org/10.1016/j.intell.2017.12.004>
- Panizzon, M. S., Lyons, M. J., Jacobson, K. C., Franz, C. E., Grant, M. D., Eise, S. A., ... Kremen, W. S. (2011). Genetic architecture of learning and delayed recall: A twin study of episodic memory. *Neuropsychology*, 25(4), 488–498. <https://doi.org/10.1037/a0022569>
- Papenberg, G., Lindenberger, U., & Bäckman, L. (2015). Aging-related magnification of genetic effects on cognitive and brain integrity. *Trends in Cognitive Sciences*, 19(9), 506–514. <https://doi.org/10.1016/j.tics.2015.06.008>
- Pedersen, D. A., Larsen, L. A., Nygaard, M., Mengel-From, J., McGue, M., Dalgård, C., ... Christensen, K. (2019). The Danish Twin Registry: An updated overview. *Twin Research and Human Genetics*, 22, 499–507. <https://doi.org/10.1017/thg.2019.72>
- Pedersen, N. L., Christensen, K., Dahl, A. K., Finkel, D., Franz, C. E., Gatz, M., ... Reynolds, C. A. (2013). IGEMS: The consortium on Interplay of Genes and Environment across Multiple Studies. *Twin Research and Human Genetics*, 16(1), 481–489. <https://doi.org/10.1017/thg.2012.110>
- Pedersen, N. L., Gatz, M., Finch, B. K., Finkel, D., Butler, D. A., Dahl Aslan, A., ... Whitfield, K. E. (2019). IGEMS: The consortium on Interplay of Genes and Environment across Multiple Studies - an update. *Twin Research and Human Genetics*, 22(6), 809–816. <https://doi.org/10.1017/thg.2019.76>
- Pedersen, N. L., McClearn, G. E., Plomin, R., Nesselroade, J. R., Berg, S., & DeFaire, U. (1991). The Swedish Adoption Twin Study of Aging: An update. *Acta Geneticae Medicae et Gemellologiae: Twin Research*, 40(1), 7–20. <https://doi.org/10.1017/S0001566000006681>
- Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science*, 3(6), 346–353. <https://doi.org/10.1111/j.1467-9280.1992.tb00045.x>
- Plassman, B. L., Steffens, D. C., Burke, J. R., Welsh-Bohmer, K. A., Newman, T. N., Drosick, D., ... Breitner, J. C. (2006). Duke Twins Study of Memory in Aging in the NAS-NRC Twin Registry. *Twin Research and Human Genetics*, 9(6), 950–957. <https://doi.org/10.1375/183242706779462381>
- Procopio, F., Zhou, Q., Wang, Z., Gidziela, A., Rimfeld, K., Malanchini, M., & Plomin, R. (2022). The genetics of specific cognitive abilities. *Intelligence*, 95. <https://doi.org/10.1016/j.intell.2022.101689>
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <https://www.R-project.org/>.
- Read, S., Pedersen, N. L., Gatz, M., Berg, S., Vuoksimaa, E., Malmberg, B., ... McClearn, G. E. (2006). Sex differences after all those years? Heritability of cognitive abilities in old age. *Journal of Gerontology: Psychological Sciences*, 61, P137–P143.
- Rey, A. (1964). *L'examen clinique en psychologie*. Presses Universitaires de France.
