

Evaluating Language Comprehension in Alzheimer's disease: the use of the Token Test

Avaliação da compreensão da linguagem na doença de Alzheimer: o uso do Token Test

Jonas Jardim de Paula¹, Laiss Bertola¹, Rodrigo Nicolato², Edgar Nunes de Moraes², Leandro Fernandes Malloy-Diniz³

ABSTRACT

Objective: To analyze the psychometric properties of the Token test (TT), a verbal comprehension test, and its applicability to the diagnosis of mild Alzheimer's disease (AD). **Methods:** One hundred and sixty participants (80 AD and 80 controls) performed the TT and a short battery of neuropsychological tests designed to evaluate general cognitive status, working memory and executive functions. Internal consistency, factor structure, correlation with other measures and group comparisons were evaluated. **Results:** The test evinced good internal consistency and a bi-factorial structure (related to comprehension and attention). Differences between AD and controls were significant, however the TT presented only moderate sensitivity and specificity for the AD diagnosis. **Conclusion:** The TT showed evidence of good psychometric properties and adequacy for characterizing comprehension deficits in AD, but it was not an appropriate test for the AD detection and diagnosis.

Key words: Token test, psychometric properties, language, Alzheimer disease.

RESUMO

Objetivo: Investigar as propriedades psicométricas do Token test (TT), um teste de compreensão verbal, e sua aplicabilidade no diagnóstico da doença de Alzheimer em fase inicial (DA). **Métodos:** Cento e sessenta participantes (80 DA e 80 controles) foram avaliados por meio do TT e de testes neuropsicológicos que avaliam a cognição geral, memória operacional e funções executivas. Avaliou-se a consistência interna, correlações com outros testes neuropsicológicos, estrutura fatorial e poder diagnóstico do TT. **Resultados:** O teste apresenta boa consistência interna e estrutura bi-fatorial (relacionada à compreensão e atenção). Foram encontradas diferenças significativas entre o desempenho de DA e dos controles. Contudo, a sensibilidade e especificidade do TT para o diagnóstico de DA foi apenas moderada. **Conclusão:** O TT apresenta boas propriedades psicométricas e mostra-se adequado para a caracterização de comprometimentos de linguagem na DA. Entretanto, nossos resultados sugerem que o teste por si só é insuficiente para a detecção e diagnóstico dessa demência.

Palavras-Chave: Token test, propriedades psicométricas, linguagem, doença de Alzheimer.

The Alzheimer's disease (AD) is the most diagnosed dementia in late life and one of the most incapacitating syndromes of elderly. AD progression is characterized by impairment in episodic memory, associated with at least one more cognitive deficit (agnosia, executive dysfunction, apraxia or aphasia). Furthermore, AD patients always present functional impairment in comparison to their previous level of performance¹.

The impairment in language abilities is a common feature of AD, even in early stages. The neuropsychological evaluation of language in AD patients usually reveals deficits in language production and comprehension². Furthermore, lexical and semantic processes are often impaired, leading to

communication deficits commonly referred to as an aphasic syndrome³. Due to the expressive role that communication (specially spoken language) plays in daily life functioning, the evaluation of these abilities is essential for diagnosis, treatment and even for cognitive assessment interpretation in AD patients. One of the most significant language impairments is the deficit in verbal comprehension abilities, including the process of accurately understanding verbal orders. Previous studies indicate that comprehension deficits are associated with poor social interaction and quality of life⁴. In this context, effective ways of evaluating comprehension skills may play a crucial role in patients' rehabilitation.

Faculdade de Filosofia e Ciências Humanas da Universidade Federal de Minas Gerais (UFMG), Belo Horizonte MG, Brazil.

¹Laboratório de Investigações Neuropsicológicas (LIN) da Universidade Federal de Minas Gerais (UFMG), Belo Horizonte MG, Brasil;

²Departamento de Saúde Mental, Faculdade de Medicina, UFMG, Belo Horizonte MG, Brasil;

³Departamento de Psicologia, Faculdade de Filosofia e Ciências Humanas, UFMG, Belo Horizonte MG, Brasil.

