



Published in final edited form as:

*Exp Aging Res.* 2016 ; 42(4): 315–328. doi:10.1080/0361073X.2016.1191838.

## Population normative data for the CERAD Word List and Victoria Stroop Test in younger- and middle-aged adults: Cross-sectional analyses from the Framingham Heart Study

Lisa D. Hankee, PsyD<sup>1,2</sup>, Sarah R. Preis, ScD<sup>2,3</sup>, Ryan J. Piers, BS<sup>1,2</sup>, Alexa S. Beiser, PhD<sup>1,2,3</sup>, Sherral A. Devine, PhD<sup>1,2</sup>, Yulin Liu, MS<sup>1,2</sup>, Sudha Seshadri, MD<sup>1,2</sup>, Philip A. Wolf, MD<sup>1,2</sup>, and Rhoda Au, PhD<sup>1,2</sup>

<sup>1</sup>Boston University School of Medicine, Boston, MA

<sup>2</sup>Framingham Heart Study/National Heart Lung and Blood Institute, Framingham, MA

<sup>3</sup>Boston University School of Public Health, Boston, MA

### Abstract

**Objective**—To provide baseline normative data on tests of verbal memory and executive function for non-demented young to middle age adults.

**Methods**—The Consortium to Establish a Registry for Alzheimer’s Disease Word List task (CERAD-WL) and Victoria Stroop Test (VST) were administered to 3362 Framingham Heart Study (FHS) volunteer participants aged 24–78 years. Analyses of the effects of age, sex and education were conducted. Normative data on traditional measures and error responses are reported for each test.

**Results**—Traditional measures were significantly associated with both age and education in this younger-aged cohort. Error responses also evidenced significant age and education effects.

**Conclusion**—These data provide a normative comparison for assessment of verbal memory and executive functioning capabilities in young adults and may be utilized as a tool for preclinical studies of disease in younger aged adults.

### Keywords

aging; cognition; mild cognitive impairment; dementia; executive functioning; memory

### Introduction

Research on Alzheimer’s disease has recently centered on identifying biomarkers in asymptomatic adults decades before onset of clinically overt symptoms. This focus on the preclinical period has led to assessment of cognitive function in young to middle age adults, often using the same neuropsychological tests that differentiated between intact versus impaired performance. A number of well-used tests, however, have been validated and

normed on small and/or older samples. Of significant need are normative values on these tests, particularly those assessing verbal memory and executive function, from younger to middle aged adults to serve as baseline measures for comparative studies with similar preclinical study samples.

The Consortium to Establish a Registry for Alzheimer's Disease Word List Memory Task (CERAD) neuropsychological battery (Morris, Mohs, Rogers et al., 1988), a relatively brief (i.e., 20-30 minutes) assessment of each cognitive domain (Morris et al., 1989), is often used to evaluate cognitive functioning in individuals thought to be at risk for Alzheimer's Disease (AD, Beeri et al., 2006). Recall memory is the hallmark early sign of AD, and the word list delayed recall has been found to effectively distinguish between individuals with dementia from those who are cognitively normal (Fillenbaum et al., 2008). Normative data are available for elderly individuals (Fillenbaum et al., 2005; Fillenbaum, Heyman, Huber et al., 2001; Collie et al., 1999) and for middle-aged to elderly members of diverse cultural and ethnic populations (e.g., Whyte et al., 2005; Welsh et al., 1994; McCurry et al., 2001; Unverzagt et al., 1996), and the reliability and utility of individual subtests of the CERAD in assessing cognitive decline have been established (Morris et al., 1993; Welsh, Butters, Hughes et al., 1992; Welsh-Bohmer & Mohs, 1997). However, no normative data has been published for adults <50 years of age.

The Victoria Stroop Test (VST; Regard, 1981), a technique for assessing executive functions (e.g., selective attention, cognitive flexibility), is another test utilized in the clinical evaluation of cognition and the assessment of cognitive impairment in a geriatric population (Strauss, Sherman, & Spreen, 2006). Studies have reported the Stroop effect to be a highly sensitive tool for differentiating elderly individuals with mild AD, mild cognitive impairment (MCI) and normal cognition (Bondi et al., 2002; Kramer et al., 2006). Executive functioning performance has also been associated with stroke risk factors and measures of brain atrophy (Debette et al., 2011; Seshadri et al., 2004) in non-demented younger- and middle-aged adults. The original norms (Regard, 1981) were published with a wide age range but of limited sample size (n=126; age range 20-94 years), and therefore may not capture the full spectrum across age, education and gender. Additional normative data are published for elderly (aged 50+) individuals (Bayard, Erkes, & Moroni, 2011) and for a relatively small (n=272) community sample of individuals aged 18-94 (Troyer, Leach, & Strauss, 2006).

For each of these tests, there is a need for normative age, sex and educational data on younger-aged adults. This need is emphasized by the recently published preclinical criteria for AD (Sperling et al., 2011). Early assessment of memory and executive functioning is essential for accurate identification of those younger aged individuals with a heightened risk for cognitive decline. In addition, these norms provide a baseline assessment of young adults, which permits longitudinal comparison of change in cognition across the lifespan.

