Title

Semantic item-level metrics relate to future memory decline beyond existing cognitive tests in older adults without dementia.

Journal

Psychology and Aging, 38(5)

Authors

Vonk, Jet
Geerlings, Mirjam
Avila-Rieger, Justina
et al.

Publication Date

2023-08-01

DOI

10.1037/pag0000747

Peer reviewed

HHS Public Access

Author manuscript

Psychol Aging. Author manuscript; available in PMC 2024 August 01.

Published in final edited form as:

Psychol Aging. 2023 August ; 38(5): 443–454. doi:10.1037/pag0000747.

Semantic item-level metrics relate to future memory decline beyond existing cognitive tests in older adults without dementia

Jet M.J. Vonk, Ph.D.[1,2], **Mirjam I. Geerlings**[2], **Justina F. Avila**[3], **Carolyn L. Qian**[3], **Nicole Schupf**[3], **Richard Mayeux**[3], **Adam M. Brickman**[3], **Jennifer J. Manly**[3]

[1]Department of Neurology, Memory and Aging Center, University of California San Francisco (UCSF), San Francisco, CA, USA

[2]Julius Center for Health Sciences and Primary Care, Department of Epidemiology, University Medical Center Utrecht and Utrecht University, Utrecht, The Netherlands

[3]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University, New York, NY, USA

Abstract

In normal aging, the cognitive domain of semantic memory remains preserved, while the domain of episodic memory declines to some extent. In Alzheimer's disease dementia, both semantic and episodic memory become impaired early in the disease process. Given the need to develop sensitive and accessible cognitive markers for early detection of dementia, we investigated among older adults without dementia whether item-level metrics of semantic fluency related to episodic memory decline above and beyond existing neuropsychological measures and total fluency score. Participants were drawn from the community-based WHICAP cohort (N=583 English speakers, mean age=76.3±6.8) followed up to 5 visits across up to 11 years. We examined the association of semantic fluency metrics with subsequent declines in memory performance using latent growth curve models covaried for age and recruitment wave. Results showed that item-level metrics (e.g., lexical frequency, age of acquisition, and semantic neighborhood density) were associated with a decline in episodic memory—even when covarying for other cognitive tests—while the standard total score was not. Moderation analyses showed that the relationship of semantic fluency metrics with memory decline did not differ across race, sex/gender, or education. In conclusion, item-level data hold a wealth of information with potential to reveal subtle semantic memory impairment, which tracks with epidosic memory impairment, among older adults without dementia beyond existing neuropsychological measures. Implementation of psycholinguistic metrics may point to cognitive tools that have better prognostic value or are more sensitive to cognitive change in the context of clinical trials or observational studies.

Corresponding author: Jet M.J. Vonk, Ph.D., Memory and Aging Center, Department of Neurology, University of California San Francisco (UCSF), San Francisco, CA, USA, Jet.Vonk@ucsf.edu.

All analysis code is available at https://github.com/jmjvonk. This study's design, hypotheses, and analyses were not pre-registered. The ideas and data appearing in the manuscript were previously disseminated during an oral presentation at the 2020 Alzheimer's Association International Conference.

Keywords

category fluency; verbal fluency; dementia; longitudinal; cohort

As individuals age, they typically undergo cognitive aging, which is a natural process that can result in a decline in various mental abilities and processes, such as memory (Liverman, Yaffe, & Blazer, 2015). Based on research in cognitive psychology and neuroscience, several principles of cognitive aging have been affirmed (e.g., Gazzaley & D'esposito, 2007; Salthouse, 2016). These principles include that cognitive decline is not uniform across all domains (i.e., some domains may be preserved or improve), different cognitive abilities decline at different rates, inter-individual variability is present in cognitive aging, and compensatory mechanisms can offset cognitive decline (e.g., Salthouse, 2004; Stern et al., 2018; Stine-Morrow & Manavbasi, 2022). For example, the cognitive domain of semantic memory—our collective understanding of facts and concepts, encompassing word meanings—remains preserved in normal aging (Venneri, Jahn-Carta, de Marco, Quaranta, & Marra, 2018; Vonk et al., 2020), while domains like episodic memory and executive control decline to some extent (Hanninen & Soininen, 1997; Salthouse, 2010). Cognitive changes associated with normal aging, however, are distinguished from cognitive changes due to neurodegeneration, such as dementia.

Alzheimer's disease (AD) is the most common type of dementia (Fratiglioni, De Ronchi, & Agüero-Torres, 1999; Zhang et al., 2021). Semantic memory is impaired early in the AD process, in parallel to episodic memory (Venneri et al., 2018; Verma & Howard, 2012). Impairment in the domain of episodic memory is strongly associated with AD as of an early stage in the disease (Greene, Baddeley, & Hodges, 1996), but episodic memory impairment is also present to some extent in normal aging (Hanninen & Soininen, 1997). In contrast, semantic memory stays relatively intact with normal aging while being susceptible to impairment during the early progression of AD (Park et al., 2002; Salthouse, 2010; Venneri et al., 2018; Vonk et al., 2020)—but see work by Nebes and colleagues (e.g., Nebes & Brady, 1990). This relationship between early semantic impairment and the likelihood of developing AD in the future is well-established (Mistridis, Krumm, Monsch, Berres, & Taylor, 2015; Verma & Howard, 2012).

Neurobiologically, semantic processing has been located to temporal and inferior parietal brain regions, particularly for verbal semantics in the left hemisphere and non-verbal semantics in the right hemisphere (Binder, Desai, Graves, & Conant, 2009; Butler, Brambati, Miller, & Gorno-Tempini, 2009; Vonk, Borghesani, et al., 2019). The nature of semantic impairment with pathological aging is typically related to brain damage in the temporal lobe and inferior parietal lobule (Grossman et al., 2002; Libon et al., 2013; Vonk, Rizvi, et al., 2019). In AD, atrophy starts early in the pathological process in medial-temporal regions, then spreads towards lateral temporal and temporal-parietal regions, and only later in the disease spreads towards the frontal lobes (Whitwell, 2010). This pattern of atrophy progression with early damage to temporal regions, the neurobiological location of semantic processing, explains the onset of semantic impairment in AD.

Semantic fluency—generating as many exemplars within a category within one minute—is the most-often used test to assess semantic processing; a relatively low total score is considered an important marker of cognitive impairment in manifest AD (Mueller et al., 2015). While this task is considered to rely heavily on semantic processing, the standard score of total number of words is also thought to reflect executive functioning abilities (Aita et al., 2018; Shao, Janse, Visser, & Meyer, 2014).

Item-level data from the semantic fluency task contain an underutilized wealth of information and provide an opportunity to isolate metrics of semantic processing from executive functioning aspects (Vonk, Flores, et al., 2019). Item-level data can be of particular interest for quantifying individual characteristics of words, as individuals with progressive semantic impairment (e.g., AD) are more likely to lose lower-frequency and less well-connected words (e.g., lynx) before more frequent and better-connected words (e.g., dog) (Lambon Ralph, Graham, Ellis, & Hodges, 1998; Vonk, Jonkers, et al., 2019).

These qualities of words can be captured through linguistic databases that quantified psycholinguistic characteristics of words. It is important to note, however, that the cultural, socioeconomic, and linguistic backgrounds of the participants from which the source data of these linguistic databases were collected, or the individuals who generated the speech or text for the source data, are typically homogenous. The linguistic database samples are often primarily collected from younger to middle-aged adults, and the racial and educational composition of the database samples, if reported, is typically White and college- educated. It is therefore unknown whether the application of these data is appropriate to evaluate language use from people who are older, not White, or not college-educated.

Examples of psycholinguistic characteristics are lexical frequency, age of acquisition, semantic neighborhood density, and lexical decision response time (RT). *Lexical frequency* indicates how often a word occurs in our daily language. *Age of acquisition* implies the age at which a word was learned. *Semantic neighborhood density* refers to the degree of co-occurrence of other words with the word. *Lexical decision RT* signifies the time in milliseconds to decide whether a word is a real and existing word or not. Theoretically, words are thought to be organized in a hierarchically organized mental lexicon with information about different aspects of a word, including meaning (conceptual/semantic level), word form (lexeme level), and syntactic characteristics (lemma level) (Bock & Levelt, 1994). For example, age of acquisition is considered to be related to the semantic level in the mental lexicon, while the length of a word is related to the lexeme level (i.e., word form) (Brysbaert, Van Wijnendaele, & De Deyne, 2000). The level to which lexical frequency is related to in the mental lexicon is debated but has been proposed to relate to both semantic and lexeme levels (Vonk, Jonkers, et al., 2019). Previous work showed that words with a higher lexical frequency, earlier age of acquisition, and larger semantic neighborhood density are processed more quickly on a lexical decision task than words at the other ends of these spectra in cognitively normal individuals (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Morrison & Ellis, 2000). The mechanism underlying the greater accessibility of words with higher frequency, higher semantic neighborhood density, and lower age of acquisition is thought to involve less cognitive effort to process and understand these words than words with reverse values on those scales (e.g., Cuetos, Barbón, Urrutia, & Domínguez,

2009; Mutter & Hashtroudi, 1987). These three individual psycholinguistic values of words as well as their lexical decision RT may therefore be related to the semantic level in the mental lexicon and capture a qualitative aspect of a word's position within the semantic network. Isolating these semantic characteristics of a word from other mental lexicon levels as well as separating semantic from executive functioning components by investigating semantic fluency at the item-level may reveal fluency metrics that are more sensitive to subsequent declines in episodic memory than the standard total number of words typically used for this task.

