Validation of the Cognitive Assessment of Later Life Status (CALLS) Instrument: A Computerized Telephonic Screen

Valerie C. Crooks*[1], Thomas D. Parsons[2], J. Galen Buckwalter[3]

[1] Department of Research and Evaluation, Kaiser Permanente Southern California, 100 S. Los Robles, Pasadena, CA 91101,USA
[2] Institute for Creative Technologies, University of Southern California, 13274 Fiji Way, Office 301, Marina del Rey, CA 90292-4019,USA
[3] eHarmony, 300 N. Lake Ave., Suite 1111, Pasadena, CA 91101, USA

* Corresponding author

E-mail addresses:
VCC: valerie.c.crooks@kp.org
TDP: tparsons@ict.usc.edu
JGB: galenbuckwalter@eharmony.com

Abstract

Background: Brief screening tests have been developed to measure cognitive performance and dementia, yet measure limited cognitive domains and often lack construct validity. Neuropsychological assessments, while comprehensive, are too costly and time-consuming for epidemiological studies. This study's aim was to develop a psychometrically valid, brief, telephone administered test of cognitive function in aging.

Methods: Using a sequential hierarchical strategy, each stage of test development did not proceed until specified criteria were met. The 30 minute Cognitive Assessment of Later Life Status (CALLS) interview and a 2.5 hour in-person neuropsychological assessment were conducted with a randomly selected sample of 211 participants 65 years and older that included equivalent distributions of men and women from ethnically diverse populations.

Results: Overall Cronbach's coefficient alpha for the CALLS test was 0.81. A principal component analysis of the CALLS tests yielded five components. The CALLS Total Score was significantly correlated with four neuropsychological assessment components. Older age and having a high school education or less was significantly correlated with lower CALLS Total Score. Females scored better overall than males. There were no score differences based on race.

Conclusion: The CALLS is a valid test that provides a unique opportunity to reliably and efficiently study cognitive function in large populations.

Background

The human and economic costs of dementia [1] point to the need for an improved understanding of age-related cognitive deficits and the prevalence of such deficits in the United States [2,3]. The distinction between normal age-related deficits in cognition from deficits indicative of incipient dementia is problematic but increasingly relevant in health care research and epidemiological studies. The identification of persons with or at risk for neurodegenerative conditions is critical for planning treatment and facilitating effective health care and life planning. Furthermore, persons with cognitive deficits are often at higher risk for other unidentified co-morbidities [4].

Current standards of good practice entail that the assessment of cognitive performance to differentiate between age-related neurocognitive deficits and neurodegenerative disorders be conducted by neurologists and by clinical neuropsychologists who use standardized neuropsychological testing [5,6,7]. Exhaustive neuropsychological batteries, however, tend to be limited to specialized medical centers. As a result, primary care physicians are the most likely observer of cognitive change among older persons. Studies have shown, however, that physicians often underreport or fail to identify problems with cognition or dementia [8,9]. Strategies to improve the differentiation of age related cognitive deficits and dementia are timely and appropriate.

A variety of brief in-person and telephonic screening tests have been developed and refined to measure cognitive performance and dementia [10-16]. The Mini-Mental State Exam (MMSE) [17] is the most widely used in-person instrument and putatively considered the "gold standard" for assessing cognition among older adults. The MMSE has shown good test-retest reliability (0.89) and inter-rater reliability (0.82) [17]. The MMSE, however, has failed to demonstrate consistent predictive ability among heterogeneous populations. It shows a high rate of false positives among individuals with low socioeconomic status and low education [18-20] and false negatives in mildly impaired persons [21-23].

Telephone screening tests of cognition have been developed. The most frequently used is the Telephone Interview of Cognitive Status (TICS) [24] and an adapted version that adds a delayed memory item, the TICS-M [25]. The TICS was originally adapted from the MMSE. The TICS and the TICS-M show high correlations with the MMSE [24,26] and equivalent sensitivity and specificity as cognitive screens [27]. However, the fact that these screens do not measure many of the cognitive domains used in a full neuropsychological battery is a major limitation. Hence, there is need for a reliable and more comprehensive tool.

To this end, a psychometrically valid, time-efficient, telephone-administered test of cognitive performance associated with aging, the Cognitive Assessment of Later Life Status (CALLS), was modeled after standardized neuropsychological batteries to overcome the limitations of screening batteries modeled on the MMSE and in person administration. The objective of this study was to validate the CALLS instrument.

Methods

The study was reviewed and approved by the Institutional Review Board of Kaiser Permanente Southern California.

Instrument Development

The CALLS instrument has undergone extensive developmental work. Applying classical psychometric theory, we have followed a sequential, hierarchical strategy for developing this test, where each stage of development does not proceed further until specified criteria are met. A brief explanation of preliminary work to the validation study follows.

