



Genetic and Environmental Influences on Verbal Fluency in Middle Age: A Longitudinal Twin Study

Daniel E. Gustavson¹ · Matthew S. Panizzon¹ · Jeremy A. Elman¹ · Carol E. Franz¹ · Asad Beck^{1,2} · Chandra A. Reynolds³ · Kristen C. Jacobson⁴ · Hong Xian⁵ · Rosemary Toomey⁶ · Michael J. Lyons⁶ · William S. Kremen^{1,7}

Received: 17 January 2018 / Accepted: 11 June 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Mounting evidence suggests that measures of phonemic fluency and semantic fluency are differentially associated with other cognitive and health phenotypes, but few studies have examined their shared and unique variance, especially using genetically-informative designs. In this study, 1464 middle-aged twins completed six fluency subtests at up to two time-points (mean age 56 and 62 years). Confirmatory factor analyses supported a two-factor solution: a General Fluency latent factor explained variation in all six subtests and a Semantic-Specific factor accounted for additional variance in semantic subtests. Both factors were explained primarily by genetic influences at both waves ($a^2=0.57-0.76$). There was considerable stability of individual differences over 6 years ($r=.90$ for General Fluency, $r=.81$ for Semantic-Specific), especially for genetic influences ($r_g=.94$ and 1.0 , respectively). These results suggest that semantic fluency can be viewed as a combination of general and semantic-specific variance, but phonemic fluency is captured entirely by the general factor.

Keywords Word fluency · Category fluency · Longitudinal design · Heritability · Twin study

Edited by Deborah Finkel.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10519-018-9910-6>) contains supplementary material, which is available to authorized users.

✉ Daniel E. Gustavson
dgustavson@ucsd.edu

- ¹ Department of Psychiatry, Center for Behavior Genetics of Aging, University of California, San Diego, 9500 Gilman Dr. (MC 0738), La Jolla, CA 92093, USA
- ² Department of Psychology, San Diego State University, San Diego, CA, USA
- ³ Department of Psychology, University of California, Riverside, Riverside, CA, USA
- ⁴ Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA
- ⁵ Department of Biostatistics, St. Louis University and Clinical Epidemiology Center, Veterans Affairs St. Louis Healthcare System, St. Louis, MO, USA
- ⁶ Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA
- ⁷ Center of Excellence for Stress and Mental Health, Veterans Affairs San Diego Healthcare System, La Jolla, CA, USA

Introduction

Neuropsychological measures of verbal fluency have been widely used in studies of cognitive aging. Verbal fluency is impaired in Alzheimer's disease, depression, and many other clinically relevant domains (for meta-analyses, see Henry and Crawford 2004a, b, 2005a, b; Henry et al. 2004). The impairments are observed regardless of whether fluency is assessed using phonemic fluency (i.e., naming words that start with a cue, such as words that begin with *F*) or semantic fluency (i.e., naming words from a given category, such as types of animals). Phonemic and semantic measures are sometimes combined into a single fluency score, but there is substantial evidence that these measures are differentially associated with neuropsychiatric conditions and other cognitive functions (Henry and Crawford 2004b; Henry et al. 2004; Stolwyk et al. 2015; van den Berg et al. 2017; Whiteside et al. 2016). For example, both types of fluency abilities are associated with vocabulary and working memory updating, but only semantic fluency is associated with lexical access speed (Shao et al. 2014). Compared with phonemic fluency, semantic fluency is more strongly impaired in Alzheimer's disease (for meta analysis, see Henry et al.

2004), and similar results were observed in a meta-analysis of schizophrenia, although the disparity was smaller (Henry and Crawford 2005b).

Although it is clear that phonemic and semantic fluency share considerable covariance (Hedden and Yoon 2006; Shao et al. 2014; Unsworth et al. 2011; Whiteside et al. 2016), a remaining question concerns whether there are unique variance components that differentiate phonemic from semantic fluency. For example, a meta-analysis of lesion studies suggested that frontal lesions are associated with similar deficits in both phonemic and semantic fluency, but that damage in temporal regions is more strongly associated with semantic fluency than phonemic fluency (Henry and Crawford 2004a). These findings have led to the suggestion that phonemic fluency relies on frontally-mediated strategic search processes, whereas semantic fluency relies on both frontally-mediated search and temporally-mediated associative processes (Henry and Crawford 2004a; Unsworth et al. 2011). Similarly, there is evidence that the number of words generated during phonemic fluency are driven by effective switching between clusters (e.g., *fright/fight/flight* to *flat/fat*), whereas semantic fluency is driven more equally by switching between clusters and generating many words within each cluster (Troyer and Moscovitch 2006).

Together, these results suggest that there is unique variance in semantic fluency above and beyond its common variance with phonemic fluency. It has also been proposed that phonemic fluency places greater demands on the executive function processes involved in frontally-mediated strategic search (Moscovitch 1994). However, it is not clear whether this reflects unique variance in phonemic fluency or just a greater degree of strategic search demand relative to semantic fluency.

Thus, it will be useful to quantify more directly whether there are any unique phonemic-specific or semantic-specific variance components in addition to their common variance. In some studies, measures of phonemic and semantic fluency were combined onto a single fluency factor (Hedden and Yoon 2006; Unsworth et al. 2011) or loaded onto the same factor (Lee et al. 2018; Whiteside et al. 2016), which is not surprising when fluency measures are included in a larger set of cognitive measures. However, it will be important to also hone in on variance solely among different fluency tests to examine whether other unique variance components can be isolated, especially if they may be particularly relevant to neuropathology and aging.

The first goal of the current study was to test five different possible models regarding the covariance among measures of phonemic and semantic fluency. These candidate models are displayed in Fig. 1. They include a model with separate but correlated phonemic and semantic factors (Fig. 1a), a model with a general factor and two specific fluency factors (Fig. 1b), models with a general factor and only one of the

two specific factors (Fig. 1c, d), and a model with only a general factor (Fig. 1e). We hypothesized that there would be a general factor and at least one specific factor given the evidence for unique variance in semantic fluency (Fig. 1c), though it was unclear whether we would also observe an additional phonemic-specific factor (Fig. 1b).

Genetic and environmental influences on verbal fluency

Because cognition is strongly heritable, a more thorough understanding of neuropsychological function should ultimately include a fuller understanding of its genetic underpinnings (Kremen et al. 2016). Moreover, given the rapid advances in gene discovery and other genetic studies, the integration of neuropsychology and genetics has become increasingly important (Kremen et al. 2016). For example, there is not necessarily a one-to-one correspondence between the phenotypic factor structure and the underlying genetic/environmental structures (Kremen et al. 2009; Vasilopoulos et al. 2012). Our work has shown that even within a cognitive domain, different tests or factors can have unique sets of genetic influences (Gustavson et al. 2018a; Panizon et al. 2014, 2015). Thus, in addition to identifying the best-fitting model that accounts for individual differences in phonemic and semantic fluency, it will be useful to quantify the extent to which genetic and environmental influences account for these sources of variance.

Despite the wide use of measures of verbal fluency, little is known about the genetic and environmental architecture of verbal fluency (Bratko 1996; Giubilei et al. 2008; Lee et al. 2012; McGue and Christensen 2001; Swan and Carmelli 2002). One study of 472 older Australian twins ($M=71$ years) reported a heritability estimate of $a^2=0.63$ for a combined measure of three phonemic fluency subtests (Lee et al. 2012). In other words, about 63% of the variance could be explained by genetic influences. These estimates were similar in smaller studies of Italian and Croatian twins ($a^2=0.52-0.62$; Bratko 1996; Giubilei et al. 2008). Moreover, the heritability of semantic fluency has been reported as 0.54 and 0.37 in older adults, mean age 68 and 80 respectively (Giubilei et al. 2008; McGue and Christensen 2001). To our knowledge, only one study has quantified the shared genetic and unique genetic/environmental variance between phonemic and semantic fluency (Lee et al. 2018); the genetic correlation was only $r_g=.28$, indicating substantial non-shared genetic influences between phonemic and semantic fluency.