The focus of this study is to provide normative data for the CERAD-WL and VST from a large, community-based study with a younger-aged cohort. We anticipate these normative data may be utilized in baseline assessment of cognitive functioning of younger adults and

facilitate the differentiation of cognitive changes attributed to normal aging from those potentially associated with a pathological disease processes.

## Methods

### Participants

Established in 1948, the Framingham Heart Study (FHS) recruited 5209 participants (the Original cohort) for a longitudinal study designed to identify common characteristics contributing to cardiovascular disease. In 1971, the biological children of the Original cohort and their spouses (the Offspring cohort) were recruited for participation (Kannel & McGee, 1979). Most recently, in 2001, a third generation of participants (Gen 3), the grandchildren of Gen 1 and children of Gen 2, was recruited for studies of genetic heritability of cardiovascular and cerebrovascular diseases (Splansky et al., 2007).

A total of 3411 Gen3 participants completed the second exam cycle in 2008-2010, which included a detailed medical history, physical examination, laboratory tests, and cognitive screening. Table 1 provides demographic information for the study sample. Participants were excluded from the analysis due to missing education status (n=10), incomplete CERAD and Victoria Stroop Test data (n=37), or needing consent by substituted judgment (n=2). Therefore, a total of 3362 participants (53% women) comprised the normative sample for this study. The CERAD word list task and Victoria Stroop Test were administered using standard administration procedures. The Institutional Review Board (IRB) at Boston University Medical Center (BUMC) approved the study protocol. Informed consent was obtained from all participants.

### CERAD-WL Administration and Scoring Procedures

The Consortium to Establish a Registry for Alzheimer's Disease Word List Memory Task (CERAD-WL) assesses the ability to learn and remember verbal information. The test consists of three learning trials, a delayed recall trial and a recognition trial (for detailed description, see: Morris et al., 1989). The examiner transcribed verbatim and in sequential order every participant response from the learning and recall trials, including erroneous responses (e.g., perseverations, intrusions).

The number of words correctly recalled was calculated for each of the three learning trials. The total traditional score for immediate verbal memory is comprised of a sum of correct responses collectively across the three learning trials. The maximum number of correct responses is 30 (i.e., 10 for each of the three trials), with higher numbers indicating better traditional learning performance. The error score measures of perseverations and intrusions were also totaled across the three learning trials. Lower numbers of error commissions indicate better error score performance.

On the delayed verbal memory trial, the traditional score was the total number of words correctly recalled after presentation of a distractor task. The maximum value for delayed recall was 10, with higher values indicating better traditional memory performance. Error score measures of perseverations and intrusions were also coded on the delayed recall trial, and fewer error commissions indicated better error score performance. An additional percent

savings score (i.e., percent retention) was calculated to determine the amount of information originally encoded that was later recalled. The CERAD percent savings was calculated by dividing the score on the delayed recall by the score on the third learning trial.

For the recognition trial, the total score was calculated as the number of correct answers (i.e., true positives + true negatives). Recognition scores ranged from 0-20, with higher values indicating better recognition performance.

### **VST Administration and Scoring Procedures**

The VST, a measure of executive functioning, is utilized in the assessment of cognitive flexibility, response inhibition and selective attention. The test consists of three successive trials administered via the use of different stimulus cards: a colored dots trial (trial 1), a non-colored words trial (trial 2) and a colored words trial (trial 3). For a detailed description of the VST, see Regard (1981).

For each trial, a traditional score (total time to completion), and error score (total number of errors) were calculated. In addition, error scores were derived to assess interference effects. Two definitions of the interference score were used. "Version 1" was calculated by dividing the time to complete the colored words trial (trial 3) by the time to complete the colored dots trial (trial 1). "Version 2" was calculated by dividing the time to complete the non-colored words trial (trial 2) by the time to complete the colored dots trial (trial 1). Faster times to completion (i.e., lower scores) are indicative of better traditional performance. Lower values for total number of errors committed are considered better error score performance.

### **Statistical Analyses**

For all variables, means and standard deviations were calculated for the total study sample as well as for each age group (i.e., <35, 35-44, 45-54, 55 years) and for each level of educational achievement (i.e., high school, some college, college degree, graduate degree). For the Percent Savings score, all scores greater than 100% were set to a maximum of 100% retention. Due to the small number of errors, variables for CERAD delayed recall errors and for VST total errors were dichotomized to 1 error versus zero errors. Analysis of variance (ANOVA) was used to compare means across age groups and educational groups. For categorical variables, a chi-square test was used to compare differences across age and education groups. The resulting sample sizes were insufficient for reporting combined normative data of age by education. A p-value of <0.05 was considered statistically significant. All statistical analyses were done using SAS version 9.2 (Cary, NC).

### **Results**

Significant effects of age and education were observed for traditional measures and several error score measures. Overall, traditional scores declined with advancing age and improved with higher levels of education up until a college degree. This linear trend did not hold for education beyond a college degree, and in several subtests, average scores did not improve from college degree to graduate degree. Error score performance, evidenced by the rate of error commission, generally showed effects for age and education, with more errors committed in the older age groups and fewer errors committed by those with higher levels of

education. Tables 2-3 contain the descriptive statistics (mean and standard deviation) for traditional variables on the CERAD-WL and the VST, stratified by age and educational attainment.