Correspondence: Jonas Jardim de Paula; Av. Antônio Carlos, 6627; 31270-901 Belo Horizonte MG - Brasil; E-mail: jonasjardim@gmail.com

Conflict of interest: There is no conflict of interest to declare.

Received 20 September 2011; Received in final form 12 January 2012; Accepted 19 January 2012

A useful neuropsychological task devised to assess comprehension is the Token test (TT), a screening tool for deficits in receptive language in aphasics⁵. It consists of a series of colored tokens in squared or round forms and two different sizes, and the patient is instructed to perform a series of commands over them, such as “touch the red circle” or “touch the small yellow square and the big red circle”. The test has shown great efficiency in detecting and characterizing comprehension impairments. It has some advantages for the appraisal of these features, such as easiness of testing, non-verbal output, different levels of complexity, low cost and quick application⁵⁻⁸.

The TT is a worldwide screening tool with psychometric studies in several countries⁸. The test presents several different versions, including a short one⁹. Similarly to other countries, there are several Brazilian studies suggesting the clinical applicability of the TT to language assessment in children⁶, adults¹⁰ and elderly^{7,11,12}. Nonetheless, there are not Brazilian studies assessing psychometric properties of the TT, such as its validity, on Brazilian elderly subjects.

Therefore, this study aimed at investigating psychometric properties of the TT short version, for example, its internal consistency and the effects of aging, formal schooling, gender and depressive symptoms in test performance, as well as analyzing TT’s construct validity, the test criteria validity for mild AD and comparing test performance between these patients and normal controls (NC).

METHODS

Sample and procedure

A total of 160 participants, 80 diagnosed with probable mild AD by the NINCDS-ADRDA criteria¹³ and 80 NC were studied

in the present work. The inclusion criteria in the second group was absence of neurological or psychiatric diseases investigated by a clinical interview conducted by the neuropsychologist, no functional impairment (evaluated via Lawton-Brody Inventory¹⁴), total score above the appropriated cutoff score on the Mini-Mental State Exam (MMSE) Brazilian Version following the formal education criteria¹⁵ and GDS-15 scores below the proposed cutoff for depression in older Brazilians¹⁶. AD patients were examined at least by one gerontologist and one clinical neuropsychologist. Diagnoses were defined by a consensus of a multidisciplinary assessment in agreement with the previous criteria. All AD participants pursued their treatment plans, which included taking cholinesterase inhibitors. The Table 1 shows the patients demographic data.

Token test short version

The short version contains 36 orders which must be performed by the patient in 20 wooden pieces (five small squares, five big squares, five small circles, five big circles, each one presented in five colors – green, red, yellow, black and white). The test is divided into six parts, in that the orders become progressively more complex. Test total score is computed by the sum of correct items (one point each).

Procedures

The participants performed the short version TT, MMSE¹⁵, semantic verbal fluency (animals, fruits) (SVF), and forward Digit Span (DS) and Corsi Blocks¹⁷ (CB), two tests designed for the evaluation of verbal and spatial working memory, respectively.

Psychometric properties

The aim of the psychometric study of the TT is to assess its applicability and potential use as a neuropsychological

Table 1. Demographic data, neuropsychological tests scores and group comparisons.

	NC (n=80)		AD (n=80)		Group comparisons		
	Mean	SD	Mean	SD	Sig. (0.05)	d	AUC
Age	73.26	9.18	75.29	6.44	0.108	-	-
Education	5.64	4.22	4.94	3.49	0.252	-	-
MMSE	26.09	3.80	20.02	4.11	0.000	1.54	0.859
GDS	2.60	1.96	4.43	3.07	0.000	0.73	-
SVF - Animals	13.29	4.06	8.42	3.51	0.000	1.29	0.823
SVF - Fruits	12.09	4.12	7.99	2.85	0.000	1.18	0.792
TT	30.01	3.36	26.60	4.62	0.000	0.86	0.724
TT - SS	11.90	2.97	9.36	3.40	0.000	0.50	0.725
Digit Span	33.95	16.28	28.25	13.12	0.016	0.39	0.607
Corsi Blocks	30.06	14.66	21.32	10.13	0.000	0.71	0.669
TT - Part 1	6.93	0.22	6.97	0.16	0.233	-	-
TT - Part 2	3.81	0.48	3.61	0.82	0.052	-	-
TT - Part 3	3.83	0.64	3.74	0.54	0.332	-	-
TT - Part 4	3.72	0.53	3.25	1.02	0.000	0.61	0.638
TT - Part 5	3.44	0.87	2.86	1.08	0.000	0.60	0.661
TT - Part 6	8.18	2.41	6.17	2.68	0.000	0.79	0.709