Item Generation Stage A range of cognitive items were identified as necessary to be included for an effective telephone screen. These comprehensive cognitive items were pilot tested with 43 elderly participants (mean age = 73.2; female = 56 percent; non-white = 41 percent) over six separate telephone testing sessions. These cognitive domains were correlated with a brief battery of in-person neuropsychological tests (Judgment of Line Orientation, Boston Naming Test, Letter Number Sequencing, Trail-Making Test Parts A and B and the California Verbal Learning Test). Four separate focus groups were held with participants to elicit feedback on the comprehension and clarity of the questions and test experience. An Expert Panel, comprised of a team from neuropsychology, psychometry, geriatrics, speech pathology, audiology and epidemiology were consulted and a core set of items with acceptable face validity, usability and preliminary convergent validity were identified. Verbal learning and memory, attention, working memory, orientation, processing speed, executive functioning, a test of auditory discrimination and a depression assessment were considered essential.

Item Selection Stage Based on the selected cognitive items, the prototype CALLS test was administered to 101 participants over the age of 64, randomly selected from the membership of Kaiser Permanente Southern California. All participants took part in two 30-minute test sessions over the telephone. The reliability of the subtests was good with coefficient alphas between 0.72 and 0.87. Principle component analysis was conducted to evaluate construct validity. Resulting components were derived from tests of immediate memory, long-term memory, working memory, executive functioning, verbal fluency, and semantic memory. Coefficient alpha for the final set of items was 0.77. The Expert Panel agreed that, those with high factor loadings be retained and those without be eliminated. A 30 minute interview was created through streamlining of instructions, use of adaptive questions and skip patterns.

Validation Stage To validate the test, each participant was given the 30-minute telephone CALLS battery and a full 2.5 hour in-person neuropsychological (NP) battery of tests. Approximately half of the final sample was given the CALLS interview first (n=108) and half were given the NP interview first (n=103). Every effort was made to ensure that the two tests were administered within a reasonably close time period without fatiguing the participant. The mean time between tests was 16.27 days (Range: from one day to 60 days).

Current Study

Sample

No participants from previous interview pools were recruited to subsequent interview pools. In the validation study, 908 men and women 65 years and older were randomly selected from the membership of Kaiser Permanente Southern California. Sampling was conducted to maximize the chances of ethnic and racial diversity. Based on geocoding, equal numbers of African –Americans, Whites, Hispanics and Asians were sought. Also equal numbers of men and women were targeted. Once the target criteria for a given ethnic or racial group were met, no more participants were recruited for that group. Each participant was recruited by initially sending a letter which described the nature of the study and provided an opt-out postcard and study brochure. Due to the time commitment and diversity goals, participants were provided with an $80.00 incentive for participation in the CALLS interview and the in-person neuropsychological assessment. A maximum of six calls were made to recruit for the study.

Of the original 908 sample, 152 were excluded due to ineligibility (125 language barrier; 10 deaths; 6 under 65; 5 illness; 3 each for severe hearing problem and relocation out of area). From an eligible pool of 756, 211 consented to participate (response rate 28 percent) to both the in-person neuropsychological battery of tests and the 30-minute CALLS battery. Each participant signed an Institutional Review Board approved consent form prior to taking the in-person neuropsychological test battery. The breakdown of the final sample selection is described in Table 1.

There were no significant mean age differences between participants (mean = 73.4 years; SD = 5.8) and non-participants (mean = 72.8 years; SD = 6.4). As shown in Table 2, the sample was evenly divided between men (49 percent) and women (51 percent) in both groups. Hispanics were about twice as likely to be non-participants (36 percent) than participants (19 percent). Asians were slightly more likely to be non-participants (28 percent) than participants (21 percent). Contrariwise, Whites were about two and a half times more likely to be participants (36 percent) than non-participants (14 percent). There were no differences in groups for African-Americans ($p<0.0001$).

As shown in Table 2, the participant group is well represented in terms of age, gender and racial and ethnic groups. The study sample is slightly better educated than the general population in these age ranges, but over one quarter have a high school education or less.

Cognitive Measures

Lay interviewers (with at least a bachelor' degree) were trained and supervised by a neuropsychologist to conduct the standardized in-person neuropsychological test battery. Lay interviewers were also trained by a neuropsychologist and supervised by project staff in conducting the standardized CALLS telephone test. Analysis of the reliability of the interviewers' performance across testing sessions revealed correlations in an acceptable range from 0.75 to 0.86.

CALLS Telephone Test The CALLS test includes many of the same cognitive items as are used in a neuropsychological battery. It also includes items that measure response time. The CALLS is a computer-assisted test that is standardized with precise scripts and cues for interviewers. The program is designed to not proceed to the next question item until a valid response is entered. Response time items are recorded and time stamped to the millisecond to ensure accuracy. Animal naming, F words, and similarities are also audio recorded for post-test scoring to ensure that all responses are entered correctly and in order given.