For the CERAD-WL, there were significant effects of age ( $p < 0.0001$ ) and education ( $p < 0.0001$ ) for each traditional measure. Based on the means, traditional performance appeared to decline with increasing age and improve with higher levels of education. The mean score across age and educational group was 22.6 (sd=3.2) words. Median scores combined across the three learning trials ranged from 22-24 words across age group and 21-23 words across education group. On average, participants recalled 6.8 (sd=1.9) of the 10 words at the delayed recall trial. The percent of information retained across the timeframe between the learning condition and the recall condition (i.e., percent savings score) was 79.8% (sd=20.7).

On the VST, significant effects were noted for both age and education ( $p < 0.0001$ ) on each traditional measure. Overall, the mean completion time (in seconds) across age and educational group was 12.5 (sd=2.7) for trial 1, 15.0 (sd=3.4) for trial 2, and 24.5 (sd=7.3) for trial 3. On average, it took participants approximately twice as long and 1.5 times as long to complete the third trial when compared to the first and second trials, respectively. The interference score reflecting the time to complete the third trial divided by the time to complete the first trial was 2.0 (sd=0.6). A second interference score, reflecting the time to complete the third trial divided by the time to complete the second trial was 1.6 (sd=0.4). Effects of age and education were evident in both interference scores, with the means of each appearing to increase with age and decrease with education. These scores are an indication that older and less educated adults, in comparison to younger-aged and more educated adults, take longer to complete the executive functioning component even after accounting for differences in completion time on the previous two trials (i.e., naming color of dots and naming color of non-color words).

Table 4 contains normative data on error score measures for the CERAD-WL and the VST, stratified by age and educational attainment. The error score variables for CERAD-WL delayed recall errors and VST errors were dichotomized into two groups: those who exhibited an error-free performance and those whose performance included one or more errors.

On the CERAD-WL, after categorizing delayed recall error commission into two groups (i.e., error-free performance vs. participants who made 1+ errors), there was a significant age effect, indicating that older participants were more likely to make at least one error ( $p = 0.0002$ ). On average, approximately half of the participants made 1 error on this delayed recall trial. There was a significant education effect on immediate recall for error commission ( $p = 0.0008$ ) such that errors were made less frequently by participants with higher levels of education.

On the VST, after dichotomizing the groups, significant effects were noted for both age and education ( $p < 0.0001$ ), in the same direction as other significant effects. More than half of participants (57%) committed at least one error.

## Discussion

This is the largest cohort study to our knowledge reporting normative data for young- and middle-aged, cognitively-healthy adults on the CERAD-WL and the VST, two widely used measures of cognitive functioning. Also somewhat novel to most normative study is the separation of total score and error performance, particularly for the CERAD-WL. The significance is that in a younger, asymptomatic population, differentiation in performance levels is more difficult to discern using traditional total scores as an outcome metric. Higher than average rates of errors, even with within normal traditional scores may reflect different risk for future cognitive decline compared to low/normal error rates.

In this young- to middle-aged adult sample, there were significant effects of both age and education on each of the CERAD-WL traditional scores (i.e., total number of words recalled on the recall trials, correct identification of words on the recognition trial, and percent savings). Age and education also significantly impacted completion times for each of the three VST trials and the VST inference measures. Gender effects were non-significant.

Overall, traditional scores declined with advancing age and improved with higher levels of education up until a college degree. This linear trend did not hold for education beyond a college degree, and in several subtests, average scores did not improve from college degree to graduate degree.

The cognitive reserve hypothesis suggests that people with higher cognitive reserve are more protected against cognitive decline associated with brain pathology than those with lower cognitive reserve (Stern, 2009; Stern, 2002). Evidence indicates that level of schooling can bolster cognitive reserve; that is, individuals with higher levels of educational attainment may see a delay in cognitive symptoms associated with AD brain damage (Ewers et al., 2013; Pernecky et al., 2006) and dementia (Schmand et al., 1997; Ott et al., 1995; Stern et al., 1994).

Our results suggest that the typical linear relationship of improved performance with increasing education may have a threshold at which additional education no longer shows protective effects. It also appears that this education attainment advantage on test performance may be more related to verbal memory compared to executive function. Given the reliance of many studies of Alzheimer's disease on changes in verbal memory as a preclinical indicator, there is a potential bias of false negatives for those more highly educated.

With regard to error score findings, there was a significant educational effect on error commission on the CERAD-WL immediate recall trial, and a significant age effect on the CERAD delayed recall trial. Age and educational effects were significant for the VST when comparing those whose performance was error free to those who committed one or more errors. The commission of errors may indicate subtle changes in cognitive function, with higher rates than normal indicative of increased risk for cognitive decline (Lamar et al., 2010; Libon et al., 2011).

We observed a different pattern in the error data for age vs. education. For CERAD-WL immediate recall, there were few commission of errors overall, and no significant differences by age. While the error rate was low, there was a significant increase in errors across education groups. For CERAD-WL delayed recall, more errors were made compared to immediate recall, but significant differences were only for age and not educational attainment. For VST, there was a linear relationship of increased errors with age and education.

