Published in final edited form as:

Curr Alzheimer Res. 2013 September ; 10(7): 776–783.

Quality of Education and Memory Test Performance in Older Men: The New York University Paragraph Recall Test Normative Data

Melissa Mathews, Ph.D.¹, Erin Abner, M.P.H.^{1,3}, Allison Caban-Holt, Ph.D.^{1,2}, Brandon C. Dennis, Psy.D.^{1,4}, Richard Kryscio, Ph.D.^{1,3}, and Frederick Schmitt, Ph.D.^{1,2,4}

¹Sanders-Brown Center on Aging, University of Kentucky, Lexington, KY, USA

²Department of Behavioral Science, University of Kentucky, Lexington, KY, USA

³Department of Statistics and Biostatistics, University of Kentucky, Lexington, KY, USA

⁴Department of Neurology, University of Kentucky, Lexington, KY, USA

Abstract

Objectives—Memory evaluation is a key component in the accurate diagnosis of cognitive disorders. One memory procedure that has shown promise in discriminating disease-related cognitive decline from normal cognitive aging is the New York University Paragraph Recall Test; however, the effects of education have been unexamined as they pertain to one’s literacy level. The current study provides normative data stratified by estimated quality of education as indexed by irregular word reading skill.

Method—Conventional norms were derived from a sample (N = 385) of cognitively intact elderly men who were initially recruited for participation in the PREADViSE clinical trial. A

Corresponding Authors Contact Information: Frederick Schmitt, Ph. D., 333B Sanders-Brown Center On Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY 40536-023, USA, fascom@uky.edu; voice: 859.257.1412 x 229; fax: 859.323.1772.

Co-Authors Full Names and Details:

- Melissa Mathews, Ph.D., 333A Sanders-Brown Center on Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY 40536-023, USA, melissa.mathews@uky.edu; voice: 859.257.1412 x 240; fax: 859.323.1772
- Erin Abner, M. P. H., 207 Sanders-Brown Center on Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY 40536-023, USA, elabne0@email.uky.edu; voice: 859.257.1412 x 489; fax: 859.323.1772
- Allison Caban-Holt, Ph.D., 333B Sanders-Brown Center on Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY, USA, 40536-023, amcaba2@email.uky.edu; voice: 859.257.1412 x 322; fax: 859.323.1772
- Brandon C. Dennis, Psy.D., Department of Neurology, University of Kentucky College of Medicine, Suite L-445, 740 S. Limestone St., Lexington, KY, USA, 40536-0284; brandondennis@uky.edu; voice: 859.257.1412 x 239; fax: 859.323.1772
- Richard Kryscio, Ph.D., 230 Sanders-Brown Center on Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY 40536-023, USA, kryscio@email.uky.edu; voice: 859.257.4064; fax: 859.323.1772

This manuscript entitled “Quality of Education and Memory Test Performance in Older Men: The New York University Paragraph Recall Test Normative Data” has not been published elsewhere in any form, neither has it been submitted for publication in any other source.

PREADViSE ClinicalTrials.gov identifier: NCT00040378

Conflict of Interest: No disclosures to report

series of multiple linear regression models were constructed to assess the influence of demographic variables on mean NYU Paragraph Immediate and Delayed Recall scores.

Results—Test version, assessment site, and estimated quality of education were significant predictors of performance on the NYU Paragraph Recall Test. Findings indicate that estimated quality of education is a better predictor of memory performance than ethnicity and years of total education. Normative data stratified according to estimated quality of education are presented.

Discussion—The current study provides evidence and support for normative data stratified by quality of education as opposed to years of education.

Key Terms

aging; literacy; memory; New York University (NYU) paragraph recall; normative data; norms; paragraph recall

INTRODUCTION

Clinical trials often use cognitive testing for many purposes such as enrollment screening, tracking disease-related cognitive declines, and estimating treatment response. Determining whether someone is experiencing actual changes in cognition relative to pre-existing levels of cognition (i.e., pre-morbid ability) requires comparing expected functioning to actual performance (1). Expected functioning typically refers to estimated average performance in a population of demographically similar individuals who are presumed cognitively normal.

The effects of age, educational achievement or literacy, and intellectual ability on standardized test performance have been extensively documented in the literature; thus, accounting for these variables is critical in assessing cognition and diagnosing cognitive decline. Cognitively normal older adults could be classified as impaired if factors related to baseline ability and demographics are not considered. Furthermore, base rates of low scores (i.e., below the 5th percentile) vary according to intellectual ability. Approximately 43–56% of healthy adult participants with low average intellectual abilities had at least one memory test score at or below the 5th percentile on standardized neuropsychological batteries, compared to 22–31% of those with average intellectual ability after being adjusted for age, education, and gender.(2, 3) Normative data are often stratified by ethnicity; however, this is possibly a spurious relationship because ethnicity-based performance differences are attenuated or disappear when educational quality (i.e., literacy) is included in the analysis.(4) Therefore, it has been suggested that literacy, more than age or educational attainment(5) and sex or ethnicity,(6, 7) is a better representation of educational quality and a more accurate predictor of cognitive performance.(5–7)