Test items include: *Volume Configuration* – 4 different volume choices for the CALLS interview; *Pitch Discrimination* – 15 paired tones (select whether tones are same or different); *Date* – Month, day, date, season and year; *Naming* – 4 questions to identify number or objects; *Three Trial Wordlist* – 12 words across 3 trials with immediate and delayed recall; *Serial Backward 7* – Subtract 7 from 100 up to 5 times; *Simple Reaction Time* – For right and left ears, respond "now" when a specific tone is heard; *Digit Span Forward* – Digits given from 3 to 7 digits; *Digit Span Backward* – Digits given from 2 to 6 digits; *Animal Naming* – 30 seconds to name animals; *F Words* – 30 seconds to name F words; *President/Vice President* – name current; *Similarities* – 4 different similarities to describe; *Tone Decision or Choice Reaction Time* – distinguish tones by responding "now" for 20 specific tone sequences (5 each high and low, 10 medium); and *Wordlist Recall* – recall all remembered, recall when cued, recall with intrusion words.

Other tests included a 20 question adaptation of the Center for the Epidemiological Study of Depression (CESD) [28] and a brief hearing survey was also conducted regarding phone use, use of amplifier, and hearing aids. Interviewers completed a feedback questionnaire to evaluate protocol adherence, hearing assessment and attitude of interviewee.

In-Person Neuropsychological Battery The in-person tests chosen to compare with the CALLS are well standardized and have acceptable reliability and validity. Verbal memory was assessed with the paragraph prose recall tests from the Wechsler Memory Scale–III: Logical

Memory I, Logical Memory II, Logical Memory Recognition (WMS III) [29] and the California Verbal Learning Test (CVLT) [30]. Nonverbal memory was assessed with the Faces I and II test from the WMS III [29]. Attention was assessed with the Digit Span Forward test from the WMS-III [29] and with the Trail-Making Test Part A [31]. Working memory was assessed with Letter-Number Sequencing and Digit Span Backward from the WMS-III [29]. Visuospatial perception was tested with the Judgment of Line Orientation (JLO) [32]. The Trail-Making Test Part B [31] and the Controlled Oral Word Association Test (Phonemic Fluency - F-A-S) [33] evaluated executive functioning. Language skills were tested with the Boston Naming Test (BNT) [34] and Animal Naming (Semantic Fluency) [35]. One hundred and ninety-seven participants were given the Mini-Mental Status Exam [17].

Additional tests conducted in-person but not analyzed here included the Wechsler Test of Adult Reading (WTAR) [36] Symptom Checklist-90 (SCL-90) [37], Geriatric Depression Scale (GDS) [38], and the Lubben Social Network Scale (LSNS) [39]. Interviewers were trained by an audiologist, and administered standard audiology tests to assess hearing during the in-person interviews.

Statistical Analyses

Descriptive statistics (t-test, chi squares) were generated for demographic characteristics of the sample (Table 2) and mean scores for CALLS (Table 3) and NP assessment data were calculated.

Internal consistency of the CALLS was evaluated by means of item analysis and measured with Cronbach’s coefficient alpha. Validity of the CALLS was assessed with respect to the MMSE (concurrent validity). Concurrent validity was measured with Pearson’s r, after verifying the linear relationship between the CALLS and MMSE.

To assess construct validity, we conducted a principal component analysis of the CALLS battery, in which the covariance structure of the dependent variables was decomposed into orthogonal components by calculating the eigenvalues and eigenvectors of the data covariance matrix [40]. The eigenvalues were used in decision-making related to the number of orthogonal components used in subsequent analyses. Eigenvectors were used for determining the relationship between the original variables and subsequent components. Principal components were extracted using roots greater than one criteria and submitted to the Varimax procedure with an oblique rotation. The eigenvectors and eigenvalues transformed the initial variable space into a novel variable set of principal components.

Given the aim of concurrent validation, the same principal component analysis strategy was applied to the neuropsychological battery. Next correlations were calculated between the items in the CALLS battery and the component scores of the NP battery.

Results

The means for each of the individual CALLS tests are displayed in Table 3. Of a possible 180 points, the CALLS TOTAL mean score for all participants was 104.4 (S.D. 19.9; range 50-150). Thirty-nine (18%) scored one standard deviation below the mean and 43 (20%) scored one standard deviation above the mean. As shown in Table 3, the CALLS scores are mormally distributed. The distribution of CALLS total scores do not present ceiling/floor effects.

Internal Consistency

As shown in Table 4, the CALLS showed a high internal consistency, as measured by Cronbach's alpha (0.81). The Cronbach's alpha for the major factors were as follows: verbal memory 0.88, processing speed 0.73, attention and mental control 0.56, language processing 0.46 and language concepts 0.18.

Concurrent Validity

The CALLS Total score correlated moderately with the MMSE Total score (Pearson's correlation, $r=0.60$). This indicated that about 36 percent of the CALLS variance could be accounted for by the MMSE. Additional analyses of the relations between the MMSE Total score and each of the CALLS domain factors revealed significant correlations: Verbal Memory ($r = 0.41$; $p<0.001$); Processing Speed ($r = 0.24$; $p<0.001$); Attention and Mental Control ($r = 0.23$; $p<0.001$); Language Processing ($r = 0.38$; $p<0.001$); and Language Concepts ($r = 0.33$; $p<0.001$).

Construct Validity

The principal component analysis resulted in five components with eigenvalues above one. These components accounted for 11 percent of the total matrix variance. The loadings are described in Table 4. The components were labeled as a) verbal memory (0.883), b) processing speed (0.731), c) attention and mental control (0.555), d) language processing (0.457), and e) language concepts (0.179).

