

APPLICABILITY OF THE CERAD NEUROPSYCHOLOGICAL BATTERY TO BRAZILIAN ELDERLY

Paulo Henrique Ferreira Bertolucci¹, Ivan Hideyo Okamoto¹, Sonia Maria Dozzi Brucki¹, Marilena Ochini Siviero¹, João Toniolo Neto², Luis Roberto Ramos²

ABSTRACT - There is a limited choice of psychometric tests for Portuguese speaking people which have been evaluated in well defined groups. A Portuguese version of CERAD neuropsychological battery was applied to a control group of healthy elderly (CG) (mean age 75.1 years/ education 7.9 years), 31 Alzheimer disease (AD) patients classified by clinical dementia rating (CDR) as CDR1 (71.4/ 9.0) and 12 AD patients CDR 2 (74.1/ 9.3). Cut-off points were: verbal fluency-11; modified Boston naming-12; Mini-mental State Examination (MMSE) -26; word list memory-13; constructional praxis-9; word recall-3, word recognition-7; praxis recall-4. There was a significant difference between CG and AD-CDR1 ($p < 0.0001$) for all tests. There was a less significant difference for constructional praxis and no difference for Boston naming. Comparison between AD-CDR1 and AD-CDR2 showed difference only for MMSE, verbal fluency, and Boston naming. The performance of CG was similar to that of a US control sample with comparable education level. These results indicate that this adaptation may be useful for the diagnosis of mild dementia but further studies are needed to define cut-offs for illiterates/ low education people.

KEY WORDS: CERAD, neuropsychology, Alzheimer disease.

Aplicabilidade da bateria neuropsicológica CERAD em idosos brasileiros

RESUMO - Existem poucos instrumentos psicométricos em português aplicados em grupos bem definidos. Uma versão em português da bateria neuropsicológica CERAD foi aplicada a um grupo controle de 85 idosos saudáveis (GC) (média de idade 75,1/ média de educação 7,9), 31 indivíduos com doença de Alzheimer (DA) classificados pela Classificação Clínica para Demência (CDR) como CDR1 (71,4/ 9,0) e 12 indivíduos com DA CDR2 (74,1/ 9,3). Os pontos de corte foram: fluência verbal - 11; teste de nomeação de Boston - 12; MEEM - 26; memória da lista de palavras - 13; praxia construtiva - 9; evocação de palavras - 3; reconhecimento de palavras - 7; evocação da praxia - 4. Houve uma diferença significativa ($p < 0,0001$) para todos os testes, exceto o de nomeação de Boston ($p < 0,368$). A comparação entre AD-CDR1 e AD-CDR2 mostrou diferença apenas para o MEEM, fluência verbal e teste de nomeação de Boston. O desempenho do GC foi semelhante ao de uma população controle americana pareada para nível educacional. Estes resultados indicam que esta adaptação pode ser útil para o diagnóstico de demência inicial, mas estudos mais detalhados devem ser realizados para determinar os pontos de corte para pessoas analfabetas ou com baixa escolaridade.

PALAVRAS-CHAVE: CERAD, avaliação neuropsicológica, doença de Alzheimer.

Well accepted criteria for the diagnosis of dementia, like the DSM-IV and the ICD-10 are based on the identification of a memory deficit, and at least one more deficit in another area of cognition. The same criteria were adopted for the diagnosis of Alzheimer's disease (AD) by the US National Institutes of Health (NINCDS-ADRDA)¹ with a sensitivity of 80%

for the diagnosis of probable AD. This led to the creation, in 1986, of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), with the objective of setting a standard for the clinical and neuropsychological assessment of AD². Eventually, standardized criteria for the neuropathological³ and imaging⁴ diagnosis were established.

Escola Paulista de Medicina, Universidade Federal de São Paulo (UNIFESP), São Paulo SP, Brasil: ¹Setor de Neurologia do Comportamento - Departamento de Neurologia e Neurocirurgia; ²Disciplina de Geriatria - Departamento de Medicina. This research was supported by CNPq (grant 522835/95-3 to PHFB and IHO) and FAPESP (grant 97 17926-1 to MOS)

Received 30 November 2000, received in final form 3 April 2001. Accepted 10 April 2001.

