Title: Scaling cognitive domains of the Montreal Cognitive Assessment (MoCA): An analysis using the Partial Credit Model

Title suitable for the running head: Scaling MoCA's cognitive domains with Partial Credit Model

Authors:

Sandra Freitas[1,2,3], Gerardo Prieto[4], Mário R. Simões [2,3] & Isabel Santana [1,5,6]

[1] Centre for Neuroscience and Cell Biology (CNC), University of Coimbra, Coimbra, Portugal

[2] Centro de Investigação do Núcleo de Estudos e Intervenção Cognitivo Comportamental (CINEICC), University of Coimbra, Coimbra, Portugal

[3] Psychological Assessment Lab., Faculty of Psychology and Educational Sciences, University of Coimbra, Coimbra, Portugal

[4] Departamento de Psicología Básica, Psicobiología y Metodología de las Ciencias del Comportamiento, Universidad de Salamanca, Salamanca, Spain

[5] Neurology Department, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal

[6] Faculty of Medicine, University of Coimbra, Coimbra, Portugal

Corresponding Author:

Sandra Freitas

Centre for Neuroscience and Cell Biology (CNC),

University of Coimbra, Largo Marquês de Pombal

3004-517 Coimbra, Portugal

E-mail: sandrafreitas0209@gmail.com

Phone: +351 239820190

Fax: +351 239822776

Funding

Conflict of Interest

None declared.

Abstract

The psychometric properties of the MoCA were examined by using the Partial Credit Model. The study sample included 897 participants who were distributed into two main subgroups: (I) the clinical group (90 patients with Mild Cognitive Impairment, 90 patients with Alzheimer's Disease, 33 patients with Frontotemporal Dementia, and 34 patients with Vascular Dementia, whose diagnoses were previously established according to a consensus that was reached by a multidisciplinary team, based on the international criteria) and (II) the healthy group (composed of 650 cognitively healthy community dwellers). The results show (i) an overall good fit for both the items and the persons' values, (ii) high variability for cognitive performance level of the cognitive domains (ranging between 1.90 and -3.35, where *Short-term Memory* was the most difficult item, and *Spatial Orientation* was the easiest item) and between the subjects on the scale, (iii) high reliability for the estimation of the persons' values, (iv) good discriminant validity and high diagnostic utility, and (v) a minimal Differential Item Functioning (DIF) effect related with of pathology, gender, age and educational level. MoCA and its cognitive domains are suitable measures to use for screening the cognitive status of cognitively healthy subjects and patients with cognitive impairment.

Introduction

In response to the need for early identification of mild states of cognitive impairment among older people, particularly in the spectrum of Mild Cognitive Impairment (MCI) and Alzheimer's disease (AD), Nasreddine and collaborators developed the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) as a brief cognitive screening test. Although it was initially designed for the global cognitive assessment of these patients, the MoCA is currently an extensively validated screening tool for many disorders, and it is translated into 36 languages and dialects (see studies in http://www.mocatest.org). The MoCA's widespread international use and its recognition as one of the best screening tests (Gauthier et al., 2011; Ismail, Rajji, & Shulman, 2009; Jacova, Kertesz, Blair, Fisk, & Feldman, 2007; Lonie, Tierney, & Ebmeier, 2009) is explained and supported by several previous studies that have consistently reported the good overall psychometric properties of MoCA and its improved sensitivity and usefulness in accurately identifying milder forms of cognitive impairment in many clinical conditions.

In fact, the MoCA has been commonly used to measure global cognitive function in both clinical and research contexts. A few studies have examined the psychometric properties of the MoCA by using item response theory (IRT), namely the Rasch model (Andrich, 1988; Rasch, 1960; Wright & Mok, 2004). The Rasch model is a psychometric method that is suitable for the analysis of neuropsychological assessment instruments (Conrad & Smith, 2004; Prieto, Contador, Tapias-Merino, Mitchell & Bermejo-Pareja, 2012; Prieto, Delgado, Perea & Ladera, 2010). In this context, Koski and collaborators (2009) verified that the MoCA can provide a quantitative estimate of cognitive function, which allows for its use in monitoring changes in global cognition over time. Afterwards, Koski and collaborators (2011)

proposed an algorithm that combined items of the Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) and the MoCA in order to improve measurement precision for cognitive ability in milder forms of impairment. Konsztowicz and collaborators (2011) developed a simplified adaptive approach to cognitive assessment (Geriatric Rapid Adaptive Cognitive Estimate - GRACE method), which reduces the test burden and allows for the evaluation of individuals across a broader range on the cognitive ability continuum, when compared to the MMSE and the MoCA. More recently, Tsai and collaborators (2012) used the IRT to demonstrate that the levels of information that are provided by each of the MoCA's domains are consistent and that the language and executive functions domains provided the highest discrimination in the identification of MCI patients. Finally, Freitas and collaborators (2014) used the Rasch model for dichotomous items and demonstrated that there was an overall good fit for both the items and the persons' values, there was high variability on the cognitive performance level, the measurements were of adequate quality, there was a good discriminant validity that provided high diagnostic value and the generalized validity of the MoCA scores was provided, according to DIF analyses.

Nasreddine and collaborators (2005) made a theoretical proposal, according to which the MoCA's 30 items could be categorized into the following cognitive domains: executive functions; visuospatial abilities; short-term memory; language; attention, concentration and working memory; and temporal and spatial orientation. Further, Freitas, Simões, Marôco and collaborators (2012) conducted a study in order to evaluate the factorial structure of the MoCA and analyze its construct related validity, which required testing several models by using a Confirmatory Factor Analysis (CFA). The six-factor model, according to the original authors' *a priori* hypothesized domains, showed the best fit in this study. However, the results also supported a second-order

factor model, with all of the first-order factors contributing to a common underlying second-order factor: "Cognition". The analysis revealed a second-order unidimensional tendency and serves as a good indicator that the MoCA, as a whole, evaluates individuals' global cognition (a positive finding, considering that the MoCA is a brief cognitive screening instrument). In fact, this study provides additional evidence that supports the idea that this screening test measures cognitive ability as a global dimension (total score), which is based on different cognitive domains that can be considered as polytomous items.

To date, as far as we know, there have been no other research studies that analyze the MoCA results by treating its cognitive domains as polytomous items and using the Rasch Partial Credit Model (PCM; Andrich, 1988), which represents a significant gap in the psychometric literature on this instrument. The use of PCM is particularly recommended for polytomous items because it allows for the estimation of the values of these items on a latent continuum, which makes it possible to infer the value of the items that indicate varying degrees of cognitive impairment. Consequently, the present study aims to further assess the psychometric characteristics of the MoCA by using the PCM for polytomous items. First, we analyzed the fit of the data to the model (including to the unidimensionality assumption), the functionality of the ratings, and the reliability values for the estimation of the items and persons. Afterwards, we performed DIF analyses in order to explore the likelihood that cognitive domains of the MoCA may work differently based on pathology, gender, age and educational level.

