

# BRBN-T validation: adaptation of the *Selective Reminding Test* and *Word List Generation*

Validação da BRBN-T: adaptação do *Selective Reminding Test* e do *Word List Generation*

Mariana Rigueiro Neves<sup>1</sup>, Ana Margarida Passos<sup>1</sup>, Aristides Ferreira<sup>1</sup>, Cláudia Sousa<sup>2</sup>, Andreia Sá<sup>2</sup>, Maria José Sá<sup>2</sup>

## ABSTRACT

**Objective:** This study aims to present the *Selective Reminding Test* (SRT) and *Word List Generation* (WLG) adaptation to the Portuguese population, within the validation of the *Brief Repeatable Battery of Neuropsychological Tests* (BRBN-T) for multiple sclerosis (MS) patients. **Method:** 66 healthy participants (54.5% female) recruited from the community volunteered to participate in this study. **Results:** A combination of procedures from Classical Test Theory (CTT) and Item Response Theory (ITR) were applied to item analysis and selection. For each SRT list, 12 words were selected and 3 letters were chosen for WLG to constitute the final versions of these tests for the Portuguese population. **Conclusion:** The combination of CTT and ITR maximized the decision making process in the adaptation of the SRT and WLG to a different culture and language (Portuguese). The relevance of this study lies on the production of reliable standardized neuropsychological tests, so that they can be used to facilitate a more rigorous monitoring of the evolution of MS, as well as any therapeutic effects and cognitive rehabilitation.

**Keywords:** neuropsychological tests, multiple sclerosis, working memory, verbal fluency, psychometrics.

## RESUMO

**Objectivo:** Este estudo tem como objectivo a adaptação do *Selective Reminding Test* (SRT) e do *Word List Generation* (WLG) para a população portuguesa, no decurso da validação da *Brief Repeatable Battery of Neuropsychological Tests* (BRBN-T) para doentes com esclerose múltipla (EM). **Método:** 66 sujeitos saudáveis (54.5% sexo feminino), foram recrutados da comunidade e voluntariaram-se para, participar neste estudo. **Resultados:** Uma combinação de procedimentos da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI) foi aplicada à análise e seleção dos itens. Para cada lista do SRT foram seleccionadas 12 palavras e para o WLG foram escolhidas 3 letras, que constituem a versão final destes testes para a população Portuguesa. **Conclusões:** A combinação da TCT e da TRI maximizou a tomada de decisão na seleção de itens na adaptação do SRT e do WLG para uma língua e cultura diferentes (o Português). A relevância deste estudo prende-se com a produção de testes de avaliação neuropsicológica fidedignos, que podem ser utilizados na monitorização rigorosa da evolução da EM, efeitos da terapêutica e reabilitação cognitiva.

**Palavras-chave:** testes neuropsicológicos, esclerose múltipla, memória de curto prazo, fluência verbal, psicometria.

Multiple sclerosis (MS) is a chronic, neurodegenerative disease whose various neurological manifestations have a multidimensional impact on patients' lives and clinical status<sup>1,2</sup>. It mostly affects young adults, aged between 20 and 45 and it has a prevalence of about 450,000 patients in Europe and around 5000 patients in Portugal<sup>3</sup>. In addition to the physical symptoms associated with MS, 40 to 70% of the patients suffer from cognitive deficits that neurological exams cannot detect<sup>4,5,6</sup>. This cognitive impairment can occur in the early stages of the disease and increase over time, having a negative impact on the MS patients' quality of life<sup>2,4,5,7</sup>.

Given the frequency and the importance of the cognitive impairment of MS, appropriate neuropsychological

assessment is an extremely relevant complementary procedure to neurological exams<sup>4,8</sup>. These results, supported by normative quality data, provide important support for clinical and therapeutic decisions, assessment of intervention efficacy and MS research<sup>1</sup>. Therefore, it is important to have instruments that are sensitive, reliable and easy to administer in order to implement neuropsychological assessment as a part of the routine care of MS patients. In recent years, brief test batteries to assess cognition in MS have been proposed and the most commonly reported in the literature is the *Brief Repeatable Battery of Neuropsychological Tests* (BRBN-T) which has been validated in several cultures and languages<sup>1,4,8</sup> and has proved to be one of the most sensitive

<sup>1</sup>Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL, Lisboa, Portugal;

<sup>2</sup>Hospital S. João, Serviço de Neurologia, Porto, Portugal.