Dr. Paulo Henrique F. Bertolucci - Disciplina de Neurologia, Escola Paulista de Medicina, UNIFESP - Rua Botucatu 740 - 04023-900 São Paulo SP - Brasil. FAX: + 55 11 575 5240 - E-mail: paulohb@sun-nepi.epm.br

CERAD criteria for neuropsychological evaluation are compatible with those of DSM and there is explicit reference, for diagnosis of probable AD, to the need of "dementia established by clinical examination and documented by the Mini-Mental State Examination, Blessed Dementia Scale or a similar test, confirmed by neuropsychological examination". There are several advantages in using a standard battery like CERAD: it is possible to compare results among different groups and studies; the examination is uniform and agreement among examiners is high²; test-retest is reliable²; being more complete than a screening test, this battery is able to detect dementia at an initial stage⁵. Finally the battery is not extensive, the time of application being around 30 minutes. With the large group of patients involved along the study, it was possible to determine the effect of variables such as education, age and gender on the performance⁶.

The following tests were chosen, to evaluate the main cognitive alterations in AD: verbal fluency (animal category), abridged Boston naming test (15 of the 60 original drawings), Mini-mental State Examination (MMSE), word list memory, with repetition, recall and recognition, and constructional praxis with copy and recall².

In former studies the battery was applied to subjects with probable AD and mild to severe disease, whose performance was compared with paired controls. This initial study² showed several advantages for the battery: the acceptance among the different centres was good, and technical difficulties in the application were not observed; the time of application was short (between 20 and 30 minutes); the reliability coefficient among examiners was satisfactory for all the tests². Subsequent studies, with additional inclusion of other patients, permitted to identify differences among subtypes of AD⁷ and modifications in the performance of the battery with disease evolution⁸.

A Brazilian version of the CERAD clinical and neuropsychological batteries was developed by the Disciplines of Neurology and Geriatrics of the Escola Paulista de Medicina in São Paulo. Compared to the Portuguese version, developed by the Department of Neurology of the University of Lisbon, the Brazilian version was remarkably similar, except for one word (grass), that has different meanings in Portugal and Brazil.

The main objective of this study was to evaluate the applicability of this Portuguese-Brazilian version in our population and the possible differences between normal aged people and probable AD.

METHOD

Compared to the English version there were two major points to be considered in the Brazilian version. For verbal memory tests, words were chosen with roughly the same meaning and extension, but some adaptation was needed. For instance, the literal translation of the word "shore" would be "costa", which has more than one meaning in Brazil, so "praia" (beach) was chosen instead. The abridged Boston naming test has items supposed to reflect decreasing word frequency. Since there is not, as far as we know, a study of word frequency in Brazilian Portuguese, we kept the original items in the same sequence. The tests were applied in the following sequence:

- verbal fluency⁹ – it is given the order "Tell me all the animals you can remember. You can say any animal". One minute is counted starting from the end of the command and the score corresponds to the number of animals reminded in this period. Proper nouns and repetitions are not counted. When animals whose gender is similar are reminded only one is scored, but when the denomination is different (e.g. horse and mare) both are scored.

- Boston naming test (abridged version)¹⁰ - 15 drawings of the Boston naming test are presented (tree, bed, whistle, flower, house, boat, toothbrush, volcano, mask, camel, harmonica, tongs, hammock, funnel, dominoes). One point is given for each correct answer without cues, with a maximum score of 15 points.

- Mini-mental State Examination¹¹ – for this research we used a Brazilian version from a study in Sao Paulo metropolitan area which established cut-offs according with education level¹². Maximum score for this test is 30 points.

- Word list memory task¹³ – ten unrelated words are read aloud (Appendix 1), one by one, by the subject (or the examiner, in case of reading difficulty) at a speed of one word every 2 seconds. Recalling is done immediately after the last word, for a maximum period of 90 seconds. The procedure is repeated, with the words in a different order, two more times. The score is obtained by the sum of the words recalled in the 3 trials, with a maximum score of 30 points.

- Constructional praxis¹⁴ - four drawings are presented, one at a time (circle, diamond, overlapping rectangles and cube), with a maximum of 2 minutes for the copy of each drawing. Scoring is done separately for each drawing, the sum of scores being 11 points maximum.

- Word list recall¹³ – immediately after the praxis test the recall of the list of words previously presented is done, for a maximum period of 90 seconds, with a maximum score of 10 points.

- Word list recognition¹⁵ - after the spontaneous recalling, the 10 words are presented mixed to 10 new words (appendix 1). To correct for a chance effect, the score is calculated as the total number of correct answers minus 10. As the maximum number of correct answers is 20, the maximum score is 10.

- Praxis recall¹⁴ - the four drawings previously copied should be reproduced spontaneously, with a maximum

Appendix 1. Word list for fixation and recall.

Manteiga	Braço	Praia
Carta	Rainha	Cabana
Poste	Bilhete	Erva
Motor		

Appendix 2. Word list for recognition.