Method

Participants and procedures

The study used 897 participants, who were distributed into two main subgroups:

I)   The Clinical Group

The clinical group included (i) 90 patients with MCI, (ii) 90 with AD, (iii) 33 with Frontotemporal dementia (FTD) and (iv) 34 with Vascular dementia (VaD). All of the participants were recruited at the Dementia Clinic of a central hospital. Each patient underwent a comprehensive neuropsychological assessment, laboratory tests that were essential in order to exclude a reversible form of dementia or significant comorbidities, imaging studies (CT or MRI and SPECT or FDG-PET), Apolipoprotein E allele genotyping and Cerebrospinal Fluid (CSF) analyses (Aβ42, Tau, and P-tau) and biomarker profiles for neurodegenerative diseases, when needed. This comprehensive evaluation that frequently included CSF and genetic studies was considered to be essential in order to exclude non-degenerative or vascular forms of cognitive decline and for the differential diagnoses of these situations. The final diagnosis was established according to a consensus that was reached by a multidisciplinary team, based on the international criteria for the (i) MCI (Petersen, 2004, 2007); (ii) probable AD (American Psychiatric Association, 2000; McKhann et al., 2011); (iii) FTD (Neary et al., 1998); and (iv) VaD (Román et al., 1993). The MCI group included patients who were classified as "amnesic MCI" (single or multidomain; Petersen, 2007) and those who had a classification of 0.5 on the CDR (*Clinical Dementia Rating* scale; Hughes, Berg, Danziger, Coben & Martin, 1982). The AD, FTD and VaD groups only included patients with mild to moderate severity (classified by CDR ≥1 and ≤ 2; and MMSE ≥ 12 points). The FTD group only considered patients who had a behavioral variant of the disease. Were excluded to this study patients with Vascular Mild Cognitive Impairment or mixed dementia (patients meeting criteria for possible Alzheimer's disease were not included in VaD group whereas cases with a progressive decline but also presenting significant vascular pathology on neuroimaging study were not included in the AD

group). Significant vascular pathology on neuroimaging was operationalized according to NINDS-AIREN criteria, namely: (1) evidence of large cortico-subcortical infarcts; (2) extensive periventricular (PVL) or subcortical white matter lesions (SCWML), affecting more than 25% of the volume of the supra-tentorial white matter (WML), based on a CT or MRI volumetric qualitative analysis; (3) uni- or bilateral thalamic strokes; (4) lacune in the head of the caudate nucleus or in the inferior genu of the internal capsule; (5) more than two lacunes. In order to exclude cases of mixed dementia, we privileged cases of VaD with large cortical infarcts or vascular strategic lesions. Additionally, only patients with a complete clinical evaluation, a well-established diagnosis at the time of data collection (according to the above international criteria) and a stable clinical condition (without significant comorbidities) were eligible for this study. At the outset, the inclusion criteria included the following: another native language than the Portuguese; schooling outside of Portugal; high dementia severity, recent pharmacotherapy changes, recent psychiatric comorbidity (clinical diagnosis within the last 6 months prior to the current neuropsychological evaluation), and significant motor or sensorial deficits that could influence neuropsychological assessment.

II)   The Healthy Group

The healthy group included 650 cognitively healthy community dwellers from all of the geographic regions of continental Portugal. All of the participants were recruited through the national health and social security services. The demographic and clinical inclusion criteria that were considered in the initial selection phase included the following: being 25 years of age and over; having Portuguese as the native language and completing schooling in Portugal; and the absence of significant motor, visual or auditory deficits that may influence performance on tests. In order to ensure that the

participants were cognitively healthy adults, the inclusion criteria also included the following: autonomy in daily activities; no history of developmental delay or learning disabilities; no history of alcoholism or substance abuse; absence of neurological or psychiatric diseases and of chronic unstable systemic disorders that may compromise cognitive function (e.g., vitamin deficits, hypothyroidism, uncontrolled diabetes, hypertensive encephalopathy, systemic infections, abstinence syndromes and delirium); and the absence of significant depressive complaints and medication that could possibly impact cognition (e.g., psychotropic or psycho-active drugs). The psychologist determined if inclusion criteria were met via clinical interview through the use of a standard questionnaire that included a complete socio-demographic questionnaire, an inventory of current clinical health status, and past habits and medical history. Regarding the older participants, collateral information was also obtained from practitioners, community center directors, close relatives, or an informant living with the participant, in order to confirm the inclusion criteria were met. After the initial selection, a set of instruments of global assessment was administered and included the following: the MMSE (Folstein et al., 1975; Guerreiro, Silva, Botelho, Leitão, Castro-Caldas & Garcia, 1994); the CDR (Hughes, Berg, Danziger, Coben & Martin, 1982; Garret, Santos, Tracana, Barreto, Sobral & Fonseca, 2008), which was only for participants who were over 49 years of age; the *Irregular Word Reading Test* (TeLPI: Teste de Leitura de Palavras Irregulares; Alves, Simões & Martins, 2009), which was used for pre-morbid intelligence estimation; the *Subjective Memory Complaints* scale (SMC; Schmand, Jonker, Hooijer & Lindeboom, 1996; Ginó, Mendes, Ribeiro, Mendonça, Guerreiro & Garcia, 2008); and the *Geriatric Depression Scale* (GDS-30; Yesavage, Brink, Rose, Lum, Huang, Adey & Leirer, 1983; Barreto, Leuschner, Santos & Sobral, 2008). Only the participants with normal performance on the MMSE

(according to Portuguese cut-off points; Guerreiro et al., 1994), a score of zero on the CDR and a GDS-30 score that was below 20 points were eligible for participation in this study. In this group, given the sample size and the geographical distribution of the participants, it was not possible to perform a neurological consultation or additional diagnostic tests, such as neuroimaging or biomarker analyses. This sample served as the basis of the MoCA normative study for the Portuguese population (Freitas, Simões, Alves & Santana, 2011), which provides more details about the recruitment process.

Each participant was assessed during a single session by one of the two psychologists with expertise in neuropsychological assessment. In the current study, the MoCA was not used as a diagnostic tool because all of the researchers who were involved in forming the consensus diagnoses were blind to the MoCA results throughout the study. Informed consent was obtained from all of the participants after the research aims, the procedures and the confidentiality requirements were fully explained by a member of the research team. For the patients who were not capable of providing informed consent, a legal representative fulfilled that requirement on their behalf. The study complies with the ethical guidelines on human experimentation that are stated in the Declaration of Helsinki and was approved by the Portuguese Foundation for Science and Technology and by the Faculty of Psychology and Educational Sciences Scientific Committee.

Measure: Montreal Cognitive Assessment (MoCA)

The MoCA was developed as a brief cognitive screening instrument for milder forms of cognitive impairment and provides a quick measure of an individual's global cognitive state. It is a one-page test with a paper-and-pencil format that requires approximately 10 to 15 minutes to administer. A manual provides explicit instructions

for administration and an objectively defined scoring system. The MoCA score is derived by adding the points of each successfully completed task, with higher scores indicating better cognitive performance. The MoCA evaluates seven indicators of cognitive functioning (Nasreddine et al., 2005; Julayanont, Phillips, Chertkow, & Nasreddine, 2013), which were scored with a variable range of ratings, based on the addition of correct answers to the dichotomous items for each cognitive domain. Thus, each cognitive domain can be considered to be a polytomous item or a "testlet". Testlets are sets of dichotomous items that are scored as a single test question (Embretson and Reise, 2000). Yen (1993) suggests that combining items into testlets minimizes the chances of local dependencies. Furthermore, the MoCA's total score refers to the raw score that does not consider the correction point for educational effects that is proposed in the original study (Nasreddine et al., 2005) because this correction point is not used for the Portuguese population (Freitas et al., 2011).

Statistical Analysis

Descriptive statistics were used to derive the sample characteristics and were computed by using the Statistical Package for Social Sciences (SPSS, version 19.0; IBM SPSS, Chicago, IL).