Thus, the second goal of the current study was to quantify the extent to which genetic and environmental influences account for the variance components identified in the first step. To do so, the candidate models of fluency from Fig. 1

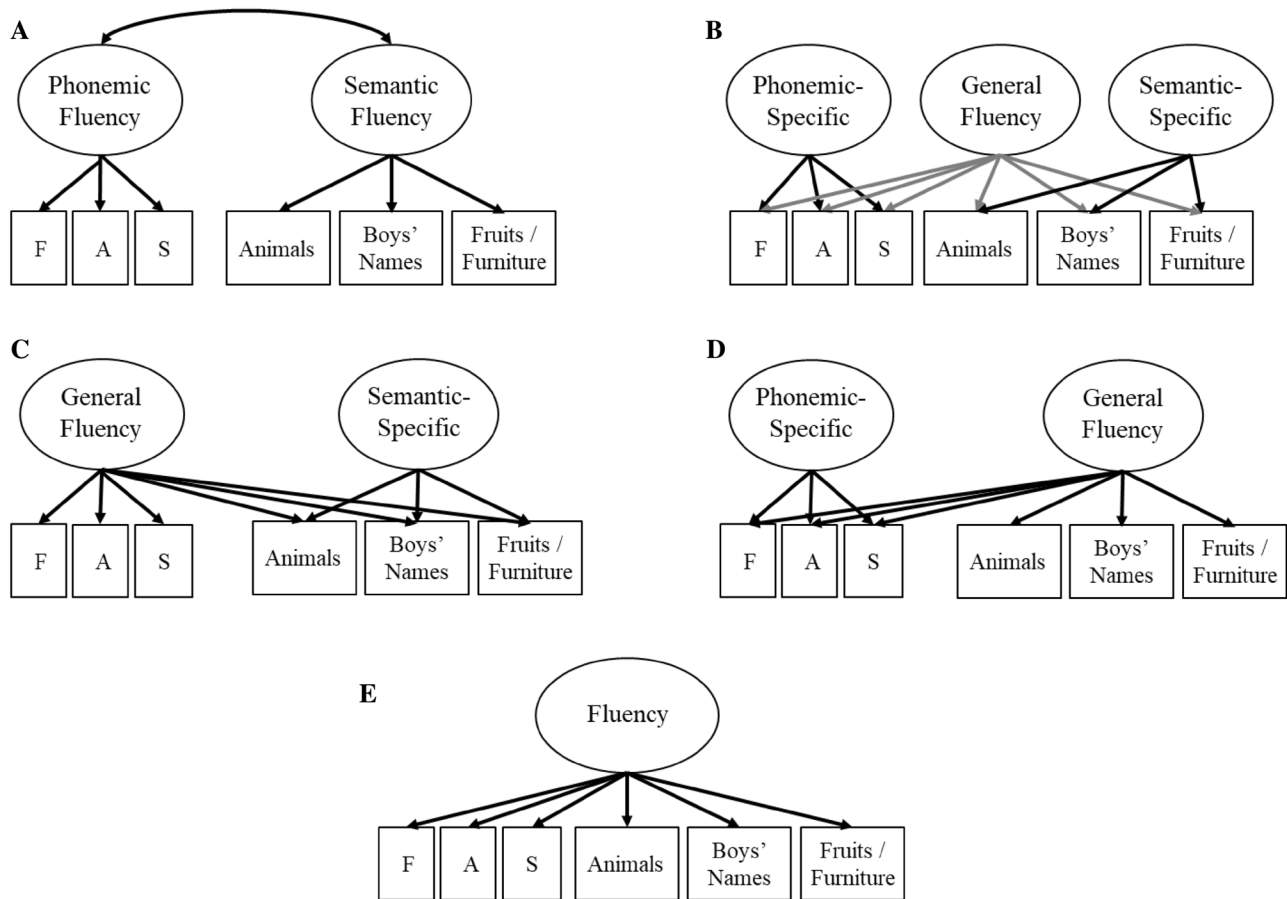


Fig. 1 Comparison of the models of verbal fluency. Model fit for each model is displayed in Table 4 for both waves of assessment. Not shown here, in all models we also decompose variation in all latent

variables (and residual variances for each subtest) into genetic and environmental influences. Ellipses indicate latent variables and rectangles indicate measured variables

were evaluated in the context of the multivariate twin model. We hypothesized that there would be substantial heritability on the fluency factors, consistent with the limited existing research. We expected a multi-factor solution to emerge, suggesting that there are multiple unique sources of genetic influences underlying different types of fluency, as we have shown for other cognitive abilities (Gustavson et al. 2018a; Panizzon et al. 2015, 2014). This is especially true for the bifactor models that most directly isolate unique sources of variance (Models B, C, and D). Furthermore, consistent with the existing estimates (e.g. Lee et al. 2018), we expected that the remaining variance in each factor would be accounted for by non-shared environmental influences rather than by shared environmental influences.

Stability of fluency in midlife

The final goal of the current study was to examine the stability of individual differences in verbal fluency across middle age and identify evidence for mean-level decline in the

fluency factors over the course of 6 years. Cognitive abilities demonstrate considerable stability of individual differences over the lifespan (Lyons et al. 2017, 2009), though they also begin to decline in middle age (Harris and Deary 2011; Rönnlund and Nilsson 2006). Recent work suggests that, to the extent that verbal fluency captures executive function processes, these abilities may already be declining by middle age (Gustavson et al. 2018b).

After identifying the best fitting models at the two waves of assessment, we combined them in a single longitudinal analysis. We hypothesized that phenotypic, genetic, and environmental influences would demonstrate substantial correlations over time, though the genetic correlations might be stronger than the environmental correlations as is common for cognitive abilities. To the extent that the data demonstrated factorial invariance over the 6-year window, we also examined evidence for mean-level decline in the latent fluency factors. We expected to observe mean-level decline in performance across all fluency tasks, resulting in a significant mean-level decline in the latent factors identified in the first step. Thus, the multi-wave nature of the study will

Table 1 Demographic characteristics of the sample

Demographic variable	N	M	SD	Range
All subjects				
Lifetime education	1483	13.80	2.11	5, 30
Occupation	1477	5.47	1.87	0, 9
Ethnicity (% white non-Hispanic)	1484	88.10	–	–
Wave 1				
Age	1290	51.08	2.44	51.08, 60.67
Depression symptoms	1283	8.40	8.22	0, 52
Wave 2				
Age	1207	61.72	2.45	56.00, 66.92
Depression symptoms	1195	7.41	8.11	0, 53

Lifetime education was the number of years of school completed. Occupation was based on the Hollingshead Four-Factor index (Hollingshead 1975), from 0 (unemployed) to 9 (major professionals). Depression symptoms were measured with the Center for Epidemiologic Studies—Depression Scale (Radloff 1977), with scores above 15 indicating risk for clinical depression

provide an internal replication of the best-fitting model of fluency. Furthermore, to the extent that mean-level decline differs between latent factors over 6 years, this would provide further evidence that these reflect unique and meaningful sources of variance.

Method

Subjects

Data analyses were based on 1464 individual male twins who participated in at least one wave of the longitudinal Vietnam Era Twin Study of Aging (VETSA) project. At wave 1 ($N=1285$), participants included 359 full MZ twin pairs, 271 full DZ twin pairs, and 25 unpaired twins. At wave 2 ($N=1193$), participants included 328 full MZ twin pairs, 231 full DZ twin pairs, and 74 unpaired twins. Most individuals participated at both sessions ($N=1014$).

All participants were recruited randomly from the Vietnam Era Twin Registry from a previous study (Tsuang et al. 2001). All individuals served in the United States military at some time between 1965 and 1975, but are generally representative of American men in their age group with respect to health and lifestyle characteristics (see Table 1 for demographic characteristics), and nearly 80% did not serve in combat or in Vietnam (Kremen et al. 2011, 2006; Schoenborn and Heyman 2009). Data for wave 1 were collected between 2003 and 2007, and data for wave 2 were collected between 2008 and 2013. The only inclusion criteria for the first wave were that twins must be between ages 51 and 59 at the time of recruitment, and that both twins in a pair agreed

to participate in the study. All twins were invited to complete the second wave of testing regardless of the participation of their co-twin.