Results demonstrate that error commission on these neuropsychological tests of verbal memory and executive functioning is evident in persons presumed cognitively intact. Error score analysis of test performance revealed that approximately half of this younger-aged, cognitively healthy sample made one or more errors on the VST and the delayed free recall trial of the CERAD wordlist.

While this study reports results from two specific tests, the results can be applied more generally to other tests purported to similarly measure verbal memory and executive function. As discussed above, use of verbal memory tests may be inherently lack sensitivity to detect preclinical changes in more highly educated persons, whereas changes in executive function may more likely be detected across a broader education spectrum. Higher than expected error responses may also provide an additional metric from which to discern preclinical performance changes, though future studies relating errors to disease risk are needed to determine if in fact these responses are clinically meaningful.

Limitations of this study include that this sample is relatively well-educated. Approximately 17% of the sample had a graduate education and an additional 36% were college graduates. The lower education range, especially those with less than a high school education, is under-represented by this study sample. In addition, this cohort is comprised of people of European descent and does not adequately reflect the diverse ethnicities of the broader population. This limits the external validity of the data, and these norms may not be generalizable to non-Caucasian populations. Also, any sample of presumably cognitively-normal individuals may be contaminated with some individuals in the very early stages of a neurodegenerative disease process (Sliwinski, Lipton, Buschke, et al., 1996; De Santi et al., 2008). While this study sample is relatively young, preclinical studies of AD suggest that pathology may be evident decades before clinical symptoms are evident (Morris, 2005; Sperling et al., 2011; Sperling, Karlawish, & Johnson, 2013). Further these normative values reflect cross-sectional performance. Longitudinal follow-up for incident changes will determine what trajectories of change are normal versus those that reflect neurodegenerative decline. Finally, the diagnostic value of these error score neuropsychological data has yet to be determined, and the clinical significance of these error score measures require longitudinal assessment.

Despite these limitations, these analyses provide normative data for a younger-aged cohort for two widely used measures of cognitive functioning. The inclusion of error score data provides the added benefit of establishing error commission as normative behavior and allows for diagnostic comparison of clinical and research samples exhibiting pre-clinical cognitive impairments. The potential of error commission is a preclinical marker in

identifying individuals at various stages of the AD continuum (i.e., preclinical cognitive changes, mild cognitive impairment, clinical dementia) remains to be determined.

## Acknowledgements

This work was supported by the Framingham Heart Study's National Heart, Lung, and Blood Institute contract (N01-HC-25195), by grants (R01-AG16495, R01-AG08122, R01-AG033040) from the National Institute on Aging, and by grant (R01-NS17950) from the National Institute of Neurological Disorders and Stroke.

The authors thank the extraordinary participants and families of the Framingham Heart Study who made this work possible. We also acknowledge the great work of all the research assistants and study staff.

## References

- Bayard S, Erkes J, Moroni C. Victoria Stroop Test: Normative data in a sample group of older people and the study of their clinical applications in the assessment of inhibition in Alzheimer's Disease. *Archives of Clinical Neuropsychology*. 2011; 26:653–661. [PubMed: 21873625]
- Beeri MS, Schmeidler J, Sano M, Wang J, Lally R, Grossman H, Silverman JM. Age, gender, and education norms on the CERAD neuropsychological battery in the oldest old. *Neurology*. 2006; 67(6):1006–1010. [PubMed: 17000969]
- Bondi MW, Serody AB, Chan AS, Ebersone-Shumate SC, Delis DC, Hansen LA, Salmon DP. Cognitive and neuropathologic correlates of Stroop Color-Word Test performance in Alzheimer's disease. *Neuropsychology*. 2002; 16(3):335–343. [PubMed: 12146681]
- Collie A, Shafiq-Antonacci R, Maruff P, Tyler P, Currie J. Norms and the effects of demographic variables on a neuropsychological battery for use in healthy ageing Australian populations. *Australian and New Zealand Journal of Psychiatry*. 1999; 33(4):568–575. [PubMed: 10483853]
- De Santi S, Pirraglia E, Barr W, Babb J, Williams S, Rogers K, Glodzik L, Brys M, Mosconi L, Reisberg B, Ferris S, de Leon MJ. Robust and conventional neuropsychological norms: Diagnosis and predictions of age-related cognitive decline. *Neuropsychology*. 2008; 22(4):469–484. [PubMed: 18590359]
- Debette S, Seshadri S, Beiser A, Au R, Himali JJ, Palumbo C, Wolf PA, DeCarli C. Midlife vascular risk factor exposure accelerates structural brain aging and cognitive decline. *Neurology*. 2011; 77:461–468. [PubMed: 21810696]
- Ewers M, Insel PS, Stern Y, Weiner MW. Cognitive reserve associated with FDG-PET in preclinical Alzheimer disease. *Neurology*. 2013; 80:1194–1201. [PubMed: 23486873]
- Fillenbaum GG, McCurry SM, Kuchibhatla M, Masaki KH, Borenstein AR, Foley DJ, Heyman A, Larson EB, White L. Performance on the CERAD neuropsychology battery of two samples of Japanese-American elders: norms for persons with and without dementia. *Journal of the International Neuropsychological Society*. 2005; 11(2):192–201. [PubMed: 15966108]
- Fillenbaum GG, Heyman A, Huber MS, Ganguli M, Unverzagt FW. Performance of elderly African American and White community residents on the CERAD Neuropsychological Battery. *Journal of the International Neuropsychological Society*. 2001; 7(4):502–509. [PubMed: 11396552]
- Fillenbaum GG, van Belle G, Morris JC, Mohs RC, Mirra SS, Davis PC, Tariot PN, Silverman JM, Clark CM, Welsh-Bohmer KA, Heyman A. Consortium to establish a registry for Alzheimer's disease (CERAD): The first twenty years. *Alzheimer's & Dementia*. 2008; 4:96–109.
- Kannel WB, McGee DL. Diabetes and cardiovascular risk factors: The Framingham Study. *Circulation*. 1979; 59:8–13. [PubMed: 758126]
- Kramer JH, Nelson A, Johnson JK, Yaffe K, Glenn S, Rosen HJ, Miller BL. Multiple cognitive deficits in amnesic mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*. 2006; 22(4):306–311. [PubMed: 16931884]
- Lamar M, Libon DJ, Ashley AV, Lah JJ, Levey AI, Goldstein FC. The impact of vascular comorbidities on qualitative error analysis of executive impairment in Alzheimer's disease. *Journal of the International Neuropsychological Society*. 2010; 16(1):77–83. [PubMed: 19835657]