The Rasch analysis was computed by using the WINSTEPS package (Linacre, 2011). The major assumptions of the Rasch model include the following: (1) unidimensionality (the attribute can be represented on a single dimension where subjects and items are conjointly located) and (2) the score values of individuals and items are expressed in a *logits* (units of measurements) scale that has interval-level properties (Conrad & Smith, 2004). The analysis that used the Rasch model of MoCA is justified by previous analyses, which have shown that the data have a good fit regarding

the hypothesis that there is a dominant, second-order factor (Freitas, Simões, Marôco et al., 2012), thus ensuring that the assumption of fundamental unidimensionality is met. Additionally, the Rasch's Principal Components Analysis (PCA) of residuals corroborates the unidimensionality of the data in this study.

The PCM is a Rasch model that is recommended for polytomous items and can be expressed by the following formula:

$$1n(P_{nik}/P_{ni(K-1)}) = B_n - D_i - F_{ki},$$

where:

P_{nik} is the likelihood that person n on item i will score value k;

$P_{ni(K-1)}$ is the likelihood that person n on item i will score the value that is immediately below k;

B_n is the parameter of person n for the variable that is measured;

D_i is the difficulty of the item;

F_{ki} is the location on the latent variable for which the values k and k-1 are equiprobable. F_{ki} is referred to as a step or threshold and can vary among the items.

According to Linacre (2002), an essential feature of a rating scale is that increasing amounts of the latent variable for a respondent correspond to increasing probabilities of the respondent being observed in higher categories of the rating scale. The requirement for this type of interpretability is that the step calibrations (F_{ki}) advance monotonically with the categories (i.e., in the PCM, the existence of increasing order between the successive category steps is a necessary condition for a suitable measure; Andrich, 2013).

A fit analysis was completed based on the following three main indicators: (I) means of the residuals (*Infit* and *Outfit*), (II) PCA of the residuals (Wilson, 1994, 2005), and (III) the DIF analysis.

The statistics *Infit* and *Outfit* allow us to quantify the means of the squared residuals (differences between the observed responses and those that are predicted by the model). *Person fit statistics* measure the extent to which a person's pattern of responses to the items corresponds to what is predicted by the model. *Item fit statistics* are used to identify items that may not contribute to a unitary scale. Misfit items need to be examined in order to determine whether a second dimension may exist. These items should be eliminated when a one-dimensional measure of the construct is required (Conrad & Smith, 2004). *Infit* and *Outfit* have an expected value of 1.0. Conventionally, misfit is considered to be moderately high if these statistics range between 1.5 and 2.0, and it is considered to be severe if the statistics are higher than 2.0 (Linacre, 2011).

Rasch's PCA of the residuals (Wilson, 1994, 2005) examines the patterns in the part of the data that is unexplained by the Rasch measures. In the PCA of the residuals we attempt to disprove the hypothesis that the residuals are random noise by finding the component that explains the largest possible amount of variance. The Linacre criteria for fundamental unidimensionality include the following: (i) a small eigenvalue for the first component of the residuals (less than 2.0) and (ii) a large percentage of the raw variance being explained by the Rasch dimension (over 50%).

A test item is said to have DIF when subjects with the *same level* of the variable that is being measured and who belong to different groups do not have the same likelihood of producing a correct answer. The presence of DIF can create adverse consequences for the validity of the scores because it reveals the inclusion of construct-irrelevant variance in the scores, given that there are factors that are not related to the attribute that is being measured but affect the responses. The presence of DIF would make it likely that factors outside of the construct that is being measured are incorrectly affecting the MoCA scores.

The hypothesis regarding the absence of DIF was tested for persons with different pathologies, genders, ages and education levels by comparing the two groups (focal and reference) of these variables (pathology: healthy / clinical; gender: male / female; age: < 65 years old / > 64 years old; education: 1-4 years / > 4 years). In this study, we employed two methods that had different characteristics (Potenza & Dorans, 1995): the Mantel-Haenszel procedure (a non-parametric method that was based on the direct scores) and a Rasch-based DIF analysis (a parametric method that was based on the values of a latent variable).

The Mantel-Haenszel (MH) method makes it possible for us to compare the answers that are provided for an item by the focal and reference groups, whose members have previously been found to have the same level of the measured attribute. The total score for the variable is used as an internal matching criterion (Holland & Thayer, 1988). The procedure is based on an analysis of the contingency tables that correspond to the different levels into which the variable has been divided. For each level j, the odds-ratio (α) is calculated:

$$\alpha = (p_{Rj}/1-p_{Rj}) / (p_{Fj}/1-p_{Fj})$$

where p_{Rj} and p_{Fj} are the odds of a correct answer for the item being provided by the reference and focal groups, respectively. There will be no DIF on a specific level j if α = 1 (the likelihood of responding to the item correctly is equal in the focal group and in the reference group). The MHα statistic reports a weighted average odds-ratio across an entire score level (Dorans & Holland, 1993; Holland & Thayer, 1988). The usual interpretation of the MHα statistic is that values that are close to 1 indicate an absence of DIF; values that are notably greater than 1 indicate a DIF in favor of the reference group; and values that are closer to 0 indicate a DIF in favor of the focal group. The null hypothesis of DIF being absent (MHα=1) can be tested by using the MHχ^2 statistic

(Holland & Thayer, 1988), which is distributed as χ^2, with one degree of freedom. Testing for the absence of DIF on a test involves multiple comparisons (at least one for each item). It is, therefore, logical to use the Bonferroni adjustment to maintain the family wise error rate (usually .05). As a result, each individual hypothesis is tested at a statistical significance level of p<.05/(the number of contrasts) (Benjamini &Hochberg, 1995).

Given that the MHα is not symmetrical, Holland and Thayer (1988) proposed a logarithmic transformation, Delta-MH ($\Delta\alpha$MH), whose values oscillate symmetrically around 0. A value of zero indicates the absence of DIF; a negative value indicates that the item favors the reference group, and a positive value indicates that the item favors the focal group. Delta-MH is obtained through the transformation $\Delta\alpha$MH = -2.35 ln (MHα). Based on the Delta-MH statistic, Zwick and Ercikan (1989) proposed classifying the DIF magnitude into three categories (adopted by the Educational Testing Service):

Type A items-negligible DIF: $\Delta\alpha$MH < |1|.

Type B items-moderate DIF: |1| $\leq$ $\Delta\alpha$MH $\leq$ |1.5|, and the MH test is statistically significant.

Type C items-large DIF: $\Delta\alpha$MH > |1.5|, and the MH test is statistically significant.

In addition to the MH method, we also used a detection method that was derived from the Rasch model (1960). The most important property of the model, which is known as *specific objectivity* (Andrich, 1988), indicates that individuals with the same ability (B) will have the same likelihood of correctly answering an item, regardless of whether they belong to groups with different pathologies, genders or ages. The DIF detection procedure in the RM is based on the Item Characteristic Curve (ICC), which is

the proportion of subjects at the same ability level who answer a given item correctly; if the item measures the same ability across groups, then, except for random variations, the same proportion is found, regardless of the nature of the group. This means that in the absence of DIF, the ICC in the different groups and the item parameter of difficulty (D) will be invariable. As a result, the hypothesis of DIF being absent was tested by calculating the difference between the estimators of the item parameter of difficulty for each group ($D_f - D_r$), thus controlling for the possible differences between the groups in the latent variable. Wright and Douglas (1976) found that differences that were lower than 0.50 logits had negligible consequences in regard to the validity of the measures. Thus, the DIF is usually considered to be substantial if the absolute difference is higher than 0.50 logits and is statistically significant. The *t*-test with the Bonferroni adjustment (Benjamini & Hochberg, 1995) was used to test the significance and is described below.

$$t = D_f - D_r / (SE_{Df}^2 + SE_{Dr}^2)^{1/2}$$

where SE_{Df} and SE_{Dr} are the standard errors of both parameters of difficulty. If any of the *t*-tests on the list has $p < .05/$(the number of *t*-tests on the list), then, the hypothesis of No DIF is rejected (Bonferroni correction).