Measures of verbal fluency

Subjects first performed the phonemic fluency subtests (*F*, *A*, and *S*), followed by two semantic fluency subtests (*Animals* and *Boys' Names*). Finally, subjects performed a category switching subtest from the Delis–Kaplan Executive Function System (D-KEFS; Delis et al. 2001) in which they were instructed to alternate between naming fruits and items of furniture. Dependent measures for each of the subtests were the correct number of exemplars named within a 60-s response window. For category switching, we used the same dependent measure, ignoring the number or accuracy of switches between categories. This was done to be most comparable to the other measures of semantic fluency and so we could have three conditions for both phonemic and semantic fluency to aid in model identification.

All measures were adjusted for age by creating residualized scores after accounting for age at that wave of assessment. This enables us to interpret the data at each wave as representing a single age (e.g., age 56 for wave 1 and age 62 for wave 2), and allows for the examination of the change in mean-level performance over 6 years.

Additionally, for the second wave only, we adjusted scores to account for the fact that many of the subjects had encountered the tasks before (Elman et al. 2018). Practice effects for each subtest were computed according to the method of Rönnlund et al. (Rönnlund and Nilsson 2006; Rönnlund et al. 2005), and utilized data from individuals who completed both waves of assessment ($N=1014$), individuals who did not return at the second wave ($N=271$), and attrition-replacement subjects randomly selected from the same twin registry who completed the test battery for the first time at the second wave ($N=179$) and were the same age as the wave 2 subjects (56–66). For each task, the practice effect calculation estimates a difference score (returnees minus attrition-replacements), and an attrition effect (returnees minus all wave 1 subjects). The practice effect is the difference score minus the attrition effect, and was subtracted from scores for all returnees. Although practice effects were nonsignificant for any given fluency subtest, it is important to correct for small and nonsignificant practice effects as ignoring these small differences will still mask the true extent of decline.

Data analysis

Analyses were conducted using the structural equation modeling package OpenMx in R (Boker et al. 2011), which accounts for missing observations using a full-information

maximum likelihood approach. Model fit was determined using $-2 \log$ -likelihood values ($-2LL$), Bayesian information criterion (BIC), the root mean square error of approximation (RMSEA), and the Tucker–Lewis Index (TLI). Good fitting models had the lowest BIC values, RMSEA values < 0.06 , and TLI values > 0.95 (Hu and Bentler 1998; Markon and Krueger 2004). Additionally, good fitting models did not fit significantly worse than a full genetic Cholesky decomposition by comparing the $-2LL$ values using χ^2 difference tests (χ^2_{diff}). We also used χ^2 difference tests to compare competing nested models. Significance of individual parameters was established with χ^2 difference tests and with likelihood-based 95% confidence intervals (95% CI).

Genetically informed models were based on the standard assumptions in twin designs. Additive genetic influences (A) are correlated at 1.0 for MZ twin pairs and 0.5 for DZ twin pairs because MZ twins share 100% and DZ twins share, on average, 50% of their alleles identical-by-descent. Non-additive/dominant genetic influences (D) are correlated at 1.0 for MZ twins and 0.25 for DZ twins. Shared environmental influences (C) are correlated at 1.0 in both MZ and DZ twins. Non-shared environmental influences (E), which, include measurement error, are set to not correlate for both MZ and DZ twin pairs. The standard twin design also assumes equal means and variances within pairs and across zygosity. These assumptions for univariate analyses apply to multivariate cases and to situations where phenotypic correlations between constructs are decomposed into their genetic (r_g) and environmental components (r_e).

The cross-twin cross-trait correlations for MZ and DZ pairs displayed in Supplemental Table S1 indicated evidence for non-additive genetic models (i.e., ADE models). However, there is little power to distinguish between additive and non-additive genetic influences in the classic twin study, even with large samples (Martin et al. 1978). Thus, here we report models with only additive genetic and non-shared environmental influences (AE models), which did not fit significantly worse than the corresponding ADE models (see Supplementary Tables S2, S3 and Figs. S1–S3, all $ps > .186$). Nevertheless, because there was some evidence for non-additive genetic influences on all latent factors, the genetics estimates presented here should be interpreted as broad-sense heritability (i.e., additive + non-additive genetic influences) rather than narrow-sense heritability (additive influences only). We also tested ACE models with shared environmental influences (see Tables S2, S3). These models did not fit as well as AE or ADE models, and there was only evidence for weak and nonsignificant residual shared environmental influences on individual subtests ($c^2 = 0.01–0.03$ on letter *F* and *S*). Shared environmental influences on latent variables were always estimated at 0.00.

The candidate models of verbal fluency are displayed in Fig. 1. First, we examined whether the phonemic and

semantic subtests load on distinct but correlated factors (Fig. 1a). Utilizing a different approach, the remaining “bifactor” models hypothesize a General Fluency latent factor that accounts for variation in all six subtests, reflecting their common variance.¹ Depending on the model, other factors account for additional variation in phonemic or semantic subtests that are not captured by the common factor (and are uncorrelated with the common factor). Additionally, we considered the possibility that the category switching subtest might be different from the other semantic fluency measures, although the dependent measure was based on the number of words generated rather than switching accuracy. Analyses indicated that modeling this subtest as separate but correlated with the other fluency latent factors did not provide a more parsimonious fit to the data than the best-fitting models identified below. Moreover, the pattern of results described below was similar even if we remove category switching entirely (see Supplemental Table S4 and Figs. S4, S5). However, it is included here because it aids in model identification.

Results

Descriptive statistics

Descriptive statistics for all fluency subtests are displayed in Table 2. The phenotypic correlation matrix at both waves of assessment is displayed in Table 3. As shown in Table 3, there were moderate phenotypic correlations between all fluency subtests within waves of assessment, $r_s = .25–.63$ for wave 1; $r_s = .25–.65$ for wave 2. Performance on a given subtest at the first wave was also moderately correlated with performance on that same subtest at the second wave ($r_s = .46–.63$), suggesting that individual differences on any given subtest were relatively stable over this 6-year window. Additionally, the correlations between semantic and phonemic subtests ($r_s = .25–.44$) were about as high as those among the semantic fluency subtests ($r_s = .31–.45$), suggesting considerable common variance across all subtests.

Indicative of age-related decline, mean-level performance decreased significantly over time (by about 0.25 *SD* for each subtest) for all subtests; $t_s(1005) < -4.91$, $ps < .012$; except the *Animals* subtest; $t(1005) = -0.26$, $p = .796$.

¹ The term “bifactor” does not actually refer to the number of factors in the model. Rather, it refers to the approach of fitting a common factor explaining variance in all the candidate indicators, and fitting specific factors as needed to capture additional variation (e.g., Semantic-Specific) not already captured by the common factor, even though in some cases there may be more than one additional factor.