**Correspondence:** Mariana Rigueiro Neves; Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL; Av. das Forças Armadas, Sala 2W8; 1649-026 Lisboa, Portugal; E-mail: [marianarigueironeves@gmail.com](mailto:marianarigueironeves@gmail.com)

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

Received 11 April 2014; Received in final form 14 May 2015; Accepted 02 June 2015.

measures for detecting cognitive impairment in MS<sup>6,9,10</sup>. As the MS cognitive deficits most frequently involve working and episodic memory, sustained attention, information processing speed and mental flexibility<sup>6,9,11</sup>, Rao et al.<sup>4</sup> found that the most sensitive tests for detecting cognitive impairment in this disease were: *Selective Reminding Test* (SRT), *10/36 Spatial Recall Test* (SPART), *Symbol Digit Modalities Test* (SDMT), *Paced Auditory Serial Addition Test* (PASAT) and *Word List Generation* (WLG)<sup>14,10</sup>. There is a consensus regarding BRBN-T psychometric properties, particularly its sensitivity and reliability in detecting cognitive impairment<sup>4,6,8</sup>. A study by Strober<sup>10</sup> and Sepulcre<sup>9</sup> found that BRBN-T has discriminative validity, as it distinguishes MS from healthy controls, with MS patients performing worse than the controls. Many studies have also showed that BRBN-T test results are correlated with multiple brain MRI measures, such as ventricle enlargement<sup>12</sup>, neocortical volume<sup>13</sup> and reveal longitudinal changes in MS<sup>5</sup>.

The surplus values associated with the use of psychological tests is that they provide a common measurement scale, and hence the potential for ascertaining the differences and similarities in unobservable aspects of interest in a population<sup>14,15,16</sup>. However, this potential cannot be fully realized, unless tests are used with comparable validity in all the populations involved<sup>14</sup>. Cross-cultural differences are considered an important issue in any research that deals with neuropsychological assessment<sup>15,16</sup>. Despite the fact that cognitive disturbances associated with brain pathology have similar manifestations across the human species, cognitive abilities measured by neuropsychological tests represent, at least in their contents, culturally learned abilities<sup>15</sup>. Human performance on neuropsychological tests is, therefore, influenced by a range of moderating variables, such as: culture, primary language and level of education<sup>14,15</sup>. For these reasons, the assessment of a different cultural group using the original neuropsychological tests, procedures and norms, results in conceptual errors in assessment<sup>14,15,16,17</sup>.

Thus, in order to routinely implement neuropsychological assessment in MS, we need to devise accurate neuropsychological tests, with determined and established norms, validated for the specific population we intend to analyze<sup>16,18</sup>. This study is part of a project whose goal is to improve the neuropsychological assessment of MS patients in Portugal, through the adaptation and validation of the BRBN-T. As *SRT* and *WLG* are tests with a strong verbal content (they use a list of words and a set of letters, respectively), special attention is needed when adapting them to other languages (e.g. Portuguese). In this case, a precise translation requires the maintenance of words frequency between the original (English) and the target language (Portuguese)<sup>19</sup>. This exploratory study aims to adapt the *SRT* and *WLG* to the Portuguese context, and for the effect a combination of procedures from Classical Test Theory

(CTT) and Item Response Theory (IRT) approaches was used in order to maximize the decision making process for words and letters selection.

## METHOD

### Subjects

Sixty-six healthy participants participated in the study and gave informed consent. The participants were recruited among relatives and friends from the medical team of the hospital and from the research team, among relatives and friends of MS patients from the MS Outpatient Clinic (Hospital de S. João) and word of mouth. All participants were aware of the experimental nature of this study. The sociodemographic characteristics of the healthy participants are presented in Table 1. The majority of the participants who performed in the battery of tests was women (54.5%) with an average age of 36,4 years old (ranging from 21 to 59) and not graduated (57.6%).