Igreja	Rainha	Café
Cabana	Manteiga	Chinelo
Dolar	Poste	Braço
Aldeia	Praia	Corda
Cinco	Bilhete	Carta
Tropa	Hotel	Erva
Montanha	Motor	

score of 11 points. The last item of the MMSE, is the copy of two overlapping pentagons. If it is spontaneously recalled by the subject it is also included in the score.

Subjects

The battery was applied to three groups, whose demographic data are shown in Table 1. There was no difference among:

Control group (CG) - this group was drawn from a cohort being studied by the Discipline of Geriatrics of the Federal University of São Paulo / Escola Paulista de Medicina. The original study included information on other variables, like social adaptation and mental health. In summary, a visit was done to every household in the catchment area of the University Hospital, in the metropolitan area of São Paulo. Of the 1.667 individuals aged 65 or over being followed since 1991, one in every 10 was randomly selected for the second stage, so that a total of 160 subjects were invited to the University Hospital, for a more detailed evaluation. Of this group 85 subjects were randomly selected for this study. To be considered a control a subject should have no memory complaints on his/her own or on his/her relatives report; no evidence of cognitive or functional decline; no abnormalities at the neurological examination, and have a normal brain CT scan

Table 1. Study population demographic data.

	CG (n=85)	AD1 (n=31)	AD2 (n=12)
Age	75.1 ± 6.2	71.4 ± 6.9	74.1 ± 9.1
Education	7.9 ± 4.8	9.0 ± 5.0	9.3 ± 6.0
Sex (M/F)	41/44	16/15	6/6

T test: age: CG-AD1 - 0.06; CG-AD2 - 0.61; AD1-AD2 - 0.30
Education: CG-AD1 - 0.28; CG-AD2 - 0.39; AD1-AD2 - 0.90

or MRI. Clinicians were kept unaware of the neuropsychological findings.

Alzheimer disease 1 (AD1) - this group was composed by 31 subjects with AD attending the Behaviour Neurology Outpatient Clinic of the University Hospital who fulfilled the NINCDS-ADRDA criteria for probable AD¹. They were at a mild stage of the disease, with a rating of 1 at the Clinical Dementia Rating (CDR)¹⁶, what would mean someone with a moderate memory loss interfering with daily activities, possibly with some degree of temporal disorientation, moderate difficulty to evaluate situations and to take decisions and difficulty to maintain social activity. It should be stressed that all these people could seem normal at a casual examination.

Alzheimer disease 2 (AD2) - this consisted of 12 subjects from the same source as AD1 scoring 2 at CDR. A rating of 2 means someone who has severe memory loss and will possibly lose any new material. There is disorientation not only for time, but also for place, and difficulty in handling problems and making judgements. These subjects can not have independent activity outside home and can only do very simple chores at home. This group was included only to compare performance between AD1 and AD2. AD2 scores were not used for the calculation of cut-offs.

All subjects went through the complete CERAD protocol which includes not only the neuropsychological battery, but also an interview on cognitive decline, vascular brain disease, Parkinson's disease, depression, use of drugs affecting the cognition, alcohol abuse, and complete physical and neurological examination.

Statistical analysis was done by using the program SPSS version 7.5. Before analysing the variance between groups, the possibility of using parametric tests was tested by verification of distribution curves skewness and kurtosis. The non normal distribution indicated the use of Mann-Whitney test for analysis of each variable comparing CG with AD1 and AD1 with AD2. A trade-off sensitivity and specificity was done for each subtest, using the statistical program MedCalc. Receiver operator characteristic (ROC) curve were performed in groups CG and AD1/AD2 to identify the better cut-off point, represented by the largest area under the curve. A t-test was used to compare different age and scholasticity groups. For the study of sex influence we used a binomial test. The confidence interval was set at 95%.

Before entering the study all subjects had an explanation on the procedures and objectives of the research. All subjects or responsible caregivers signed an informed consent approved by the University Federal de São Paulo Ethics Committee. All proceedings were approved by the Universidade Federal de São Paulo Ethics Committee on Research in Humans.

RESULTS

The comparison among the three groups with regard to demographic variables (Table 1) did not

Table 2. Mean and SD for the sub-tests of CERAD battery.