Table 2 Descriptive statistics

Task	<i>N</i>	<i>M</i>	<i>SD</i>	Range	Skewness	Kurtosis
Wave 1 (age 56)						
Letter F	1277	12.28	4.09	1–29	0.29	–0.02
Letter A	1277	11.15	3.90	1–29	0.41	0.34
Letter S	1277	13.48	4.32	1–31	0.25	0.08
Animals	1275	19.20	4.43	6–39	0.26	0.26
Boys' names	1276	19.09	4.48	6–40	0.34	0.58
Fruits/furniture	1277	12.75	2.55	4–22	–0.01	0.30
Wave 2 (age 62)						
Letter F	1189	11.68	4.04	0.82–28.00	0.32	0.14
Letter A	1189	10.39	3.87	1.44–26.00	0.31	–0.06
Letter S	1189	12.70	4.31	0.00–28.83	0.24	–0.08
Animals	1189	19.11	4.51	5.08–35.08	0.17	0.11
Boys' names	1189	18.32	4.50	4.31–37.31	0.23	0.47
Fruits/furniture	1188	12.30	2.58	2.71–21.71	0.02	0.38

In all analyses, these dependent measures were standardized residual scores after removing the effect of age on each measure, but the unadjusted scores are presented here. However, the wave 2 scores reported here reflect the adjustments for practice effects for individuals who participated at both waves of assessment

Table 3 Phenotypic correlations between fluency subtests at both waves of assessment

Measure	1	2	3	4	5	6	7	8	9	10	11	12
Wave 1												
1. Letter F	1											
2. Letter A	0.60	1										
3. Letter S	0.63	0.61	1									
4. Animals	0.33	0.31	0.35	1								
5. Boys' names	0.38	0.44	0.44	0.42	1							
6. Fruits/furniture	0.25	0.30	0.28	0.34	0.31	1						
Wave 2												
7. Letter F	0.61	0.55	0.58	0.29	0.37	0.23	1					
8. Letter A	0.54	0.60	0.58	0.30	0.38	0.24	0.62	1				
9. Letter S	0.58	0.55	0.63	0.28	0.42	0.26	0.63	0.65	1			
10. Animals	0.32	0.31	0.32	0.52	0.35	0.33	0.34	0.35	0.35	1		
11. Boys' names	0.30	0.37	0.39	0.38	0.58	0.27	0.37	0.44	0.41	0.45	1	
12. Fruits/furniture	0.25	0.25	0.24	0.30	0.29	0.46	0.25	0.27	0.28	0.38	0.34	1

All correlations are significant ($p < .05$)

Models of fluency at age 56 and 62

First, we fit factor models of the verbal fluency subtests within each wave of assessment based on the a priori models displayed in Fig. 1. The results are displayed in Table 4 for wave 1 (top) and wave 2 (bottom). The best fitting models are displayed in Fig. 2. In these models at both waves, a General Fluency latent factor accounted for variation in all six fluency subtests and a Semantic-Specific fluency factor accounted for some additional variation in semantic fluency subtests that was not already captured by the common factor (Model C in Fig. 1).

As shown in Table 4, these models had better fit statistics than the competing models displayed in Fig. 1, with

one caveat. At wave 1, there was some evidence that Model B (both phonemic-specific and semantic-specific variation) fit better than Model C (only semantic-specific variation). Model B had a worse (higher) BIC value but slightly better RMSEA and TLI values compared to Model C, and it fit significantly better than Model C, $\chi^2_{\text{diff}}(5) = 11.88$, $p = .036$. Despite these conflicting fit indices, we chose to reject Model B, in part because the wave 2 data also supported Model C for all fit statistics. Moreover, in the longitudinal model described in the following section, the Phonemic-Specific factor in Model B at wave 1 could be dropped from the model without any significant decrement in fit, $\chi^2_{\text{diff}}(5) = 4.78$, $p = .443$. Nevertheless, this alternate model for wave 1 is displayed in the supplement (Model

Table 4 Model comparisons for models of verbal fluency within each wave

	-2LL	df	BIC	RMSEA	TLI	vs. Saturated Cholesky		
						diff (LL)	df	<i>p</i>
Wave 1 (mean age 56)								
Saturated Cholesky (AE)	19070.10	7611	-31,406	0.020	0.992			
A. Two correlated factors	19104.90	7631	-31,504	0.025	0.988	34.80	20	0.021
B. Bifactor—semantic-specific and phonemic-specific	19086.37	7626	-31,489	0.020	0.992	16.27	15	0.364
C. Bifactor—common fluency and semantic-specific	19098.25	7630	-31,504	0.023	0.990	28.15	19	0.081
D. Bifactor—common fluency and phonemic-specific	19101.51	7630	-31,501	0.024	0.989	31.41	19	0.036
E. Single factor	19229.82	7634	-31,399	0.052	0.949	159.72	23	<.001
Wave 2 (mean age 62)								
Saturated Cholesky (AE)	17602.42	7085	-29,385	0.000	1.000			
A. Two correlated factors	17630.74	7105	-29,490	0.000	1.000	28.32	20	0.102
B. Bifactor—semantic-specific and phonemic-specific	17619.20	7100	-29,468	0.000	1.000	16.78	15	0.332
C. Bifactor—common fluency and semantic-specific	17623.53	7104	-29,490	0.000	1.000	21.11	19	0.331
D. Bifactor—common fluency and phonemic-specific	17635.00	7104	-29,478	0.000	1.000	32.58	19	0.027
E. Single factor	17810.00	7108	-29,330	0.052	0.946	207.58	23	<.001

Model fit statistics for the models tested at wave 1 (mean age 56) and wave 2 (mean age 62), corresponding to the models displayed in Fig. 1. The final three columns show the model comparison from the saturated Cholesky model of the genetic/environmental relations between the six fluency subtests (at each wave). The best fitting models are displayed in bold

AE model with only genetic (A) and non-shared environmental influences (E), -2LL negative two times the log likelihood, BIC Bayesian information criterion, RMSEA root mean square error of approximation, TLI Tucker–Lewis Index

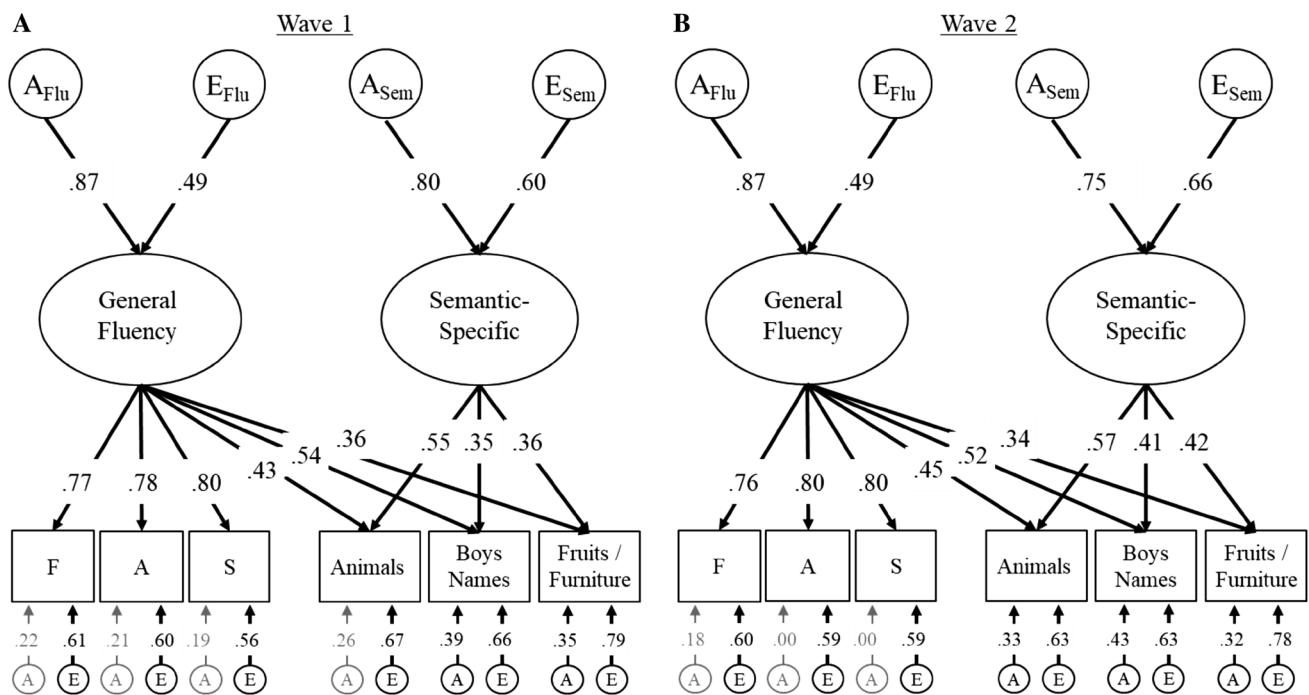


Fig. 2 Best fitting models of the fluency data at both waves of assessment. AE factors represent the genetic (A) and non-shared environmental influences on the latent fluency variables. Ellipses indicate latent variables and rectangles indicate measured variables. Signifi-

cant factor loadings are displayed with black text and lines (*p* < .05). Variation explained by latent factors can be computed by squaring the factor loadings