- Libon DJ, Bondi MW, Price CC, Lamar M, Eppig J, Wambach DM, Nieves C, Delano-Wood L, Giovannetti T, Lippa C, Kabasakalian A, Cosentino S, Swenson R, Penney DL. Verbal serial list learning in mild cognitive impairment: a profile analysis of interference, forgetting, and errors. *Journal of the International Neuropsychological Society*. 2011; 17(5):905–914. [PubMed: 21880171]
- McCurry SM, Gibbons LE, Uomoto JM, Thompson ML, Graves AB, Edland SD, Bowen J, McCormick WC, Larson EB. Neuropsychological test performance in a cognitively intact sample of older Japanese American adults. *Archives of Clinical Neuropsychology*. 2001; 16(5):447–459. [PubMed: 14590159]
- Morris JC. Early-stage and preclinical Alzheimer disease. *Alzheimer Disease and Associated Disorders*. 2005; 19:163–165. [PubMed: 16118535]
- Morris JC, Edland S, Clark C, Galasko D, Koss E, Mohs R, van Belle G, Fillenbaum G, Heyman A. The consortium to establish a registry for Alzheimer’s disease (CERAD). Part IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer’s disease. *Neurology*. 1993; 43(12):2457–2465. [PubMed: 8255439]
- Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C, CERAD investigators. The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD). *Neurology*. 1989; 39:1159–1165. [PubMed: 2771064]
- Morris JC, Mohs R, Rogers H, Fillenbaum G, Heyman A. CERAD clinical and neuropsychological assessment of Alzheimer’s disease. *Psychopharmacological Bulletin*. 1988; 24:641–651.
- Ott A, Breteler MM, van Harskamp F, Claus JJ, van der Cammen TJ, Grobbee DE, Hofman A. Prevalence of Alzheimer’s disease and vascular dementia: Association with education. The Rotterdam Study. *BMJ*. 1995; 310:970–973. [PubMed: 7728032]
- Perneckzy R, Drzezga A, Diehl-Schmid J, Schmid G, Wohlschlager A, Kars S, Grimmer T, Wagenpfeil S, Monsch A, Kurz A. Schooling mediates brain reserve in Alzheimer’s disease: findings of fluoro-deoxy-glucose-positron emission tomography. *Journal of Neurology, Neurosurgery & Psychiatry*. 2006; 77:1060–1063.
- Regard, M. Stroop Test: Victoria Version, Manual of Instructions and Norms. 1981.
- SAS Institute Inc.. SAS@ 9.2 Enhanced Logging Facilities. SAS Institute Inc.; Cary, NC: 2008.
- Schmand B, Smit J, Geerlings M, Lindeboom J. The effects of intelligence and education on the development of dementia: A test of the brain reserve hypothesis. *Psychological Medicine*. 1997; 27:1337–1344. [PubMed: 9403905]
- Seshadri S, Wolf PA, Beiser A, Elias MF, Au R, Kase CS, D’Agostina RB, DeCarli C. Stroke risk profile, brain volume and cognitive function: The Framingham Offspring Study. *Neurology*. 2004; 63(9):1591–1599. [PubMed: 15534241]
- Sliwinski M, Lipton RB, Buschke H, Stewart W. The effects of preclinical dementia on estimates of normal cognitive functioning in aging. *Journal of Gerontology Series B: Psychological Sciences and Social Sciences*. 1996; 51(4):217–225.
- Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Simers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorod M, Wagster MV, Phelps CH. Toward defining the preclinical stages of Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimer’s Dementia*. 2011; 7(3):280–292.
- Sperling RA, Karlawish J, Johnson KA. Preclinical Alzheimer disease – the challenges ahead. *Nature Reviews Neurology*. 2013; 9(1):54–58.
- Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D’Agostino RB, Fox CS, Larson MG, Murabito JM, O’Donnell CJ, Vasan RS, Wolf PA, Levy D. The third generation cohort of the National Heart, Lung, and Blood Institute’s Framingham Heart Study: Design, recruitment, and initial examination. *American Journal of Epidemiology*. 2007; 165:1328–1335. [PubMed: 17372189]
- Stern Y. Cognitive reserve. *Neuropsychologia*. 2009; 47:2015–2028. [PubMed: 19467352]
- Stern Y. What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*. 2002; 8:448–460. [PubMed: 11939702]