	CG (n = 85)	AD1 (n = 31)	AD2 (n = 12)
Verbal fluency	15.6 ± 3.9	10.3 ± 3.3	6.4 ± 3.4*°
Boston naming	13.1 ± 1.7	12.4 ± 1.8	9.2 ± 2.9°
MMSE	27.8 ± 2.2	23.0 ± 2.8	14.8 ± 3.0*°
Word list memory	18.0 ± 4.1	11.0 ± 4.1	7.4 ± 2.6 *
Constructional praxis	9.0 ± 1.9	7.7 ± 2.1	5.6 ± 2.5
Word list recall	5.5 ± 2.2	2.1 ± 2.1	0.2 ± 0.4 *
Word list recognition	9.0 ± 1.7	5.9 ± 2.8	3.4 ± 2.2 *
Praxis recall	6.0 ± 3.3	1.7 ± 2.2	0.3 ± 0.6 *

Mann-Whitney analysis of variance: * significant difference between CG and AD1 (p < 0.001); ° significant difference between AD1 and AD2 (p < 0.001)

show significant differences. Since distribution did not fit a normal curve, Mann-Whitney test was used to compare means. A preliminary within-group analysis was done, comparing performance between both sexes and between subjects under or over 75 year. No difference was shown and these groups were collapsed for further analysis. A similar comparison was done for education level, comparing those under or over 8 years of education. Except for constructional praxis and praxis recall, all tests showed no difference between the 2 groups. Means and SD for each group and sub-test are shown in Table 2. Analysis of variance and multiple comparisons were done between CG, AD1 and AD2, showing a significant difference for all sub-tests, except Boston naming and constructional praxis. Differences between AD1 and AD2 were less impressive, with a p higher than 0.001 for word list memory (0.008), list recall (0.003), praxis (0.007), and praxis recall (0.054).

We then proceed to a trade-off for sensitivity and specificity for the determination of the best cut-off points, through ROC curves, considering as a cut-off the point comprising the largest area under the

Table 3. Median and range for the sub-tests of CERAD battery.

	CG	AD1	AD2
Verbal fluency	16 (3-25)	10 (5-18)	6 (2-14)
Boston naming	14 (8-15)	12 (9-15)	9 (6-13)
MMSE	29 (20-30)	24 (17-28)	15 (10-19)
Word list memory	18 (5-26)	10 (4-21)	7 (3-11)
Constructional praxis	10 (4-11)	7 (2-11)	5 (2-10)
Word list recall	5 (1-10)	2 (0-7)	0 (0-1)
Word list recognition	10 (1-10)	6 (0-10)	4 (0-7)
Praxis recall	7 (0-11)	0 (0-7)	0 (0-2)

curve. Table 4 shows the cut-off considering specificity and sensitivity for the separation of normal elderly and AD. All considered, MMSE had the best balance between sensitivity and specificity, though other sub-tests were more specific. The cut-off of 26 is not the same of the original report (24 points - ref. 11), but is in agreement with our previous research on a Brazilian sample of people with more than 8 years of education¹². Not surprisingly the memory tests – word list memory, word list recall, praxis recall, and associate verbal learning - were also efficient. For language tests, while verbal fluency was relatively specific, the naming test was not so efficient, sensitivity being the lowest among all sub-tests. The lowest specificity was for the constructional praxis test, which also had a low sensitivity. Our data are in accordance with those of a study including people with less than 12 years of education¹⁷, except for a higher score in verbal fluency (14.4) and a lower score (8.8) in praxis. Table 3 shows that there was a wide range variation for all tests.

DISCUSSION

This research was useful in identifying some neuropsychological tests suitable for application at large

Table 4. Cut-off scores, sensitivity and specificity for sub-tests of CERAD battery.

CI	Cut-off	sens.	spec.	area under ROC curve	Standard error	95%
Verbal fluency	11	73.8	87.1	0.877	0.03	0.807-0.929
Boston naming	12	61.9	69.4	0.699	0.04	0.611-0.711
MMSE	26	97.6	75.3	0.942	0.02	0.885-0.975
Word list memory	13	85.7	87.1	0.905	0.02	0.840-0.950
Constructional praxis	9	81.0	51.8	0.729	0.04	0.643-0.804
Word list recall	3	74.2	82.4	0.858	0.03	0.781-0.916
Word list recognition	7	76.2	87.1	0.875	0.03	0.804-0.927
Praxis recall	4	87.1	67.1	0.857	0.03	0.780-0.915

in our population. The MMSE mean of 23.0 for AD1 is an indication that this is a sample at a mild stage of dementia. This is important because this is the setting in which the diagnosis of dementia is more difficult, and where objective examination of cognitive functions would be more useful. This being a cross-sectional study, care should be taken about conclusions on disease evolution, but comparing between AD1 and AD2 there was difference for MMSE, verbal fluency and naming only. This could indicate that, after the initial decline, at the beginning of disease, scores remain relatively stable or that there is a stepwise progression. Our results could be compared with those of a larger CERAD sample⁸, in which a nonlinear rate of decline was found for some measures. It is not clear if the rate of progression is slower for less or more severely impaired subjects, and data supporting the first¹⁷ and second¹⁸ hypothesis were reported. Our study design is not appropriate to address this issue, which could be clarified by a prospective study.