Additionally, clusters of words within semantic subcategories and switching between clusters are two components of verbal fluency that capture the sequential order in which words were produced. Cluster size is the number of words from a subcategory produced consecutively, and switches represent the number of times one shifts between clusters. Clusters are considered to reflect the organization of concepts in the semantic network and switches the executive functioning ability to shift between clusters (Eng, Vonk, Salzberger, & Yoo, 2019; Troyer, Moscovitch, & Winocur, 1997). Individuals with AD produce smaller clusters and less switches on a semantic fluency task than cognitively normal individuals (Gomez & White, 2006; Troyer, Moscovitch, Winocur, Leach, & Freedman, 1998).

This study investigated whether behavioral markers of semantic access predict memory decline, by examining the relation of several item-level metrics of semantic fluency to future episodic memory decline—a symptom of clinical progression to AD—in older adults without dementia. We hypothesized that item-level semantic fluency metrics would be related to future memory decline over and above scores of other standard cognitive tests. We previously applied psycholinguistic characteristics from a lexical frequency database on semantic fluency performance of Black American older adults (Vonk, Flores, et al., 2019). However, it remains undetermined whether the application of different psycholinguistic characteristics as cognitive metrics is equally appropriate across race, ethnicity, and levels of education. It is imperative to know if these metrics have equivalent ability to identify cognitive decline early in populations that are not historically included in dementia research. Therefore, we investigated whether race and education moderated the relationship between novel item-level metrics and memory decline.

Methods

Transparency and Openness

We define how we determined our sample size, report the gender and racial distribution of the sample, and describe all data exclusions, manipulations, and all measures in the study. Data and research materials are available upon reasonable request to the WHICAP Publications Committee. IRB approval is required for all proposed analyses, and a Data Use Agreement (DUA) or Material Transfer agreement needs to be established. Due to legal reasons as described in the DUA, Vonk et al. have agreed to retain control over the data and to not disclose, release, sell, rent, lease, loan, or otherwise grant access to the data to any third party, except Authorized Persons, without the prior written consent of The Trustees of Columbia University in the City of New York. Researchers interest in using WHICAP data are encouraged to submit a WHICAP Data Request/Manuscript Proposal to the WHICAP

Publications Committee. Data and research materials requests should be submitted at https://cumc.co1.qualtrics.com/jfe/form/SV_6x5rRy14B6vpoqN.

All analysis code is available at https://github.com/jmjvonk (Vonk, 2023). Data were prepared using SPSS Version 27 (IBM Corp, 2020). Descriptive analyses were performed using R version 3.6.0 (R Core Team, 2021), with packages haven (Wickham & Miller, 2018), xlsx (Dragulescu, Dragulescu, & Provide, 2020), dplyr (Wickham, Francois, Henry, & Müller, 2019), furniture (Barrett & Brignone, 2017), ggplot2 (Wickham, 2016), GGally (Schloerke, Crowley, & Cook, 2018), and ggpubr (Kassambara & Kassambara, 2020). Longitudinal analyses (i.e., growth curve models) were performed using Mplus Version 8 (Muthén & Muthén, 2019). This study's design, hypotheses, and analyses were not pre-registered.

Participants

Participants were part of the Washington Heights/Inwood Columbia Aging Project (WHICAP; Columbia University). WHICAP was designed to examine the epidemiology of cognitive aging and dementia in older adults from multiple racial, ethnic, and educational backgrounds living in the community in Northern Manhattan (Stern et al., 1992; Tang et al., 2001). Within WHICAP, three independent cohorts were recruited in 1992, 1999, and 2009 (N = 6842), and followed over time with assessments approximately every 18 to 30 months for up to 25 years. Participants completed cognitive, functional, and health measures in their preferred language, i.e., English (n = 3973) or Spanish (n = 2845) (24 missing values for language of administration). Diagnosis of all-cause dementia at each evaluation was determined via consensus case conference following standard research criteria based on neurological, neuropsychological, functional, medical, and psychiatric evaluation (American Psychiatric Association, 1987; McKhann et al., 1984).

While verbal fluency was administered as part of the neuropsychological battery to all participants, the exact words that were generated (i.e., item-level fluency data) has only been entered for a subset of individuals at one or more of their visits at the time of this analysis, based on a draw of charts from the archived paperwork related to archival location. WHICAP charts are archived in three different locations (one on-site and two off-site). We were able to enter item-level data from all charts that were archived on-site; these data tend to include more individuals from the latest 2009 recruitment wave (84%) resulting in a shorter follow-up time than the average of all three recruitment waves. Individuals were included in this study if item-level fluency data were entered for them; for this analysis, baseline was defined as the first visit with item-level fluency data available. Individuals were excluded if they were not tested in English (because we lacked the capacity to enter, score, and analyze item-level fluency data in Spanish), if they were diagnosed with dementia at baseline, if they were missing episodic memory scores at baseline, or if they had a race/ethnicity other than non-Hispanic White or non-Hispanic Black; because we aimed to perform moderation analyses, we restricted the data for this study to race/ethnicity groups with enough item-level data available to perform these analyses (i.e., model convergence).

A total number of 583 participants were included; a flowchart of participant selection is presented in Figure 1. Data availability allowed us to model up to five visits (mean number

of visits = 2.9±1.4) across a period of up to 11 years. Retention rates decreased at each follow-up, with n = 455 (78%) at visit 2, n = 335 at visit 3 (74%), n = 201 (60%) at visit 4, and n = 97 (48%) at visit 5. Among all individuals, those lost because of death were n = 45 at visit 2, n = 30 at visit 3, n = 10 at visit 4, and n = 3 at visit 5. This study was approved by the Columbia University Medical Center Institutional Review Board (Study title: "Preclinical markers of Alzheimer's disease using psycholinguistic semantic measures", protocol number AAAS9419) and all participants provided written informed consent.

Sociodemographic Grouping Variables

We investigated the distribution of the total and item-level semantic fluency metrics by age, sex/gender, race, education, APOE ε4 status, and subjective cognitive complaints. APOE ε4 is to date the strongest genetic risk factor for developing Alzheimer's disease (Sienski et al., 2021). Age at baseline was operationalized in years since date of birth until date of baseline visit, including decimals. Sex/gender was classified by asking individuals to identify as male or female—however, it is ambiguous whether individuals reported their biological sex or their gender identity, hence we refer to this variable as 'sex/gender' (Avila et al., 2019). Education was self-reported and operationalized as years of formal education (0–20 years). Race and ethnicity in WHICAP was self-reported based on the 1990 US Census guidelines (United States Bureau of the Census, 1991). APOE genotyping was performed with a slight modification from the protocol described by Hixson and Vernier (1990); individuals were categorized as APOE ε4 positive or negative based on the presence of at least one ε4 allele. Subjective cognitive complaints were measured as no complaints versus any complaints on a set of questions for memory complaints, which included questions about memory difficulties in general as well as for specific areas such as names.

Neuropsychological Assessments

An extensive neuropsychological battery was administered, including tests of episodic memory, language, and visuospatial ability (Siedlecki, Honig, & Stern, 2008; Stern et al., 1992). Memory was assessed with the Selective Reminding Test (Buschke & Fuld, 1974). An episodic memory domain composite score was previously developed in WHICAP by standardizing the total recall, delayed recall, and delayed recognition scores of the Selective Reminding Test using the larger WHICAP sample's means and standard deviations at first visit, and averaging these z-scores to create a composite score (Zahodne, Schofield, Farrell, Stern, & Manly, 2014). Language tests included letter and semantic fluency tests, 15-item Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983), Similarities subtest from the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981), and Repetition and Comprehension subtests of the Boston Diagnostic Aphasia Evaluation (Goodglass & Kaplan, 1983). Visuospatial ability was assessed with Recognition and Matching tasks on the Benton Visual Retention Test (Benton, 1955), Rosen Drawing Test (W. Rosen, 1981), and Identities and Oddities subtest from the Mattis Dementia Rating Scale (Mattis, Bellak, & Karasu, 1976).