Using the same principal component analysis strategy, the Neuropsychological Battery yielded six components similar to the CALLS components. As shown in Table 5, the CALLS total score had largely moderate correlations (all statistically significant) with each of the neuropsychological tests. The strongest correlations were with verbal memory and language processing. Weaker correlations tended to be with visuospatial (e.g. JLO) and non-verbal items (e.g. Facial Recognition). Further, the CALLS total score correlated with four neuropsychological testing components: Verbal Memory ($r =0.42$; $p<.0001$), Language Processing ($r = 0.44$; $p<.0001$), Episodic Memory for Contextual Information ($r =0.22$; $p < 0.0016$), and Attention and Mental Control ($r = 0.29$; $p<.0001$). The CALLS did not correlate with visuospatial processing or non-verbal memory.

The CALLS battery also produced expected results for age and education. Older age was significantly correlated with lower CALLS scores ($r = -0.35$; $p<0.0001$). Those with the highest education scored better on the total CALLS test than those with lower education.

Women scored higher than men on the CALLS total score ($p = 0.0285$), perhaps due to the higher proportion of verbal memory on the test. No CALLS score differences were found based on race or ethnicity.

Moderate intercorrelations were found between simple reaction time and language processing items such as animal naming (left ear 0.28, $p<0.0001$; right ear 0.33, $p<0.001$), F words (left ear 0.21, $p<0.0027$; right ear 0.19, $p<0.0048$) and word list trial 1 (left ear 0.25, $p<0.003$; right ear 0.27, $p<0.0001$). There were no significant correlations of the CALLS Total or individual scores with depression.

Discussion

The results of the current validation study suggest that the CALLS instrument is a valid measure for assessing cognitive function in an aging population. The linear correlation between the CALLS and the MMSE (Pearson $r=0.60$; $p<0.05$) revealed a moderate level of concurrent validity, despite different administration modalities (in-person administered versus telephone-administered). Additional analyses of the relations between the MMSE Total score and each of the CALLS individual test items revealed significant correlations.

The CALLS Total score was found to be strongly related to verbal memory, language processing, attention and mental control, and episodic memory for contextual information. It was not associated with visuospatial or non-verbal factors from the neuropsychological battery. The majority of findings regarding the effect of age and education on the cognitive outcome were consistent with previous screens and all results were in expected directions. These findings indeed suggest that the CALLS can be effectively used in place of standard in-person neuropsychological evaluations in situations where the CALLS would be more practical or where the standard in-person evaluations would be impractical to administer.

While further analytical work is required to assess the norms and predictive capacity of the CALLS, the potential clinical utility of the CALLS is reflected in its ability to perform as well as other tests or procedures. For example, the CALLS is well suited for assessing aspects measured by the MMSE, as well as some domains not well assessed by the MMSE. Additionally, the CALLS battery's 12-item word list with immediate and delayed conditions is significantly associated with the neuropsychological battery's verbal memory component. Similarly a strong association exists between the neuropsychological battery's verbal fluency and the CALLS version of semantic (animal naming) and phonemic (F words) fluency. Additionally, there was a noteworthy association between the CALLS version and the neuropsychological battery's version of digit span tests (forward and backward). The fact that these findings reveal such robust associations gives credence to the assertion that the CALLS battery validly measures these cognitive domains.

The CALLS battery provides unique measures of reaction time and processing speed. As a part of the cognitive progression, speed of processing is well documented to decline with age [41,42]. Moreover, the enhanced accuracy of timing assessment in the CALLS may make it more suitable for identifying deficits, especially when reduced processing speed and reaction time include delayed onset of responses and increased decision making times (i.e., reduced information processing speed). Further, in non-computerized assessments there are some cases, in which uncontrolled error margins between stimulus onset and actual stimulus display may result in the modeling of "noise" rather than veridical information [43].

There were also moderate intercorrelations of reaction times with verbal memory and language processing items in the CALLS battery and very small intercorrelations with the language processing items in the neuropsychological tests. These findings suggest the possible relationship of processing speed in retrieval of words from memory. They further suggest that failure to remember words in these tests may be more a function of slow speed in recalling words than of loss of verbal memory.

Simple reaction time can also be a measure that distinguishes cognitively healthy and dementia groups [44]. The addition of response time choices found in the CALLS battery enhances the complexity of the response time measures and may increase sensitivity to screen for early dementia [44-46]. The addition of the shortened Center for Epidemiologic Studies Depression scale also provides a screen for depression, which is also known to slow processing speed.

Language concepts including word naming and similarities offers a simple test of concept formation and verbal expression. Each of these was correlated with the verbal memory components, and they uniquely address the ability to demonstrate abstraction ability and ease of retrieval of accurate words.