- Reynolds, C. A., & Finkel, D. (2015). A meta-analysis of heritability of cognitive aging: Minding the “missing heritability” gap. *Neuropsychology Review*, 25(1), 97–112. <https://doi.org/10.1007/s11065-015-9280-2>
- Reynolds, C. A., Gatz, M., & Pedersen, N. L. (2002). Individual variation for cognitive decline: Quantitative methods for describing patterns of change. *Psychological Aging*, 17(2), 271–287. <https://doi.org/10.1037/0882-7974.17.2.271>
- Royall, D. R., & Palmer, R. F. (2012). Getting past “g”: Testing a new model of dementing processes in persons without dementia. *The Journal of Neuropsychiatry and Clinical Neuropsychology*, 24(1), 37–46. <https://doi.org/10.1176/appi.neuropsych.11040078>
- Sachdev, P. S., Lammell, A., Trollor, J. N., Lee, T., Wright, M. J., Ames, D., ... Schofield, P. R. (2009). A comprehensive neuropsychiatric study of elderly twins: The Older Australian Twins Study. *Twin Research and Human Genetics*, 12(6), 573–582. <https://doi.org/10.1375/twin.12.6.573>
- Sachdev, P. S., Lee, T., Wen, W., Ames, D., Batouli, A. H., Bowden, J., ... OATS Research Team. (2013). The contribution of twins to the study of cognitive ageing and dementia: The Older Australian Twins Study. *International Review of Psychiatry (Abingdon, England)*, 25(6), 738–747. <https://doi.org/10.3109/09540261.2013.870137>
- Skytte, A., Christiansen, L., Kyvik, K. O., Bodker, F. L., Hvidberg, L., Petersen, I., ... K. (2013). The Danish twin registry: Linking surveys, national registers, and biological information. *Twin Research and Human Genetics*, 16, 104–111. <https://doi.org/10.1017/thg.2012.77>
- Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., ... Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 280–292. <https://doi.org/10.1016/j.jalz.2011.03.003>
- Storandt, M., Botwinick, J., Danziger, W. L., Berg, L., & Hughes, C. P. (1984). Psychometric differentiation of mild senile dementia of the Alzheimer type. *Archives of Neurology*, 41(5), 497–499. <https://doi.org/10.1001/archneur.1984.04050170043013>
- Swan, G. E., Reed, T., Jack, L. M., Miller, B. L., Markee, T., Wolf, P. A., ... Carmelli, D. (1999). Differential genetic influence for components of memory in aging adult twins. *Archives of Neurology*, 56(9), 1127–1132. <https://doi.org/10.1001/archneur.56.9.1127>
- Turkheimer, E., Beam, C. R., Sundet, J. M., & Tambs, K. (2017). Interaction between parental education and twin correlations for cognitive ability in a Norwegian conscript sample. *Behavior Genetics*, 47(5), 507–515. <https://doi.org/10.1007/s10519-017-9857-z>
- United Nations Educational, Scientific, and Cultural Organization. (2011). International Standard Classification of Education. Retrieved from <https://web.archive.org/web/20130124032233/http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf>
- Vlahou, C. H., Kosmidis, M. H., Dardagani, A., Tsotsi, S., Giannakou, M., Giakoulidou, A., ... Pontikakis, N. (2013). Development of the Greek Verbal Learning Test: Reliability, construct validity, and normative standards. *Archives of Clinical Neuropsychology*, 28(1), 52–64. <https://doi.org/10.1093/arclin/acn099>
- Wechsler, D. (1945). *Manual for the Wechsler Memory Scale*. The Psychological Corporation.
- Wechsler, D. (1987). *Manual for the Wechsler Memory Scale*. rev., WMS-R. The Psychological Corporation.
- Wechsler, D. (1997). *Manual for the Wechsler Memory Scale-III (WMS-III)*. The Psychological Corporation.
- Welsh, K. A., Breitner, J. C., & Magruder-Habib, K. M. (1993). Detection of dementia in the elderly using telephone screening of cognitive status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 6(2), 103–110.
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York, NY: Guilford Press.
- Whitfield, K. E. (2013). A registry of adult African American twins: The Carolina African American Twin Study of Aging. *Twin Research and Human Genetics*, 16(1), 476–480. <https://doi.org/10.1017/thg.2012.79>
- Whitfield, K. E., Brandon, D. T., Wiggins, S., Vogler, G., & McClearn, G. (2003). Does intact pair status matter in the study of African American twins? The Carolina African American Twin Study of Aging. *Experimental Aging Research*, 29(4), 407–423. <https://doi.org/10.1080/03610730303699>
- Wilson, J. A., Zozula, C., & Gove, W. R. (2011). Age, period, cohort and educational attainment: The importance of considering gender. *Social Science Research*, 40(1), 136–149. <https://doi.org/10.1016/j.ssresearch.2010.09.011>
- Wilson, R. S., Barral, S., Lee, J. H., Leurgans, S. E., Foroud, T. M., Sweet, R. A., ... Bennett, D. A. (2011). Heritability of different forms of memory in the late onset

- Alzheimer's disease family study. *Journal of Alzheimer's Disease: JAD*, 23(2), 249–255. <https://doi.org/10.3233/JAD-2010-101515>
- Zaheed, A. B., Sharifian, N., Kraal, A. Z., Sol, K., Manly, J. J., Schupf, N., ... Zahodne, L. B. (2021). Mediators and moderators of the association between perceived stress and episodic memory in diverse older adults. *Journal of the International Neuropsychological Society: JINS*, 27(9), 883–895. <https://doi.org/10.1017/S1355617720001253>
- Van der Sluis, S., Posthuma, D., & Dolan, C. V. (2012). A note on false positives and power in G×E modelling of twin data. *Behavior Genetics*, 42(1), 170–186. <https://doi.org/10.1007/s10519-011-9480-3>