

# Verbal fluency tests reliability in a Brazilian multicentric study, ELSA-Brasil

Valéria Maria de Azeredo Passos<sup>1</sup>, Luana Giatti<sup>1</sup>, Sandhi Maria Barreto<sup>1</sup>, Roberta Carvalho Figueiredo<sup>1</sup>, Paulo Caramelli<sup>1</sup>, Isabela Benseñor<sup>2</sup>, Maria de Jesus Mendes da Fonseca<sup>3</sup>, Nagela Valadão Cade<sup>4</sup>, Alessandra Carvalho Goulart<sup>2</sup>, Maria Angélica Nunes<sup>5</sup>, Márcia Guimarães de Mello Alves<sup>6</sup>, Ana Angélica Martins da Trindade<sup>7</sup>

## ABSTRACT

Verbal Fluency Tests (VFT) are commonly used to assess executive functions and language in clinical and epidemiological studies. These tests were included to assess cognitive function and predictors to its decline in ELSA-Brasil (Longitudinal Study of Adults' Health), a cohort that investigates incidence and predictors of chronic diseases among 15,000 civil servants from six public educational institutions. **Objective:** To investigate the reliability of VFT scoring by supervisors of ELSA research centers, who independently judged 120 category (animals) and 120 phonemic (F letter) tests. **Method:** Their scores were compared to a reference standard score obtained by independent judgment of two experts. Intraclass correlation coefficient ratings reliability and Bland-Altman plot examined patterns of ratings disagreement. **Results:** Scores were very similar among ELSA centers and a high level of agreement was observed between each center and the reference standard. **Conclusion:** The high consistency of VFT scores confirms reliability and validity of the test and assures quality for its use in multicenter studies.

**Key words:** verbal fluency test, reliability, multicenter study, cohort.

## Reprodutibilidade dos escores em testes de fluência verbal em estudo multicêntrico brasileiro

## RESUMO

**Introdução:** Testes de fluência verbal (TFV) são muito utilizados para verificação de linguagem e função executiva em estudos clínico-epidemiológicos. Estes testes foram incluídos no ELSA-Brasil (Estudo Longitudinal de Saúde do Adulto), coorte que investiga incidência e fatores prognósticos de doenças crônicas em 15.000 servidores públicos de seis instituições de ensino. **Objetivo:** Investigar concordância dos escores dos TFV aferidos por supervisores dos centros de pesquisa, que interpretaram independentemente 120 TFV para categorias (animais) e 120 TFV fonêmica (letra F). **Método:** Os escores foram comparados a padrão-referência, definido por julgamento independente de dois especialistas. O índice de correlação intraclasse mediu concordância entre os escores e o gráfico de Bland-Altman espelhou padrões de discordância. **Resultados:** Os escores foram muito similares, com alto grau de concordância entre cada centro e o padrão-referência. **Conclusão:** A alta consistência dos escores revela importância de mecanismos de garantia de qualidade e normatização em estudos multicêntricos e permite ao ELSA avançar na investigação cognitiva.

**Palavras-Chave:** testes de fluência verbal, concordância, estudo multicêntrico, coorte.

## Correspondence

Valéria Maria de Azeredo Passos  
Av. Alfredo Balena 110  
30130-100 Belo Horizonte MG - Brasil  
E-mail: vpassos@medicina.ufmg.br

## Support

Decit-Ministry of Health FINEP and CNPq. SMB and VMAP are research fellows of the National Research Council (CNPq, grant nº 300908/95 and 300159/99-4). Luana Giatti is research fellow of CAPES

## Conflict of interest

The authors report no conflict of interest

Received 27 October 2010  
Received in final form 10 May 2011  
Accepted 17 May 2011

<sup>1</sup>Faculdade de Medicina da Universidade Federal de Minas Gerais, Belo Horizonte MG, Brazil; <sup>2</sup>Faculdade de Medicina da Universidade de São Paulo, São Paulo SP, Brazil; <sup>3</sup>Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Rio de Janeiro RJ, Brazil; <sup>4</sup>Universidade Federal do Espírito Santo, Vitória ES, Brazil; <sup>5</sup>Faculdade de Medicina, Programa de Pós Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul, Porto Alegre RS, Brazil; <sup>6</sup>Instituto de Saúde da Comunidade da Universidade Federal Fluminense, Niterói RJ, Brazil; <sup>7</sup>Centro de Investigação ELSA-Bahia, Salvador BA, Brazil.

The pronounced and fast population ageing of low and middle-income countries is a reality. As a consequence, cognitive decline and dementia are now a leading cause of disability worldwide<sup>1</sup>.

Cognitive tests are used as a proxy of intrinsic and complex neurological mechanisms. In this way, studies on cognitive assessment face a great range of challenges, from determining the most sensitive and specific diagnostic tests to verifying the applicability and comparability of results among different population and countries, with different languages, cultures and levels of education. Normative data by age, gender and educational level have been provided for some of these cognitive tests in Brazil, generally obtained by cross-sectional studies<sup>2,3</sup>.