Inclusion criteria for the study were: i) age 20 or older, ii) Portuguese as their native language, iii) attended school in Portugal, iv) completed primary school. Exclusion criteria for the study were: i) neurological disorder, ii) serious head injury, iii) major psychiatric illness, iv) history of alcohol or drug abuse, v) learning disability, vi) other major medical illnesses and vii) regular dosage of antidepressants or anxiolytics. Every person had the necessary vision, audition and motor skills to complete the testing.

Data collection was previous approved by the Portuguese Data Protection Authority and by the ethical committee of Hospital de São João.

### Neuropsychological tests procedures

The participants were examined by two well-trained neuropsychologists. The tests used in this exploratory study were administrated in a standardized way and in a fixed order: (i) *SRT*<sup>20</sup>, (ii) *WLG*<sup>4</sup>, (iii) *Symbol Search*<sup>21</sup>, (iv) *Coding*<sup>21</sup> and (v) Delayed recall of the SRT. The administration of all of the tests took approximately 25 minutes.

(i) *Selective Reminding Test (SRT)*: measures verbal learning and memory through a learning task in a paradigm of multiple trials. The list includes 12 words which the examiner reads at a rate of one word per two seconds. The participant

**Table 1.** Sociodemographic characteristics of the healthy participants undergoing *Selective Reminding Test* and *Word List Generation* Portuguese adaptation (N = 66).

	Mean	SD
Age	36.4	12
Education	13.76	4.2
Gender (%M - %F)	(45.5% - 54.5%)	
Residential Area	34.8% North; 37.9% Center; 27.3% Lisbon Metropolitan Area	

is instructed to recall all 12 words. In every consecutive trial only the words that are missed in the preceding one are given to the participant. After 15 minutes, the participant is asked to recall the 12-word list. The SRT distinguishes between short-term and long-term memory components and also examines the consistency of retrieval from long-term memory. The test was scored according to the rules previously published<sup>4,20</sup>.

(ii) *Word List Generation (WLG)*: evaluates the spontaneous production of words and mental flexibility, when given a letter from the alphabet and within a limited amount of time (one minute) (phonemic verbal fluency task). The scoring procedures consist of totaling the correct responses<sup>4</sup>.

(iii) *Symbol Search* and (iv) *Coding*: two subtests from the *Wechsler Adult Intelligence Scale-III* that evaluate information processing speed<sup>21</sup>.

### Translation and adaptation of SRT and WLG

Many authors maintain that it is more suitable to translate words and expressions from the original test into those with the same frequency of use in the target language (in this case, Portuguese)<sup>19</sup>, as an effort to approximate the level of difficulty of the words in the two languages<sup>16,19,22</sup>.

During the translation process, the Basic Vocabulary of Portuguese<sup>23</sup> was used by two independent linguists to translate the lists of words according to the norms for word frequency in English and Portuguese<sup>16,22</sup>. Following this, the translations were analyzed by a group of neuropsychologists that constructed experimental versions for this study (18 words were tested for each list). The selection of letters followed a similar procedure: within the original forms of WLG 5 letters which present equivalent levels of frequency and difficulty in the Portuguese lexicon were tested in this study.

### Item analysis and selection

Item analysis and selection was based on a set of criteria established according to the CTT and IRT. The decision for the inclusion of IRT parameters, which is very useful when validating and studying neuropsychological test batteries<sup>24</sup>, is related to the advantages associated with this approach (e.g. people and item results are calibrated on a common scale)<sup>12,25</sup>. *Item difficulty index (P)* and *item discrimination index (D)* were included to select the words for the final lists in SRT. *P* refers to the proportion of participants that responds correctly to the item ( $P = R / T$ ; R = number of correct responses to the item; T = total number of responses). The results range from 0 (all participants fail) and 1 (all participants succeed)<sup>17</sup>. Garret's typology was used as a guideline regarding the *P* values to select the twelve words for each list: 50% of the words between 0.25 and 0.75, 25% below 0.25 and 25% above 0.75<sup>17</sup>.