Semantic Fluency Metrics

Semantic fluency was administered for the category ‘animals.’ Participants were instructed as follows: “I want to see how many different animals you can name. Any animals will do; they can be from the farm, the jungle, the ocean, or house pets. They can begin with any letter. Go as quickly as possible. Begin.” Participants were allowed 60 seconds to generate words; responses were recorded on paper by the examiner.

We analyzed the standard metric of semantic fluency, i.e., the total number of correct items generated within 60 seconds, as well as four item-level metrics related to semantic properties (lexical frequency, age of acquisition, semantic neighborhood density, lexical decision RT) and two metrics based on the sequence in which the items were generated (cluster size and switches). These measures are also summarized in Supplementary Table S1. Psycholinguistic metrics of lexical frequency, age of acquisition, semantic neighborhood density, and lexical decision RT are available from existing and freely accessible linguistic corpora; spreadsheets with the metrics used in this study are also available for download at https://github.com/jmjvonk (Vonk, 2023).

Lexical frequency was derived from SUBTLEXus, based on 51 million words from American English film (from 1900–2007) and television subtitles (timeframe undefined) (Brysbaert & New, 2009) (available at: https://www.ugent.be/pp/experimentele-psychologie/en/research/documents/subtlexus). The cultural, socioeconomic, and linguistic backgrounds of the script writers has not been summarized. Age of acquisition was derived from norms for over 50,000 English words, a subjective measure in which participants were asked to place on a timeline the age at which they believe they learned each word (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) (available at http://crr.ugent.be/archives/806). These age of acquisition ratings were collected via Amazon Mechanical Turk through 1,960 US-based web surfers; 99.3% had English as a first language with 1.8% bi- or multilingual, the majority of respondents had at least some college education, and race/ethnicity was not reported (Kuperman et al., 2012). Lexical decision RT was obtained from the English Lexicon Project, which measured visual lexical decision performance of over 40,000 words in 816 university students (Balota et al., 2007) (available at https://elexicon.wustl.edu/); the cultural, socioeconomic, and linguistic backgrounds of the university students have not been reported in detail. Semantic neighborhood density was obtained from the co-occurrence matrix of 287,000 word pairs by Shaoul and Westbury (2010) via the English Lexicon Project (Balota et al., 2007). The sequential metrics of clusters and switches were derived using the Semantic Network and Fluency Utility (SNAFU), a procedure of automatically identifying clusters and switches in semantic fluency (Zemla, Cao, Mueller, & Austerweil, 2020) (available at: https://github.com/AusterweilLab/snafu-py).

The lexical frequency, age of acquisition, and semantic neighborhood density data were validated by how well they predicted human processing latencies (i.e., lexical decision RT from the English Lexicon Project). The age of acquisition data were additionally validated by correlating them to other age of acquisition ratings from smaller samples (Kuperman et al., 2012), and in a later study by correlating them to age of acquisition norms from testing children’s knowledge of word meanings at various ages (Brysbaert & Biemiller, 2017). The

semantic neighborhood density computational model was additionally internally validated on random subsets of the lexicon. The SNAFU tool has been validated by comparing results to five trained human coders using multiple datasets (Zemla et al., 2020). The cultural, socioeconomic, and linguistic backgrounds of the human coders has not been reported.

Item-level data was entered into a database and prepared for analyses following the rules described in Vonk, Flores, et al. (2019). For words without an available value, lexical frequency was imputed with a log value equal to an occurrence of .5 in 51 million words (0.2% values imputed) following Kuperman et al. (2012), an age of acquisition of 12 given the slower growth of vocabulary beyond age 12 (Beitchman et al., 2008) (1.2% values imputed), and semantic neighborhood density (1.0% values imputed) and lexical decision RT (1.0% values imputed) of their mean database values of .422 and 784.1, respectively. The metrics for lexical frequency and semantic neighborhood density were multiplied by −1 such that higher values reflected better performance.

Statistical Analysis

The sample's characteristics and the distribution of the total and item-level semantic fluency metrics by age, sex/gender, race, education, APOE ε4 status, and subjective cognitive complaint were analyzed with descriptive statistics, general linear models, chi-square tests, and Pearson's correlation coefficients in R version 3.6.0 (R Core Team, 2021).

Univariate latent growth curve models in Mplus Version 8 (Muthén & Muthén, 2019) were used to investigate the association between various semantic fluency metrics at baseline as determinants, and change in episodic memory composite score across up to five visits as an outcome. Models were estimated with maximum likelihood with standard error approximation by first-order derivatives. All models included random intercepts and random slopes for time; time was parameterized as years since baseline and modelled with individually-varying time scores to incorporate different intervals between visits. All fluency metrics were standardized for comparability of parameter estimates and covariates were centered.

As a base, we built unconditional latent growth curve models to compare model fit of linear versus quadratic change over time using Bayesian Information Criterion (BIC). With lower values indicating a better fit, BIC values indicated that linear models (2991.667) provided a better fit than those including a quadratic effect (3012.234). Because baseline was defined as the first visit with item-level fluency data available, and many individuals had received cognitive testing prior to this item-level fluency baseline meaning we did not include the potential change due to practice effect between their very first and second visits in our dataset, we did not include practice effects in our models.

A first set of conditional models were fitted with age and recruitment wave as covariates (Models A). We did not covary for sex/gender or race because we have previously shown that sex/gender differences in cognition vary as a function of race/ethnicity (Avila et al., 2019). Adding these variables as covariates would remove the effect of each variable assuming that there is no interaction.

Subsequently, extended conditional models additionally adjusted for performance on all other administered neuropsychological tests through a latent cognition factor with its mean fixed at 0 and variance at 1 (Models B). For the item-level metrics, these extended models were estimated with and without including the total score of semantic fluency as a separate covariate to analyze the association of the item-level metrics with cognitive change over and above other neuropsychological tests as well as the total score (Models C). As the psycholinguistic item-level metrics are very strongly correlated with each other and therewith suggest to represent the same underlying construct of semantic memory, we additionally analyzed these four metrics as a latent factor in relationship to memory performance for the main Models A, B, and C. This analysis allows for a more robust test of the predictive relationship between individual differences in semantic processing and subsequent declines in episodic memory.

Sensitivity analyses were performed to investigate potential selection bias due to missing data attributable to death and loss-to-follow-up (i.e., informative censoring) (Wang, Shen, & Boye, 2012), as cognitive performance of individuals who die or drop-out during follow-up is typically poorer than that of individuals who remain alive and keep participating in the study (Kurland, Johnson, Egleston, & Diehr, 2009). The survival process was jointly modeled with the longitudinal process as a discrete-time survival model in the overall sample for two scenarios: the survival model included a latent hazard function representing the conditional probability of 1) death (but not other reasons for drop-out) at a specific visit given survival and no drop-out at previous visits, or 2) death or drop-out at a specific visit given survival and no drop-out at previous visits. This latent hazard function was regressed on the latent growth parameters (intercept and slope) to adjust the trajectory estimates.

Moderation models were performed based on Model A using a classic interaction approach for education (continuous moderator) and a multiple-group modeling approach across race and sex/gender (categorical moderators), which allows for non-invariance in variances, covariances, and other parameters across groups. In the multiple-group modeling approach, we compared the intercept and slope estimates across the groups within each fluency metric by testing linear restrictions on the parameters (i.e., interactions), with White participants and female participants as the reference groups, respectively.

Results

Sample Characteristics and Distribution of Fluency Metrics

Participants generated on average 16.0±5.3 words during the semantic fluency task; 99.0% generated at least five words, 88.5% generated at least ten words, and 56.1% generated at least 15 words. Demographic information and performance on the different semantic fluency metrics are presented in Table 1 for the overall sample, and in Supplementary Table S2 across groups of sex/gender, race, APOE ε4 status, and subjective cognitive complaints. The distribution of fluency characteristics across these groups is also visualized in Supplementary Figure S1. Correlations of age and years of education with the different semantic fluency metrics and accompanying scatterplots are displayed in Supplementary Figure S2. Lower total number scores were associated with older age, lower education, being racialized as Black, and more subjective cognitive complaints. The

four psycholinguistic metrics of higher lexical frequency, earlier age of acquisition, higher semantic neighborhood density, and faster lexical decision time were consistently associated with older age, lower education, being female, being racialized as Black, and—with the exception of age of acquisition—more subjective cognitive complaints.

All semantic fluency metrics were correlated with baseline memory performance (Figure 2). Strength of correlations can be described as very weak (.00–.19), weak (.20–.39), moderate (.40–.59), strong (.60–.79), and very strong (.80–1.0) (Evans, 1996). The correlations of psycholinguistic metrics (i.e., lexical frequency, age of acquisition, semantic neighborhood density, and lexical decision RT) with each other were very strong. The correlations of the sequential metrics (i.e., clusters and switches) with the psycholinguistic metrics were weak to moderate.