- Stern Y, Gurland B, Tatemichi TK, Tang MX, Wilder D, Mayeux R. Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA*. 1994; 271:1004–1010. [PubMed: 8139057]
- Strauss, E.; Sherman, S.; Spreen, O. A compendium of neuropsychological tests: Administration, norms and commentary. 3rd Ed.. Oxford University Press; New York: 2006.
- Troyer AK, Leach L, Strauss E. Aging and response inhibition: Normative data for the Victoria Stroop Test. *Aging, Neuropsychology, and Cognition*. 2006; 13:20–35.
- Unverzagt FW, Hall KS, Torke AM, Rediger JD, Mercado N, Gureje O, Osuntokun BO, Jenrie HC. Effects of age, education and gender on CERAD neuropsychological test performance in an African American sample. *The Clinical Neuropsychologist*. 1996; 10(2):180–190.
- Welsh KA, Butters N, Hughes JP, Mohs RC, Heyman A. Detection and staging of dementia in Alzheimer's disease. Use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer's Disease. *Archives of Neurology*. 1992; 49:448–452. [PubMed: 1580805]
- Welsh K, Butters N, Mohs RC, Beekly D, Edland S, Fillenbaum G, Heyman A. CERAD. Part V: A normative study of the neuropsychological battery. *Neurology*. 1994; 44:609–614. [PubMed: 8164812]
- Welsh-Bohmer KA, Mohs RC. Neuropsychological assessment of Alzheimer's disease. *Neurology*. 1997; 49:S11–S13. [PubMed: 9310507]
- Whyte SR, Cullum CM, Hynan LS, Lacritz LH, Rosenberg RN, Weiner MF. Performance of elderly Native Americans and Caucasians on the CERAD Neuropsychological Battery. *Alzheimer Disease and Associated Disorders*. 2005; 19(2):74–78. [PubMed: 15942324]

**Table 1**

Study sample characteristics.

	<b>N=3362</b>
Age, mean (SD)	46.5 (8.7)
Age, y, n (%)	
<35	314 (9.3)
35-44	1020 (30.3)
45-54	1401 (41.7)
55	627 (18.7)
Sex, n (%)	
Women	1782 (53.0)
Men	1580 (47.0)
Educational Level, n (%)	
HS Diploma	518 (15.4)
Some College	1057 (31.4)
College Degree	1222 (36.4)
Graduate Degree	565 (16.8)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

CERAD-WL and VST traditional data stratified by age group.

<i>Age (years)</i>		<i>&lt;35</i>	<i>35-44</i>	<i>45-54</i>	<i>55</i>	<i>Total</i>	<i>P-value</i>
CERAD-WL Learning <sup>1</sup>							
	Mean (SD)	23.6 (2.8)	23.1 (2.9)	22.4 (3.2)	21.8 (3.4)	22.6 (3.2)	<0.0001
	5%	19.0	18.0	17.0	16.0	17.0	
	25%	22.0	21.0	20.0	20.0	21.0	
	Median	24.0	23.0	23.0	22.0	23.0	
	75%	26.0	25.0	25.0	24.0	25.0	
	95%	28.0	28.0	27.0	27.0	27.0	
CERAD-WL Delayed Recall							
	Mean (SD)	7.4 (1.7)	7.0 (1.8)	6.6 (1.9)	6.4 (1.9)	6.8 (1.9)	<0.0001
	5%	4.0	4.0	3.0	3.0	3.0	
	25%	6.0	6.0	5.0	5.0	6.0	
	Median	8.0	7.0	7.0	7.0	7.0	
	75%	9.0	8.0	8.0	8.0	8.0	
	95%	10.0	10.0	9.0	9.0	10.0	
CERAD-WL Percent Savings <sup>2</sup>							
	Mean (SD)	83.7 (15.3)	80.0 (17.2)	77.7 (18.1)	76.2 (19.3)	79.7 (17.9)	<0.0001
	5%	55.6	50.0	44.4	37.5	44.4	
	25%	75.0	66.7	66.7	66.7	66.7	
	Median	87.5	80.0	77.8	77.8	80.0	
	75%	100	100	90.0	88.9	90.0	
	95%	100	100	100	100	100	
CERAD-WL Recognition Correct							
	Mean (SD)	19.7 (0.7)	19.6 (0.8)	19.5 (0.9)	19.4 (1.0)	19.6 (0.9)	<0.0001
	5%	19.0	18.0	18.0	17.0	18.0	
	25%	20.0	20.0	19.0	19.0	19.0	
	Median	20.0	20.0	20.0	20.0	20.0	
	75%	20.0	20.0	20.0	20.0	20.0	
	95%	20.0	20.0	20.0	20.0	20.0	
CERAD-WL Recognition Correct = 20							
	N (%)	241 (78.5)	742 (75.3)	935 (69.7)	398 (65.7)	2316 (71.5)	<0.0001
VST Trial 1 Completion Time							
	Mean (SD)	12.1 (2.4)	12.2 (2.6)	12.6 (2.7)	13.2 (2.7)	12.5 (2.7)	<0.0001
	5%	9.0	8.9	9.2	9.6	9.1	
	25%	10.5	10.5	10.9	11.3	10.7	
	Median	11.7	11.8	12.1	12.8	12.1	
	75%	13.2	13.4	13.9	14.5	13.8	
	95%	16.0	17.2	17.7	18.1	17.5	
VST Trial 2 Completion Time							
	Mean (SD)	13.5 (2.8)	14.1 (2.9)	15.3 (3.3)	16.8 (3.9)	15.0 (3.4)	<0.0001