Conventional test-norming methods may also introduce noise into the diagnostic procedure. Conventional norms are the result of cross-sectional studies and exclude participants based on cognitive issues of sufficient severity to be detected on screening instruments. However, these samples almost certainly include participants with preclinical dementia or subtle disease-related changes in the brain who are still performing cognitively within normal limits but below their pre-morbid capacity.(8–10) Conventional norms likely underestimate individuals' normal cognitive performance, overestimate variability in cognitively healthy

older adults, and have reduced diagnostic sensitivity, especially for brain diseases that manifest subtly.⁽¹⁰⁾ As a result, impaired performance may be considered average and average performance might be considered above average in some cases. Robust norming procedures (8, 11, 12) follow individuals longitudinally to determine who transitions from cognitively normal to impaired. These ‘converters’ presumably reflect pre-clinical cases at baseline and are removed from the sample since they might artificially reduce average baseline scores in an aging cohort. In other words, one may still perform in the average range while, relative to premorbid ability, they are actually performing well below their own baseline. Of note, the preclinical stages of dementia may be present for many years prior to diagnosis,⁽¹³⁾ thus, robust normative studies following individuals for periods of 5–10 years may still include individuals with preclinical disease. Evidence indicates that robust norms improve case finding ability and more accurately describe normal cognition in older adults. (8, 11, 12)

Normative data from memory measures are often used both clinically and in research settings to determine level of memory ability at screening and baseline assessments and to track longitudinal changes in memory performance. The New York University Paragraph Recall Test (NYU PRT) (14), a measure of episodic prose recall, has been found to outperform other cognitive and biomarker predictors of decline to dementia.^(14, 15) Discrepancies in cognitive, functional, and psychological performance between normally aging older adults and adults with mild cognitive impairment (MCI) have been described in cross-sectional studies.⁽¹⁶⁾ Scores for immediate and delayed paragraph recall also revealed the greatest discrepancies between these clinical groups with delayed recall being the most powerful predictor of transition to dementia as compared to all other demographic and cognitive variables and clinician ratings on a global deterioration scale in a longitudinal study examining cognitive predictors of transition from cognitively normal or MCI to dementia as defined by a Global Deterioration Scale (GDS)⁽¹⁷⁾ score of 3 indicating a diagnosis of MCI and 4 indicating a diagnosis of dementia. A cut score < 7 segments recalled yielded the greatest accuracy for predicting transition to dementia or probable Alzheimer's disease (AD) an average of four years after testing⁽¹⁴⁾.

Despite its apparent clinical utility, as demonstrated in several studies^{(11),(14, 15)} and a clinical trial,⁽¹⁶⁾ the use of normative data as opposed to uncorrected raw scores has been sparse. Furthermore, normative data currently available for the NYU PRT have not fully investigated the effects of ethnicity or estimated pre-morbid ability. The purpose of this presentation is to provide normative data from a population of older adult men. The current study examines the effects of ethnicity, years of education, and quality of education (i.e., literacy) upon immediate and delayed recall performance and presents normative data accounting for the relevant influencing factors based on a clinical trial cohort.

MATERIAL AND METHODS

Participants

The current normative information is based upon data collected for the Prevention of Alzheimer's Disease by Vitamin E and Selenium (PREADViSE) trial. This trial is a companion study to the Selenium and Vitamin E Cancer Prevention Trial (SELECT), a

large-scale, multi-site, phase III randomized prostate cancer clinical trial funded by the National Cancer Institute.(18) The PREADViSE protocol was administered through 130 clinical research sites in the United States, Canada, and Puerto Rico that also participated in the SELECT trial. Over 7,500 men were recruited into the PREADViSE study between 2002 and 2008. Design of the PREADViSE project has been described elsewhere (19, 20) and is only briefly described here.

The SELECT trial recruited men aged 55 and up; however AD incidence rates are low below the age of 65. Thus, PREADViSE enrolled 7,547 men age 62 or older (60 if African-American due to higher AD incidence rates in this group (21)). Exclusion criteria included neurological and medical conditions affecting cognition or physical ability to participate in evaluation, severe mental illness, diagnoses of depression or anxiety in the four months prior to evaluation, history of substance abuse in the last 24 months, head injury with loss of consciousness (LOC) greater than 30 minutes, and current use of memory-enhancing medications. Additionally, participants were required to have normal memory functioning at baseline as determined by a score >4 (out of 8 possible points) on the Memory Impairment Screen (MIS; described below).(22)

A subset of the 7,547 participants ($n = 563$) from eight clinical sites across the United States also served as a normal aging group (NAG) validation subsample for the primary MIS screen and completed the NYU PRT as part of the Long Memory and Thinking Screen (LMTS),(19) an extended cognitive assessment that included the Consortium to Establish a Registry for Alzheimer's Disease (CERAD)(23) assessments along with additional measures of processing speed and executive function: Ruff 2 & 7,(24) Trail Making,(25) WAIS-III Digit Symbol Coding,(26) and phonemic verbal fluency.(27) In addition to the exclusion criteria given above, participants in the NAG validation group were excluded from this analysis if they scored <26 on the Mini-Mental State Exam (28) or performed poorly enough on the cognitive battery during any follow-up assessment that a diagnosis of MCI or dementia was suspected. These exclusion criteria were added to the NAG validation subsample to ensure that the group was indeed cognitively normal. Exclusion of participants who were cognitively normal at baseline but showed evidence of impairment on follow-up was also necessary to create a robust normative sample.