AD: Alzheimer’s disease; NC: normal controls; MMSE: Mini-Mental State Exam; GDS: geriatric depression scale; SVF: semantic verbal fluency; TT: short version Token test; TT-SS: short version Token test – scaled scores; d: Cohen’s d; AUC: area under curve; Sig: significance level; SD: standard deviation.

logical measure of verbal comprehension. The first step was the analysis of the test internal consistency, which was calculated by the Cronbach's alpha for the 36 items. It was expected from moderate to high internal consistency by this method, since test orders and its difficulty level broadly vary along the test. The second step was to assess the association between the TT performance and age (years), schooling (years) and depressive symptoms measured by the geriatric depression scale (GDS)¹⁸, using Person's correlation. Construct validity was first analyzed by partial correlations (controlled by age, education and depressive symptoms) between the TT and other neuropsychological tests (MMSE, SVF, DS and CB). Then, a principal components analysis with varimax rotation was performed, in order to evaluate the TT latent structure. Considering the bias from the two procedures, pointed by Delis et al.¹⁹, it is seen that they were performed independently in the controls, in the clinical group and also in the mixed sample. Since this procedure may be more useful for construct validity when performed among homogeneous samples, only patients with mild AD were included in this study. Differences in the factor structure possibly represent changes in performance in the evaluated construct, which may be more evident when the patterns of variance and correlation between different populations are analyzed. The confidence level for all the analyses was defined at 0.05.

Group comparisons

The AD patients were compared to healthy older adults matched by age, gender and educational background (total years of formal schooling) by independent sample's *t*-tests, and the effect sizes were computed by the Cohen's *d*. For these group comparisons, the TT raw scores and also the scaled scores proposed by Moreira et al.⁷ in the test normalization for Brazilian older adults were used. Scaled scores had the advantage of gender, age and educational control, offering a more unbiased result. The receiver characteristic operator curve (ROC) analysis was used to evaluate the sensitivity and specificity of the TT for AD diagnosis when compared to NC. The confidence level for all the analyses was 0.05.

RESULTS

The 36 items of the TT show from moderate to high internal consistency in controls ($\alpha=0.746$), AD patients ($\alpha=0.809$) and the whole sample ($\alpha=0.837$).

In controls, the TT score correlated with education ($r=0.533$; $p<0.001$; $r^2=0.28$) and the GDS score ($r=-0.311$; $p<0.005$; $r^2=0.10$), but in AD no significant correlations were found between the TT and demographic data. Considering all the participants, significant correlations were found only between the TT and education ($r=0.324$; $p<0.001$; $r^2=0.10$).

Considering the association between the TT and other neuropsychological measures, a significant partial correlation was found only with the MMSE ($r=0.398$; $p<0.001$; $r^2=0.16$), and trends with DS ($r=0.198$; $p=0.084$; $r^2=0.04$) and the Clock Drawing Test ($r=0.217$; $p=0.058$; $r^2=0.05$) in the NC were found. In AD patients, the TT was related to the MMSE total score ($r=0.343$; $p<0.001$; $r^2=0.12$) and to CB ($r=0.290$; $p=0.011$; $r^2=0.08$), and trends were found with DS ($r=0.203$; $p=0.076$; $r^2=0.04$) and SFV "fruits" ($r=0.215$; $p=0.060$; $r^2=0.05$). When the performance of all the participants was considered, the TT score was correlated with all the other neuropsychological measures, MMSE ($r=0.505$; $p<0.001$; $r^2=0.26$), SVF "Animals" ($r=0.331$; $p<0.001$; $r^2=0.11$), SVF "fruits" ($r=0.317$; $p<0.001$; $r^2=0.10$), DS ($r=0.220$; $p=0.006$; $r^2=0.05$) and CB ($r=0.334$; $p<0.001$; $r^2=0.11$).