As it can be concluded from Table 3, there is an overlap of scores between the three groups, except for MMSE and word list recall when comparing CG and AD2. This means that for individual diagnosis, when a mild stage of dementia is being considered, the isolated performance in any of the sub-tests should be taken with care. The overlap is largely the result of similar performance between low education CG and higher education AD1. The same pattern was observed in a previous study with MMSE¹² and our impression was that test sensitivity and specificity could be increased by using different cut-offs for different education groups. This may also be the case with CERAD battery.

The sensitivity and specificity of memory recall should not be a surprise in a dementing illness. The verbal fluency test, very easily applied, has shown to be useful in this population. It should again be stressed the bias of education, and the cut-off of 11, which is very close to that of a community-based study in the metropolitan area of São Paulo, designed to establish norms for the general population, which included middle-aged and young adults¹⁹. This similarity of results could be an indication that our control group is representative of the general population, but a more detailed analysis would be necessary. Though the naming test performance was close to that of the American population², it should be noted that the drawings were chosen according to word frequency in American English, but this could not be the same for Brazilian Portuguese.

With the results of this research, the usefulness of MMSE as a screening test is confirmed. The CERAD battery seems to be a quick and valid set of tests for the diagnosis of dementia, which could be further shortened by the suppression of sub-tests like memory recognition without loss of efficiency. It should still be reminded that there was an effect of education, which was verified for this battery, by the classification of CG in groups with less or more than 12 years of education⁶. The same would probably happen with our population, for which stratification for education levels should be more detailed. To increase the efficiency of this battery, a study of this type will be necessary.

REFERENCES

- McKhann G, Drachman D, Folstein MF, Katzman R, Price D, Stadlan 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.
- Morris JC, Heyman A, Mohs RC, 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;39:1159-1165.
- Mirra SS, Heyman A, McKeel D, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* 1991;41:479-486.
- Davis PC, Gado M, Albert M. CERAD protocol for assessing neuroimaging (MRI) findings in Alzheimer's disease. *Neurology* 1988;38 (Supp 1):43 (abst).
- Welsh KA, Butters N, Hughes J, Mohs R, Heyman A. Detection of abnormal memory in mild cases of Alzheimer's disease using CERAD neuropsychological measures. *Arch Neurol* 1991;48:278-281.
- Welsh KA, Butters N, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology* 1994;44:609-614.
- Koss E, Edland S, Fillenbaum G, et al. Clinical and neuropsychological differences between patients with earlier and later onset of Alzheimer's disease. A CERAD analysis. Part XII. *Neurology* 1996;46:136-141.
- Morris JC, Edland S, Clark C, et al. 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;43:2457-2465.
- Isaacs B, Kennie AT. The set test as an aid to the detection of dementia in old people. *Brit J Psych* 1973;123:467-470.
- Kaplan EF, Goodglass H, Weintraub S. The Boston Naming Test. 2nd ed. Philadelphia: Lea & Febiger, 1983.
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psych Res* 1975;12:189-198.
- Bertolucci PHF, Brucki SMD, Campacci SR, Juliano Y. O Mini-exame do Estado Mental em uma população geral. Impacto da escolaridade. *Arq Neuropsiquiatr* 1994;52:1-7.
- Atkinson RC, Shiffrin RM. The control of short-term memory. *Scientific American*, 1971;221:82-90.
- Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psych*, 1984;141:23-24.
- Mohs RC, Kim Y, Johns CA. Assessing change in Alzheimer's disease: memory and language tests. In Poon LW (ed). *Handbook for clinical memory assessment of older adults*. Washington: American Psychological Association 1986;149-155.
- Hughes CP, Berg L, Danziger W, Cohen LA, Martin RL. A new clinical scale for the staging of dementia. *Brit J Psych* 1982;140:566-572.
- Botwinick J, Storandt M, Berg L. A longitudinal study of dementia of the senile type. *Arch Neurol* 1986;43:1124-1127.
- Katzman R, Brown T, Thal LJ. Comparison of rate of progression in four independent studies of patients with Alzheimer's disease. *Ann Neurol* 1988;24:384-389.
- Brucki SMD, Malheiros SMF, Okamoto IH, Bertolucci PHF. Dados normativos sobre o teste de fluência verbal categoria animais em nosso meio. *Arq Neuropsiquiatr* 1997;55:57-61.