<i>Age (years)</i>		<i>&lt;35</i>	<i>35-44</i>	<i>45-54</i>	<i>55</i>	<i>Total</i>	<i>P-value</i>
	5%	9.9	10.4	11.2	11.8	10.8	
	25%	11.7	12.1	13.1	14.0	12.8	
	Median	13.2	13.6	14.9	16.2	14.5	
	75%	14.4	15.4	16.8	18.6	16.6	
	95%	18.6	19.4	21.0	23.8	21.3	
VST Trial 3 Completion Time							
	Mean (SD)	20.3 (5.3)	22.2 (6.1)	25.3 (7.8)	28.6 (8.7)	24.5 (7.8)	<0.0001
	5%	13.7	14.4	16.0	18.2	15.1	
	25%	16.4	18.0	20.3	22.6	19.4	
	Median	19.5	21.1	23.8	27.0	23.0	
	75%	22.6	25.3	28.6	33.3	28.0	
	95%	29.3	33.4	39.9	44.0	39.0	
VST Interference Score Version 1 <sup>3</sup>							
	Mean (SD)	1.7 (0.4)	1.9 (0.5)	2.0 (0.6)	2.2 (0.6)	2.0 (0.6)	<0.0001
	5%	1.2	1.2	1.3	1.4	1.3	
	25%	1.4	1.5	1.6	1.8	1.6	
	Median	1.7	1.8	2.0	2.1	1.9	
	75%	1.9	2.1	2.3	2.5	2.2	
	95%	2.4	2.7	3.1	3.4	3.1	
VST Interference Score Version 2 <sup>4</sup>							
	Mean (SD)	1.5 (0.3)	1.6 (0.3)	1.7 (0.4)	1.7 (0.4)	1.6 (0.4)	<0.0001
	5%	1.1	1.1	1.2	1.2	1.2	
	25%	1.3	1.4	1.4	1.4	1.4	
	Median	1.5	1.5	1.6	1.6	1.6	
	75%	1.6	1.7	1.8	1.9	1.8	
	95%	2.1	2.2	2.4	2.5	2.3	

Note:

<sup>1</sup>Sum across three learning trials.

<sup>2</sup>Maximum set to 100.

<sup>3</sup>Completion time: Trial 3/Trial 1.

<sup>4</sup>Completion time: Trial 3/Trial2.

**Table 3**

CERAD-WL and VST traditional data stratified by education

<i>Education</i>		<i>High School Diploma</i>	<i>Some College</i>	<i>College Degree</i>	<i>Graduate Degree</i>	<i>Total</i>	<i>P-value</i>
CERAD-WL Learning <sup>1</sup>							
	Mean (SD)	21.2 (3.1)	22.2 (3.2)	23.3 (3.0)	23.3 (3.0)	22.6 (3.2)	<0.0001
	5%	16.0	17.0	18.0	18.0	17.0	
	25%	19.0	20.0	21.0	21.0	21.0	
	Median	21.0	23.0	23.0	23.0	23.0	
	75%	23.0	24.0	25.0	25.0	25.0	
	95%	26.0	27.0	28.0	28.0	27.0	
CERAD-WL Delayed Recall							
	Mean (SD)	6.2 (1.9)	6.6 (1.9)	7.1 (1.8)	7.0 (1.9)	6.8 (1.9)	<0.0001
	5%	3.0	3.0	4.0	4.0	3.0	
	25%	5.0	5.0	6.0	6.0	6.0	
	Median	6.0	7.0	7.0	7.0	7.0	
	75%	8.0	8.0	8.0	8.0	8.0	
	95%	9.0	9.0	10.0	10.0	10.0	
CERAD-WL Percent Savings <sup>2</sup>							
	Mean (SD)	75.4 (19.1)	77.3 (18.4)	80.5 (16.7)	80.2 (18.0)	78.7 (17.9)	<0.0001
	5%	40.0	42.9	50.0	44.4	44.4	
	25%	62.5	66.7	70.0	66.7	66.7	
	Median	77.8	77.8	80.0	83.3	80.0	
	75%	88.9	90.0	100	100	90.0	
	95%	100	100	100	100	100	
CERAD-WL Recognition Correct							
	Mean (SD)	19.4 (1.1)	19.5 (0.9)	19.6 (0.8)	19.6 (0.9)	19.6 (0.9)	0.0005
	5%	17.0	18.0	18.0	18.0	18.0	
	25%	19.0	19.0	19.0	19.0	19.0	
	Median	20.0	20.0	20.0	20.0	20.0	
	75%	20.0	20.0	20.0	20.0	20.0	
	95%	20.0	20.0	20.0	20.0	20.0	
CERAD-WL Recognition Correct = 20							
	N (%)	332 (66.1)	718 (70.5)	871 (73.9)	395 (73.3)	2316 (71.5)	0.002
VST Trial 1 Completion Time							
	Mean (SD)	13.4 (2.9)	12.7 (2.7)	12.3 (2.5)	12.1 (2.5)	12.5 (2.7)	<0.0001
	5%	9.9	9.2	9.0	8.8	9.1	
	25%	11.5	10.9	10.5	10.4	10.7	
	Median	12.8	12.2	11.8	11.9	12.1	
	75%	14.6	13.9	13.5	13.3	13.8	
	95%	18.8	17.8	17.0	16.8	17.5	
VST Trial 2 Completion Time							