The accuracy of the item-person estimations was assessed by using the standard error of the parameters, or the Person Separation Reliability (PSR) and the Item Separation Reliability (ISR) statistics (Wright and Mok, 2004). The PSR and ISR statistics (range: 0-1) are similar to the classical reliability coefficient (the quotient between the true variance and the observed variance). In order to achieve a suitable measure, a value above 0.70 is recommended.

Results

Sample Characterization

A total of 897 participants were recruited and considered to be eligible for this study. The characteristics (variables that were described: sample size, gender, age, educational level and MoCA score) of the total sample and more details about all of the subgroups are presented in Table 1.

(Insert Table 1 about here)

Item Categories

According to Andrich (2013), the threshold calibrations between the adjacent categories of an item need to be ordered monotonically. This is a necessary condition for a suitable measure. Accordingly, we firstly analyze this property for each of the seven cognitive domains (corresponding to the seven cognitive domains, as described by the authors). We found that the steps (F_{ki}) between the successive categories of two cognitive domains (*Short-term Memory* and *Temporal Orientation*) were not ordered. A possible solution would be to aggregate adjacent categories until an effective and parsimonious set is obtained. Thus, a new analysis was run after collapsing the initial six categories into five for *Short-term Memory* and the initial five categories into three for *Temporal Orientation*. According to the quality criteria that were proposed by Linacre (2002), the new categories functioned correctly. After setting the category number as described, a new data analysis was run in order to quantify model fit, unidimensionality, item estimates and person parameters and to analyze differential item functioning. Initial and recoded values of the cognitive domains are included in Table 2.

(Insert Table 2 about here)

Unidimensionality

First, the unidimensionality of the data were analyzed by using a PCA of the residuals. The results show that the data fulfill the criteria of Linacre (2011) in regard to upholding the assumption of fundamental unidimensionality: the percentage of variance that is explained by the measures (68.4%) is higher than 50%, and the eigenvalue of the first component of the residuals (1.5) is less than 2.0, which indicates that a second dominant factor does not exist.

Item Analysis

In this analysis, a total of seven cognitive domains was considered: (i) *Executive Functions*; (ii) *Visuospatial Abilities*; (iii) *Short-term Memory*; (iv) *Language*; (v) *Attention, Concentration and Working Memory*; (vi) *Temporal Orientation;* (vii) *Spatial Orientation*. The ratings that were applied in order to evaluate the persons' performance on each cognitive domain show appropriate metric functionality (i.e., all of the items in the steps between the successive categories were monotonically and increasingly ordered). *Infit* and *Outfit* statistics demonstrate the validity of the items and the persons' values. The statistics for item fit are presented in Table 3. It can be observed that the *Infit* values range between .77 and 1.38, and the *Outfit* values range between .59 and 1.31. Linacre (2011) recommends that values between .50 and 1.50 indicate a suitable fit, while values higher than 2.0 reveal a severe misfit. Accordingly, all of the cognitive domains show an adequate fit that is within the optimal range for a good measurement. The cognitive domains show high variability for the cognitive performance level, ranging between -3.35 and 1.90 (SD = 1.59), as illustrated in Figure 1. The cognitive domains with a higher level of difficulty are for *Short-term Memory* and *Executive Functions*, while at the other extreme, *Spatial Orientation* domain are the easiest. From the perspective of classical test theory, it can be concluded that most of the cognitive

domains have proper discrimination (RiX >.30). In fact, the average of the item-test correlations is .68. The PCM allows us to estimate the likelihood of a person obtaining the maximum possible score on a cognitive domain (for example, a 5 on *Short-term Memory*). In Table 3, the column Pcg provides the likelihood that a person with the mean score of the clinical group (.04) will obtain the maximum score on each cognitive domain. Only the *Orientation* domains indicate a medium to high likelihood (*Spatial Orientation* = .95 and *Temporal Orientation* = .52). This means that subjects with a score that corresponds to the "average clinical group" are more likely to correctly resolve these cognitive domains, whereas they have a very low or null probability of correctly solving the other five cognitive domains. On the other hand, the Phg column shows the likelihood that a person with the mean score of the healthy group (2.55) will obtain the maximum score on each cognitive domain. The *Short-term Memory* (.20), *Attention, Concentration and Working Memory* (.32) and *Executive Functions* (.38) indicate a relatively low probability, whereas the *Orientation* domains indicate the highest probability (*Spatial Orientation* = 1.00 and *Temporal Orientation* = .95). Finally, the Pcg-Phg column displays the differences in the likelihoods that an average person (considering both groups) will attain a perfect cognitive domain score. The cognitive domains that are marked with an asterisk are the ones that better discriminate the differences between the clinical and healthy groups. The *Short-term Memory* and *Spatial Orientation* were the cognitive domains that worst discriminated the differences between the clinical and healthy groups.

(Insert Table 3 about here)

(Insert Figure 1 about here)

Differential Item Functioning

It is essential that the screening instruments lack DIF because we are only able to compare the performance of the different groups (healthy vs. clinical) if the MoCA has the same metric properties (Prieto et al., 2012; Freitas et al., 2014). DIF analyses were performed in order to investigate the likelihood that the cognitive domains of the MoCA may work differently as a function of pathology, gender, age or educational level. An item showed an appreciable DIF if the following criteria were met: (a) a difference that was greater than .50 logits and a statistically significant difference (Bonferroni's correction) among the difficulty parameters of the reference group and the focal group (Prieto et al., 2010; $P_{Bonferroni} = .05/7 = .0071$) and (b) a Delta MH value that was classified as C, which is consistent with the criteria of the Educational Testing Service (Padilla, Hidalgo, Benítez & Gómez-Benito, 2012; C in logits: size>.64). According to these criteria, the cognitive domains show no age-related or gender-related DIF. Only the *Short-term Memory* domain presents pathology-related DIF, and *Executive Functions* and *Temporal Orientation* exhibit DIF that is associated with education. Regarding the domains with education-related DIF, the *Executive Functions* domain favors the higher education group (>4 years), and the *Temporal Orientation* domain benefits the lower education group (1-4 years).

Person Analysis

The descriptive statistics of the participants' scores ($n = 897$) are summarized in Table 4. Both the classic scores and the logits values show a high variability between the subjects on the scale. The mean of the logits (1.86) suggests that the average cognitive performance of the study sample is high, which is a consequence of the large number of cognitively healthy participants in the sample (72%). Only 9.93% of the

participants indicated a severe model misfit. The person values have been estimated with high reliability in regard to both hits (α=.83) and logits (PSR =.81).

(Insert Table 4 about here)

The means of the scores of pathology, gender, educational level and the age subgroups are provided in Table 5. The MoCA scores reveal good discriminant validity with high diagnostic utility, which is expressed by the large and significant differences that are observed between the control and clinical groups. The effect size (Cohen's d; Cohen, 1988) ranges between 1.47 (in the MCI group) and 2.76 (in the AD group). Moreover, with the exception of the comparisons between the FTD and VaD groups and between the AD and FTD groups, statistically significant differences of the means between all of the other groups were observed. As expected due to previous studies (Freitas et al., 2011; Freitas, Simões, Alves & Santana, 2012; Freitas, Prieto, Simões & Santana, 2014), there were no significant differences in cognitive performance between genders. On the other hand, the cognitive performance mean of the participants with a higher educational level (> 4 years) was significantly higher than the mean of the group with a primary educational level (1 to 4 years; t=17.293; df=895; p<.01; d=1.21). Finally, statistically significant differences of the means between the age subgroups were observed; the older participants (older than 64 years old) had a lower performance than the younger participants (t=16.874; df=895; p<.01; d=1.14).