*D* refers to the degree in which an item differentiates in the same direction as the global test, ranging from -1.0 to + 1.0. The formula used to calculate this index is  $D = \frac{X1 - X}{SD} \times \sqrt{\frac{n1 - n}{n0(n - 1)}}$  (X1 - mean of the test scores

of the participants that succeed on the item; X - overall mean on the global test; SD - standard deviation of the results; n - total number of participants; n<sub>1</sub> - number of participants that succeed on the item; n<sub>0</sub> - number of participants that failed on the item)<sup>16,17</sup>. The discriminatory power - asymmetry coefficient and kurtosis coefficient - and normality - Kolmogorov-Smirnov Test - were also analyzed considering the CTT approach.

As IRT measures item adjustment - *infit* and *outfit mean square* - and the range of the *person measures estimates* were used. In this case, the items are analyzed in order to know if they fit the model ( $p < 0.01$ ) or, on the contrary, if they present excessive *infit* and *outfit* measures<sup>12</sup>. The range of *person measure estimates* represents the amplitude of the difficulty of the participants (between the best and the worst performance on the LOGIT measures)<sup>18,26</sup>.

### Data analysis

The statistical analyses were performed with IBM PASW 18 software - descriptive statistics, t-test, asymmetry coefficients, kurtosis coefficients and Kolmogorov-Smirnov test - and with Winsetps software for IRT analyses (item adjustment - *infit* and *outfit* mean squares - and *person measure estimate* range).

## RESULTS

### Selective Reminding Test results

Three different lists were tested. Since each list was tested on only 22 participants, it was decided not to use IRT parameters. Therefore, the criteria for SRT analyses were the *item difficulty index (P)* and *item discrimination index (D)*. The SRT index chosen as the basis of these analyses was *Consistent Long Term Retrieval (CLTR)* as prior research shows that CLTR is one of the most sensitive measures for recent memory<sup>4</sup>.

First, it was established that word selection would be based on the *P* values according to Garret's typology<sup>17</sup>. However, when there were words with the same *P*, the chosen word would be the one that presented the best *D* values (i.e., high correlated with other items).

Table 1 shows the results on *D* and *P* for each word in keeping with Garret's typology<sup>17</sup>. Thus, the words chosen (ü) and their level of difficulty (low, medium or high) are indicated according to Garret's typology<sup>17</sup>. It is important to note that, after the selection of the 12 words, they were randomized in order to constitute the final version of SRT for the Portuguese language. This procedure was followed for the other two remaining lists.

### Word List Generation results

Five letters were tested - P, A, D, F and S - in the total sample (N = 66). In this case, the selection respected CTT and

IRT criteria. The goal was to select three letters for the final version of the *WLG* test in Portuguese.

In the first stage, and after a literature review, it was concluded that if the effect of the variables “age” and “gender” on phonemic verbal fluency performance is not very consensual, the same is not true for “educational level”. Many studies have analyzed the influence of this variable on the performance of phonemic verbal fluency and a robust effect regarding the level of education has been demonstrated: the level of education is a strong predictor of the performance on this task, with increasing educational achievement being associated with higher fluency scores<sup>27,28,29</sup>. Thus, the influence of the level of education on the participants’ performance for each letter was analyzed: D was the only one that did not differentiate between participants with high levels of education and participants with low levels of education (see Table 2). Therefore, D was excluded from the next analyses.

For each letter – P, A, F and S - item adjustment and the range of *person measures estimate* (IRT), as well as discriminatory power and normality (CTT) were analyzed. The goal was to select the three letters that presented best results for each parameter and then choose the three, which had the best performance for all of them. Table 3 shows the results for each letter for each of the parameters analyzed.