The principal component analysis reveals a two-component structure in controls, in AD patients and in the whole sample (Table 2). The consistency of the components was the same to the three groups, with approximated Eigenvalues and total variance explained. The first component seems to be more related to verbal comprehension, while the second one seems to be more related to attention. For further exploration, the TT was used in another principal component analysis, this time also containing the MMSE total score, SFV total words, DS and CB results. These results are exposed on Table 3.

The control group and AD patients mean and standard deviation of age, education, depressive symptoms, as well as the test scores for each neuropsychological measure, significance of the *t*-tests, effect sizes and area under curve were exposed on Table 1. The cutoff raw score in the TT of 27/28

Table 2. The components structure of the Token test short version.

	NC		AD		Mixed	
	1	2	1	2	1	2
TT - Part 1	0.295	0.693	0.008	0.868	0.081	0.798
TT - Part 2	0.739	0.291	0.631	0.051	0.651	0.122
TT - Part 3	0.270	0.790	0.217	0.775	0.057	0.804
TT - Part 4	0.681	0.458	0.734	0.060	0.747	0.177
TT - Part 5	0.790	0.023	0.822	0.092	0.832	0.002
TT - Part 6	0.701	0.112	0.776	0.169	0.794	0.016
Eigenvalue	2.364	1.255	2.449	1.209	2.397	1.185
% Variance	37.91	22.39	37.67	23.27	38.58	21.1

NC: normal controls; AD: Alzheimer's disease; TT: short version Token test.

Table 3. Factor structure of the neuropsychological tests.

	NC			AD			Mixed		
	1	2	3	1	2	3	1	2	3
MMSE	0.898	0.008	0.103	0.865	0.037	-0.046	0.830	0.283	0.117
SVF - Animals	0.186	0.862	0.065	0.150	0.858	0.050	0.282	0.855	0.103
SVF - Fruits	0.120	0.811	0.245	0.082	0.890	0.024	0.215	0.871	0.167
Digit Span	0.046	0.122	0.915	0.144	0.009	0.954	0.079	0.110	0.938
Corsi Blocks	0.392	0.217	0.691	0.649	0.042	0.422	0.505	0.218	0.603
TT	0.765	0.241	0.196	0.684	0.249	0.124	0.820	0.212	0.158
Eigenvalue	2.738	1.107	0.967	2.204	1.335	0.944	3.082	1.135	0.935
% Variance	45.629	16.111	14.934	36.728	22.247	14.058	51.364	15.587	12.100

AD: Alzheimer's disease; NC: normal controls; MMSE: Mini-Mental State Exam; SVF: semantic verbal fluency; TT: short version Token test.

(case/non-case) shows the best balance in sensitivity and specificity (0.725/0.637), and the cutoff 9/10, the best for scaled scores (0.650/0.675). The AUCs of the TT can be classified as weak/moderate.

DISCUSSION

The TT shows from moderate to high internal consistency, confirming our prediction. This finding indicates that the test items had a good correlation pattern, however not being redundant. We didn't find studies in Brazil which analyzed the test internal consistency, notwithstanding our results are similar to those from Gallardo et al.²⁰ and Spellacy and Spreen²¹. In controls, the test was significantly correlated with education and depressive symptoms, but not with age. Nevertheless, these correlations were not significant in the AD group. The moderate correlation between education and test performance was congruent with Moreira et al.⁷ findings, except by the lack of association with age, since the author and other studies^{11,12} have found that data. The influence of depressive symptoms in the TT is not well documented, but the weak correlation found in our study indicates that depression may be associated with worst performance, as pointed by Brown et al.²².