The CALLS battery has a number of limitations. The CALLS battery requires the use of a telephone and there are no equivalent visuospatial or non-verbal tasks conducted. While there were modest yet significant correlations between the CALLS and the Trail Making Test (Parts A and B), as well as between the CALLS and Facial Recognition I and II, there was a lack of association with the full components. Given the fact that visuospatial deficits (problems with drawing, constructions, and orientation in their own surrounding) are among the earliest manifestations of Alzheimer’s disease [47,48], the CALLS battery is faced with an important limitation. On the other hand, the lack of a visuospatial component in the CALLS battery may also be helpful in situations where a neuropsychological evaluation or screen needs to be administered to persons with severe visual deficits and specific physical disabilities.

While the sample had fairly equal representation for gender, age, and ethnicity, there were few with less than a high school education. Lower education participants are generally more difficult to recruit and tend to have lower scores on cognitive tests. The small numbers in this group may have affected the distribution and results of the CALLS scores. The generalizability of results also can be affected by the relatively low response rate.

A further possible limitation of the CALLS battery is that it is not adapted for subjects where English is not their primary language. This resulted in 125 or 14 percent fewer possible subjects. While no CALLS score differences were found in the current study among ethnic and racial groups, it is possible that inclusion of these subjects would have altered that finding. Future studies should include a translated version of the CALLS for use with persons whose language is other than English.

An additional limitation is that the current study's data was insufficient to evaluate the validity of the CALLS battery for application to a sample inclusive of individuals with cognitive impairment ranging from mild (mild cognitive impairment) to severe (dementia). Although we did not specifically exclude anyone in our random sample based on cognitive status, we expect that the majority of our sample was cognitively unimpaired. Future studies should examine inclusion of patients affected by mild cognitive impairment, whether progressing or not to dementia. Hence, the CALLS battery should be applied to the study of prodromic cognitive deficits [49].

Despite these limitations, the CALLS battery has a number of strengths. Studies have shown that telephone testing of participants at home is not only reliable [50,51] but that screening at home rather than in the clinician's office may actually improve the performance of elderly subjects on these cognitive tests [52]. Further, the CALLS test provides a mechanism for the participant to select a hearing level comfortable to them that ensures appropriate volume for the test. One of the best features of the CALLS is its unique ability to measure simple and choice response times for each participant. Moreover, the thirty minutes required for the CALLS battery is more efficient and time preserving than most standard in-person neuropsychological evaluations. The two to four hour time period needed for face-to-face administration make such tests expensive and logistically unsuitable in most clinical and research settings. This is even more apparent with epidemiological studies. In addition to reduction of fatigue and increased accessibility, the CALLS battery reduces the need for expensive professional staff and locations.

Conclusions

In summary, the CALLS battery was found to be a brief, yet comprehensive standardized cognitive assessment tool with robust correlations to the more time-consuming and costly in-person neuropsychological battery. The test was scrupulously pre-tested and hierarchically staged to ensure that each step followed psychometrically valid procedures. These results show that multiple domains of cognitive functioning can be reliably assessed over the telephone. The CALLS instrument is a valid test that provides a unique opportunity to efficiently study cognitive function in large populations.

Competing Interests

None of the authors have competing personal or financial interests in the interpretation or presentation of the data.

Authors' Contributions

VCC & JGB contributed to all aspects of design, analyses and implementation and interpretation of study, and drafts, revisions and critical review of paper.

TDP contributed to analyses and interpretation of study, and drafts, revisions and critical review of paper.

All have given final approval of this submission.

Acknowledgements

The authors acknowledge the invaluable assistance and contributions of Vanessa Kapinus, Aura Barreda & Jesus Gomez.
This study was funded in part by the U.S. Public Health Service, National Institute of Health (NIA AG14745).

References

1. Langa KM, Larson EB, Wallace RB, Fendrick AM, Foster NL, Kabeto MU, Weir DR, Willis RJ, Herzog AR: **Out-of-pocket health care expenditures among older Americans with dementia.** *Alzheimer Dis Assoc Disord* 2004, **18**:90-98.

2. Fitzpatrick AL, Kuller LH, Ives DG, Lopez OL, Jagust W, Breitner JC, Jones B, Lyketsos C, Dulberg C: **Incidence and prevalence of dementia in the Cardiovascular Health Study.** *J Am Geriatr Soc* 2004, **52**:195–204.

3. Kukull WA, Higdon R, Bowen JD, McCormick WC, Teri L, Schellenberg GD, van Belle G, Jolley L, Larson EB: **Dementia and Alzheimer disease incidence: a prospective cohort study.** *Arch Neurol* 2002, **59**: 1737–1746.

4. Brody KK, Maslow K, Perrin N, Crooks V, Dellapenna R, Kuang D**: Usefulness of a single item in a mail survey to identify persons with possible dementia; a new strategy for finding high-risk elders.** *Dis Mgt* 2005,**8**:59-72.

5. Doody RS, Stevens JC, Beck C, Dubinsky RM, Kaye JA, Gwyther L, Mohs RC, Thal LJ, Whitehouse PJ, DeKosky ST, Cummings JL: **Practice parameter: management of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology.** *Neurology* 2001*,* **56:**1154–1166.

6. Lezak, MD: *Neuropsychological Assessment, Third Edition*. New York: Oxford University Press;1995

7. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadian EM: **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* 1984, **34**: 939–944.