The results in Table 4 show that the letters do not match concerning both asymmetry and the kurtosis coefficient [-0.5, 0.5]. P, F and S presented the best asymmetry coefficients and P, A and S presented the best kurtosis coefficients. Regarding normality, P, A and F were the letters that presented a normal distribution with K-S values  $\geq 0.05$ . Additionally, performance in phonemic verbal fluency was

assessed to see if it fitted the model ( $p < 0.01$ ) or if there were items with excessive *infit* and *outfit mean square residuals*<sup>30</sup>. According to Linacre<sup>12</sup>, the letters that best fitted the model ( $p < 0.01$ ) are P, F and S. Regarding the range of *person measures estimate*, it was found that all letters had a wide range between the person with the highest and the person with the lowest performance score. However, the distribution was narrower for the letter F ( $-2.76 < D_i < 0.97$ ). Thus, as a result of these analyses, the letters selected for the Portuguese version of the *WLG* test were P, F and S.

## DISCUSSION

Over the last years, neuropsychological assessment of MS-related cognitive impairment has received growing attention in the routine care of these patients<sup>5,6,8,9</sup>. Unfortunately, in many clinical settings neuropsychological batteries are not well-validated and this can compromise the usefulness of the results to MS patients care. The relevance of our study lies on the illustration of the key aspects from psychometric theory that must be considered when adapting neuropsychological tests to a new population, so that they can be considered reliable tools to assess cognition, in this case for MS patients. In this study, a combination of procedures from Classical Test Theory and Item Response Theory was applied to item analysis and selection. The decision for this combination was taken in order to draw the best of both approaches, enrich the decision making process, and refine the results of *SRT* and *WLG* adaptation, within the validation of BRBN-T to the Portuguese population of MS patients. In our opinion, the combination of these two approaches represents a contribution to a systematization of procedures, which still appear in a hardly consensual way in the literature, and that can contribute for future work.

Three lists were drawn up for *Selective Reminding Test* (Portuguese version), in order to guarantee that alternative versions are available for repeated testing in the routine care of MS patients (one of the BRBN-T main advantages is to have multiple versions available for situations in which the plan is to repeat testing over time)<sup>1,8</sup>. Like in other studies with *SRT*<sup>19</sup>, the translation rule concerning the maintenance of words frequency and difficulty levels between the

**Table 2.** The Item Difficulty Index and Item Discrimination Index for each letter of List A.

List A	Item Difficulty Index (P)	Item Discrimination Index (D)	Decision	Difficulty level
Rua (Street)	0.86	0.53	✓	Low
Célula (Cell)	0.77	0.48	✗	---
Bem (Good)	0.45	0.52	✓	Medium
Cela (Prison cell)	0.64	0.43	✓	Medium
Curso (Course)	0.27	0.27	✓	High
Aldeia (Village)	0.50	0.38	✓	Medium
Vestido (Dress)	0.77	0.70	✓	Low
Motor (Motor)	0.50	0.12	✗	---
Bar (Pub)	0.55	0.36	✗	---
Jogo (Game)	0.36	0.43	✓	Medium
Material (Material)	0.32	0.37	✗	---
Tecido (Fabric)	0.59	0.31	✓	Medium
Júri (Jury)	0.23	0.37	✓	High
Exemplo (Example)	0.27	0.27	✓	High
Festa (Party)	0.32	0.23	✗	---
Fábrica (Factory)	0.55	0.45	✓	Medium
Exército (Army)	0.86	0.58	✓	Low
Mãe (Mother)	0.77	0.52	✗	---

**Table 3.** The influence of the level of education on the phonemic verbal fluency test (Descriptive statistics and t-test).

	- 12 years (n = 38)		+ 12 years (n = 28)		t	p-value
	Mean	SD	Mean	SD		
P	11.29	4.07	13.54	4.38	-2.146	0.036
A	9.24	3.90	11.43	3.06	-2.465	0.016
D	8.92	3.67	10.00	3.67	-1.180	0.242
F	9.82	3.35	11.82	3.39	-2.391	0.020
S	10.13	2.86	11.79	3.14	-2.227	0.029

SD: standard deviation.