Of the 563 NAG participants enrolled, 178 participants were excluded for: triggering a second tier screen during the course of the NAG study ($n = 40$), baseline Mini-Mental State Exam (MMSE) score below 26 ($n = 32$), baseline CERAD T-score(29) below 36 ($n = 15$), NAART estimated full-scale IQ (eFSIQ) less than 90 ($n = 5$), tests administered in Spanish ($n = 1$), and participant-reported history of depression ($n = 56$) or cerebrovascular accident ($n = 11$). Finally, participants over age 80 ($n = 18$) were excluded because they were inadequately represented for normative estimation. Thus, the normative sample consists of 385 cognitively and neurologically intact male volunteers ranging in age from 61–80 from eight PREADViSE study sites.

Procedures

All procedures were conducted with the approval of the University of Kentucky Institutional Review Board. Data were collected at baseline and yearly follow-up over a period of up to

five years. During the initial baseline visit, all 7,547 PREADViSE participants completed a neurological and medical history questionnaire and the MIS. Participants were followed on a yearly basis during which they completed the MIS and provided updates pertaining to medical history and family history of dementia. Any PREADViSE participant scoring 4 on the MIS at baseline and all NAG validation subsample participants completed the LMTS. Participants were considered impaired on the LMTS if they received a CERAD T-score of 30 or less (i.e., 2 SDs below the normal mean) or a score of 24 or less on the MMSE. Participants who 'failed' the LMTS were referred for a medical workup to evaluate for AD. Potential diagnoses included MCI-mixed or amnesic, multiple or single domain, AD, or other dementia.

Measures

Memory Impairment Screen (MIS). (22)—The MIS is a brief four-item delayed free and cued recall measure that maximizes encoding through controlled learning and the use of cues for recall. Specifically, the target words are presented on a sheet of paper that is read aloud by the participant. Afterwards, participants select the item matching a given category. Participants are also asked to recall the words following a 2–3 minute interference task. For items not freely recalled, participants are provided the categorical cue to aid recall. Scores range from 0–8 with 2 points given for free recall, 1 point given for cued recall, and a cut score 4 indicating a higher risk of dementia(22) and suggesting further evaluation is needed. Alternate forms were used on every other screen to reduce potential practice effects.

Consortium to Establish a Registry for Alzheimer's Disease (CERAD). (23)—The CERAD utilizes five previously established cognitive tests: Animal Naming, Boston Naming Test, Mini-Mental State Exam, Constructional Praxis, and Word List Memory. A normative total score combining the CERAD subtests was developed to allow users to characterize severity of cognitive impairment and discriminate between normal aging and pathological aging, including discriminating between AD and MCI.(29) Scores are tabulated by summing the subtest scores, excluding the MMSE. The maximum CERAD total raw score is 100. Corrections according to age, education, and gender are used to derive a T-score for standardization.

NYU Delayed Paragraph Recall Test. (14, 30)—The NYU Delayed Paragraph Recall Test is a verbal memory measure that is one component of the Guild Memory Test.(30) It consists of a short paragraph presented verbally and measures encoding of contextual information. Participants are asked to recall the paragraph *verbatim* following immediate and 5-minute delays. No credit is given for gist responses. Scores range from 0–21 on both immediate and delayed portions of the test. Alternate forms were used to avoid practice effects on re-testing; thus, approximately half of the sample received Version A and half received Version B at baseline.

North American Adult Reading Test. (31)—The NAART is a measure of premorbid intellectual ability, and it was used in the current study to provide an estimate of intellectual ability and quality of education. It is similar to other measures of literacy used to estimate quality of education (e.g., Wide Range Achievement Test, Reading subtest(32)) as it is a

measure of single-word reading. The NAART consists of 61 words with irregular spellings that participants are required to pronounce correctly. WAIS-R Full-scale IQ estimates (eFSIQ) are obtained; however, performance is highly correlated with WAIS-R Verbal IQ, (31) (33–38) which is heavily dependent upon crystallized knowledge and educational achievement (i.e., quality of education). The NAART is also generally resistant to age, gender, and ethnicity.(39–42) NAART eFSIQ scores range from 80–128; however, the current analysis is based upon those participants scoring 90–128 as only a small number of participants (n = 5) scored 80–89.

Statistical Analysis

A series of multiple linear regression models were constructed to assess the influence of the following variables on mean NYU Paragraph Immediate and Delayed Recall scores: age at assessment, education level (less than high school, high school, and more than high school), ethnicity, NAART eFSIQ level (90–109.9, 110), test version, and study site. Two-way interactions between age and education level, age and ethnicity, ethnicity and education, age and version, education and version, and ethnicity and version were also tested. Significant covariates were retained based on a statistical significance level of $\alpha = 0.05$. All models were fit using PROC GLM in SAS 9.3®.

RESULTS

Participants (mean age = 69.2 ± 5.1 years) were primarily Caucasian (66.1%); however, PREADViSE has a notably large minority sample with nearly 30% of the normative study population being African-American. The participants in this sample were highly educated with 52.4% having earned a college degree or higher (See Table 1). Participants were similar in age across eFSIQ levels ($t=0.47$, 383 df, $p=0.64$), ethnicity ($F(3,381)=0.22$, $p=0.89$), and education level ($F(2,382)=1.23$, $p=0.29$). Mean Immediate and Delayed Recall scores were independent of both ethnicity and education level, which were dropped from the full models. Of note, ethnicity was a significant predictor of performance until study site was entered into the model. It is not expected that ethnicity would influence memory over and above the effects of education or literacy; thus, site was retained in the model. No two-way interactions were significant.