Estimates of Fluency Metrics in Relation to Memory Decline

Table 2 presents the unconditional growth model, which estimated an initial level (i.e., intercept) of memory performance of B = .469, 95% confidence interval (CI) [.406, .533], with a linear decline (i.e., slope) of B = −.042 [−.054, −.029] units per year.

In Models A—adjusted for age and recruitment wave—for the overall sample, each fluency metric was associated with baseline memory performance (i.e., intercept), and all fluency metrics except clusters and switches (i.e., sequence) were associated with the slope of memory decline (Table 2 and Figure 3). A latent factor that combined the four psycholinguistic features was also related to the intercept (B = .185 [.127, .243]) and slope (B = .022 [.010, .035]) of memory decline.

In Models B, which additionally included adjustment for all available neuropsychological tests, total score and switches remained associated with baseline memory performance (Table 2 and Figure 3). In these models, the psycholinguistic metrics except lexical decision RT (95% CI ranging between .001 and .032 across the three metrics) but not the total score (B = .011 [−.002, .024]) or sequential metrics (95% CI ranging between −.012 and .020 across the two metrics) were associated with the slope of memory decline. A latent factor that combined the four psycholinguistic features was also related to the slope of memory decline (B = .017 [.001, .034]) after adjusting for all available neuropsychological tests (not intercept (B = −.041 [−.109, .026]). The strength of this relationship for the latent factor with memory slope only slightly diminished after additionally adjusting for total fluency score (Model C; B = .015 [−.002, .032]). The relationship of lexical frequency with memory decline remained even when additionally adjusting for total fluency score (Models C; Table 2 and Figure 3).

A joint unconditional growth model that included a latent hazard function showed that risk of death was not related to lower initial levels of memory performance (B = −.598 [−1.305, .109]), or to slope (B = −9.501 [−22.667, 3.664]). It should be noted that the estimate for slope had an extremely wide confidence interval due to a limited number of deaths particularly at later follow-ups: n = 45 (7.7%) at visit 2, n = 30 at visit 3 (5.1%), n = 10 (1.1%) at visit 4, and n = 3 (.5%) at visit 5. These confidence intervals indicate high dispersion and less confidence about the true effect of risk of death on memory slope.

The risk of death as well as other loss-to-follow-up was related to lower initial levels of memory performance (B = −.386 [−.670, −.101]), and to slope (B = −4.143 [−8.091, −.196]). However, the inclusion of a hazard function for either scenario of loss-to-follow-up in the growth curve model did not change the estimates of the different semantic fluency metrics in relation to memory decline (detailed for loss-to-follow-up due to death in Table 2).

Moderation analyses by sex/gender showed that the effect of fluency metric on memory slope did not differ between male and female participants for any of the metrics in Models A: total number (ΔB = .009 [−.015, .034]), lexical frequency (ΔB = .006 [−.019, .031]), age of acquisition (ΔB = .008 [−.017, .032]), semantic neighborhood density (ΔB = .013 [−.012, .038]), lexical decision RT (ΔB = .002 [−.022, .026]), clusters (ΔB = .010 [−.019, .039]), or switches (ΔB = −.006 [−.034, .021]). Moderation analyses by race showed that the effect of fluency metric on memory slope did not differ between White and Black participants for any of the metrics in Models A: total number (ΔB = −.005 [−.030, .020]), lexical frequency (ΔB = .009 [−.016, .034]), age of acquisition (ΔB = .004 [−.023, .031]), semantic neighborhood density (ΔB = .009 [−.015, .034]), lexical decision RT (ΔB = .014 [−.011, .038]), clusters (ΔB = −.005 [−.032, .021]), or switches (ΔB = .009 [−.017, .036]). Moderation analyses by education showed that the memory slope estimates were not moderated by education for any of the fluency metrics either: total number (ΔB = −.000 [−.003, .003]), lexical frequency (ΔB = .002 [−.002, .005]), age of acquisition (ΔB = .002 [−.001, .005]), semantic neighborhood density (ΔB = .002 [−.002, .005]), lexical decision RT (ΔB = .002 [−.002, .006]), clusters (ΔB = .000 [−.004, .004]), or switches (ΔB = .000 [−.003, .004]).

Discussion

The current study demonstrated a robust and consistent relationship of three different psycholinguistic metrics—lexical frequency, age of acquisition, and semantic neighborhood density—as well as their latent factor with memory decline over and above other cognitive tests among English-speaking Black and White older adults. The standard total animal fluency score was not related to future memory decline when adjusting for other cognitive tests, nor were the item-level indices of clusters and switches. None of the relationships between item-level fluency metrics and future memory decline were moderated by sex/gender, race, or educational attainment.

Semantic memory reflects our shared knowledge of facts and concepts, including the meanings of words; these concepts are organized in a semantic network that becomes vulnerable during the AD process (Venneri et al., 2018). The link between early semantic impairment and future AD diagnosis is well-established (Mistridis et al., 2015; Verma & Howard, 2012). However, relationships of risk factors, biomarkers, and future diagnosis of AD are typically demonstrated with longitudinal semantic decline, as semantic measures can lack sensitivity cross-sectionally prior to the onset of cognitive impairment (Papp et al., 2016; Pike et al., 2011). This study identified novel semantic metrics that were related to future cognitive decline among individuals without dementia. The strong correlation among the four psycholinguistic item-level suggests their shared representation of the semantic

memory aspects of a stable network structure and automatized processes for accessing that network.

Prior studies showed that individuals with mild cognitive impairment and dementia have a lower number of clusters and switches (Pakhomov & Hemmy, 2014; Raoux et al., 2008; V. M. Rosen et al., 2005), but these metrics are not predictive of future AD in individuals without dementia (Raoux et al., 2008). Correspondingly, our results did not show a relationship between clusters or switches and the rate of subsequent memory decline. The absence of an effect for switches may be explained by this metric's relation to executive functioning abilities, mainly linked to frontal brain regions (Alvarez & Emory, 2006), which get affected by AD only in a later stage of the disease process. In contrast, clusters are considered to reflect the organization of concepts in the semantic network (Troyer et al., 1997). One explanation may be that while the size of the clusters has been argued to reflect semantic organization (Troyer et al., 1997), our results suggest that this metric may be more closely aligned with executive control.

In contrast to sequential metrics, our previous work demonstrated that psycholinguistic metrics could cross-sectionally distinguish non-demented APOE ε4 carriers from non-carriers (Vonk, Flores, et al., 2019). Similarly in the current study, the psycholinguistic metrics of higher lexical frequency, earlier age of acquisition, higher semantic neighborhood density, and faster lexical decision RT were related to steeper subsequent memory decline. In other words, a tendency to generate words with qualities that are 'easier' to process was related to steeper future memory decline, a cognitive marker of AD. The psycholinguistic metrics analyzed in this study seem to capture and quantify processes we observe in individuals with progressive semantic impairment: less prominent words are lost before more commonly used words (Lambon Ralph et al., 1998; Vonk, Jonkers, et al., 2019). As such, these psycholinguistic metrics may capture a qualitative aspect of a word's position within the semantic network, through a (partial) link to the semantic level in the mental lexicon. These results demonstrate the value of conducting more detailed, item-level analysis of animal fluency, beyond total score, by employing a psycholinguistic approach.

The psycholinguistic metrics did not differ in their prediction of memory decline in English-speaking Black and White older adults, male and female participants, or lower and higher educated individuals. One reason to test for differences across racial and educationial groups is a concern for differential sensitivity in cognitive screening tools (Glymour, Weuve, & Chen, 2008). Differential sensitivity of a test across subgroups may lead to incorrect classification of individuals as cognitively impaired, which threatens the validity of a test (Pedraza & Mungas, 2008). If psycholinguistic metrics are going to be applied as a cognitive sreening tool, the metrics should have equivalent ability to identify early cognitive change in populations that have been historically excluded in dementia research. The absence of moderation effects of race, sex/gender, and education in our results suggest that these psycholinguistic metrics may be generalizable across these populations.

It is worth pointing out that the linguistic databases used to derive the psycholinguistic features were established from samples that are very different from the group of community-based older adults in this study. To start, the linguistic database samples were primarily

collected from younger to middle-aged adults. Moreover, the racial and educational composition of the database samples, if reported, was primarily White and college-educated. For example, lexical decision RT in the English Lexicon Project is collected from students across seven universities; while the racial/ethnic distribution among these participants is not reported (Balota et al., 2007), the data available for enrolled student populations at these institutions ranges from 3.23–15.5% Black or African American students (data obtained from datausa.io based on public US Government data). These considerations are important because culture and life experiences influence qualitive aspects of tests of verbal fluency, such as relative salience and clusters (Eng et al., 2019; Medin, Waxman, Woodring, Ross, & Winkler-Rhoades, 2010). Since these features were derived from predominantly high-educated and White samples, psycholinguistic metric values may be under- or over-estimated in people who are not represented in the linguistic databases.