The different pattern of correlations between the TT, the socio-demographic data and other cognitive measures, according to groups, can be analyzed through the concept of cognitive reserve. Lower age and formal education are associated with cognitive reserve, which may delay the expression of clinical features of AD²³. In this perspective, the correlation between age, education and cognitive performance in healthy older adults may be seen as an expression of cognitive reserve. However, after the dementia onset, these two features may not influence cognitive performance, since cognitive reserve is reduced or depleted. Our findings suggest that indicators of cognitive reserve may influence the period of onset of dementia, but, when the syndrome is stabilized, the moderate effect of this variable may not influence performance on the TT.

The highest correlations found between the TT and other cognitive tests in all the three groups were those with the MMSE. Our results are similar to Swihart et al.²⁴, although this author had found even higher correlations. Considering the total sample, the TT also correlates moderately with DS, CB, SVF "animals" and SVF "fruits". The correlation between the TT and verbal fluency was well established in our previous study in a children sample⁶. However, in normal aged controls (NC) no significant correlations were found besides the MMSE's. This result may account for the independence of the construct evaluated by the TT when compared to tasks of working memory and executive functions. In AD patients, a significant correlation was found between the TT and CB, suggesting that nonverbal working memory may play a role on the test performance.

The TT component structure indicates that two major factors explain most of test variance in the three groups. The first component, named Complex Verbal Comprehension, had greater factor loadings in the TT parts two, four, five and six, while factor 2, Simple Verbal Comprehension, showed greater loadings in parts one and three. Surprisingly, the second part of the TT loaded on the component one. The seven first items of the test screen the ability to correctly identify each stimulus (shape, size and color), while, in the second part, a comprehension process is more evident, demanding a different behavior. However, in the third part, while the commands are still simple and the subject has comprehended the test structure, the attentional process is more evident, changing only when the commands become more complex.

Our results are different from those presented by Hula²⁵, according to whom the TT first five components loaded on one factor and the last part loaded on another. Mansur et al.¹⁰, exploring the speech and language disturbances in adults, found that only the first part of the TT was not able to differentiate controls from aphasic patients. According to these findings, our results can also be interpreted in another way. Assuming that the third part of the test was expected to load on the Complex Verbal Comprehension factor, we can suppose that the existence of two similar commands, in the parts two and three, enables the occurrence of a priming

effect which can act like a facilitator for the comprehension process. Also, differences in the test version and in the studied population may influence this finding inconsistency.

When the principal component analysis was performed with the TT and other neuropsychological measures, a three component structure was the most adequate for all the groups. The three factors seem to represent a general-cognition/comprehension, verbal fluency and working memory components, respectively. As expected, due to the previous correlation, the MMSE and the TT loaded on the same component in the three groups. These results differ from prior findings using the same tests in a heterogeneous cognitively impaired sample, suggesting a more characteristic pattern of mild AD¹². Interestingly, the forward CB loaded partially on the first factor in the AD and the mixed group, indicating an association of spatial working memory and test performance in these, probably because the tests demand visuospatial processing for a motor output. This result and the lack of correlation between verbal working memory performance (measured by the DS) and the TT suggest that the worst performance of mild AD patients would be more associated with the motor output processing and programming rather than with the verbal input processing and comprehension. Future studies may verify the contribution of verbal and spatial working memory on test performance.

The comparisons between AD patients and normal aged controls indicated that AD patients' performance on the TT differs in the Complex Verbal Comprehension component (except for the second part despite its strong tendency to significance), but not in the Simple Verbal Comprehension one. Differences in test score were significant and show great (raw scores) or moderate (scaled scores) effect sizes, suggesting a robust dissociation of test performance when the two groups are compared. However, the ROC analysis indicates only low to moderate sensitivity and specificity. In this regard, as expected, the TT may not be a robust neuropsychological test for the discrimination of mild AD, since the purpose of the test is to detect aphasia. The classical semiology of this syndrome is found in less than 60% of the cases²⁶ and the expected degree of impairment in those usually is not found in AD early stages. Our findings are supported by other studies using the instrument²³. The progression of AD neuropathology and cortex diffuse degeneration usually implies a less specific pattern of cognitive impairment. Since language system is a complex cognitive network distributed in several different neuroanatomical regions, usually defined as superior, medial and inferior temporal and frontal gyrus, supramarginal gyrus, angular gyrus, insula, cingulate gyrus and subcortical structures like the ventrolateral and pulvinar nuclei of the thalamus²⁷, the disease can affect various of its components, making it difficult to correctly track a language impairment pattern. Considering the perspective of Neuronal Ensembles²⁸, critical components of a