**Table 4.** Results for the letters P, A, F and S from the phonemic verbal fluency test (CCT and IRT parameters).

	Classical Test Theory			Item Response Theory		
	Discriminatory power		Normality	Item adjustment		Person measure estimate range
	Asymmetry coefficient	Kurtosis coefficient	K-S Test	Infit mean square	Outfit mean square	Measure Range
P	0.44	0.13	0.17	0.96	0.99	(min: -2.48; max: 2.34)
A	0.93	-0.52	0.20	0.93	0.96	(min: -2.78; max: 1.22)
F	-0.33	-0.61	0.20	0.95	0.98	(min: -2.76; max: 0.97)
S	0.45	-0.19	0.04	0.97	0.98	(min: -2.62; max: 1.20)

K-S: Kolmogorov-Smirnov Test; CCT: Classical Test Theory; IRT: Item Response Theory.

original (English) and the target language (Portuguese) was applied - a fundamental step when adapting neuropsychological tests where verbal content has a special relevance<sup>19,22</sup>. After this stage, the most suitable strategy for selecting the words to be included on a list that intend to assess verbal learning and memory, seemed to be the performance of an *item difficulty index* and an *item discrimination index* analyses. This procedure was also conducted in other studies<sup>21, 31</sup> and it allows a comprehensive analyses of each item. In this paper, only one of the three lists which were constructed is presented (Table 1). Although, these item analyses and selection procedures were applied to the other two, and each one of the lists combines twelve words with different levels of difficulty: easy, medium and difficult<sup>17</sup>, as an effort to draw up three lists of equivalent difficulty. These adaptation procedures will allow future reliable cross-cultural studies with other countries that also have normative data on BRBN-T.

Regarding *Word List Generation*, discriminatory power and normality analysis (CTT) and item adjustment (IRT) were analyzed. However, in the first stage of these analyses, and knowing that the variable level of education is highly correlated with the performance on verbal fluency tasks<sup>27,28,29</sup>, we analyzed the influence of this variable in our results. Finding that D was the only letter that did not differentiate participants with high levels of education from those with low levels of education ( $p < 0.005$ ), we excluded it from the next set of analyses. P, A, F and S were then analyzed according to the criteria previously established (CTT plus IRT measures), and the letters P, F and S were those that showed the best performance for all of the analyzed parameters - better asymmetry and kurtosis coefficients, non-significant K-S test, adequate infit and outfit values, higher variability of person measures range. In fact, considering that Rasch measures overcome some limitations found in the CTT literature (e.g., sample size characteristics)<sup>18</sup>, the results suggested adequate infit and outfit values for the tested letters, as well as ample intervals that show some distance between participants' performances. Thereby, P, F and S were chosen to constitute the final version of WLG in the Portuguese language. Although "A" presented good results in the Portuguese context, we choose P, F and S, which showed more suitable scores for this context according to the predefined criteria. In our opinion, these results provide

important information for future studies with Portuguese-speaking subjects in terms of which letters should be used to produce words in a phonemic verbal fluency task. Likewise, the information which was obtained through the application of this methodology goes far beyond what would be achieved if only descriptive statistics were applied<sup>32</sup>.

Despite the contribution of the present study to the psychometric procedures applied to neuropsychological assessment, there are some limitations that must be pointed out: the fact that is a convenience sample constituted only for healthy subjects who volunteered to participated in this test adaptation, although this neuropsychological battery is targeted to MS patients<sup>33</sup>, and the number of participants, particularly the number of participants per list of words (N = 22), restricted the use of IRT procedures to word analyses and selection.

Although more studies need to be conducted to improve comparability of the Portuguese forms of SRT and WLG, as well to enhance their robustness to ensure clinical efficacy, the results of this study are promising and might, at least, suggest some guidelines for future work on the translation and adaptation of neuropsychological tests. Additional research would be important to demonstrate the clinical usefulness of these forms and it would be very interesting to administer these Portuguese SRT and WLG forms to other Portuguese-speaking populations, culturally and linguistically different, in order to compare the results and thus, cross-validate them.