(Insert Table 5 about here)

Discussion

The MoCA (Nasreddine et al., 2005) is a brief cognitive screening instrument that has shown sensitivity and diagnostic utility in the detection of milder forms of cognitive impairment that are associated with various clinical conditions. Several

studies have consistently reported the good overall psychometric properties of MoCA, with a few using item response theory (IRT) or, more specifically, the Rasch model (Freitas et al., 2014; Konsztowicz et al., 2011; Koski et al., 2009; Koski et al., 2011; Tsai et al., 2012). However, as far as we know, there have been no other studies that analyse the MoCA results while considering its seven indicators of cognitive function that were proposed by the authors [(i) *Executive Functions*; (ii) *Visuospatial Abilities*; (iii) *Short-term Memory*; (iv) *Language*; (v) *Attention, Concentration and Working Memory*; (vi) *Temporal Orientation*; (vii) *Spatial Orientation*; Nasreddine et al., 2005; Julayanont, Phillips, Chertkow, & Nasreddine, 2013] as polytomous items by using the PCM. The PCM allows us to estimate the values of these polytomous items on the latent continuum of cognitive function. This makes it possible to scale the seven cognitive domains along the continuum and infer its values in order to identify varying degrees of cognitive impairment. As the first step of the analysis of this study, we investigate the quality of the categories of each cognitive domain by emphasizing the monotonic ordering of the steps between the adjacent categories. We found that two cognitive domains (*Short-term Memory* and *Temporal Orientation*) were not monotonically ordered (Andrich, 2013). Consequently, these cognitive domains were recoded (collapsing six into five categories for *Short-term Memory* and five categories into three for *Temporal Orientation*), and the new categories revealed an increased order of Andrich's thresholds between the adjacent categories. The overall aim of the study was to further assess the psychometric characteristics of the MoCA by using the PCM for polytomous items regarding the evaluation of data fit and reliability values for the estimation of the items and persons. We also ran DIF analyses in order to explore whether MoCA's cognitive domains may work differently as a function of pathology, gender, age or educational level.

The results of this study show an overall good fit for both the items and the persons' values. The item analysis reveals the psychometric adequacy of the seven cognitive domains of the MoCA because they are within the optimal range for a good measurement. No domain presents a severe misfit. The high variability in the cognitive performance level is an additional indicator of the validity of the MoCA domains. The cognitive domains show high variability in the cognitive performance level. The cognitive domain with the highest level of difficulty is the *Short-term Memory* (1.90), followed by the domains of *Executive Functions* (1.47), *Attention, Concentration and Working Memory* (.52), *Visuospatial Abilities* (.16), *Language* (-.02), *Temporal Orientation* (-.68) and finally, as the easiest cognitive domain, *Spatial Orientation* (-3.35). This means that a person needs a high level of cognitive function in order to obtain a high score on the *Short-term Memory* and *Executive Functions* domains. Otherwise, a high level of cognitive function is not required in order to attain good performance on *Spatial Orientation*; this means that a low score on this domain reveals severe cognitive impairment. It is also possible to estimate that a person with a score that corresponds to the "average clinical group" score is more likely to obtain a maximum score on the *Spatial Orientation* (.95) and *Temporal Orientation* (.52) domains, whereas this person has a very low or null probability of correctly resolving the other five cognitive domains. On the other hand, a person with a score that corresponds to the "average healthy group" score has a high probability of attaining a perfect score on both of the *Orientation* domains, but that person also has a low probability of obtaining a maximum score on the *Executive Functions* (.38), *Attention, Concentration and Working Memory* (.32) and *Short-term Memory* (.20) cognitive domains. The cognitive domains that worst discriminate the differences between the clinical and healthy groups are on the domains of *Short-term Memory* and *Spatial*

Orientation. In fact, both of the groups have a high likelihood of obtaining a perfect cognitive domain score on *Spatial Orientation*, whereas regarding the *Short-term Memory* domain, both of the groups have a low likelihood of correctly resolving the tasks and attaining the maximum score.

Regarding the results of the person analysis, suitable psychometric characteristics of the measures can be observed. Both the classic scores and the logits values indicate high variability between the subjects. The person values have been estimated with high reliability for both the hits and the logits. The MoCA scores reveal good discriminant validity and high diagnostic utility, which is shown by the significant differences that are found between the means of the control and clinical groups and additionally between the clinical groups with different diagnoses. The exceptions include the comparison between the FTD and VaD groups, which are groups with small sizes (33 and 34 patients, respectively), and between the AD and FTD groups. This close proximity of the MoCA scores of these three diagnostic groups is expected because there are three dementia groups whose patients display symptoms of mild to moderate severity (classified with CDR ≤ 2 and MMSE ≥ 12). As expected, according to previous studies that use the Portuguese population (Freitas et al., 2011; Freitas, Simões, Alves & Santana, 2012; Freitas et al., 2014), statistically significant differences in cognitive performance are observed between the age groups and between the educational level groups, with the worst performances being observed among the participants with an older age (older than 64 years old) or a lower education level (1-4 years), whereas the mean cognitive performance was similar for both genders.

DIF analyses were conducted in order to explore the possibility that the cognitive domains of the MoCA may work differently as a function of pathology, gender, age or educational level. The MoCA's domains show no age-related or gender-

related DIF. Only the *Short-term Memory* domain reveals pathology group-related DIF. The difficulty parameter of the *Short-term Memory* domain is higher in the clinical group (2.61) than in healthy group (1.74), possibly due to the characteristics of the clinical group (73% of the patients having amnesic impairment: 90 patients with amnesic-MCI and 90 patients with AD). The *Executive Functions* and *Temporal Orientation* domains exhibit education-related DIF; the *Executive Functions* domain favors the higher education group (>4 years), and the *Temporal Orientation* domain benefits the lower education group (1-4 years). This balanced phenomenon is known as bias cancellation because at the overall test score level, the respective biases may cancel each other out (Drasgow, 1987; Nandakumar, 1993; Roznowski, 1987). Therefore, it is reasonable to assume that the cognitive domains with education-related DIF do not spuriously change the differences in MoCA scores.

To explore the efficiency of cognitive domains to differentiate between normal and clinical group (called "impact") we analyzed the differences between the probabilities of obtaining the maximum possible score in each domain of the individuals with a score that corresponds to the average of the both groups (healthy and clinical). It was possible to observe that all cognitive domains showed a high impact to differentiate between the healthy and the clinical group, with the exception for *Short-term Memory* (higher level of difficult) and *Spatial Orientation* domains (lower level of difficult). Regarding the *Spatial Orientation* domain, an individual with a score that corresponds to the "average clinical group" score has a high probability (.95) of obtaining the highest score in this task, similar to an individual with a "average healthy group" score (1.00). On the contrary, on the *Short-term Memory* domain, an individual with a score that corresponds to the "average healthy group" score has a low probability (.20) of

obtaining the highest score in this task, while an individual with a "average clinical group" score has a null probability (.00).

The results of this study demonstrate (i) the overall good fit of the items and persons' values, the fundamental unidimensionality of the data and the appropriate metric functionality of the ratings for all of the cognitive domains, (ii) high variability in the cognitive performance level of the cognitive domains and between the subjects on the scale, (iii) high reliability in the estimation of person values, (iv) good discriminant validity and high diagnostic utility, and (v) minimal DIF effects. Thus, the study demonstrates that the MoCA and its items/cognitive domains are suitable measures for the brief screening of the global cognitive status of cognitively healthy subjects and patients with cognitive impairment.