Age at assessment, while statistically significant for both Immediate ($F[1,374]=6.34$, $p=0.0122$) and Delayed Recall ($F[1,374]=6.46$, $p=0.0115$), had little clinical significance: the difference in adjusted mean scores between the youngest (at age 61) and oldest participants (at age 80) was just over 1 point for both Immediate ($\beta = 1.12$) and Delayed Recall ($\beta = 1.26$). Thus, age was eliminated from the model for the sake of improved clinical utility. NYU PRT Version, eFSIQ level, and study site were statistically significant contributors to performance. One site (a VA facility in the upper Midwest) was largely responsible for the site effect; thus, to verify that study site differences were not due to scoring errors or differences in test administration, a random sample of 40 NYU PRT protocols (20 each of Versions A & B) was selected for audit. The intraclass correlation coefficient (ICC) among the four auditors and multiple original administrators was >0.95 for both immediate and delayed recall subtests across both versions. Individuals at this site also

had significantly higher eFSIQ than participants at other sites in addition to providing a significant portion of the sample overall (n=104) and being predominantly African-American (80.4%). Given that performance on literacy measures is often related to ethnicity as a function of educational quality, eFSIQ by ethnicity was examined and the outlier site was compared to the remaining sites. eFSIQ at the outlier site was significantly higher than all other sites combined for both Caucasians (eFSIQ: 120.2 versus 111.0, $t=3.94$, $df=252$, $p=0.0001$) and African-Americans (eFSIQ: 111.7 versus 105.8, $t=3.12$, $df=110$, $p=0.002$).

NYU PRT version, eFSIQ level, and study site were retained in the final models for both Immediate and Delayed Recall. The purpose of normative data is to provide information regarding performance while controlling for or otherwise accounting for influencing factors (e.g., age, pre-morbid ability). This is often done by stratifying results according to those factors, as we have done here with eFSIQ. However, since average performance was influenced by site and it is both difficult and inefficient to produce site-specific results for the variety of existing cognitive assessments, we opted to include site as a covariate (with the outlier site as the reference category) but present no site specific results. Adjusted means and standard deviations from the regression models (Table 2) were used to construct the normative tables (Tables 3 and 4). Given that NAART eFSIQ scores do not fall below 80 and that our sample comprised largely average or above-average functioning individuals, our estimated IQ ranges were also limited to those performing in the average-to-above average range. Raw scores were converted to their z-score equivalent, which varied depending upon estimated IQ range and the version of the NYU PRT administered. A typical indicator of impairment or poor performance is any score falling 1.5 standard deviations below the mean (i.e., a z-score of -1.5 suggests impairment). (43)

DISCUSSION

The current study provides previously unavailable normative data for the NYU PRT and accounts for significant factors in task performance in this sample: site, test version, and estimated quality of education. Norms accounting for these variables have the advantage of improving diagnostic decision-making in both clinical and research contexts. As stated previously, determining whether one is performing within the normal range requires comparing observed performance to expected performance. It is well-established that demographic characteristics such as age, education, and baseline intellectual ability influence expected performance on numerous neuropsychological measures.(1–3, 44) Failure to consider such variables may result in misdiagnosing some individuals as cognitively impaired while failing to detect impairment in others. Normative data accounting for these factors may improve diagnostic accuracy by improving one's ability to determine what constitutes "expected" performance. In a research context, participants are often divided into groups comprising "normal" and "impaired" participants based upon neuropsychological test performance; however, without appropriate norms, it is likely that a certain percentage of participants in each group may be misclassified. For example, in our sample, NYU PRT Delayed Recall performance ≤ 5 is considered impaired for individuals with eFSIQ's ≥ 120 (i.e., high educational quality); however, a score ≤ 3 is required for classification of impairment in participants whose eFSIQ is between 90–109.9 (i.e., average educational quality).

Notably, our findings depart from the overall means of normal controls found in some studies (16, 45) and from the suggested overall cut-score < 7 to distinguish those who transitioned to dementia from those who did not.(14) Although these studies accounted for age and education in predicting *decline*, they did not apply corrections to or explore the factors influencing performance on the NYU PRT. Thus, the previously obtained means and the suggested cut score of six may be inflated due to the nature of the samples. Cut scores for impairment based on WAIS Vocabulary scaled scores have been previously suggested, (46) and it is noteworthy that they reflect shifting criteria for impairment based on intellectual ability. For example, a delayed recall score of 2.5 was considered impaired for someone with a low average-to-average vocabulary score whereas a score of 6 was indicative of impairment for participants whose scores were in the upper half of the average range. Although this clearly illustrates the importance of accounting for ability level in determining whether a score is considered impaired, differences between versions was not explored. Given our findings, there are significant score differences based upon version administered.

This study also has several methodological strengths relevant to obtaining a nationally representative sample of normally aging individuals. The study population, although limited to men, is sufficiently large ($N = 385$) and ethnically and geographically diverse. Participants were men from eight different study sites across the country, thus results should be typical of the average older American male. The study sample was also comprised of approximately 30% African-Americans, a minority sample size that is not typical of most normative studies. It should perhaps be re-stated that this sample is representative of a diverse group of American men. Additionally, generalizability to individuals outside the United States may not be warranted given that the literacy measure was also adapted for a North American population.