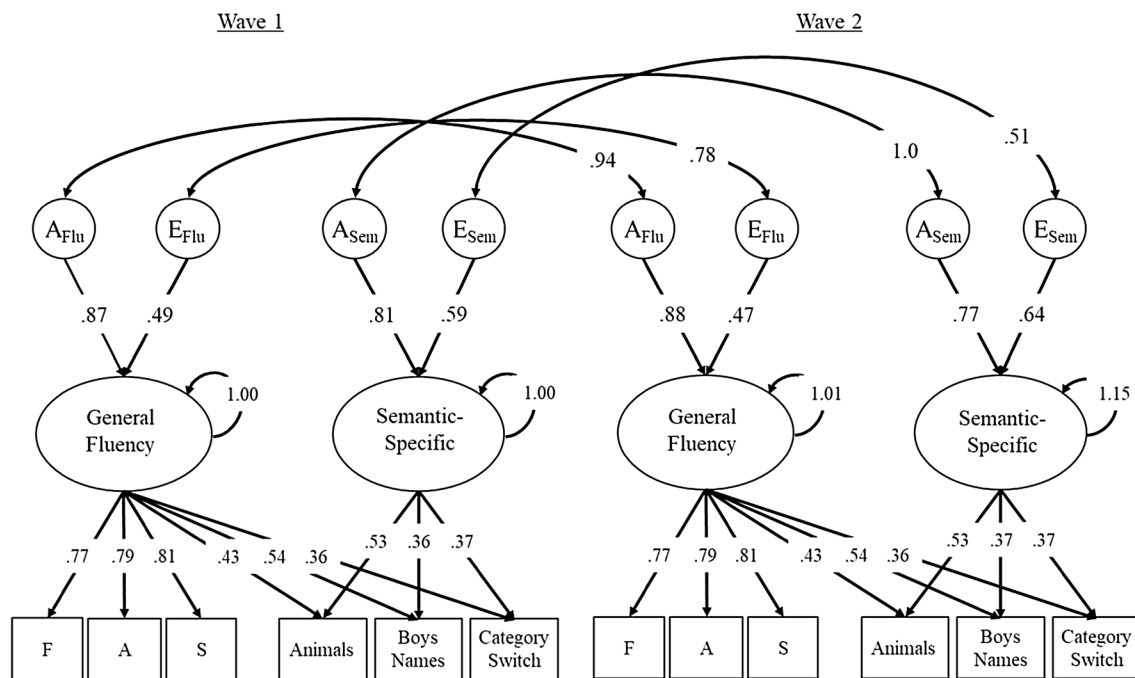


Fig. 3 Longitudinal model of verbal fluency between wave 1 (mean age 56) and wave 2 (mean age 62). AEs represent genetic (A) and non-shared environmental (E) influences on the fluency latent variables. Ellipses represent latent variables and rectangles represent measured variables. Curved arrows pointing from a latent variable to itself represent variances of that latent variable. Significant factor loadings and correlations are displayed with black text and lines

($p < .05$). Latent variable means are not shown but fixed at 0 for General Fluency and Semantic-Specific at wave 1 and estimated at -0.22 and -0.01 (respectively) at wave 2. AE latent variables and all factor loadings are standardized at both waves. Residual AE paths on individual subtest are not displayed, but are similar to those cross-sectional association presented in Fig. 2 (see Supplement Table S5)

B, Fig. S2). The supplement also displays the correlated-factor model (Model A, Fig. S3) to compare these findings with traditional conceptualizations of phonemic fluency and semantic fluency as unique but correlated abilities.

Genetic and environmental results

As shown in Fig. 2, genetic influences accounted for 76% of the variation in the General Fluency factor at the first wave; heritability, or $a^2 = 0.76$, 95% CI [0.69, 0.82]. Non-shared environmental influences accounted for the remaining 24% of the variation, $e^2 = 0.24$, 95% CI [0.18, 0.31]. Similar results were observed for the Semantic-Specific factor, in which 64% of the variance could be explained by genetic influences, $a^2 = 0.64$, 95% CI [0.43, 0.84]. The remaining 36% was explained by non-shared environmental influences, $e^2 = 0.36$, 95% CI [0.16, 0.57].

At wave 2, the results for the General Fluency factor were nearly identical to wave 1. Genetic influences accounted for 76% of the variation, $a^2 = 0.76$, 95% CI [0.69, 0.82], and non-shared environmental influences accounted for the remaining 24%, $e^2 = 0.24$, 95% CI [0.18, 0.31]. The Semantic-Specific factor continued to be explained mostly by genetic influences, though the estimate was slightly

lower than at the first wave, $a^2 = 0.57$, 95% CI [0.38, 0.74]. Non-shared environmental influences explained the remaining 43% of the variance, $e^2 = 0.43$, 95% CI [0.26, 0.62]. At both waves, residual non-shared environmental influences were significant on all subtests (residual $e^2 = 0.31$ – 0.62), and residual genetic influences were significant for *Animals* (wave 1 only), *Boys' Names*, and *Fruits/Furniture* subtests (residual $a^2 = 0.11$ – 0.18).

Longitudinal model of fluency between mean age 56 and 62

Next, we combined the bifactor models of fluency at waves 1 and 2 into a single longitudinal model. This model is displayed in Fig. 3 and fit well, $-2LL = 35315.06$, $df = 14,735$, $BIC = -62,407$, $RMSEA = 0.021$, $TLI = 0.986$. In this model, genetic and environmental correlations between latent factors were not estimated directly, but computed from the statistically equivalent Cholesky decomposition. It was also necessary to include Cholesky paths from residual genetic/environmental factors between waves 1 and 2 for the same task (e.g., letter *F* at wave 1 to letter *F* at wave 2) to account for additional correlations within each subtest

over time (see Supplement Table S5 for residual genetic and environmental influences and correlations between them).

The model displayed in Fig. 3 displays strict invariance. That is, the factor loadings, intercepts, and residual variances for each subtest were equated over time. This model fit significantly worse than the model with configural invariance (i.e., factor loadings, intercepts, and residual variances freed across time), $\chi^2_{\text{diff}}(18) = 34.28, p = .012$, but this was likely due to the large sample size and high power to detect small deviations in observed versus predicted correlations and means, rather than poor model fit (for a similar example, see Gustavson et al. 2018b). For example, this strict invariance model had the lowest BIC value compared to the models with configural invariance, weak factorial invariance, or strong invariance, suggesting it was the best at balancing parsimony and fit. Individual differences results and RMSEA values were nearly identical across these models.

As shown in Fig. 3, individual differences in the General Fluency and Semantic-Specific factors were highly stable over time (estimated phenotypic $r = .90$ and $.81$, respectively). For the General Fluency factor, the genetic correlation was $r_g = .94$, 95% CI [0.90, 0.97], suggesting that the genetic influences were nearly identical between age 56 and 62. Non-shared environmental influences were also strongly correlated, $r_e = .78$, 95% CI [0.65, 0.90], but that new non-shared environmental influences also explain variance in the General Fluency factor at age 62.

For the Semantic-Specific factor, genetic influences were perfectly correlated over time, even though genetic influences explained a slightly smaller portion of the total variance in Semantic-Specific at the second wave, $r_g = 1.0$, 95% CI [0.88, 1.0]. Non-shared environmental influences were only moderately correlated over time, $r_e = .51$, 95% CI [0.21, 0.78].

This longitudinal model also provides some information about cognitive decline over time. As we expected, performance in the General Fluency factor declined by 0.22 *SD* (based on the mean and *SD* at wave 1), $d = -0.22$, 95% CI [-0.27, -0.17]. However, there was no change in mean-level performance for the Semantic-Specific factor, $d = -0.01$, 95% CI [-0.11, 0.09]. There was no evidence for change in variance of the latent variables at the second wave, 1.00, 95% CI [0.92, 1.10] for General Fluency, 1.15, 95% CI [0.94, 1.41] for Semantic-Specific.