8. Chodosh J, Petitti DB, Elliott M, Hays RB, Crooks VC, Wenger N: **Physician recognition of cognitive impairment: evaluating the need for improvement.** *J Am Geriatr Soc* 200*4,* **52**:1051-1059.

9. Valcour VG, Masaki KH, Curb JD, Blanchette PL: **The detection of dementia in the primary care setting.** *Arch Intern Med* 2001*,* **161**:1238-1239.

10. Crooks VC, Petitti DB, Robins SB, Buckwalter JG: **Cognitive domains associated with performance on the telephone interview for cognitive status-modified**. *Am J Alzheimers Dis Other Demen* 2006, **21**: 45-53.

11. deJager CA, Budge MM, Clarke R: **Utility of TICS-M for the assessment of cognitive function in older adults.** *Int J Geriatr Psychiatry* 2003*,* **18**:318-24.

12. Herzog AR, Wallace RB: **Measures of cognitive functioning in the AHEAD Study.** *J Gerontol B Psychol Sci Soc Sci* 1997*,* **52B:**37-48.

13. Hill J, McVay JM, Walter-Ginzburg A, Mills CS, Lewis J, Lewis BE, Fillit H: **Validation of a brief screen for cognitive impairment (BSCI) administered by telephone for use in the Medicare population.** *Dis Manag* 2005, **8**:223-234.

14. Kawas C, Karagiozis H, Resau L, Corrada M., Brookmeyer R: **Reliability of the Blessed Telephone Information-Memory-Concentration Test.** *J Geriatr Psychiatry Neurol* 1995, **8:** 238-242.

15. Knopman DS, Knudson D, Yoes ME, Weiss DJ: **Development and standardization of a new telephonic cognitive screening test: the Minnesota Cognitive Acuity Screen (MCAS).** *Neuropsychiatry Neuropsychol Behav Neurol* 2000*,* **13**:286-296.

16. Roccaforte WH, Burke WJ, Bayer BL, Wengel SP: **Reliability and validity of the Short Portable Mental Status Questionnaire administered by telephone**. *J Geriatrc Psychiatry Neurol* 1994,**7**:33-38.

17. Folstein MF, Folstein SE, McHugh PR:**"Mini-Mental State". A practical method for grading the cognitive state of patients for the clinician.** *J Psychiatr Res* 1975, **12**:189-198.

18. Fountoulakis KN, Tsolaki M, Mohs RC, Kazis A: **Epidemiological dementia index: a screening instrument for Alzheimer's disease and other types of dementia suitable for use in populations with low education level.** *Dement Geriatr Cogn Disord* 1998*,* **9**:329-338.

19. Lindesay J, Jagger C, Mylnik-Szmid A, Sinorwala A, Peet S, Moledina F: **The Mini-Mental State Examination (MMSE) in the elderly immigrant Gujarati population in the United Kingdom.** *Int J Geriatr Psychiatry* 1997, **12**:1155-1167.

20. Tombaugh TN, Mcintyre NJ: **The Mini-Mental Status Examination: A comprehensive review**. *J Am Geriatr Soc* 1992, **40**:922-935.

21. Nelson A, Fogel BS, Faust D: **Bedside cognitive screening instruments: a critical assessment.** *J Nerv Ment Dis* 1986, **174**:73-83.

22. Sabe L, Jason L, Juejati M: **Sensitivity and specificity of the Mini-Mental State Exam in the diagnosis of dementia.** *Behav Neurol* 1993, **6**:207-210.

23. Swirsky ST, Field HL, Mitchell DR:**Longitudinal diagnosis of memory disorders**. *Age Ageing* 1992, **21**:393-397.

24. Brandt J, Spencer M, Folstein, M: **The telephone interview for cognitive status.** *Neuropsychiatry Neuropsychol Behav Neurol* 1988, **1:**111-117.

25. Welsh KA, Breitner JCS, Magruder-Habib KM: **Detection of dementia in the elderly using telephone screening of cognitive status.** *Neuropsychiatry Neuropsychol Behav Neurol* 1993*,* **6**: 103-110.

26. Plassman BL, Newman TT, Welsh KA, Helms M, Breitner JCS: **Properties of the telephone Interview for Cognitive Status: application in epidemiological and longitudinal studies.** *Neuropsychiatry Neuropsychol Behav Neurol* 1994*,* **7**:235-241.

27. Ferrucci L, Del Lungo I, Guralnik JM, Bandinelli S, Benvenuti E, Salani B, Lamponi M, Ubezio C, Benvenuti F, Baroni A: **Is the telephone interview for cognitive status a valid alternative in oersons who cannot be evaluated by the Mini Mental State Examination?** *Aging* 1988, **10**:332-338.

28. Ensel W M : Measuring depression: **The CES-D scale. In N. Lin, A. Dean & W. Ensel (Ed.)**. *Social support, life events, and depression*. New York: Academic Press;1986

29. The Psychological Corporation: *WAIS-III:WMS-III: Technical Manual*. San Antonio: Harcourt Brace;1997.

30. Delis D, Kramer J, Kaplan E, Ober B: *California Verbal Learning Test: Adult Version Manual*. San Antonio: Harcourt Brace;1987.