In our opinion, the methodology that is used is the main strength point of this study, including (i) rigorous MoCA administration, with no inter-rater variability (all of the participants were assessed by one of two expert neuropsychologists); (ii) a control group with subjects who were recruited from the community and were well-characterized as cognitively healthy adults (in order to ensure the cognitive health of the control participants, we established strict inclusion criteria that were checked in the clinical interview and the neuropsychological assessment, and for older participants, confirmatory information was also obtained by a general practitioner, the community center director and/or an informant); (iii) well-validated clinical samples (diagnoses were established by a multidisciplinary team that used the standard criteria, based on a full investigation); and (iv) the homogeneity of the clinical groups (patients with misclassification and more advanced dementia cases were excluded).

However, there are a few limitations that should be recognized in the study: (i) the control sample size and geographical distribution of the participants did not allow us

to perform a neurological consultation or additional diagnostic tests (such as neuroimaging), which would have further ensured the normal cognitive status of the participants; (ii) the sample sizes of the FTD and VaD groups were reduced, which did not allow for a more detailed analysis that considered the clinical groups separately; (iii) in this study age and education mean differences between healthy and clinical groups were not controlled, since this would result in a significant reduction of the sample size, obviating the proposed analysis with the PCM model; this is an interesting topic to address in future studies; (iv) due to the large extension of the comprehensive neuropsychological assessment protocols, a measure of effort was not included and according to this, sub-optimal engagement was not considered as an exclusion criterion for the study; and (v) due to the lack of other studies that used the Partial Credit Model to examine the MoCA's cognitive domains, it was not possible to compare the results that were obtained in this study to those of any other study.

In conclusion, the results of the study are consistent with extensive research that has demonstrated the good psychometric properties and high utility of the MoCA for the screening of global cognitive status, which adds to the evidence of the suitability of the MoCA's cognitive domains.

References

Alves, L., Simões, M. R., & Martins, C. (2009). *Teste de Leitura de Palavras Irregulares (TeLPI)* [Reading Word Irregular Test]. Coimbra, Portugal: Serviço de Avaliação Psicológica da Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra [Psychological Assessment Department, Faculty of Psychology and Educational Sciences, University of Coimbra].

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. – Text Revised). Washington, DC: American Psychiatric Association Press.

Andrich, D. (1988). *Rasch models for measurement*. London: Sage.

Andrich, D. (2013). Na expanded derivation of the threshold stucture of the polytomous Rasch model that dispels any "threshold disorder controversy". *Educational and Psychological Measurement, 73*(1), 78-124.

Barreto, J., Leuschner, A., Santos, F., & Sobral, M. (2008). Escala de Depressão Geriátrica [Geriatric Depression Scale]. In Grupo de Estudos de Envelhecimento Cerebral e Demências [Study Group on Brain Aging and Dementia] (Ed.), *Escalas e testes na demência* [Scales and tests in dementia] (pp. 69-72). Lisbon: GEECD.

Benjamini Y., & Hochberg Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57*(1), 289-300.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Conrad, K. J., & Smith, E. V. (2004). International conference on objective measurement: Applications of Rasch analysis in health care. *Medical Care, 42,* 1-6.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.

Drasgow, F. (1987). Study of measurement bias of two standardized psychological tests. *Journal of Applied Psychology, 72*, 19–29.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: LEA.

Folstein, M., Folstein, S., & McHugh, P. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Freitas, S., Prieto, G., Simões, M.R., & Santana, I. (2014). Psychometric properties of the Montreal Cognitive Assessment (MoCA): An analysis using the Rasch model. *The Clinical Neuropsychologist, 28(1),* 65-83.

Freitas, S., Simões, M. R., Alves, L., & Santana, I. (2011). Montreal Cognitive Assessment (MoCA): Normative study for the Portuguese population. *Journal of Clinical and Experimental Neuropsychology*, *33*(9), 989-996.

Freitas, S., Simões, M. R., Alves, L., & Santana, I. (2012). Montreal Cognitive Assessment (MoCA): Influence of sociodemographic and health variables. *Archives of Clinical Neuropsychology, 27*, 165-175.

Freitas, S., Simões, M. R., Marôco, J., Alves, L., & Santana, I. (2012). Construct validity of the Montreal Cognitive Assessment (MoCA). *Journal of International Neuropsychology Society*, *18,* 242-250.

Garret, C., Santos, F., Tracana, I., Barreto, J., Sobral, M., & Fonseca, R. (2008). Avaliação Clínica da Demência [Clinical Dementia Rating Scale]. In Grupo de Estudos de Envelhecimento Cerebral e Demências [Study Group on Brain Aging and Dementia] (Ed.), *Escalas e testes na demência* [Scales and tests in dementia] (pp. 17-32). Lisbon: GEECD.

Gauthier, S., Patterson, C., Gordon, M., Soucy, J., Schubert, F., & Leuzy, A. (2011). Commentary on "Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease": A Canadian perspective. *Alzheimer's & Dementia, 7*, 330-332.

Ginó, S., Mendes, T., Ribeiro, F., Mendonça, A., Guerreiro, M., & Garcia, C. (2008). Escala de Queixas de Memória [Memory Complaints Scale]. In Grupo de Estudos de Envelhecimento Cerebral e Demências [Study Group on Brain Aging and Dementia] (Ed.), *Escalas e testes na demência* [Scales and tests in dementia] (pp. 117-120). Lisbon: GEECD.

Guerreiro, M., Silva, A. P., Botelho, M., Leitão, O., Castro-Caldas, A., & Garcia, C. (1994). Adaptação à população portuguesa da tradução do Mini Mental State Examination [Adaptation of the Mini Mental State Examination translation for the Portuguese population]. *Revista Portuguesa de Neurologia, 1*, 9.

Holland, P., & Thayer, D. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. I. Braun (Eds.), *Test Validity* (pp.129-145). Hillsdale, NJ: LEA.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry, 140*, 566-572.

Ismail, Z., Rajji, T. K., & Shulman, K. I. (2010). Brief cognitive screening instruments: An update. *International Journal of Geriatric Psychiatry, 25*(2), 111-120.

Jacova, C., Kertesz, A., Blair, M., Fisk, J. D., & Feldman, H. H. (2007). Neuropsychological testing and assessment for dementia. *Alzheimer's Dementia, 3*(4), 299-317.

Julayanont, P., Phillips, N., Chertkow, H., & Nasreddine, N. (2013). Montreal Cognitive Assessment (MoCA): Concept and clinical review. In A. J. Larner (Ed.), *Cognitive Screening Instruments* (pp.111-151). London: Springer.

Konsztowicz, S., Xie, H., Higgins, J., Mayo, N., & Koski L. (2011). Development of a method for quantifying cognitive ability in the elderly through adaptive test administration. *International Psychogeriatrics, 23*(7), 1116-1123.

Koski, L., Xie, H., & Finch, L. (2009). Measuring cognition in a geriatric outpatients clinic: Rasch analysis of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology, 32*(3), 151-160.

Koski, L., Xie, H., & Konsztowicz, S. (2011). Improving precision in the quantification of cognition using the Montreal Cognitive Assessment and the Mini-Mental State Examination. *International Psychogeriatrics, 23*(7), 1107-1115.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*, 85–106.

Linacre, J. M. (2011). *A user's guide to WINSTEPS & MINISTEPS: Rasch model computer programs*. Chicago, IL: Winsteps.com.