Strengths of this study include an analysis of selection bias by implementing a joint model to account for loss to follow-up due to death. The use of multiple psycholinguistic metrics allowed us to demonstrate the robustness and consistency of the effect of item-level metrics in relation to future cognitive decline. Limitations include that WHICAP participants are recruited from northern Manhattan in New York City, which restricts national and international generalizability. Moreover, we did not include the Hispanic participants in WHICAP, most of whom are Spanish-speaking Caribbean emigrants, because we lacked the capacity to enter, score, and analyze item-level fluency data in Spanish. Our next step is to form partnerships to analyze these data from Spanish-speakers, and compare the psycholinguistic metrics as measures of cognition across languages. Furthermore, the restricted time of follow-up due to the availability of item-level data limited us to investigate incident AD as a clinical outcome; we anticipate that ongoing follow-up of WHICAP participants, as well as continued entry of item-level data, will allow us to analyze this outcome in the future. This future direction would also include determining the sensitivity and specificity of psycholinguistic metrics to predict progression to AD, since the current study does not show that the psycholinguistic metrics are sensitive to AD pathology. The continued entry of item-level data will include data from the other two off-site archival locations of charts, which would more comprehensively represent the three recruitment waves of the WHICAP cohort and increase the average follow-up time of the sample.

This study showed that semantic attributes are leading indicators of memory impairment, and suggests that there may be subtle degradation of the semantic network that appears before any obvious declines in memory or executive control in individuals with AD. The preclinical phase of dementia is a crucial time for potential intervention, prevention, and timely diagnosis for patients and caregivers. Current standard cognitive measures are not sensitive to the earliest stages of dementia, including AD (Loewenstein, Curiel, Duara, & Buschke, 2018). Moreover, standard cognitive measures were designed and validated on White, Western, English-speaking, and well-educated people, and often do not show comparable measurement in minoritized or marginalized groups—particularly for the domain of language (Avila, Arce Rentería, et al., 2019). As such, there is a critical need for sensitive and accessible *cognitive* tools for the preclinical phase of dementia that are fast to administer, do not require access to neuroimaging, genotyping, or other biomarkers of AD or related dementias, and are validated among people who have been historically excluded

from dementia research. Implementation of psycholinguistic metrics may point to cognitive tools that have better prognostic value or may be more sensitive to cognitive change in the context of clinical trials or observational studies that target the preclinical phase of dementia.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute on Aging (NIA) K99/R00 award R00AG066934 (PI: J.M.J. Vonk), Alzheimer Nederland Fellowship WE.15-2018-05 (PI: J.M.J. Vonk), and NWO/ZonMw Veni Grant project number 09150161810017 (PI: J.M.J. Vonk). Data collection and sharing for this project was supported by the Washington Heights-Inwood Columbia Aging Project (WHICAP, PO1AG07232, R01AG037212, RF1AG054023) funded by the NIA. This publication was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Number UL1TR001873. This manuscript has been reviewed by WHICAP investigators for scientific content and consistency of data interpretation with previous WHICAP Study publications. We acknowledge the WHICAP study participants and the WHICAP research and support staff for their contributions to this study.

References

Aita SL, Beach JD, Taylor SE, Borgogna NC, Harrell MN, & Hill BD (2018). Executive, language, or both? An examination of the construct validity of verbal fluency measures. Applied Neuropsychology: Adult.

Alvarez JA, & Emory E (2006). Executive function and the frontal lobes: a meta-analytic review. Neuropsychology review, 16(1), 17–42. [PubMed: 16794878]

American Psychiatric Association. (1987). Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition. Washington, DC: American Psychiatric Press Inc,.

Avila JF, Vonk JMJ, Verney SP, Witkiewitz K, Rentería MA, Schupf N, Mayeux R, & Manly JJ (2019). Sex/gender differences in cognitive trajectories vary as a function of race/ethnicity. Alzheimer's & Dementia, 15(12), 1516–1523.

Balota DA, Cortese MJ, Sergent-Marshall SD, Spieler DH, & Yap M (2004). Visual word recognition of single-syllable words. Journal of Experimental Psychology: General, 133(2), 283–316. [PubMed: 15149254]

Balota DA, Yap MJ, Hutchison KA, Cortese MJ, Kessler B, Loftis B, Neely JH, Nelson DL, Simpson GB, & Treiman R (2007). The English lexicon project. Behavior Research Methods, 39(3), 445–459. [PubMed: 17958156]

Barrett TS, & Brignone E (2017). Furniture for Quantitative Scientists. R J, 9(2), 142.

Beitchman JH, Jiang H, Koyama E, Johnson CJ, Escobar M, Atkinson L, Brownlie E, & Vida R (2008). Models and determinants of vocabulary growth from kindergarten to adulthood. Journal of Child Psychology and Psychiatry, 49(6), 626–634. [PubMed: 18341544]

Benton AL (1955). The Benton Visual Retention Test. In. New York: The Psychological Corporation.

Binder JR, Desai RH, Graves WW, & Conant LL (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral cortex, 19(12), 2767–2796. [PubMed: 19329570]

Bock K, & Levelt W (1994). Language production: Grammatical encoding. In Gernsbacher MA (Ed.), Handbook of Psycholinguistics (pp. 945–984). San Diego, CA: Academic Press.

Brysbaert M, & Biemiller A (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. Behavior research methods, 49(4), 1520–1523. [PubMed: 27659480]

Brysbaert M, & New B (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods, 41(4), 977–990. [PubMed: 19897807]

Brysbaert M, Van Wijnendaele I, & De Deyne S (2000). Age-of-acquisition effects in semantic processing tasks. Acta Psychologica, 104(2), 215–226. [PubMed: 10900706]

Buschke H, & Fuld PA (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. Neurology, 24, 1019–1025. [PubMed: 4473151]

Butler CR, Brambati SM, Miller BL, & Gorno-Tempini M-L (2009). The neural correlates of verbal and non-verbal semantic processing deficits in neurodegenerative disease. Cognitive and behavioral neurology: official journal of the Society for Behavioral and Cognitive Neurology, 22(2), 73. [PubMed: 19506422]

Cuetos F, Barbón A, Urrutia M, & Domínguez A (2009). Determining the time course of lexical frequency and age of acquisition using ERP. Clinical Neurophysiology, 120(2), 285–294. [PubMed: 19101202]

Dragulescu AA, Dragulescu MAA, & Provide R (2020). Package 'xlsx'. Cell, 9(1).

Eng N, Vonk JM, Salzberger M, & Yoo N (2019). A cross-linguistic comparison of category and letter fluency: Mandarin and English. Q J Exp Psychol (Hove), 72(3), 651–660. doi:10.1177/1747021818765997 [PubMed: 29512423]

Evans JD (1996). Straightforward statistics for the behavioral sciences. Pacific Grove, CA: Thomson Brooks/Cole Publishing Co.

Fratiglioni L, De Ronchi D, & Agüero-Torres H (1999). Worldwide prevalence and incidence of dementia. Drugs & aging, 15(5), 365–375. [PubMed: 10600044]

Gazzaley A, & D'esposito M (2007). Top-down modulation and normal aging. Annals of the New York Academy of Sciences, 1097(1), 67–83. [PubMed: 17413013]

Glymour MM, Weuve J, & Chen JT (2008). Methodological challenges in causal research on racial and ethnic patterns of cognitive trajectories: measurement, selection, and bias. Neuropsychology review, 18(3), 194–213. [PubMed: 18819008]

Gomez RG, & White DA (2006). Using verbal fluency to detect very mild dementia of the Alzheimer type. Arch Clin Neuropsychol, 21(8), 771–775. doi:10.1016/j.acn.2006.06.012 [PubMed: 17011743]

Goodglass H, & Kaplan D (1983). The assessment of aphasia and related disorders (2nd ed.). Philadelphia: Lea and Febiger.

Greene JD, Baddeley AD, & Hodges JR (1996). Analysis of the episodic memory deficit in early Alzheimer's disease: evidence from the doors and people test. Neuropsychologia, 34(6), 537–551. [PubMed: 8736567]

Grossman M, Koenig P, DeVita C, Glosser G, Alsop D, Detre J, & Gee J (2002). The neural basis for category-specific knowledge: an fMRI study. NeuroImage, 15(4), 936–948. [PubMed: 11906234]

Hanninen T, & Soininen H (1997). Age-associated memory impairment. Normal aging or warning of dementia? Drugs and Aging, 11, 480–489. [PubMed: 9413705]

Hixson JE, & Vernier DT (1990). Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. Journal of lipid research, 31(3), 545–548. [PubMed: 2341813]

IBM Corp. (2020). IBM SPSS Statistics for Windows, Version 27. Armonk, NY: IBM Corp.