A major advantage of longitudinal over cross-sectional designs is that within-person cognitive changes can be directly measured. In Brazil, cognitive function and predictors to its decline is one of the interests of ELSA-Brasil (Escudo Longitudinal de Saúde do Adulto, Longitudinal Study of Adult's Health), an ongoing cohort study designed to investigate the incidence and predictors of chronic non-communicable diseases, including cognitive decline and dementia, among 15,000 federal civil servants from six public educational and research institutions<sup>4</sup>. Three previously validated cognitive tests to the Brazilian population were used in this study. The memory test from the Consortium to Establish a Registry for Alzheimer's disease evaluates immediate and delayed recall of a list of 10 words and the score is based on the total number of recalled words in each trial<sup>5</sup>. In the Trail Making Test B, executive function is evaluated by asking the participants to sequentially connect 26 circles containing numbers and letters; and the total score is given by time spent to correctly complete the test. Participants are also asked to perform semantic and phonemic Verbal Fluency Tests (VFT), where they have to say as many words as possible from a previous specified semantic category (animals) or beginning with a given letter (F) within one minute<sup>6</sup>.

The performance in VFT needs to be evaluated by a trained investigator, in order to count the words excluding repetitions and intrusions. Besides the psychometric properties (reliability and validity) of these tests, multicenter studies also have to concern about the homogeneity of their interpretation. The present study intends to assess the interpreter reliability of VFT scoring by six different supervisors of the ELSA research centers.

## METHOD

In order to obtain the quality assurance of ELSA cognitive tests, all efforts were made before data collection. A written manual with a detailed description of the procedures was used to train and certify the research staff. The feasibility and efficiency of the procedures were carried-out through pretests and a pilot study<sup>7</sup>.

During data collection, the participant responses to VFT were taped and written down. Periodic supervision and evaluation of the interviewers' performance were done, in order to maintain quality control. A written manual was provided with all items that would be considered as errors: intrusions (inappropriate words for the given category or word) and perseverations (same word repeated) and general category errors (when a participant said a general and also a specific item, e.g. bird and nightingale)<sup>8</sup>.

For this study, the six ELSA supervisors independently judged 120 category (animals) and 120 phonemic (F letter) tests, 40 from each center. The results of the scores were compared to a reference standard score obtained by two experts, a geriatrician (VMAP) and a neurologist (PC).

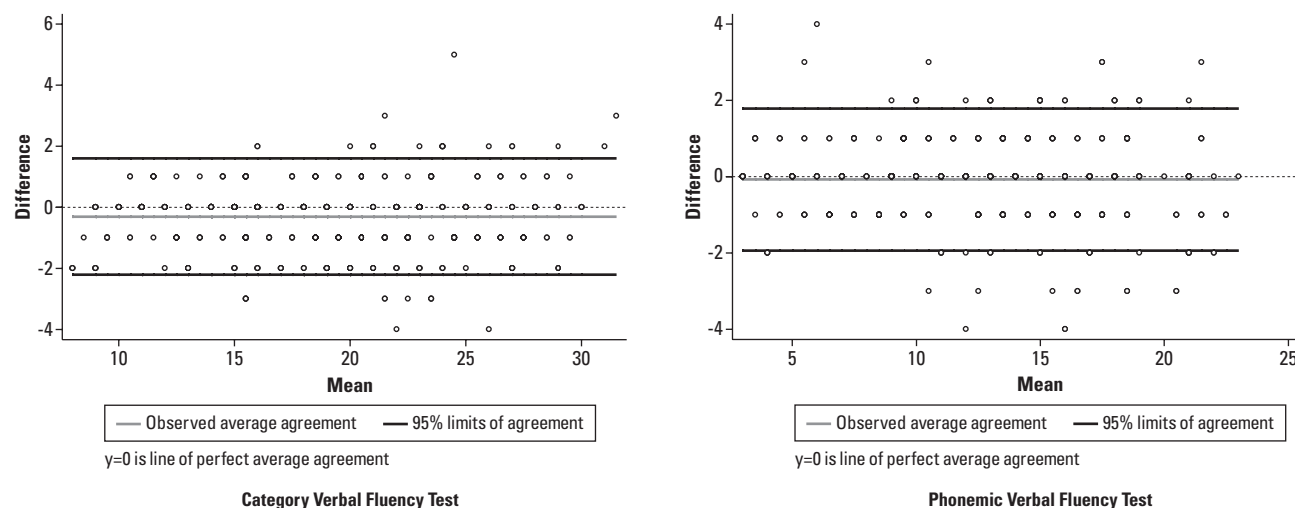
Data entry was carried out using the EPIDATA 3.1 software. The statistical analysis was carried out through the STATA version 10.0 software. The Intraclass Correlation Coefficient (ICC) was used to measure the reliability of ratings and Bland-Altman plot was included to examine the patterns of ratings disagreement<sup>7,9</sup>.