Excluding participants who may have been showing evidence of cognitive decline also allowed us to create more robust norms. By following participants longitudinally and determining who declined to a diagnosis of MCI or AD or were likely at-risk for these diseases, we were able to enhance the normative data by excluding them from the analysis in an effort to remove the effect of undetected pathology from the normative sample.(8–10) This is important because pathological aging may start many years before individuals fall below the “norm” on cognitive testing,(47) and robust norms have been shown to improve the detection of incident dementia at baseline.(8, 11, 12) The normative data provided here may be valuable in the early determination of who may be most at-risk for developing cognitive impairment or dementia. The current study also provided more evidence for the necessity of non-ethnicity based normative data. Similar to many previous studies, (5–7, 48, 49) we found that memory performance was significantly predicted by estimated quality of education whereas ethnicity and years of education were not significant predictors of performance.

Site effects are not typically reported or corrected for in most studies; however, we found one site in our sample to have an extreme influence on the overall means for both immediate and delayed paragraph recall subtests. It was statistically different from the remaining seven sites, which differed little from one another, thus we accounted for the effect of this site in

our analysis. Of note, adjusting for the effect of site versus removing the site from the analysis resulted in nearly identical results. Furthermore, a data auditing process was completed in which a sample of the NYU PRT protocols from all sites were reviewed with more extensive review of the outlier site and no scoring inaccuracies were revealed. It is possible that there are geographic differences in performance that need to be accounted for in all normative data; however, this hypothesis has not been studied extensively and would require more in-depth exploration in order to draw definitive conclusions regarding the reason for this effect.

The current study provides new normative data for the NYU PRT and highlights the influence of estimated quality of education on the interpretation of performance. Our data clearly show that cut-scores suggestive of impairment vary by estimated quality of education, indicating that reliance on a single, “one-size-fits-all” cut-score may reduce sensitivity and specificity in detecting cognitive impairment. Future studies examining whether norms stratified by estimated quality of education increase case-finding ability or improve the ability to predict decline from normal aging to MCI or dementia will be necessary to determine the true utility of this measure as a diagnostic instrument in both clinical and research settings. Furthermore, we also discovered an important site effect in our data which may suggest that despite our efforts to standardize protocol administration and participant recruitment, “normal” performance also varies by geographic location. Forthcoming normative studies may wish to examine and report on whether site-effects are observed in their samples in order to further improve the accuracy of normative data, especially in clinical trials.

Acknowledgments

Funding:

This research is supported by grants from the National Institute on Aging (AG019241) and P30AG028383. SELECT is supported by a NCI/DCP grant (CA37429).

The SELECT/PREADViSE-Normal Aging sites that contributed participant data for this manuscript: Baptist Regional Medical Center, Jacksonville, FL, Elizabeth Pantoja, Troy H. Guthrie, Jr., M.D. (PI); George Washington University, Washington, DC: Maria L. Caparas, Linda Witkin, Beverly Bentley, Richard J. Katz, M.D. (PI); Grand Rapids Community Clinical Oncology Program, Grand Rapids, MI: Nora Galvin, Joan Long, Lori Pearl-Kraus, Martin Bury, M.D. (PI); Harbor-UCLA, Torrance, CA, Johane Ferreras, Rowan Chlebowski, M.D., Ph. D. (PI); Jesse Brown VA Medical Center, Chicago, IL, Bharathi Reddivari, Thomas E. Ladd, M.D. (PI); Swedish Medical Center, Seattle, WA, Sarah Fanizzi, Kris Huget, Gary E. Goodman, M.D. (PI); UCSD, La Jolla, CA, Eugene Narsete, Sue Stockton, Theodore Ganiats, M.D. (PI); Upstate Carolina Community Clinical Oncology Program, Spartanburg, SC, Elizabeth McCullough, Jonaka Young, James D. Bearden, M.D. (PI)

References

1. Duff K, Patton D, Schoenberg MR, Mold J, Scott JG, Adams RL. Age- and education-corrected independent normative data for the RBANS in a community dwelling elderly sample. *The Clinical neuropsychologist*. 2003 Aug; 17(3):351–366. PubMed PMID: 14704885. Epub 2004/01/06. eng. [PubMed: 14704885]
2. Brooks BL, Iverson GL, White T. Substantial risk of "Accidental MCI" in healthy older adults: base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society : JINS*. 2007 May; 13(3):490–500. PubMed PMID: 17445298. Epub 2007/04/21. eng. [PubMed: 17445298]