specific language process (like the comprehension of one of the TT orders) demand a restricted pattern of activation from circumscribed neurons in different brain regions. A lesion in one of these or in the connection fibers between them may disrupt the process, making it difficult to specify the impairment location.

There are, however, some considerations about the impaired performance of AD patients in the TT. Based on the cognitive model of single word processing proposed by Morton and Patterson²⁹, the verbal comprehension of any single word would depend on the accurate perception of the auditory stimulus, being processed at the auditory input lexicon and analyzed by the semantic system. Since the TT uses instructions in different levels of complexity, a greater involvement of the verbal working memory and an accurate syntactical analysis of each sentence are also required. According to the language model proposed by Shalom and Poeppel³⁰, the language processing can be divided in three main processes: analyzing, memorizing, synthesizing, which, in turn, can be divided into different brain regions: parietal cortex, temporal cortex and frontal cortex. All the three processes seem to be related to the TT performance, since the test orders must be analyzed, interpreted based on phonological, syntactical and semantic processes, and synthesized in a motor output (which demands visuospatial analysis, planning skills, motor coordination and spatial working memory). This series of complex cognitive processes demanded for the test execution may increase its sensitivity for language deficits, but implies a lack of specificity, since circumscribed deficits in each task component may be at least partially responsible for impaired performance.

Limitations of the present study should be acknowledged: the sample size, the difficulty to control patients' medications and the absence of another instrument to measure language comprehension to be used as a parameter to validate the TT construct. Together, our results indicate that, while language impairment may be an important feature of AD, the TT should not be used alone in the evaluation process. A comprehensive neuropsychological exam is required to assess other cognitive abilities that may play a critical role in this test performance. Nonetheless, the test has presented robust psychometric properties and specific performance profiles which help to differentiate healthy older adults from mild AD patients.

In conclusion, the TT showed good evidences of psychometric properties and it was a robust instrument in the characterization of language impairments in mild AD patients, but it was not a sensitive and specific tool for detecting the disease. However, the TT is a multi-factorial test, and specific language impairments must be evaluated with some more specific instruments. Moreover, the scaled scores proposed by the most recent Brazilian normalization seem to be a more conservative approach for clinical practice.