Kaplan EF, Goodglass H, & Weintraub S (1983). The Boston Naming Test, 2nd Edition. Philadelphia, PA: Lea & Febiger.

Kassambara A, & Kassambara MA (2020). Package 'ggpubr'. R package version 0.1, 6.

Kuperman V, Stadthagen-Gonzalez H, & Brysbaert M (2012). Age-of-acquisition ratings for 30,000 English words. Behavior Research Methods, 44(4), 978–990. [PubMed: 22581493]

Kurland BF, Johnson LL, Egleston BL, & Diehr PH (2009). Longitudinal data with follow-up truncated by death: match the analysis method to research aims. Statistical science: a review journal of the Institute of Mathematical Statistics, 24(2), 211. [PubMed: 20119502]

Lambon Ralph MA, Graham KS, Ellis AW, & Hodges JR (1998). Naming in semantic dementia—what matters? Neuropsychologia, 36(8), 775–784. [PubMed: 9751441]

Libon DJ, Rascovsky K, Powers J, Irwin DJ, Boller A, Weinberg D, McMillan CT, & Grossman M (2013). Comparative semantic profiles in semantic dementia and Alzheimer's disease. Brain, 136(8), 2497–2509. [PubMed: 23824492]

Liverman CT, Yaffe K, & Blazer DG (2015). Cognitive aging: Progress in understanding and opportunities for action.

Mattis S, Bellak L, & Karasu TB (1976). Mental Status examination for organic mental syndrome in the elderly patient. New York: Grune & Stratton.

McKhann GM, Drachman D, Folstein M, Katzman R, Price D, & Stadlan EM (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology, 34(7), 939–944. [PubMed: 6610841]

Medin D, Waxman S, Woodring J, Ross N, & Winkler-Rhoades N (2010). Naming the animals that come to mind: Effects of culture and experience on category fluency. Journal of Cognition and Culture, 10(1–2), 205–220.

Mistridis P, Krumm S, Monsch AU, Berres M, & Taylor KI (2015). The 12 Years Preceding Mild Cognitive Impairment Due to Alzheimer's Disease: The Temporal Emergence of Cognitive Decline. J Alzheimers Dis, 48(4), 1095–1107. doi:10.3233/jad-150137 [PubMed: 26402083]

Morrison CM, & Ellis AW (2000). Real age of acquisition effects in word naming and lexical decision. British Journal of Psychology, 91(2), 167–180. [PubMed: 10832512]

Mueller KD, Koscik RL, LaRue A, Clark LR, Hermann B, Johnson SC, & Sager MA (2015). Verbal Fluency and Early Memory Decline: Results from the Wisconsin Registry for Alzheimer's Prevention. Arch Clin Neuropsychol, 30(5), 448–457. doi:10.1093/arclin/acv030 [PubMed: 26025231]

Muthén L, & Muthén B (2019). Mplus. The comprehensive modelling program for applied researchers: user's guide.

Mutter SA, & Hashtroudi S (1987). Cognitive effort and the word frequency effect in recognition and lexical decision. The American journal of psychology, 93–116. [PubMed: 3592027]

Nebes RD, & Brady CB (1990). Preserved organization of semantic attributes in Alzheimer's disease. Psychology and Aging, 5(4), 574. [PubMed: 2278683]

Pakhomov SVS, & Hemmy LS (2014). A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the nun study. Cortex, 55, 97–106. doi:10.1016/j.cortex.2013.05.009 [PubMed: 23845236]

Papp KV, Mormino EC, Amariglio RE, Munro C, Dagley A, Schultz AP, Johnson KA, Sperling RA, & Rentz DM (2016). Biomarker validation of a decline in semantic processing in preclinical Alzheimer's disease. Neuropsychology, 30(5), 624–630. doi:10.1037/neu0000246 [PubMed: 26595826]

Park DC, Lautenschlager G, Hedden T, Davidson NS, Smith AD, & Smith PK (2002). Models of visuospatial and verbal memory across the adult life span. Psychology and Aging, 17, 299–320. [PubMed: 12061414]

Pedraza O, & Mungas D (2008). Measurement in Cross-Cultural Neuropsychology. Neuropsychology Review, 18(3), 184–193. [PubMed: 18814034]

Pike KE, Ellis KA, Villemagne VL, Good N, Chetelat G, Ames D, Szoeke C, Laws SM, Verdile G, Martins RN, Masters CL, & Rowe CC (2011). Cognition and beta-amyloid in preclinical Alzheimer's disease: data from the AIBL study. Neuropsychologia, 49(9), 2384–2390. doi:10.1016/j.neuropsychologia.2011.04.012 [PubMed: 21529702]

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. In. Vienna, Austria.

Raoux N, Amieva H, Le Goff M, Auriacombe S, Carcaillon L, Letenneur L, & Dartigues J-F (2008). Clustering and switching processes in semantic verbal fluency in the course of Alzheimer's disease subjects: Results from the PAQUID longitudinal study. Cortex, 44(9), 1188–1196. [PubMed: 18761132]

Rosen VM, Sunderland T, Levy J, Harwell A, McGee L, Hammond C, Bhupali D, Putnam K, Bergeson J, & Lefkowitz C (2005). Apolipoprotein E and category fluency: evidence for reduced semantic access in healthy normal controls at risk for developing Alzheimer's disease. Neuropsychologia, 43(4), 647–658. doi:10.1016/j.neuropsychologia.2004.06.022 [PubMed: 15716154]

Rosen W (1981). The Rosen Drawing Test. Bronx, NY: Veterans Administration Medical Center.

Salthouse TA (2004). What and when of cognitive aging. Current directions in psychological science, 13(4), 140–144.

Salthouse TA (2010). Selective review of cognitive aging. Journal of the International neuropsychological Society, 16(5), 754–760. [PubMed: 20673381]

Salthouse TA (2016). Theoretical perspectives on cognitive aging: Psychology Press.

Schloerke B, Crowley J, & Cook D (2018). Package 'GGally'. Extension to 'ggplot2.'See, 713.

Shao Z, Janse E, Visser K, & Meyer AS (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. Frontiers in Psychology, 5(772), 1–10. doi:10.3389/fpsyg.2014.00772 [PubMed: 24474945]

Shaoul C, & Westbury C (2010). Exploring lexical co-occurrence space using HiDEx. Behavior Research Methods, 42(2), 393–413. [PubMed: 20479171]

Siedlecki KL, Honig LS, & Stern Y (2008). Exploring the structure of a neuropsychological battery across healthy elders and those with questionable dementia and Alzheimer's disease. Neuropsychology, 22, 400–411. [PubMed: 18444718]

Sienski G, Narayan P, Bonner JM, Kory N, Boland S, Arczewska AA, Ralvenius WT, Akay L, Lockshin E, & He L (2021). APOE4 disrupts intracellular lipid homeostasis in human iPSC-derived glia. Science translational medicine, 13(583), eaaz4564. [PubMed: 33658354]

Stern Y, Andrews H, Pittman J, Sano M, Tatemichi T, Lantigua R, & Mayeux R (1992). Diagnosis of dementia in a heterogeneous population. Development of a neuropsychological paradigm-based diagnosis of dementia and quantified correction for the effects of education. Archives of Neurology, 49(5), 453–460. [PubMed: 1580806]

Stern Y, Arenaza-Urquijo EM, Bartrés-Faz D, Belleville S, Cantilon M, Chetelat G, Ewers M, Franzmeier N, Kempermann G, & Kremen WS (2018). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. Alzheimer's & Dementia, 1–7. doi:10.1016/j.jalz.2018.07.219

Stine-Morrow EA, & Manavbasi IE (2022). Beyond "use it or lose it": the impact of engagement on cognitive aging. Annual Review of Developmental Psychology, 4, 319–352.

Tang MX, Cross P, Andrews H, Jacobs DM, Small S, Bell K, Merchant C, Lantigua R, Costa R, Stern Y, & Mayeux R (2001). Incidence of Alzheimer's disease in African-Americans, Caribbean Hispanics and Caucasians in northern Manhattan. Neurology, 56, 49–56. [PubMed: 11148235]

Troyer AK, Moscovitch M, & Winocur G (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. neuropsychology, 11(1), 138. [PubMed: 9055277]

Troyer AK, Moscovitch M, Winocur G, Leach L, & Freedman M (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. Journal of the International Neuropsychological Society, 4(2), 137–143. [PubMed: 9529823]

United States Bureau of the Census. (1991). Census of Population and Housing 1990. Summary tape file 1, technical documentation prepared by the Bureau of the Census. Washington, D.C.: Bureau of the Census.