Discussion

In a large longitudinal twin study, we examined verbal fluency at two times in middle age. Results suggested that it was best to view fluency as explained by two latent constructs. A General Fluency factor accounted for variation across all six subtests and a Semantic-Specific factor

accounted for additional variance in semantic fluency subtests not captured by General Fluency. Genetic influences accounted for the majority of the variation in both latent factors, and these genetic influences were highly correlated across the 6-year interval. In contrast, non-shared environmental influences explained about one-quarter to one-third of the variance in both factors, and demonstrated moderate correlations over time. Mean-level performance declined over time only for the General Fluency factor. These results provide a new framework for viewing semantic fluency as a combination of general and semantic-specific variance, both of which have unique genetic underpinnings and may be declining at different rates.

Implications for verbal fluency

The first goal of the study was to examine the best-fitting model of individual differences in phonemic and semantic fluency. The bifactor model displayed in Fig. 2 had the most parsimonious fit to the data. A strength of this bifactor approach is that we could isolate this common variance from additional variance unique to semantic fluency and estimate the genetic/environmental influences on both factors. Because the common factor accounts for variation in all six individual fluency measures, this factor most likely represents general fluency abilities associated with vocabulary and the updating and inhibition executive functions (Shao et al. 2014; Whiteside et al. 2016). Given that there was no Phonemic-Specific factor, this common factor also taps phonemic processing abilities that may aid in performance in semantic subtests as well (e.g., generating boys' names or animals that start with the same letter or rhyme). The Semantic-Specific factor may be associated with similar cognitive processes as the common factor, but this variance may also reflect other related processes such as episodic memory or lexical access speed (Shao et al. 2014; van den Berg et al. 2017).

Although Model C had the best fit, the other candidate models displayed in Fig. 1 also fit well. The two-correlated factor model of Phonemic Fluency and Semantic Fluency (Model A) had acceptable fit, but was not as parsimonious as the bifactor models. This may have been observed because semantic subtests had somewhat different factor loadings on the General Fluency and Semantic-Specific factors identified in Fig. 2 (i.e., *Boys' Names* was better explained by General Fluency and *Animals* by Semantic-Specific). One potential explanation may be related to findings that the *Boys' Names* subtest is more strongly related to Parkinson's disease than the *Animals* subtest (Fine et al. 2011), suggesting differences in reliance on lexical strategies or sub-processes even within semantic fluency (e.g., strategic search versus semantic organization, phonological versus semantic clustering). Indeed, adding an additional factor loading to Model A from

Phonemic Fluency to *Boys' Names* improves model fit (wave 1: factor loading = 0.22, $\chi^2_{\text{diff}} = 9.53$, $p = .002$; wave 2: factor loading = 0.19, $\chi^2_{\text{diff}} = 9.73$, $p = .002$) but this was not predicted *a priori* and still did not result in a better fit than Model C. Moreover, the bifactor model provides a clearer isolation of common and specific variance across fluency subtests that may be of greater use in further examination of normal and impaired functioning than the traditional view of phonemic and semantic fluency as correlated factors.

A remaining question also concerns whether there is evidence for phonemic-specific variance. In the current study, there was some evidence for a Phonemic-Specific factor in the first wave alone, but not all model fit indices agreed that this factor was necessary. Moreover, when we tried to include this factor in the longitudinal model (Fig. 3) it could be dropped without a significant decrement in fit. If there is phonemic-specific variance, it is unclear why it would disappear by the second wave, especially considering the stability of individual differences across this 6-year interval. Furthermore, this result is consistent with previous theoretical proposals that phonemic fluency relies more strongly on strategic retrieval processes in the frontal lobe (consistent with larger factor loadings on the General Fluency factor) rather than unique processes that are completely unrelated to semantic fluency (Moscovitch 1994). Nevertheless, these findings will need to be replicated before making strong conclusions about the lack of a Phonemic-Specific factor.

Interestingly, although these and similar measures of verbal fluency have been widely used, there has been relatively little work examining their underlying genetic and environmental influences (Antila et al. 2007; Bratko 1996; Lee et al. 2012, 2018; Swan and Carmelli 2002), especially using latent variable models. Our results suggest that genetic influences accounted for most of the variation in both fluency factors, as has been found for other cognitive abilities measured in middle age (Gustavson et al. 2018b; Panizzon et al. 2015). These genetic influences likely represent the contribution of hundreds or thousands of individual genetic effects. Continuing to examine these associations at the level of latent variables will be especially useful in future research, not only because these models can isolate variance in general fluency variation from semantic-specific variation, but also because this method isolates genetic/environmental variance in the latent constructs from subtest-specific influences. Thus, heritability estimates tend to be larger at the level of latent factors than for univariate measures (Antila et al. 2007; Bratko 1996; Lee et al. 2012; Swan and Carmelli 2002).

The remaining variation in both latent fluency constructs was captured by non-shared environmental influences, and not at all by shared environmental influences. The lack of any shared environmental influences is not surprising given the relatively weak contribution of shared environmental

influences to many cognitive abilities (Friedman et al. 2008; Kremen et al. 2011; Lee et al. 2012; Panizzon et al. 2015). Although these environmental influences were examined at the latent construct level, and should therefore be free from measurement error, all subtests came from the same test (D-KEFS) so it is possible that these environmental influences reflect some test-specific variance or situation-specific variance from the testing environment.

In fact, the results were more consistent with ADE models (including non-additive genetic influences) rather than an ACE models (with shared environmental influences). Supplemental analyses (Table S2, Figs. S1–S3) indicated that non-additive genetic influences accounted for the majority of the heritability of both fluency latent factors. However, even with this large sample we had little power to detect significant differences between additive and non-additive genetic influences (Martin et al. 1978). Although it may be important to consider this distinction between additive and non-additive genetic influences in future work, collapsing additive and non-additive genetic influences here had no impact on the non-shared environmental estimates.

Implications for cognitive aging

These results are also relevant to age-related decline in cognition. There is a steady decline in cognitive performance beginning as early as middle age (Kremen et al. 2014), and verbal fluency is especially relevant given its associations with heritable neuropsychological disorders (Henry and Crawford 2004a, b, 2005a, b; Henry et al. 2004). Of particular importance, the distinction between General Fluency and Semantic-Specific factors is consistent with findings that semantic fluency is more strongly impaired in Alzheimer's disease than phonemic fluency. The Semantic-Specific factor identified here, and its genetic underpinnings, may therefore be useful in future gene-finding efforts regarding Alzheimer's disease and related dementias.

We also expected some decline in performance over this 6-year window in middle age. Consistent with this hypothesis, mean-level performance in the General Fluency factor declined on average by about 0.22 *SD* (compared to wave 1), suggesting some small to moderate decline in fluency across the sample after accounting for the effects of repeated exposure. In contrast, mean-level performance did not decline for the Semantic-Specific factor. Together, these results suggest that general abilities supporting verbal fluency are declining as early as age 56, and perhaps earlier, but that the additional abilities supporting semantic fluency are not as susceptible to age-related decline until at least the mid-60s, providing further evidence that these abilities are unique and differentially related to cognitive aging.

It will be useful to examine how stability and change in the fluency factors identified here are associated with the

decline in other cognitive abilities in this same age range. Research using this sample suggests that individual differences in episodic memory show similar 6-year genetic/environmental correlations (Panizzon et al. 2015), but that executive function ability demonstrates both a stronger cross-time correlation ($r = .97$) and a sharper decline in mean-level performance ($d = -0.60$; Gustavson et al. 2018b). Verbal fluency has been characterized as an executive function ability, and is positively correlated with the updating and inhibition executive functions (Shao et al. 2014). However, recent work has suggested that both phonemic and semantic fluency may be better indicators of vocabulary and language processing than executive function ability (Whiteside et al. 2016). Future work should examine how this model of verbal fluency fits in with existing models of executive function that also emphasize shared versus unique variance (Friedman et al. 2008; Gustavson et al. 2018a, b; Miyake and Friedman 2012), especially in the context of cognitive aging, as these unique variance components may have differential rates of decline and predictive ability.