3. Brooks BL, Iverson GL, Holdnack JA, Feldman HH. Potential for misclassification of mild cognitive impairment: a study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International Neuropsychological Society : JINS*. 2008 May; 14(3):463–478. PubMed PMID: 18419845. Epub 2008/04/19. eng. [PubMed: 18419845]
4. Manly JJ, Schupf N, Tang MX, Stern Y. Cognitive decline and literacy among ethnically diverse elders. *J Geriatr Psych Neur*. 2005 Dec; 18(4):213–217. PubMed PMID: WOS:000233433100007. English.
5. Byrd DA, Jacobs DM, Hilton HJ, Stern Y, Manly JJ. Sources of error on visuoperceptual tasks: Role of education, literacy, and search strategy. *Brain and Cognition*. 2005; 58:251–257. [PubMed: 15963375]
6. Manly JJ, Byrd DA, Touradji P, Sanchez D, Stern Y. Literacy and cognitive change among ethnically diverse elders. *Int J Psychol*. 2004; 39(39):47–60.
7. Manly JJ, Jacobs DM, Touradji P, Small SA, Stern Y. Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society : JINS*. 2002 Mar; 8(3):341–348. PubMed PMID: 11939693. Epub 2002/04/10. eng. [PubMed: 11939693]
8. Holtzer R, Goldin Y, Zimmerman M, Katz M, Buschke H, Lipton RB. Robust norms for selected neuropsychological tests in older adults. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*. 2008 Sep; 23(5):531–541. PubMed PMID: 18572380. Pubmed Central PMCID:2610426. Epub 2008/06/24. eng. [PubMed: 18572380]
9. Ritchie LJ, Frerichs RJ, Tuokko H. Effective normative samples for the detection of cognitive impairment in older adults. *The Clinical neuropsychologist*. 2007; 21:863–874. [PubMed: 17853155]
10. Sliwinski M, Lipton RB, Buschke H, Stewart W. The effects of preclinical dementia on estimates of normal cognitive functioning in aging. *The journals of gerontology Series B, Psychological sciences and social sciences*. 1996 Jul; 51(4):P217–P225. PubMed PMID: 8673642. Epub 1996/07/01. eng.
11. De Santi S, Pirraglia E, Barr W, Babb J, Williams S, Rogers K, et al. Robust and conventional neuropsychological norms: Diagnosis and prediction of age-related cognitive decline. *Neuropsychology*. 2008; 22:469–484. [PubMed: 18590359]
12. Pedraza O, Lucas JA, Smith GE, Petersen RC, Graff-Radford NR, Ivnik RJ. Robust and expanded norms for the Dementia Rating Scale. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*. 2010 Aug; 25(5):347–358. PubMed PMID: 20427376. Pubmed Central PMCID:2904669. Epub 2010/04/30. eng. [PubMed: 20427376]
13. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. 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 : the journal of the Alzheimer's Association*. 2011 May; 7(3):280–292. PubMed PMID: 21514248. Pubmed Central PMCID:3220946. Epub 2011/04/26. eng.
14. Kluger A, Ferris SH, Golomb J, Mittelman MS, Reisberg B. Neuropsychological prediction of decline to dementia in nondemented elderly. *J Geriatr Psych Neur*. 1999 Winter;12(4):168–179. PubMed PMID: 10616864. Epub 2000/01/05. eng.
15. Fleisher AS, Sowell BB, Taylor C, Gamst AC, Petersen RC, Thal LJ. Clinical predictors of progression to Alzheimer disease in amnesic mild cognitive impairment. *Neurology*. 2007 May 8; 68(19):1588–1595. PubMed PMID: 17287448. Epub 2007/02/09. eng. [PubMed: 17287448]
16. Grundman M, Petersen RC, Ferris SH, Thomas RG, Aisen PS, Bennett DA, et al. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of neurology*. 2004 Jan; 61(1):59–66. PubMed PMID: 14732621. Epub 2004/01/21. eng. [PubMed: 14732621]
17. Reisberg B, Ferris SH, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. *The American journal of psychiatry*. 1982 Sep; 139(9):1136–1139. PubMed PMID: 7114305. Epub 1982/09/01. eng. [PubMed: 7114305]
18. Klein EA, Lippman SM, Thompson IM, Goodman PJ, Albanes D, Taylor PR, et al. The selenium and vitamin E cancer prevention trial. *World journal of urology*. 2003 May; 21(1):21–27. PubMed PMID: 12756490. Epub 2003/05/21. eng. [PubMed: 12756490]