31. War Department Adjutant General's Office: **Army Individual Test Battery: Manual of Directions and Scoring.** Washington D.C.: War Department Adjutant General's Office; 1944.

32. Benton A, Varney N, Hamsher K: **Visuospatial judgment: A clinical test.** *ArchNeurol* 1978, **35:**364-367.

33. Spreen O, Benton A: *Neurosensory Center Comprehensive Examination for Aphasia (NCCEA)*. Victoria, British Columbia: University of Victoria Neuropsychology Laboratory;1977

34. Kaplan E, Goodglass H, Weintraub S: *Boston Naming Test*. Philadelphia: Lea & Febiger;1982

35. Goodglass H, Kaplan E : *The Assessment of Aphasia*. Philadelphia: Lea & Febiger;1983.

36. The Psychological Corporation: *Wechsler Test of Adult Reading™ (WTAR™) Third Edition, WMS-III, WTAR Manual*, Toronto, CAN: Harcourt;2001

37. Derogatis LR: *Symptom Checklist-90- Revised.* Minneapolis, Minnesota:Pearson Assessments; 1996

38. Yesavage JA, Brink TL, Lum O, Huang V, Adey M, Leirer VO: **Development and validation of a geriatric depression scale: a preliminary report**. *J Psychiatr Res* 1983, **17**:37-49.

39. Lubben J, Gironda M, Lee A: **Refinements to the Lubben Social Network Scale: The LSNS-R.** *The Behavioral Measurements Letter* 2001, **7**:2-11.

40. Gorsuch Richard L: *Factor Analysis*. Hillsdale, NJ: Erlbaum;1983.

41. Hertzog C, Schear JM: **Psychometric considerations in testing the older person. In T. Hunt & C.J. Lindley (Ed.)**. *Testing older adults: A reference guide for geropsychological assessments.* Austin, Texas: Pro-ed;1989

42. Salthouse TA, Babcock RL, Shaw RJ: **Effects of adult age on structural and operational capacities in working memory.** *Psychol Aging* ***1991*****, 6**:118-127.

43. Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE: **Computerized neurocognitive scanning: I. Methodology and validation in healthy people.** *Neuropsychopharmacology* 2001, **25**:766–776.

44. Hofman M, Seifritz E, Krauchi K, Hock C, Hampel H, Neugebauer A, Muller-Spahn F: **Alzheimer's disease, depression and normal ageing: merit of simple psychomotor and visuospatial tasks.** *Int J Geriatr Psychiatry* 2000, **15**:31-39.

45. Ferris S, Crook T, Sathananthan G, Gershon S: **Reaction time as a diagnostic measure in senility.** *J Am Geriatr Soc* 1976, **24**:529-533.

46. Teng EL, Chui HC, Saperia D: **Senile dementia: Performance on a neuropsychological test battery**. *Recent Advances in Cardiovascular Disease* 1990, **11**:27-34.

47. Mendez MF, Cummings JL : *Dementia- A Clinical Approach, 3rd edition.* Philadelphia, PA: Butterworth-Heinemann (Elsevier);2003

48. Smith MZ, Esiri M M, Barnetson L, King E, Nagy Z: **Constructional apraxia in Alzheimer's disease: An association with occipital lobe pathology and accelerated cognitive decline.** *Dement Geriatr Cogn Disord* 2001, **12**:281–288.

49. Lines CR, McCarroll KA, Lipton RB, Block GA: **Prevention of Alzheimer's In Society's Elderly Study Group. Telephone screening for amnesic mild cognitive impairment.** *Neurology* 2003, **60**:261–266.

50. Debling D, Amelang M, Hasselbach P, Sturmer T: **Assessment of cognitive status in the elderly using telephone interviews.** *Z Gerontol Geriatr* 2005*,* **38**:360-367.

51. Wilson RS, Bennett DA: **Assessment of cognitive decline in old age with brief tests amenable to telephone administration.** *Neuroepidemiology* 2005, **25**:19-25.

52. Shievitz AL, Tudiver F, Araujo A, Sanghe P, Boyle E: **Do elderly people score better on cognitive tests at home?** *Can Fam Physician* 1998**, 44**:1652-1656.

Table 1: Selection of Sample

	Eligible Sample 756	Percent 100
Bad phone numbers or maximum calls	48	6.35
Declined by postcard	158	20.9
Declined by telephone	333	44.0
Incomplete interviews	6	0.85
Completed interviews	211	27.9

Note: For all analyses N = 211.