Limitations

First, this sample comprises only men, so these findings may not generalize across sex. Second, although the sample is representative of American men of their age, we were not able to examine whether findings generalized across ethnicity. Third, these results may not generalize to clinical populations (Delis et al. 2003). Nevertheless, this unscreened twin sample is not free from individuals with psychiatric or other diagnoses, suggesting that the heritability estimates should be unbiased. Fourth, we assessed verbal fluency with both phonemic and semantic measures, but it would be useful to examine the extent to which nonverbal fluency tasks (Baldo et al. 2001) draw on the General Fluency factor but may also have unique variance components. This would also help determine the extent to which there is truly a General Fluency factor or whether there are general verbal and nonverbal fluency factors.

Fifth, it would have been best to include a third semantic fluency subtest in place of the *Fruits/Furniture* subtest that also involved some additional switching demands. However, as described above, the exclusion of this subtest did not affect the pattern of results. Finally, in all bifactor models, the confidence intervals were wider on “specific” factors compared to the General Fluency factor. We have observed this phenomenon with similar bifactor models of executive function (Gustavson et al. 2018a, b). Thus, although bifactor models are useful in isolating unique variance specific to semantic fluency, and provided a better fit than the traditional correlated-factors model in the current study (Model A), it is still difficult to estimate semantic-specific fluency with great precision. Despite these limitations, we used a large

longitudinal twin study to elucidate the multivariate nature and complex genetic/environmental architecture of these different types of verbal fluency at key timepoints when verbal fluency is beginning to exhibit age-related decline.

Summary and conclusions

Although measures of verbal fluency are widely used in studies of cognition and aging, little is known about the differential processes underlying phonemic versus semantic fluency and their genetic/environmental etiology. The results here suggest that variance in phonemic and semantic fluency are explained by general fluency abilities and semantic-specific abilities (but not phonemic-specific abilities), and that over half of their variation can be accounted for by genetic influences. Both abilities demonstrate high correlations over time, at least across a 6-year interval in middle age, and only the general factor appears to decline between the late fifties to early sixties. Given the relevance of verbal fluency to mental and physical health, it will be useful to examine how the fluency factors identified here account for variance in other heritable neuropsychiatric conditions, and how they continue to change over the course of aging.

Acknowledgements This research was supported by Grants AG050595, AG018386, AG018384, AG022381, and AG047903 from the National Institutes of Health. The content of this manuscript is the responsibility of the authors and does not represent official views of NIA/NIH, or the Veterans’ Administration. Numerous organizations provided invaluable assistance in the conduct of the VET Registry, including: U.S. Department of Veterans Affairs, Department of Defense; National Personnel Records Center, National Archives and Records Administration; Internal Revenue Service; National Opinion Research Center; National Research Council, National Academy of Sciences; the Institute for Survey Research, Temple University. The authors gratefully acknowledge the continued cooperation of the twins and the efforts of many staff members.

Funding This research was supported by Grants AG050595, AG018386, AG018384, AG022381, and AG047903 from the National Institutes of Health.

Compliance with ethical standards

Conflict of interest Daniel E. Gustavson, Matthew S. Panizzon, Jeremy A. Elman, Carol E. Franz, Asad Beck, Chandra A. Reynolds, Kristen C. Jacobson, Hong Xian, Rosemary Toomey, Michael J. Lyons, and William S. Kremen declare that they have no conflicts of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all participants included in the study.

References

- Antila M, Tuulio-Henriksson A, Kieseppa T, Soronen P, Palo OM, Paunio T, Haukka J, Partonen T, Lonnqvist J (2007) Heritability of cognitive functions in families with bipolar disorder. *Am J Med Genet B* 144B:802–808. <https://doi.org/10.1002/ajmg.b.30538>
- Baldo JV, Shimamura AP, Delis DC, Kramer J, Kaplan E (2001) Verbal and design fluency in patients with frontal lobe lesions. *J Int Neuropsychol Soc* 7:586–596
- Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P (2011) OpenMx: an open source extended structural equation modeling framework. *Psychometrika* 76:306–317. <https://doi.org/10.1007/s11336-010-9200-6>
- Bratko D (1996) Twin study of verbal and spatial abilities. *Personality Individ Differ* 21:621–624. [https://doi.org/10.1016/0191-8869\(96\)00091-8](https://doi.org/10.1016/0191-8869(96)00091-8)
- Delis DC, Kaplan E, Kramer JH (2001) Delis-Kaplan executive function system (D-KEFS). Psychological Corporation, San Antonio
- Delis DC, Jacobson M, Bondi MW, Hamilton JM, Salmon DP (2003) The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: lessons from memory assessment. *J Int Neuropsychol Soc* 9:936–946. <https://doi.org/10.1017/S1355617703960139>
- Elman JA, Jak AJ, Panizzon MS, Tu X, Chen T, Reynolds CA et al (2018) Under-diagnosis of MCI in longitudinal studies: the problem of not accounting for practice effects. *Alzheimer's Dement*. <https://doi.org/10.1016/j.dadm.2018.04.003>
- Fine EM, Delis DC, Paul BM, Filoteo JV (2011) Reduced verbal fluency for proper names in nondemented patients with Parkinson's disease: a quantitative and qualitative analysis. *J Clin Exp Neuropsychol* 33:226–233. <https://doi.org/10.1080/13803395.2010.507185>
- Friedman NP, Miyake A, Young SE, Defries JC, Corley RP, Hewitt JK (2008) Individual differences in executive functions are almost entirely genetic in origin. *J Exp Psychol* 137:201–225. <https://doi.org/10.1037/0096-3445.137.2.201>
- Giubilei F, Medda E, Fagnani C, Bianchi V, De Carolis A, Salvetti M, Sepe-Monti M, Stazi MA (2008) Heritability of neurocognitive functioning in the elderly: evidence from an Italian twin study. *Age Ageing* 37:640–646. <https://doi.org/10.1093/ageing/afn132>
- Gustavson DE, Panizzon MS, Franz CE, Friedman NP, Reynolds CA, Jacobson KC, Xian H, Lyons MJ, Kremen WS (2018a) Genetic and environmental architecture of executive functions in midlife. *Neuropsychology* 32:18–30. <https://doi.org/10.1037/neu0000389>
- Gustavson DE, Panizzon MS, Elman JA, Franz CE, Reynolds CA, Jacobson KC, Friedman NP, Xian H, Toomey R, Lyons MJ, Kremen WS (2018b) Stability of genetic and environmental influences on executive functions in midlife. *Psychol Aging* 33:219–231. <https://doi.org/10.1037/pag0000230>
- Harris SE, Deary IJ (2011) The genetics of cognitive ability and cognitive ageing in healthy older people. *Trends Cogn Sci* 15:388–394. <https://doi.org/10.1016/j.tics.2011.07.004>
- Hedden T, Yoon C (2006) Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology* 20:511–528. <https://doi.org/10.1037/0894-4105.20.5.511>
- Henry JD, Crawford JR (2004a) A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology* 18:284–295. <https://doi.org/10.1037/0894-4105.18.2.284>
- Henry JD, Crawford JR (2004b) Verbal fluency deficits in Parkinson's disease: a meta-analysis. *J Int Neuropsychol Soc* 10:608–622. <https://doi.org/10.1017/S1355617704104141>
- Henry JD, Crawford JR (2005a) A meta-analytic review of verbal fluency deficits in depression. *J Clin Exp Neuropsychol* 27:78–101. <https://doi.org/10.1080/138033990513654>
- Henry JD, Crawford JR (2005b) A meta-analytic review of verbal fluency deficits in schizophrenia relative to other neurocognitive deficits. *Cogn Neuropsychiatry* 10:1–33. <https://doi.org/10.1080/13546800344000309>
- Henry JD, Crawford JR, Phillips LH (2004) Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia* 42:1212–1222. <https://doi.org/10.1016/j.neuropsychologia.2004.02.001>
- Hollingshead AB (1975) Four factor index of social status. Yale University, New Haven
- Hu LT, Bentler PM (1998) Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol Methods* 3:424–453. <https://doi.org/10.1037/1082-989x.3.4.424>
- Kremen WS, Thompson-Brenner H, Leung YM, Grant MD, Franz CE, Eisen SA, Jacobson KC, Boake C, Lyons MJ (2006) Genes, environment, and time: the Vietnam era twin study of aging (VETSA). *Twin Res Human Genet* 9:1009–1022. <https://doi.org/10.1375/183242706779462750>
- Kremen WS, Jacobson KC, Panizzon MS, Xian H, Eaves LJ, Eisen SA, Tsuang MT, Lyons MJ (2009) Factor structure of planning and problem-solving: a behavioral genetic analysis of the Tower of London task in middle-aged twins. *Behav Genet* 39:133–144. <https://doi.org/10.1007/s10519-008-9242-z>
- Kremen WS, Panizzon MS, Xian H, Barch DM, Franz CE, Grant MD, Toomey R, Lyons MJ (2011) Genetic architecture of context processing in late middle age: more than one underlying mechanism. *Psychol Aging* 26:852–863. <https://doi.org/10.1037/a0025098>
- Kremen WS, Moore CS, Franz CE, Panizzon MS, Lyons MJ (2014) Cognition in middle adulthood. In: Finkel D, Reynolds CA (eds) *Behavior genetics of cognition across the lifespan*. Springer, New York, pp 105–134
- Kremen WS, Panizzon MS, Cannon TD (2016) Genetics and neuropsychology: a merger whose time has come. *Neuropsychology* 30:1–5. <https://doi.org/10.1037/neu0000254>
- Lee T, Mosing MA, Henry JD, Trollor JN, Ames D, Martin NG, Wright MJ, Sachdev PS, OATS Research Team (2012) Genetic influences on four measures of executive functions and their covariation with general cognitive ability: the older Australian twins study. *Behav Genet* 42:528–538. <https://doi.org/10.1007/s10519-012-9526-1>
- Lee T, Thalamuthu A, Henry JD, Trollor JN, Ames D, Wright MJ, Sachdev PS, OATS Research Team (2018) Genetic and environmental influences on language ability in older adults: findings from the older Australian Twins Study. *Behav Genet*. <https://doi.org/10.1007/s10519-018-9897-z>
- Lyons MJ, York TP, Franz CE, Grant MD, Eaves LJ, Jacobson KC, Schaie KW, Panizzon MS, Boake C, Xian H, Toomey R (2009) Genes determine stability and the environment determines change in cognitive ability during 35 years of adulthood. *Psychol Sci* 20:1146–1152. <https://doi.org/10.1111/j.1467-9280.2009.02425.x>
- Lyons MJ, Panizzon MS, Liu W, McKenzie R, Bluestone NJ, Grant MD, Franz CE, Vuoksimaa EP, Toomey R, Jacobson KC, Reynolds CA (2017) A longitudinal twin study of general cognitive ability over four decades. *Dev Psychol* 53:1170–1177. <https://doi.org/10.1037/dev0000303>
- Markon KE, Krueger RF (2004) An empirical comparison of information-theoretic selection criteria for multivariate behavior genetic models. *Behav Genetics* 34:593–610. <https://doi.org/10.1007/s10519-004-5587-0>
- Martin NG, Eaves LJ, Kearsley MJ, Davies P (1978) The power of the classical twin study. *Heredity* 40:97–116. <https://doi.org/10.1038/hdy.1978.10>
- McGue M, Christensen K (2001) The heritability of cognitive functioning in very old adults: evidence from Danish twins aged 75 years and older. *Psychol Aging* 16:272–280. <https://doi.org/10.1037/0882-7974.16.2.272>