<i>Education</i>		<i>High School Diploma</i>	<i>Some College</i>	<i>College Degree</i>	<i>Graduate Degree</i>	<i>Total</i>	<i>P-value</i>
	Mean (SD)	17.1 (4.3)	15.4 (3.4)	14.3 (2.9)	14.2 (2.8)	15.0 (3.4)	<0.0001
	5%	11.5	11.1	10.5	10.4	10.8	
	25%	14.3	13.0	12.4	12.4	12.8	
	Median	16.5	14.8	13.8	14.0	14.5	
	75%	18.8	17.1	15.7	15.6	16.6	
	95%	24.6	21.7	19.7	18.8	21.3	
VST Trial 3 Completion Time							
	Mean (SD)	28.7 (10.2)	25.4 (7.7)	22.7 (6.3)	22.8 (6.1)	24.5 (7.8)	<0.0001
	5%	17.0	16.3	14.7	14.5	15.1	
	25%	22.2	20.2	18.2	18.7	19.4	
	Median	26.6	24.0	21.6	21.8	23.0	
	75%	34.0	28.7	25.8	26.3	28.0	
	95%	45.6	40.6	34.5	34.4	39.0	
VST Interference Score Version 1 <sup>3</sup>							
	Mean (SD)	2.2 (0.7)	2.0 (0.6)	1.9 (0.5)	1.9 (0.5)	2.0 (0.6)	<0.0001
	5%	1.4	1.3	1.2	1.2	1.3	
	25%	1.7	1.6	1.5	1.6	1.6	
	Median	2.0	1.9	1.8	1.9	1.9	
	75%	2.5	2.3	2.1	2.2	2.2	
	95%	3.4	3.2	2.9	2.8	3.1	
VST Interference Score Version 2 <sup>4</sup>							
	Mean (SD)	1.7 (0.5)	1.7 (0.4)	1.6 (0.3)	1.6 (0.3)	1.6 (0.4)	<0.0001
	5%	1.1	1.2	1.2	1.2	1.2	
	25%	1.4	1.4	1.4	1.4	1.4	
	Median	1.6	1.6	1.5	1.6	1.6	
	75%	1.9	1.8	1.8	1.8	1.8	
	95%	2.5	2.4	2.2	2.2	2.3	

Note:

<sup>1</sup> Sum across three learning trials.

<sup>2</sup> Maximum set to 100.

<sup>3</sup> Completion time: Trial 3/Trial 1.

<sup>4</sup> Completion time: Trial 3/Trial 2.

**Table 4**

Error score measures by age and education group.

<i>Error Score Measure</i>						
	<i>Age (years)</i>				<b>Total</b>	<b>P-value</b>
	<b>&lt;35</b>	<b>35-44</b>	<b>45-54</b>	<b>55</b>		
CERAD-WL IR errors, mean (SD)	4.6 (5.7)	4.5 (6.3)	4.7 (6.2)	4.7 (6.4)	4.6 (6.2)	0.89
1 CERAD-WL DR errors, n (%)	132 (43.0)	416 (42.2)	641 (47.8)	311 (51.3)	1500 (46.3)	0.0002
1 VST errors, n (%)	135 (44.1)	507 (50.9)	834 (60.7)	409 (67.4)	1885 (57.4)	<0.0001
	<i>Education (degree)</i>				<b>Total</b>	<b>P-value</b>
	<b>High School Diploma</b>	<b>Some College</b>	<b>College Degree</b>	<b>Graduate Degree</b>		
CERAD-WL IR errors, mean (SD)	4.0 (5.5)	4.4 (6.0)	4.8 (6.2)	5.4 (7.1)	4.6 (6.2)	0.0008
1 CERAD-WL DR errors, n (%)	219 (43.6)	488 (47.9)	534 (45.3)	259 (48.1)	1500 (46.3)	0.30
1 VST errors, n (%)	358 (71.3)	633 (61.5)	611 (50.8)	283 (51.5)	1885 (57.4)	<0.0001

Note: IR=immediate recall. DR=delayed recall