References

1. American Psychiatric Association (APA). Diagnostic and statistical manual of the mental disorders (DSM-IV). 4th ed. Washington; 1994.
2. Altmann LJP, Kempler D, Andersen ES. Speech errors in Alzheimer's disease: reevaluating morphosyntactic preservation. *J Speech Lang Hear Res* 2001;44:1069-1082.
3. Aronoff JM, Gonnerman LM, Almor A, Arunachalam S, Kempler D, Andersen ES. Information content versus relational knowledge: semantic deficits in patients with Alzheimer's disease. *Neuropsychologia* 2006;44:21-35.
4. Hilari K, Byng S, Lamping DL, Smith SC. Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): evaluation of acceptability, reliability, and validity. *Stroke* 2003;34:1944-1950.
5. De Renzi E, Vignolo LA. The token test: a sensitive test to detect receptive disturbances in aphasics. *Brain* 1962;85:665-678.
6. Malloy-Diniz LF, Bentes RC, Figueiredo PM, et al. Normalização de uma bateria de tests para avaliar las habilidades de comprensión del lenguaje, fluidez verbal y denominación en niños brasileños de 7 a 10 años: resultados preliminares. *Rev Neurología* 2007;44:275-280.
7. Moreira L, Schlottfeldt CG, de Paula JJ, et al. Estudo normativo do Token Test versão reduzida: dados preliminares para uma população de idosos brasileiros. *Rev Psiquiatr Clin* 2011;38:97-101.
8. Strauss E, Sherman EMS, Spreen O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. New York: Oxford University Press; 2006.
9. De Renzi E, Faglioni P. Normative data and screening power of a shortened version of the Token Test. *Cortex* 1979;14:41-49.
10. Mansur LL, Radanovic M, Rüegg D, Mendonça LIZ, Scaff M. Descriptive study of 192 adults with speech and language disturbances. *São Paulo Med J* 2002;120:170-174.
11. Carvalho SA, Barreto SM, Guerra HL, Gama ACC. Oral language comprehension assessment among elderly: a population based study in Brazil. *Prev Med* 2009;49:541-545.
12. De Paula JJ, Schlottfeldt CG, Moreira L, et al. Psychometric properties of a brief neuropsychological protocol for use in geriatric. *Rev Psiquiatr Clin* 2010;37:251-260.
13. McKhann G, Drachman D, Folstein M, 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. *Neurology* 1984;34:939-944.
14. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969;9:179-186.
15. Brucki SMD, Nitrini R, Caramelli P, Bertolucci PHF, Okamoto IH. Sugestões para o uso do Mini-Exame do estado mental no Brasil. *Arq Neuropsiquiatr* 2003;61:777-781.
16. Paradelo EMP, Lourenço RA, Veras RP. Validation of geriatric depression scale in a general outpatient clinic. *Rev Saude Publica* 2005;39:918-923.
17. Kessels RPC, van den Berg E, Ruins C, Brands AM. The backward span of the Corsi Block-Tapping task and its association with the WAIS-III Digit Span. *Assessment* 2008;15:426-434.
18. Almeida OP, Almeida SA. Short version of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. *J Geriatr Psychiatry* 1999;14:858-865.
19. Delis DC, Jacobson M, Bondi MW, Hamilton JM, Salmon DP. The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: lessons from memory assessment. *J Int Neuropsychol Soc* 2003;9:936-946.
20. Gallardo G, Guàrdia J, Villaseñor T, McNeil MR. Psychometric data for the revised token test in normally development Mexican children's ages 4-12 years. *Arch Clin Neuropsychol* 2011;26:225-234.
21. Spellacy F, Spreen O. A short form of the Token test. *Cortex* 1969;5:391-397.
22. Brown RG, Scott LC, Bench CJ, Dolan RJ. Cognitive function in depression: its relationship to the presence and severity of intellectual decline. *Psychol Med* 1994;24:829-847.
23. Whalley L, Deary I, Appleton CL, Starr JM. Cognitive reserve and the neurobiology of aging. *Ageing Res Rev* 2004;3:369-382.
24. Swihart AA, Panisset M, Becker JT, Beyer JT, Beyer JR, Boller F. The Token Test: Validity and diagnostic power in Alzheimer's disease. *Developmental Neuropsychology* 1989;5:71-80.
25. Hula WD, Doyle PJ, McNeil MR, Mikolic JM. Rasch modeling of revised token test performance: validity and sensitivity to change. *J Speech Lang Hear Res* 2006;49:27-46.
26. Marshall JC. The description and interpretation of aphasic language disorder. *Neuropsychologia* 1986;24:5-24.
27. Kolb B, Whishaw I. *Fundamentals of Human Neuropsychology*. 5th ed. New York: Worth Publishers; 2003:483-515.
28. Pulvermüller F. A brain perspective on language mechanism: from discrete neuronal ensembles to serial order. *Progr Neurobiol* 2002;67:55-111.
29. Morton J, Patterson K. A new attempt at an interpretation, or, an attempt at a new interpretation. In: Coltheart M, Patterson K, Marshall JC (Eds). *Deep dyslexia*. London: Routledge Kegan Paul; 1980:91-118.
30. Shalom DB, Poeppel D. Functional anatomic models of language: assembling the pieces. *Neuroscientist* 2008;14:119-127.