- Miyake A, Friedman NP (2012) The nature and organization of individual differences in executive functions: four general conclusions. *Curr Dir Psychol Sci* 21:8–14. <https://doi.org/10.1177/0963721411429458>
- Moscovitch M (1994) Cognitive resources and dual-task interference effects at retrieval in normal people: the role of the frontal lobes and medial temporal cortex. *Neuropsychology* 8:524–534. <https://doi.org/10.1037/0894-4105.8.4.524>
- Panizzon MS, Vuoksima E, Spoon KM, Jacobson KC, Lyons MJ, Franz CE, Xian H, Vasilopoulos T, Kremen WS (2014) Genetic and environmental influences on general cognitive ability: is g a valid latent construct? *Intelligence* 43:65–76. <https://doi.org/10.1016/j.intell.2014.01.008>
- Panizzon MS, Neale MC, Docherty AR, Franz CE, Jacobson KC, Toomey R, Xian H, Vasilopoulos T, Rana BK, McKenzie R, Lyons MJ (2015) Genetic and environmental architecture of changes in episodic memory from middle to late middle age. *Psychol Aging* 30:286–300. <https://doi.org/10.1037/pag0000023>
- Radloff LS (1977) The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1:385–401
- Rönnlund M, Nilsson LG (2006) Adult life-span patterns in WAIS-R Block Design performance: cross-sectional versus longitudinal age gradients and relations to demographic factors. *Intelligence* 34:63–78. <https://doi.org/10.1016/j.intell.2005.06.004>
- Rönnlund M, Nyberg L, Bäckman L, Nilsson LG (2005) Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. *Psychol Aging* 20:3–18. <https://doi.org/10.1037/0882-7974.20.1.3>
- Schoenborn CA, Heyman KM (2009) Health characteristics of adults aged 55 years and over: United States, 2004–2007. *Nat Health Stat Rep* 16:1–31
- Shao Z, Janse E, Visser K, Meyer AS (2014) What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Front Psychol* 5:772
- Stolwyk R, Bannirchelvam B, Kraan C, Simpson K (2015) The cognitive abilities associated with verbal fluency task performance differ across fluency variants and age groups in healthy young and old adults. *J Clin Exp Neuropsychol* 37:70–83. <https://doi.org/10.1080/13803395.2014.988125>
- Swan GE, Carmelli D (2002) Evidence for genetic mediation of executive control: a study of aging male twins. *J Gerontol B* 57:P133–143
- Troyer AK, Moscovitch M (2006) Cognitive processes of verbal fluency tasks. In: Poreh AM (ed) *The quantified process approach to neuropsychological assessment*. Taylor & Francis, New York, pp 143–160
- Tsuang MT, Bar JL, Harley RM, Lyons MJ (2001) The Harvard twin study of substance abuse: what we have learned. *Harv Rev Psychiatry* 9:267–279. <https://doi.org/10.1093/hrp/9.6.267>
- Unsworth N, Spillers GJ, Brewer GA (2011) Variation in verbal fluency: a latent variable analysis of clustering, switching, and overall performance. *Q J Exp Psychol* 64:447–466. <https://doi.org/10.1080/17470218.2010.505292>
- van den Berg E, Jiskoot LC, Grosveld MJH, van Swieten JC, Papma JM (2017) Qualitative assessment of verbal fluency performance in frontotemporal dementia. *Dement Geriatr Cogn Disord* 44:35–44. <https://doi.org/10.1159/000477538>
- Vasilopoulos T, Franz CE, Panizzon MS, Xian H, Grant MD, Lyons MJ, Toomey R, Jacobson KC, Kremen WS (2012) Genetic architecture of the delis-kaplan executive function system trail making test: evidence for distinct genetic influences on executive function. *Neuropsychology* 26:238–250. <https://doi.org/10.1037/a0026768>
- Whiteside DM, Kealey T, Semla M, Luu H, Rice L, Basso MR, Roper B (2016) Verbal fluency: language or executive function measure? *Appl Neuropsychol* 23:29–34. <https://doi.org/10.1080/23279095.2015.1004574>