Venneri A, Jahn-Carta C, de Marco M, Quaranta D, & Marra C (2018). Diagnostic and prognostic role of semantic processing in preclinical Alzheimer's disease. Biomark Med, 12(6), 637–651. doi:10.2217/bmm-2017-0324 [PubMed: 29896968]

Verma M, & Howard RJ (2012). Semantic memory and language dysfunction in early Alzheimer's disease: a review. International journal of geriatric psychiatry, 27(12), 1209–1217. [PubMed: 22298328]

Vonk JMJ (2023). Vonk Lab GitHub.

Vonk JMJ, Borghesani V, Battistella G, Younes K, DeLeon J, Welch A, Hubbard HI, Miller ZA, Miller BL, & Gorno-Tempini ML (2019). Verbal semantics and the left dorsolateral anterior temporal lobe: A longitudinal case of bilateral temporal degeneration. Aphasiology. doi:10.1080/02687038.2019.1659935

Vonk JMJ, Bouteloup V, Mangin JF, Dubois B, Blanc F, Gabelle A, Ceccaldi M, Annweiler C, Krolak-Salmon P, & Belin C (2020). Semantic loss marks early Alzheimer's disease-related neurodegeneration in older adults without dementia. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 12(1), e12066.

Vonk JMJ, Flores RJ, Rosado D, Qian C, Cabo R, Habegger J, Louie K, Allocco E, Brickman AM, & Manly JJ (2019). Semantic network function captured by word frequency in nondemented APOE ε4 carriers. Neuropsychology, 33(2)(2), 256–262. doi:10.1037/neu0000508 [PubMed: 30489116]

Vonk JMJ, Jonkers R, Hubbard HI, Gorno-Tempini ML, Brickman AM, & Obler LK (2019). Semantic and lexical features of words dissimilarly affected by non-fluent, logopenic, and semantic primary progressive aphasia. Journal of the International Neuropsychological Society, 25(10), 1011–1022. [PubMed: 31511121]

Vonk JMJ, Rizvi B, Lao PJ, Budge M, Manly JJ, Mayeux R, & Brickman AM (2019). Letter and category fluency performance correlates with distinct patterns of cortical thickness in older adults. Cerebral Cortex, 29(6), 2694–2700. doi:10.1093/cercor/bhy138 [PubMed: 29893804]

Wang P, Shen W, & Boye ME (2012). Joint modeling of longitudinal outcomes and survival using latent growth modeling approach in a mesothelioma trial. Health Services and Outcomes Research Methodology, 12(2–3), 182–199. [PubMed: 22773919]

Wechsler D (1981). Wechsler Adult Intelligence Scale-Revised. New York, NY: The Psychological Corporation.

Whitwell JL (2010). Progression of atrophy in Alzheimer's disease and related disorders. Neurotoxicity research, 18(3), 339–346. [PubMed: 20352396]

Wickham H (2016). ggplot2: elegant graphics for data analysis: Springer.

Wickham H, Francois R, Henry L, & Müller K (2019). dplyr: A Grammar of Data Manipulation. R package version 0.4.3.

Wickham H, & Miller E (2018). Haven: Import and export'SPSS','stata'and'SAS'files. R package version, 1(0).

Zahodne LB, Schofield PW, Farrell MT, Stern Y, & Manly JJ (2014). Bilingualism does not alter cognitive decline or dementia risk among Spanish-speaking immigrants. Neuropsychology, 28(2), 238. [PubMed: 24188113]

Zemla JC, Cao K, Mueller KD, & Austerweil JL (2020). SNAFU: The Semantic Network and Fluency Utility. Behavior Research Methods, 1–19.

Zhang X-X, Tian Y, Wang Z-T, Ma Y-H, Tan L, & Yu J-T (2021). The epidemiology of Alzheimer's disease modifiable risk factors and prevention. The journal of prevention of Alzheimer's disease, 8(3), 313–321.

Public Significance Statement

This study found that alternative scores of semantic fluency (an existing cognitive test), based on each item instead of the total score, can provide additional information to reveal subtle cognitive impairment among older adults without dementia beyond existing cognitive measures. These findings are important because they show the value that item-level information may have in the search for sensitive and accessible cognitive markers for early detection of dementia.