The study was approved by the Institutional Review

**Table.** Descriptive variables of Verbal Fluency Tests among the six research centers and the reference score in ELSA-Brasil.

ELSA Centers	Category Verbal Fluency Test (Animals)				Phonemic Verbal Fluency Test (F letter)			
	Percentiles				Percentiles			
	25	50	75	ICC *(CI 95%*)	25	50	75	ICC (CI 95%*)
1	15	18.5	22	0.982 (0.975-0.988)	9.5	14	16	0.982 (0.975-0.988)
2	14.5	18	22	0.979 (0.970-0.985)	9	13	15	0.979 (0.970-0.985)
3	14.5	18	22	0.987 (0.981-0.991)	9	13	15	0.987 (0.981-0.991)
4	15	18	22	0.986 (0.979-0.990)	9	13	15	0.986 (0.979-0.990)
5	15	18	22	0.980 (0.971-0.986)	9	13	16	0.980 (0.971-0.986)
6	14.5	18	22.5	0.986 (0.980-0.990)	9	12.5	15	0.986 (0.980-0.990)

ICC: intraclass correlation coefficient.



**Figure.** Agreement between scores obtained by supervisors and the reference standard scores obtained by experts.

Board of all institutions participating in the study and all participants signed an informed consent.

## RESULTS

Scores of VFT were very similar among ELSA centers and a high level of agreement was observed between each of the six centers and the reference standard (Table).

Data were also analyzed using the Bland-Altman plot (Figure). The mean difference is small for both tests, and 95% of scores are located between the mean and two standard deviations. However, it can be observed a small underestimation of the category scores determined by supervisors.

## DISCUSSION

VFT are commonly used to assess executive functions and language in clinical and epidemiological studies. Previous Brazilian studies have investigated normative, reliability and validity issues<sup>2,3,8,10</sup> but, to the best of our knowledge, this is the first study to evaluate interrater agreement.

Most measurements in human sciences involve measurement error, but judgments made by humans are specially plagued by this problem<sup>7,9</sup>. Since measurement error can affect analysis and interpretation of data, the validity of the inferences made from results of observational studies depends on the accuracy of its methods and procedures.

The high consistency of VFT scores among the ELSA research centers shows the importance of the imple-

mented measures of quality assurance and control. Besides that, this result allows ELSA cognitive investigation to proceed upon its main objective, to identify social, psychological and biological predictors of cognitive decline in a Brazilian adult population. Its urges all multicentric studies to evaluate interrater agreement in order to guarantee the quality of data analyze.

## REFERENCES

1. World Health Organization. Mental and neurological disorders: executive summary. [http://www.who.int/whr/1997/media\\_centre/executive\\_summary1/en/index14.html](http://www.who.int/whr/1997/media_centre/executive_summary1/en/index14.html) Accessed 12/10/2010.
2. Sosa AL, Albanese E, Prince M, et al. Population normative data for the 10/66 Dementia Research Group cognitive test battery from Latin America, India and China: a cross-sectional survey. *BMC Neurology* 2009;22:800-811.
3. Caramelli P, Carthery-Goulart MT, Porto CS, Charchat-Fichman Helenice, Nitrini R. Category fluency as a screening test for Alzheimer disease in illiterate and literate patients. *Alzheimer Dis Assoc Disord* 2007;21:65-67.
4. Estudo Longitudinal de Saúde do Adulto. Available at <http://www.elsa.org.br/> Accessed 10/10/2010.
5. Bertolucci PH, Okamoto IH, Brucki SM, Siviero MO, Toniolo Neto J, Ramos LR. Applicability of the CERAD neuropsychological battery to Brazilian elderly. *Arq Neuropsiquiatr* 2001;1:59-65.
6. Lezak MD, Howieson DB, Loring DW. *Neuropsychological assessment*. 4 ed. New York, NY: Oxford University Press, 2004.
7. Szklo M, Nieto FJ. *Quality Assurance and Control*. In: Szklo Moyses, Nieto FJ (Eds). *Epidemiology beyond the basics*. Aspen Publishers, USA, 2007: 297-348.
8. Machado TH, Fichman HC, Santos EL, et al. Normative data for healthy elderly on the phonemic verbal fluency task - FAS. *Dement Neuropsychol* 2009;3:55-60.
9. Shrout ES & Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-428.
10. Brucki SMD and Rocha MSG. Category fluency test: effects of age, gender and education on total scores, clustering and switching in Brazilian Portuguese-speaking subjects. *Braz J Med Biol Res* 2004;37:1771-1777.