Table 2: Demographic Characteristics of Sample

	Participants		Non-Participants		
Demographic Category	**Frequency** 211	**Percent** 100%	**Frequency** 697	**Percent** 100%	**p**
GENDER					**n.s.**
Female	107	51%	348	50%	
Male	104	49%	349	50%	
RACE/ETHNICITY					**<.0001**
Asian/Pacific Islander	44	21%	194	28%	
African-Americans	48	23%	155	22%	
Hispanic	40	19%	251	28%	
White	76	36%	97	14%	
Other	3	1%	0	0	
AGE					**n.s.**
65-69	67	32%	262	38%	
70-74	59	28%	196	28%	
75-79	52	25%	132	19%	
80-84	22	10%	68	10%	
85+	11	5%	39	6%	
EDUCATION*					**NA**
< HS Graduate	14	7%	--	--	
HS Graduate	43	20%	--	--	
Some College	81	39%	--	--	
College+	72	34%	--	--	

Note: *With the exception of Education, the sample size for all analyses was N = 211. For Education, one case was missing and the resulting N = 210. NA = Education not available for non-participants.

Table 3: Descriptive Statistics of CALLS Interview

Test	Mean Score	Standard Deviation	Range
Verbal Memory			
Word List Trial 1	4.6	1.8	0-9
Word List Trial 2	6.4	2.1	1-12
Word List Trial 3	7.4	2.0	2-12
Word List Recognition	9.2	2.0	3-12
Word List Cued Recall	6.4	2.4	0-11
Word List Delayed Recall	5.7	2.8	0-11
Processing Speed			
Choice Reaction Time Decile	5.5	2.9	1-10
Simple Reaction Time Decile – Right ear	5.5	2.9	1-10
Simple Reaction Time Decile – Left ear	5.4	2.9	1-10
Attention and Mental Control			
DATE Score	3.6	0.9	0-5
Digit Span Backward - span	4.4	1.3	0-6
Digit Span Forward – span	6.2	0.9	4-7
Language Processing			
President/VP Naming	3.0	1.0	0-4
F words -30 sec.	7.4	3.2	0-16
Animal Naming -30 sec.	11.5	3.4	0-23
Serial 7s	3.4	1.7	0-5
Language Concepts			
4 Word Naming	3.9	0.4	2-4
Similarities	4.8	2.1	0-7

Note: For all analyses N=211.

Table 4: Factor Loadings of CALLS Test

Rotated Component Matrix(a)

	Verbal Memory	Processing Speed	Attention and Mental Control	Language Processing	Language Concepts
Alpha	**0.883**	**0.731**	**0.555**	**0.457**	**0.179**
Word List Trial 1	0.592				
Word List Trial 2	0.767				
Word Recognition	0.774				
Word List Trial 3	0.792				
Word List Cued Recall (Total)	0.829				
Word List Delayed Free Recall	0.834				
Choice RT Score Decile		0.559			
Simple RT Right Score Decile		0.865			
Simple RT Left Score Decile		0.893			
DATE Score			0.373		
Digit Span Backward (Span)			0.784		
Digit Span Forward (Span)			0.804		
President/VP Naming				0.799	
F Words (30 Seconds)				0.504	
Animal Naming (30 seconds)				0.456	
Serial 7's				0.496	
4 Word Naming					0.745
Similarities					0.732

Note: The extraction method was a Principal Component Analysis. The rotation method was Varimax with Kaiser Normalization. Rotation converged in 6 iterations. The CALLS Total coefficient alpha = 0.809. Choice RT = Choice Reaction Time; Simple RT = Simple Reaction Time.

Table 5: Correlations between CALLS Total Score and Neuropsychological Components

Components-Items	**r**	**p**
Verbal Memory Component	**0.42**	**<0.0001**
CVLT Trials 1-5	0.57	<0.0001
CVLT Trial 1 List A	0.36	<0.0001
CVLT Trial 5	0.49	<0.0001
CVLT List B	0.37	<0.0001
CVLT Short Delay Free Recall	0.49	<0.0001
CVLT Short Delay Cued Recall	0.44	<0.0001
CVLT Long Delay Free Recall	0.47	<0.0001
CVLT Long Delay Cued Recall	0.49	<0.0001
CVLT Recognition	0.36	<0.0001
Language Processing Component	**0.44**	**<0.0001**
Wechsler Test of Adult Reading	0.38	<0.0001
Animal Naming	0.53	<0.0001
F Word Total	0.51	<0.0001
A Word Total	0.57	<0.0001
S Word Total	0.49	<0.0001
Episodic Memory for Contextual Information Component	**0.22**	**0.0016**
Logical Memory I	0.41	<0.0001
Logical Memory II	0.37	<0.0001
Logical Memory Recognition	0.35	<0.0001
Boston Naming Test	0.40	<0.0001
Attention & Mental Control Component	**0.29**	**<0.0001**
Digit Span Forward	0.27	<0.0001
Digit Span Backward	0.38	<0.0001
Letter-Number Sequencing	0.38	<0.0001
Judgment of Line Orientation	0.18	0.0091
Visuospatial Processing Component	**-0.13**	**n.s.**
Trail Making Test Part A (time in seconds)	-0.33	<0.0001
Trail Making Test Part B (time in seconds)	-.040	<0.0001
Non-Verbal Memory Component	**0.09**	**n.s**
Faces I Immediate Recognition	0.19	0.0067
Faces II Delayed Recognition	0.26	0.0002

Note: For all analyses N=211. CVLT = California Verbal Learning Test.