19. Caban-Holt, A.; Schmitt, FA.; Runyons, CR.; Kryscio, RJ.; Mendiondo, MS.; Healey, MS., et al. Studying the effects of vitamin E and selenium for Alzheimer's disease prevention: the PREADVISE model. In: Vellas, B.; Fitten, L.; Winblad, B.; Feldman, H.; Grundman, M.; Giacobini, E.; Kurz, A., editors. Research and Practice in Alzheimer's Disease. Vol. 11. Paris: Serdi Publisher; 2006. p. 124-130.
20. Kryscio RJ, Mendiondo MS, Schmitt FA, Markesbery WR. Designing a large prevention trial: statistical issues. *Statistics in medicine*. 2004 Jan 30; 23(2):285–296. PubMed PMID: 14716729. Epub 2004/01/13. eng. [PubMed: 14716729]
21. Tang MX, Cross P, Andrews H, Jacobs DM, Small S, Bell K, et al. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. *Neurology*. 2001 Jan 9; 56(1):49–56. PubMed PMID: 11148235. Epub 2001/01/10. eng. [PubMed: 11148235]
22. Buschke H, Kuslansky G, Katz M, Stewart WF, Sliwinski MJ, Eckholdt HM, et al. Screening for dementia with the memory impairment screen. *Neurology*. 1999 Jan 15; 52(2):231–238. PubMed PMID: 9932936. Epub 1999/02/05. eng. [PubMed: 9932936]
23. Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989 Sep; 39(9):1159–1165. PubMed PMID: 2771064. Epub 1989/09/01. eng. [PubMed: 2771064]
24. Ruff RM, Evans RW, Light RH. Automatic detection vs controlled search: a paper-and-pencil approach. Perceptual and motor skills. 1986 Apr; 62(2):407–416. PubMed PMID: 3503245. Epub 1986/04/01. eng. [PubMed: 3503245]
25. Reitan RM, Wolfson D. Category test and trail making test as measures of frontal lobe functions. *Clinical Neuropsychologist*. 1995; 9:50–56.
26. Wechsler, D. Wechsler Memory Scale-Third Edition 3ed. San Antonio, TX: Psychological Corporation; 1997.
27. Benton, AL.; Hamsher, Kd; Sivan, AB. Multilingual Aphasia Examination 3rd ed. San Antonio, TX: Psychological Corporation; 1994.
28. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*. 1975 Nov; 12(3):189–198. PubMed PMID: 1202204. Epub 1975/11/01. eng. [PubMed: 1202204]
29. Chandler MJ, Lacritz LH, Hynan LS, Barnard HD, Allen G, Deschner M, et al. A total score for the CERAD neuropsychological battery. *Neurology*. 2005 Jul 12; 65(1):102–106. PubMed PMID: 16009893. Epub 2005/07/13. eng. [PubMed: 16009893]
30. Gilbert JG, Levee RF, Catalano FL. A preliminary report on a new memory scale Perceptual and motor skills. 1968; 27:2 7–2 7 8.
31. Blair JR, Spreen O. Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical neuropsychologist*. 1989; 3:129–136.
32. Wilkinson, GS.; Robertson, GJ. Wide Range Achievement Test 4: Professional Manual. Lutz, FL: Psychological Assessment Resources; 2006.
33. Griffin SL, Mindt MR, Rankin EJ, Ritchie AJ, Scott JG. Estimating premorbid intelligence: comparison of traditional and contemporary methods across the intelligence continuum. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*. 2002 Jul; 17(5):497–507. PubMed PMID: 14592003. Epub 2003/11/01. eng. [PubMed: 14592003]
34. Johnstone B, Callahan CD, Kapila CJ, Bouman DE. The comparability of the WRAT-R reading test and NAART as estimates of premorbid intelligence in neurologically impaired patients. *Arch Clin Neuropsych*. 1996; 11(6):513–519. PubMed PMID: 14588456. Epub 1996/01/01. eng.
35. Raguette ML, Campbell DA, Berry DTR, Schmitt FA, Smith GT. Stability of intelligence and intellectual predictors in older persons. *Psychological Assessment*. 1996; 2:154–160.
36. Spreen, O.; Strauss, E. A compendium of neuropsychological tests: administration, norms, and commentary 2nd ed ed. New York: Oxford University Press; 1998.
37. Uttl B. North American adult reading test: Age norms, reliability, and validity. *J Clin Exp Neuropsych*. 2002 Dec; 24(8):1123–1137. PubMed PMID: ISI:000180797100013. English.

38. Wiens AN, Bryan JE, Crossen JR. Estimating Wechsler-R FSIQ from the National Adult Reading Test-Revised in Normal Subjects. *Clinical Neuropsychologist*. 1993 Jan; 7(1):70–84. PubMed PMID: ISI:A1993KX48400007. English.
39. Anstey KJ, Luszcz MA, Giles LC, Andrews GR. Demographic, health, cognitive, and sensory variables as predictors of mortality in very old adults. *Psychology and aging*. 2001 Mar; 16(1):3–11. PubMed PMID: 11302365. Epub 2001/04/17. eng. [PubMed: 11302365]
40. Boekamp JR, Strauss ME, Adams N. Estimating premorbid intelligence in African-American and white elderly veterans using the American version of the National Adult Reading Test. *J Clin Exp Neuropsych*. 1995 Oct; 17(5):645–653. PubMed PMID: 8557806. Epub 1995/10/01. eng.
41. Grober E, Sliwinski M. Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *J Clin Exp Neuropsych*. 1991 Nov; 13(6):933–949. PubMed PMID: 1779032. Epub 1991/11/01. eng.
42. Storandt M, Stone K, Labarge E. Deficits in Reading Performance in Very Mild Dementia of the Alzheimer-Type. *Neuropsychology*. 1995 Apr; 9(2):174–176. PubMed PMID: ISI:A1995RF68100004. English.
43. Kryscio RJ, Schmitt FA, Salazar JC, Mendiondo MS, Markesbery WR. Risk factors for transitions from normal to mild cognitive impairment and dementia. *Neurology*. 2006 Mar 28; 66(6):828–832. PubMed PMID: 16567698. Epub 2006/03/29. eng. [PubMed: 16567698]
44. Welsh-Bohmer KA, Ostbye T, Sanders L, Pieper CF, Hayden KM, Tschanz JT, et al. Neuropsychological performance in advanced age: influences of demographic factors and Apolipoprotein E: findings from the Cache County Memory Study. *The Clinical neuropsychologist*. 2009 Jan; 23(1):77–99. PubMed PMID: 18609337. Pubmed Central PMCID: 2761675. Epub 2008/07/09. eng. [PubMed: 18609337]
45. de Leon MJ, George AE, Golomb J, Tarshish C, Conwit A, Kluger A, et al. Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease. *Neurobiol Aging*. 1997; 18:1–11. [PubMed: 8983027]
46. Crook T, Gilbert JG, Ferris S. Operationalizing memory impairment for elderly persons: the Guild Memory Test. *Psychol Rep*. 1980 Dec; 47(3 Pt 2):1315–1318. PubMed PMID: 7220760. Epub 1980/12/01. eng. [PubMed: 7220760]
47. Schmitt FA, Davis DG, Wekstein DR, Smith CD, Ashford JW, Markesbery WR. "Preclinical" AD revisited: neuropathology of cognitively normal older adults. *Neurology*. 2000 Aug 8; 55(3):370–376. PubMed PMID: 10932270. Epub 2000/08/10. eng. [PubMed: 10932270]
48. Dotson VM, Kitner-Triolo MH, Evans MK, Zonderman AB. Effects of race and socioeconomic status on the relative influence of education and literacy on cognitive functioning. *Journal of the International Neuropsychological Society : JINS*. 2009 Jul; 15(4):580–589. PubMed PMID: 19573276. Pubmed Central PMCID:2722437. Epub 2009/07/04. eng. [PubMed: 19573276]
49. Dotson VM, Kitner-Triolo M, Evans MK, Zonderman AB. Literacy-based normative data for low socioeconomic status African Americans. *The Clinical neuropsychologist*. 2008 Dec; 22(6):989–1017. PubMed PMID: 18609322. Epub 2008/07/09. eng. [PubMed: 18609322]