Lonie, J. A., Tierney, K. M., & Ebmeier, P. (2009). Screening for mild cognitive impairment: A systematic review. *International Journal of Geriatric Psychiatry, 24*, 902-915.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R. Jr, Kawas, C. H., … Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia 7*(3), 263–269.

Nandakumar, R. (1993). Simultaneous DIF Amplification and Cancellation: Shealy-Stouts´s Test for DIF. *Journal of Educational Measurement, 30* (4), 293-311.

Nasreddine, Z., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for Mild Cognitive Impairment. *American Geriatrics Society*, *53*(4), 695-699.

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., … Benson, D. F. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology, 51*, 1546–1554.

Padilla, J. L., Hidalgo, M. D., Benítez, I. & Gómez-Benito, J. (2012). Comparison of three software programs for evaluating DIF by means of the Mantel-Haenszel procedure: EASY-DIF, DIFAS and EZDIF. *Psicológica, 33*, 135-156.

Petersen, R. C. (2007). Mild cognitive impairment. *Continuum Lifelong Learning in Neurology, 13*(2), 15–38.

Petersen, R. C. (2004). Mild Cognitive Impairment as a diagnostic entity. *Journal of Internal Medicine,256*(3), 183-194.

Potenza, M., & Dorans, N. (1995). DIF assessment for politomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement,19*, 23-37.

Prieto, G., Delgado, A. R., Perea, M. V., & Ladera, V. (2010). Scoring neuropsychological tests using the Rasch model: An illustrative example with the Rey-Osterreith Complex Figure. *The Clinical Neuropsychologist, 24*, 45-56.

Prieto, G., Contador, I., Tapias-Merino, E., Mitchell, A. J., & Bermejo-Pareja, F. (2012). The Mini-Mental-37 Test for dementia screening in the Spanish population: An analysis using the Rasch model. *The Clinical Neuropsychologist, 26*(6), 1003-1018.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research. (Expanded edition, 1980. Chicago: University of Chicago Press).

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4,* 207-230.

Román, G. C., Tatemichi, T., Erkinjuntti, T., Cummings, J. L., Masdeu, J. C., Garcia, J. H., … Scheinberg, P. (1993). Vascular dementia: Diagnostic criteria for research studies: report of the NINDS-AIREN International Workshop. *Neurology, 43,* 250–260.

Roznowski, M. (1987). Use of tests manifesting sex differences as measure of intelligence: Implications for measurement bias. *Journal of Applied Psychology, 72*, 480-483.

Schmand, B., Jonker, C., Hooijer, C., & Lindeboom, J. (1996). Subjective memory complaints may announce dementia. *Neurology, 46*(1), 121-125.

Tsai, C., Lee, W., Wang, S., Shia, B., Nasreddine, Z., & Fuh, J. (2012). Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: Validation of the Taiwanese version of the MoCA and an item response theory analysis. *International Psychogeriatrics, 24*(4), 651-658.

Wilson, M. (1994). *Objective measurement II: Theory into practice*. Norwood, NJ: Ablex.

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Wright, B. D., & Douglas, G. A. (1976). Rasch item analysis by hand. Research Memorandum No. 21, Statistical Laboratory, Department of Education, University of Chicago.

Wright, B. D., & Mok, M. M. (2004). An overview of the family of Rasch measurement models. In E. V. Smith Jr & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 1-24). Maple Grove, MN: JAM Press.

Yen, W.M. (1993). Scaling performance assessment: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187-213.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., &Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research, 17*(1), 37-49.

Zwick, R., & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement, 26*, 55-66.

Table 1. *Study population characteristics for the total sample and subgroups*

	N/n	**Gender** f (%)	**Age** M ± SD [Min.-Max.]	**Education** M ± SD [Min.-Max.]	**MoCA** M ± SD [Min.-Max.]
Total Sample	897	541 (60.3)	60.3 ± 15.4 [25-91]	7.6 ± 4.6 [1-27]	21.7 ± 6.5 [2-30]
Healthy Group	650	408 (62.8)	55.8 ± 15.1 [25-91]	8.2 ± 4.7 [1-27]	24.7 ± 3.7 [15-30]
Clinical Group	247	133 (53.8)	72.0 ± 8.2 [46-91]	6.2 ± 4.1 [1-20]	13.8 ± 5.6 [2-25]
MCI	90	55 (61.1)	70.5 ± 8.0 [46-91]	6.5 ± 4.6 [1-20]	18.3 ± 3.9 [10-25]
AD	90	52 (57.8)	74.2 ± 8.2 [54-91]	6.2 ± 4.1 [1-17]	10.1 ± 4.4 [2-21]
FTD	33	14 (42.4)	68.4 ± 7.0 [55-79]	6.4 ± 3.8 [3-15]	12.2 ± 4.8 [4-24]
VaD	34	12 (35.3)	73.2 ± 7.9 [51-86]	5.0 ± 2.8 [2-15]	13.0 ± 4.6 [5-24]

Note. f = feminine gender; M = Mean; SD = Standard Deviation; Min. = Minimum value; Max. = Maximum value; MoCA = Montreal Cognitive Assessment (maximum score = 30); Health Group: all of the cognitively healthy participants; Clinical Group: all of the patients with MCI, AD, FTD and VaD; MCI = Mild Cognitive Impairment; AD = Alzheimer's Disease; FTD = Frontotemporal dementia; VaD = Vascular dementia.

Table 2. *Items and ranges of the ratings of the MoCA*

Variables (polytomous items)	Items	Initial Range of Values[1]	Final Range of Values[2]
Executive Functions	trail making B adapted, phonemic fluency and two verbal abstraction tasks	[0 – 4]	[0 – 4]
Visuospatial Abilities	three-dimensional cube copy and a clock-drawing task (contour, numbers and hands)	[0 – 4]	[0 - 4]
Short-term Memory	delayed recall of the five words	[0 – 5]	[0 – 4] (0 = 0; 1 and 2 = 1; 3 = 2; 4 = 3; 5 = 4)
Language	naming three animals and the repetition of two sentences	[0 – 5]	[0 – 5]
Attention, Concentration and Working Memory	digits forward and backward, a sustained attention task and a serial subtraction task	[0 – 8]	[0 – 8]
Temporal Orientation	date, month, year and day	[0 – 4]	[0 – 2] (0 and 1 = 0; 2 and 3 = 1; 4 = 2)
Spatial Orientation	place and city	[0 – 2]	[0 – 2]
Range of Total Scores		[0 – 32]	[0 – 29]

Note.[1]Range of the values, according to the aggregation of the correct answers to the dichotomous items that are included in each domain, as described by the authors; [2]Range of the values, considering the recoding of the categories that was performed in the present study.

Table 3. *Item values of MoCA*

Item	RiX	D(logits)	SE	Infit	Outfit	XMax	Pcg	Phg	Pcg-Phg
Executive Functions	.81	1.47	.04	.77	.76	4	.00	.38	.38*
Visuospatial Abilities	.77	.16	.04	.82	.80	4	.01	.58	.57*
Short-term Memory	.72	1.90	.04	1.38	1.31	5	.00	.20	.20
Language	.75	-.02	.04	1.09	1.03	5	.01	.46	.45*
Attention, Concentration and Working Memory	.82	.52	.03	1.13	1.10	6	.00	.32	.32*
Temporal Orientation	.61	-.68	.08	.91	.85	4	.52	.95	.43*
Spatial Orientation	.31	-3.35	.19	1.03	.59	2	.95	1.00	.05
Mean	--	*.00*	*.07*	*1.02*	*.92*	--	--	--	--
SD	--	*1.59*	*.05*	*.19*	*.22*	--	--	--	--

Note. M = Mean; SD = Standard Deviation; RiX = Item-test correlation; D = Difficulty (logits); SE = Standard Error of D; XMax = maximum score; Pcg = Probability of Maximum Score to B = .04 (mean of clinical group); Phg = Probability of Maximum Score to B = 2.55 (mean of healthy group); Pcg – Phg * = Highest discriminative items.

