Multiple sclerosis (MS) is a chronic, neurodegenerative disease whose various neurological manifestations have a multidimensional impact on patients’ lives and clinical status. 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. In addition to the physical symptoms associated with MS, 40 to 70% of the patients suffer from cognitive deficits that neurological exams cannot detect. 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.

Given the frequency and the importance of the cognitive impairment of MS, appropriate neuropsychological assessment is an extremely relevant complementary procedure to neurological exams. These results, supported by normative quality data, provide important support for clinical and therapeutic decisions, assessment of intervention efficacy and MS research. 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 and has proved to be one of the most sensitive.
measures for detecting cognitive impairment in MS. As the MS cognitive deficits most frequently involve working and episodic memory, sustained attention, information processing speed and mental flexibility, Rao et al. 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). There is a consensus regarding BRBN-T psychometric properties, particularly its sensitivity and reliability in detecting cognitive impairment. A study by Strober and Sepulcre 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 ventricles enlargement, neocortical volume and reveal longitudinal changes in MS.

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. However, this potential cannot be fully realized, unless tests are used with comparable validity in all the populations involved. Cross-cultural differences are considered an important issue in any research that deals with neuropsychological assessment. 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. Human performance on neuropsychological tests is, therefore, influenced by a range of moderating variables, such as: culture, primary language and level of education. For these reasons, the assessment of a different cultural group using the original neuropsychological tests, procedures and norms, results in conceptual errors in assessment.

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. 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). 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.

**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 were 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 administered in a standardized way and in a fixed order: (i) SRT, (ii) WLG, (iii) Symbol Search, (iv) Coding 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 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).**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>36.4</td>
<td>12</td>
</tr>
<tr>
<td>Education</td>
<td>13.76</td>
<td>4.2</td>
</tr>
<tr>
<td>Gender (M - F)</td>
<td>(45.5% - 54.5%)</td>
<td></td>
</tr>
<tr>
<td>Residential Area</td>
<td>34.8% North; 37.9% Center; 27.3% Lisbon Metropolitan Area</td>
<td></td>
</tr>
</tbody>
</table>
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.[20]

(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.

(iii) Symbol Search and (iv) Coding: two subtests from the Wechsler Adult Intelligence Scale-III that evaluate information processing speed[21].

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 that evaluate information processing speed[21]. The scoring procedures consist of totaling the correct responses.

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).

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[4].

First, it was established that word selection would be based on the P values according to Garret’s typology[17]. 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[17]. Thus, the words chosen (ii) and their level of difficulty (low, medium or high) are indicated according to Garret’s typology[17]. 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
I was unable to complete the task as the content was not provided. Please provide the content so that I can convert it to a plain text representation.
anal-item difficulty index and an item discrimination index for 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 and it allows a comprehensively analysis 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, 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, we analyzed the influence of this variable in our results. Founding 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), 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.

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, 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. 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 Céfalo-Raquidiano and Biogen Idec.