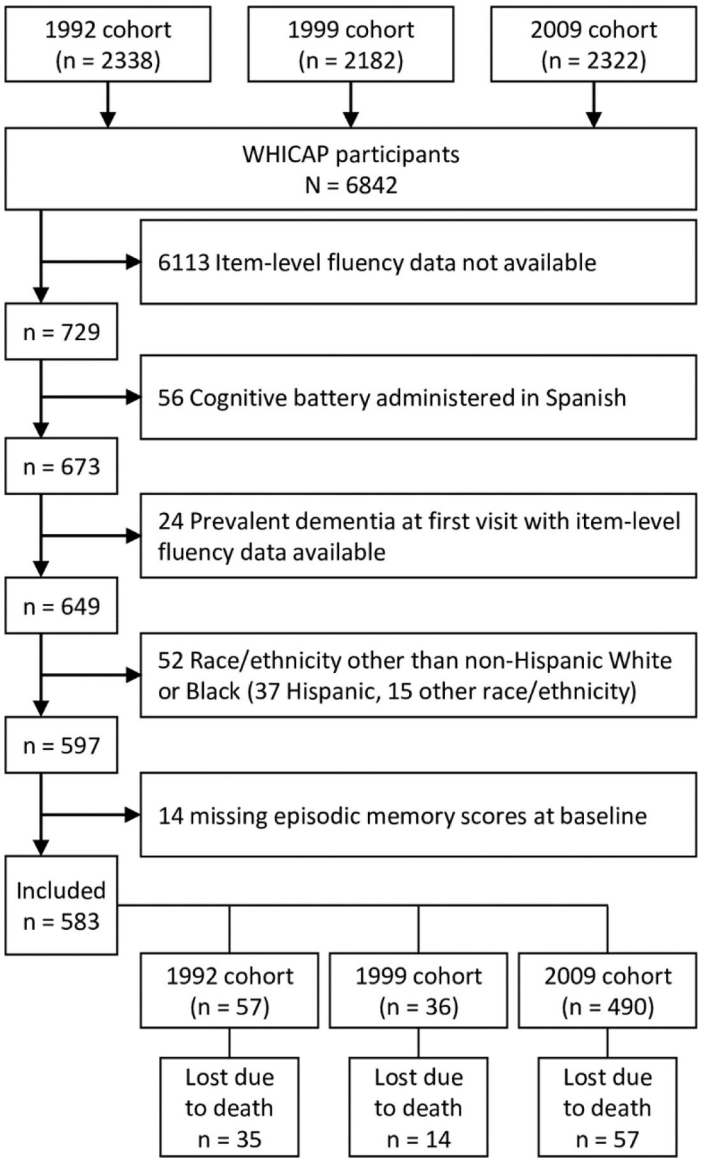

Figure 1.
Flowchart participant selection

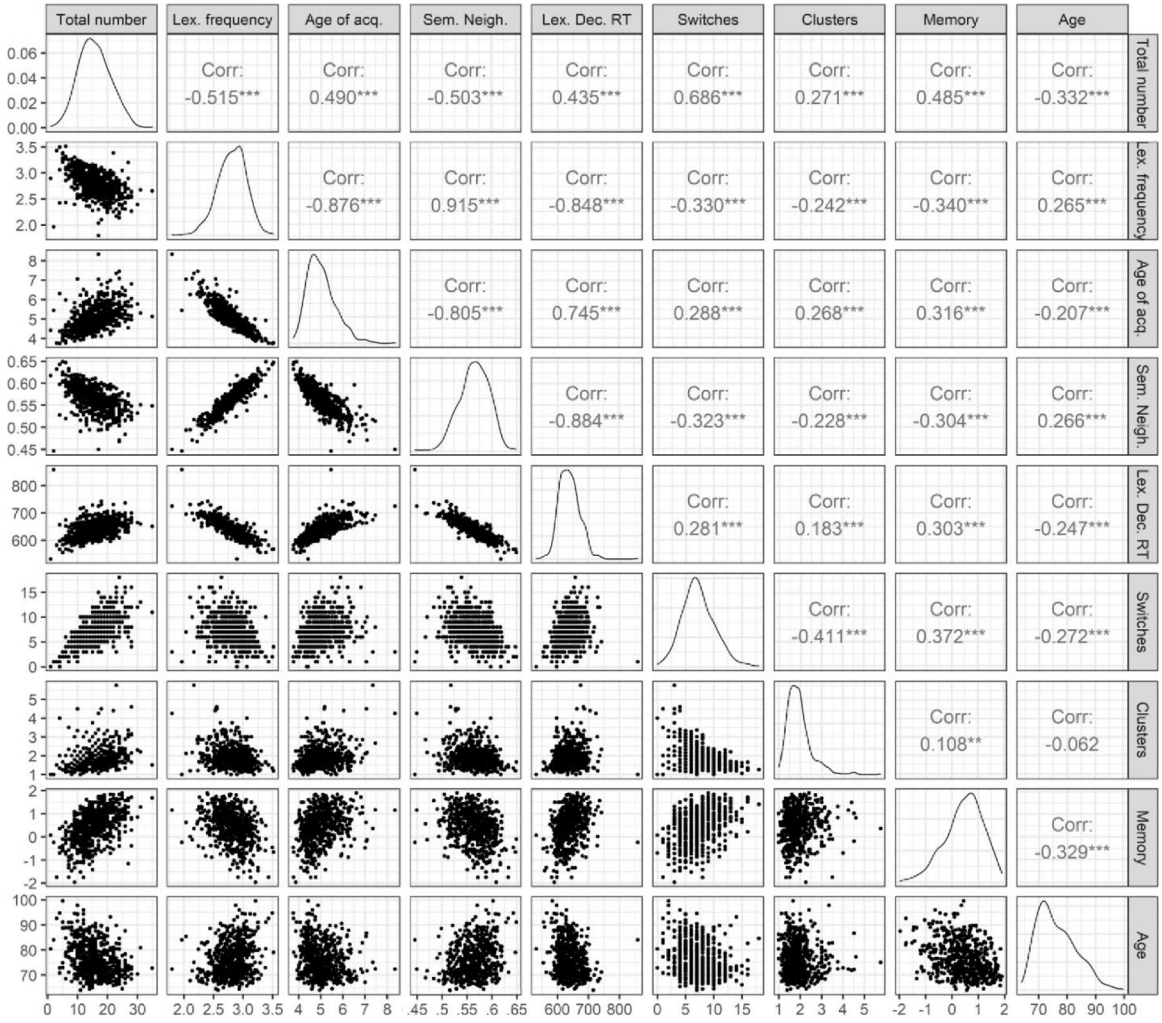

Figure 2.

Scatterplot (lower triangle) and correlation matrix (upper triangle) of the semantic fluency metrics, memory performance, and age at baseline, including density plots (diagonal); *p < .05, **p < .01, ***p < .001

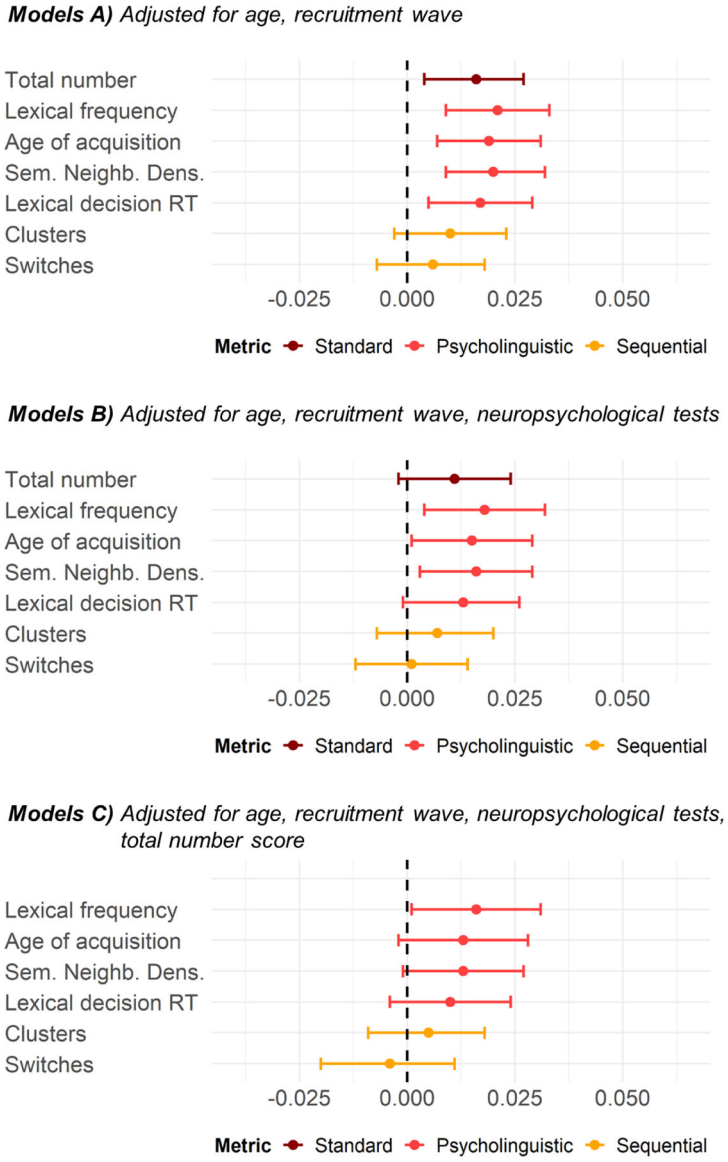

Figure 3.

Estimates of the seven fluency metrics in relation to slope of memory decline

Table 1.

Participant characteristics

	Overall sample (N = 583)
Age	76.3 (6.8, 63.9–99.7)
Sex/gender (women)	370 (63.5%)
Race (Black)	333 (57.1%)
Education in years	13.8 (3.4, 1–20)
APOE status (ε4+)	175 (30%)
SCC (Yes)	433 (74.3%)
Time in study	4.3 (3.1, 0–10.8)
Number of visits	2.9 (1.4, 1–5)
Memory baseline (WHICAP z-score)	0.46 (0.72, −1.96–1.9)
Fluency metrics	
Total number	15.7 (5.3, 1–35)
Lexical frequency	2.81 (0.2, 1.8–3.5)
Age of acquisition	5.0 (0.7, 3.8–8.3)
Sem. Neighb. Dens.	0.56 (0.0, 0.4–0.6)
Lexical decision RT	638.4 (33.0, 531.3–858.8)
Clusters	1.9 (0.6, 1–5.8)
Switches	7.3 (2.9, 0–18)
Neuropsychological battery	
Letter fluency	11.97 (4.6, 1–27)
Boston Naming Test	14.02 (1.5, 7–15)
Similarities	15.75 (6.9, 0–28)
Repetition	7.74 (0.6, 3–8)
Comprehension	5.58 (0.7, 1–6)
Recognition	7.95 (1.7, 1–10)
Matching	9.16 (1.3, 0–10)
Rosen Drawing Test	14.7 (1.3, 5–16)
Identities and Oddities	2.93 (1.1, 0–5)

Note. Measures are m (SD; range) for continuous variables and n (%) for categorical variables; m = mean, sd = standard deviation, SCC = subjective cognitive complaints

Table 2.

Intercept and slope estimates of the different semantic fluency metrics in relation to future memory decline over time and the sensitivity analysis to account for loss-to-follow-up due to death

		Total number	Lexical frequency	Age of acquisition	Sem. Neighb. Dens.	Lexical decision RT	Clusters	Switches
Model A (adjusted for age and recruitment wave)								
Overall sample	intercept	.309 [.256, .361]*	.179 [.126, .231]*	.173 [.122, .224]*	.154 [.101, .206]*	.153 [.107, .200]*	.062 [.004, .119]*	.224 [.167, .281]*
	slope	.016 [.004, .027]*	.021 [.009, .033]*	.019 [.007, .031]*	.020 [.009, .032]*	.017 [.005, .029]*	.010 [−.003, .023]	.006 [−.007, .018]
Sensitivity analysis	intercept	.309 [.256, .362]*	.179 [.126, .232]*	.174 [.121, .226]*	.154 [.101, .208]*	.155 [.108, .202]*	.063 [.004, .121]*	.224 [.166, .282]*
	slope	.016 [.005, .028]*	.021 [.008, .033]*	.018 [.006, .030]*	.019 [.007, .031]*	.016 [.004, .028]*	.010 [−.004, .023]	.006 [−.006, .019]
Model B (Model A + other neuropsychological tests)								
Overall sample	intercept	.159 [.100, .219]*	.031 [−.026, .088]	.023 [−.029, .076]	.019 [−.036, .074]	.020 [−.037, .076]	.004 [−.051, .058]	.107 [.048, .167]*
	slope	.011 [−.002, .024]	.018 [.004, .032]*	.015 [.001, .029]*	.016 [.003, .029]*	.013 [−.001, .026]	.007 [−.007, .020]	.001 [−.012, .014]
Model C (Model B + total number score)								
Overall sample	intercept	NA	−.007 [−.067, .052]	−.012 [−.066, .042]	−.021 [−.077, .036]	−.006 [−.062, .051]	−.025 [−.079, .030]	.043 [−.026, .113]
	slope	NA	.016 [.001, .031]*	.013 [−.002, .028]	.013 [−.001, .027]	.010 [−.004, .024]	.005 [−.009, .018]	−.004 [−.020, .011]

Note. Estimates with 95% confidence interval;

* p < .05; NA = not applicable