Table 1

Overall demographic characteristics of participants in the normal aging group (NAG)

	Age	Education		Ethnicity	
		Grade School or Some High School n (%)	14 (3.6%)	African-American n (%)	112 (29.1%)
Range	61–80				
Midpoint Age	68	Completed High School n (%)	82 (21.3%)	Hispanic n (%)	10 (2.6%)
Mean Age	69.2 ± 5.1	Some College n (%)	87 (22.6%)	Caucasian n (%)	255 (66.2%)
		Completed College n (%)	60 (15.6%)	Other/Unknown n (%)	8 (2.1%)
		Some Graduate School , or Completed Graduate School n (%)	142 (37.0%)		

Table 2

NYU paragraph recall adjusted means and standard deviations by test version and estimated full-scale IQ performance

	Version A			Version B		
	Average (n = 85)	Above Average (n = 120)	p	Average (n = 74)	Above Average (n = 106)	p
IR	6.2 (5.6, 6.8)	7.6 (7.1, 8.1)	0.0004	5.8 (5.3, 6.3)	6.6 (6.2, 7.1)	0.018
DR	7.9 (7.3, 8.4)	9.3 (8.8, 9.8)	0.0002	7.1 (6.5, 7.7)	8.1 (7.6, 8.6)	0.017

Note. IR = NYU PRT Immediate Recall; DR = NYU PRT Delayed Recall

Table 3

NYU paragraph immediate recall: raw score to Z-score equivalents

Raw Score	Version A		Version B	
	Average FSIQ (n = 85)	Above Average FSIQ (n = 120)	Average FSIQ (n = 74)	Above Average FSIQ (n = 106)
20	5.01	4.42	6.31	5.82
19	4.65	4.06	5.86	5.38
18	4.28	3.70	5.42	4.95
17	3.92	3.35	4.98	4.51
16	3.56	2.99	4.54	4.08
15	3.19	2.64	4.09	3.64
14	2.83	2.28	3.65	3.21
13	2.47	1.93	3.21	2.77
12	2.10	1.57	2.77	2.34
11	1.74	1.21	2.32	1.90
10	1.37	0.86	1.88	1.47
9	1.01	0.50	1.44	1.03
8	0.65	0.15	1.00	0.60
7	0.28	-0.21	0.55	0.17
6	-0.08	-0.57	0.11	-0.27
5	-0.44	-0.92	-0.33	-0.70
4	-0.81	-1.28	-0.77	-1.14
3	-1.17	-1.63	-1.22	-1.57
2	-1.53	-1.99	-1.66	-2.01
1	-1.90	-2.35	-2.10	-2.44
0	-2.26	-2.70	-2.54	-2.88

Table 4

NYU paragraph delayed recall: raw score to Z-score equivalents

Raw Score	Version A		Version B	
	Average FSIQ (n = 85)	Above Average FSIQ (n = 120)	Average FSIQ (n = 74)	Above Average FSIQ (n = 106)
20	4.35	3.75	4.83	4.36
19	4.00	3.40	4.45	4.00
18	3.64	3.05	4.08	3.63
17	3.28	2.70	3.70	3.26
16	2.92	2.35	3.33	2.89
15	2.56	2.00	2.96	2.53
14	2.20	1.65	2.58	2.16
13	1.85	1.30	2.21	1.79
12	1.49	0.95	1.83	1.42
11	1.13	0.60	1.46	1.06
10	0.77	0.25	1.08	0.69
9	0.41	-0.11	0.71	0.32
8	0.05	-0.46	0.33	-0.05
7	-0.30	-0.81	-0.04	-0.42
6	-0.66	-1.16	-0.42	-0.78
5	-1.02	-1.51	-0.79	-1.15
4	-1.38	-1.86	-1.16	-1.52
3	-1.74	-2.21	-1.54	-1.89
2	-2.10	-2.56	-1.91	-2.25
1	-2.46	-2.91	-2.29	-2.62
0	-2.81	-3.26	-2.66	-2.99