A challenge for cross-cultural neuropsychology still is the development of uniformed and standardized procedures for tests translation and adaptation within different countries, so that the influence of cultural differences on neuropsychological test scores can have a minor impact<sup>14,15,16,17</sup>. In the Portuguese context, it is absolutely necessary to produce and deliver validated and standardized neuropsychological tests to those working in clinical practice, so that they can implement structured routines of neuropsychological assessment in the care of neurological patients.

### Acknowledgements

The authors thank to *Associação Líquido Céfaló-Raquidiano* and *Biogen Idec*.

## References

1. Solari A, Mancuso L, Motta A, Mendozzi L, Serrati C. Comparison of two brief neuropsychological batteries in people with multiple sclerosis. *Mult Scler*. 2002;8(2):169-76. doi:10.1191/1352458502ms780oa
2. Bilgi E, Zdemir HH, Bingol A, Bulut S. Evaluation of the effects of group psychotherapy on cognitive function in patients with multiple sclerosis with cognitive dysfunction and depression. *Arq Neuropsiquiatr*. 2015;73(2):90-5. doi:10.1590/0004-282X20140144
3. Sá MJ. *Neurologia clínica: compreender as doenças neurológicas*. Porto: Ed. Universidade Fernando Pessoa; 2009.
4. Rao SM, Leo GJ, Bernardin L, Unverzagt F. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology*. 1991;41(5):685-91. doi:10.1212/WNL.41.5.685
5. Amato MP, Ponziani G, Siracusa G, Sorbi S. Cognitive dysfunction in early-onset of multiple sclerosis: a reappraisal after 10 years. *Arch Neurol*. 2001;58(10):1602-6. doi:10.1001/archneur.58.10.1602
6. Brooks JB, Borela MC, Fragoso YD. Assessment of cognition using the Rao's Brief Repeatable Battery of Neuropsychological Tests on a group of Brazilian patients with multiple sclerosis. *Arq Neuropsiquiatr*. 2011;69(6):887-91. doi:10.1590/S0004-282X2011000700007
7. Costa DC, Sá MJ, Calheiros JM. The effect of social support on the quality of life of patients with multiple sclerosis. *Arq Neuropsiquiatr*. 2012;70(2):108-13. doi:10.1590/S0004-282X2012000200007
8. Aupperle RL, Beatty WW, Shelton FN, Gontkovsky, ST. Three screening batteries to detect cognitive impairment in multiple sclerosis. *Mult Scler*. 2002;8(5):382-9. doi:10.1191/1352458502ms832oa
9. Sepulcre J, Vanotti S, Hernández R, Sandoval G, Cáceres F, Garcea O et al. Cognitive impairment in patients with multiple sclerosis using the Brief Repeatable Battery-Neuropsychological test. *Mult Scler*. 2006;12(2):187-95. doi:10.1191/1352458506ms1258o
10. Strober L, Englert J, Munschauer F, Weinstock-Guttman B, Rao S, Benedict, RH. Sensitivity of conventional memory tests in multiple sclerosis: comparing the Rao Brief Repeatable Neuropsychological Battery and the Minimal Assessment of Cognitive Function in MS. *Mult Scler*. 2009;15(9):1077-84. doi:10.1177/1352458509106615
11. Guimarães J, Sá MJ. Cognitive dysfunction in multiple sclerosis. *Front Neurol*. 2012;3:74. doi:10.3389/fneur.2012.00074.
12. Christodoulou C, Krupp LB, Liang Z, Huang W, Melville P, Roque C et al. Cognitive performance and MR markers of cerebral injury in cognitively impaired MS patients. *Neurology*. 2003;60(11):1793-8. doi:10.1212/01.WNL.0000072264.75989.B8
13. Amato MP, Bartolozzi ML, Zipoli V, Portaccio E, Mortilla M, Guidi L et al. Neocortical volume decrease in relapsing-remitting MS patients with mild cognitive impairment. *Neurology*. 2004;63(1):89-93. doi:10.1212/01.WNL.0000129544.79539.D5
14. Raykov T. A method for testing group differences of scale validity in multiple population studies. *Br J Math Stat Psychol*. 2005;58(1):173-84. doi:10.1348/000711005X38005
15. Ardilla A. Directions of research in cross-cultural neuropsychology. *J Clin Exp Neuropsychol*. 1995;17(1):143-50. doi:10.1080/13803399508406589
16. American Educational Research Association (AERA); American Psychological Association (APA); National Council on Measurement in Education (NCME). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 1999.
17. Almeida L, Freire T. *Metodologia de investigação científica em psicologia e educação*. 5a ed. Braga: Psiquilíbrios; 2003.
18. Ferreira AI, Almeida L, Prieto G. The role of processes and contents in human memory: an item response theory approach. *J Cogn Psychol*. 2011;23(7):873-85. doi:10.1080/20445911.2011.584692
19. Campo P, Morales M, Juan-Malpartida M. Development of two spanish versions of the verbal selective reminding test. *J Clin Exp Neuropsychol*. 2000;22(2):279-85. doi:10.1076/1380-3395(200004)22:2;1-1;FT279
20. Buschke H, Fuld PA. Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology*. 1974;24(11):1019-25. doi:10.1212/WNL.24.11.1019
21. Wechsler D. *WAIS-III: Manual da escala de inteligência de Wechsler para adultos - Terceira Edição*. Lisboa: CEGOC-TEA; 2008.
22. Hambleton R. K. *Adapting educational and psychological tests for cross-cultural assessment*. New Jersey: Lawrence Erlbaum; 2005.
23. *Universidade de Lisboa, Centro de Linguística. Português fundamental, vocabulário e gramática*. Lisboa: Instituto Nacional de Investigação Científica; 1984. Tomo 1, Vocabulário.
24. Prieto G, Delgado AR, Perea MV, Ladera V. Scoring neuropsychological tests using the Rasch model: an illustrative example with the Rey-Osterrieth Complex Figure. *Clin Neuropsychol*. 2010;24(1):45-56. doi:10.1080/13854040903074645
25. Magno C. Demonstrating the differences between classical tests theory and item response theory using derived test data. *Int J Educ Psychol Assess*. 2009;1(1):1-11.
26. Bond TG, Fox CM. *Applying the Rasch model*. Mahwah, NJ: Lawrence Erlbaum; 2007.
27. Hughes DL, Bryan J. Adult age differences in strategy use during verbal fluency test performance. *J Clin Exp Neuropsychol*. 2002;24(5):642-54. doi:10.1076/jcen.24.5.642.1002
28. Mathuranath PS, George A, Cherian PJ, Alexander A, Sarma SG, Sarma, PS. Effects of age, education and gender on verbal fluency. *J Clin Exp Neuropsychol*. 2003;25(8):1057-64. doi:10.1076/jcen.25.8.1057.16736
29. Oberg G, Ramirez M. Cross-linguistic meta-analysis of phonological fluency: normal performance across cultures. *Int J Psychol*. 2006;41(5):342-7. doi:10.1080/00207590500345872
30. Ferreira AI, Almeida LS, Prieto G. Construction of a memory battery for computerized administration, using item response theory. *Psychol Rep*. 2012;111(2):585-609. doi:10.2466/03.04.PRO.111.5.585-609
31. Wechsler D. *WMS-III: Manual da escala clínica de memória de Wechsler - Terceira Edição* [Lisboa: CEGOC-TEA, 2008.
32. Senhorini M, Amaro Júnior E, Ayres A, Simone A, Busatto GF. Phonemic fluency in Portuguese-speaking subjects in Brazil: ranking of letters. *J Clin Exp Neuropsychol*. 2006;28(7):1191-200. doi:10.1080/13803390500350969
33. Reynolds CR, Suzuki LA. Bias psychological assessment: an empirical review and recommendations. In: Weiner IB, Graham JR, Naglieri JA. *Handbook of psychology*. 2nd ed. New Jersey: Wiley; 2012. p. 82-113.