Table 4. *Descriptive Statistics of the Participants' Scores*

Statistic	X	B	SE B
Mean	20.2	1.86	.60
SD	6.5	1.69	.34
Min.	2	-3.52	.42
Max.	30	5.49	1.84
Alfa	.83	--	--
PSR	--	.81	--
% D	9.93	--	--

Note. X = Classical score (sum of the correct answers); B = Rasch person values (logits); SE B = Standard Error of the Rasch values; M = Mean; SD = Standard Deviation; Min. = Minimum value; Max. = Maximum value; α = Cronbach alpha coefficient of the scores (classic reliability); PSR = Rasch reliability (Person Separation Reliability); % D = Percentage of the subjects with severe misfit (*Outfit* or *Infit*>2).

Table 5. *Score Means of the Persons' Groups According to Pathology, Gender, Education or Age*

Group	*n*	M X	SD X	M B	SD B	PSR
Pathology						
Healthy	650	23.0$_a$	4.1	2.55$_a$	1.25	.61
MCI	90	16.7$_b$	3.9	1.02$_b$	.78	.57
AD	90	9.3$_c$	4.3	-.74$_c$	1.13	.69
FTD	33	10.8$_{cd}$	4.8	-.36$_{cd}$	1.28	.74
VaD	34	11.7$_d$	4.5	-.08$_d$	1.06	.71
Gender						
Male	356	20.2$_a$	6.6	1.88$_a$	1.72	.81
Female	541	20.1$_a$	6.4	1.85$_a$	1.67	.81
Education						
1-4 years	416	16.5$_a$	5.9	.91$_a$	1.33	.83
> 4 years	481	23.3$_b$	5.2	2.68$_b$	1.53	.68
Age						
<65	474	23.1$_a$	5.1	2.64$_a$	1.51	.68
>64	423	16.8$_b$	6.3	1.00$_b$	1.43	.84

Note. B means with different subscripts differ significantly ($p<.05$). M X = Mean of Classic score; SD X = Standard Deviation of Classic score; M B = Mean of Rasch values (logits); SD B = Standard Deviation of Rasch values (logits); PSR = Rasch reliability (Person Separation Reliability).
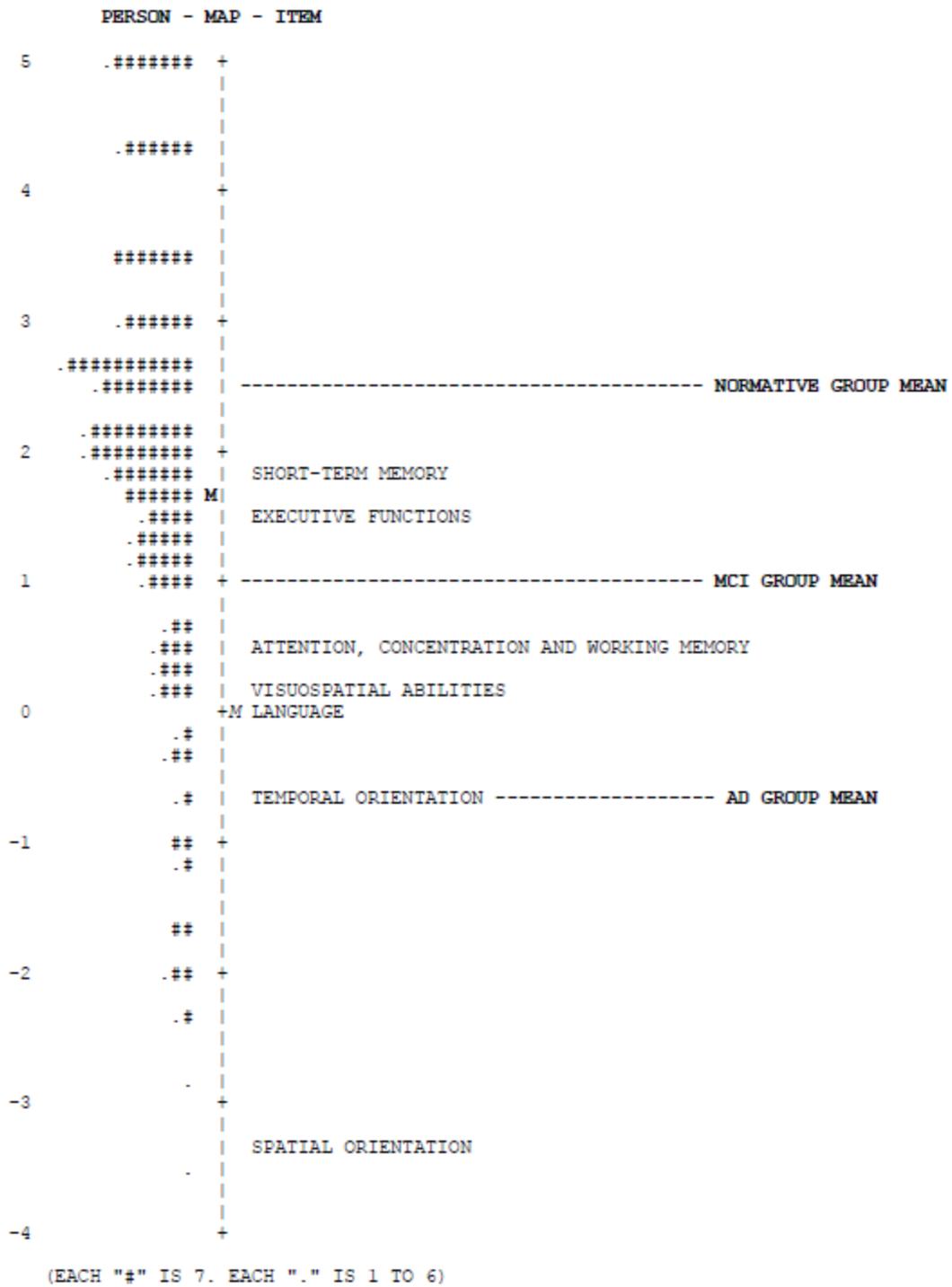
```
PERSON - MAP - ITEM
 5        .#######  +
                    |
                    |
                    |
          .######   |
 4                  +
                    |
                    |
         #######    |
                    |
                    |
 3        .######   +
      .############  |
        .########   | ----------------------------------- NORMATIVE GROUP MEAN
                    |
        .#########  |
 2    .#########   +
        .#######   |   SHORT-TERM MEMORY
         ###### M|
         .####    |   EXECUTIVE FUNCTIONS
         .#####   |
         .#####   |
 1        .####   + ----------------------------------- MCI GROUP MEAN
                  |
          .##     |
          .###    |   ATTENTION, CONCENTRATION AND WORKING MEMORY
          .###    |
          .###    |   VISUOSPATIAL ABILITIES
 0               +M LANGUAGE
           .#     |
          .##     |
                  |
           .#     |   TEMPORAL ORIENTATION -------------------- AD GROUP MEAN
                  |
-1         ##     +
           .#     |
                  |
                  |
          ##      |
                  |
-2        .##     +
                  |
           .#     |
                  |
                  |
           .      |
-3                +
                  |
                  |   SPATIAL ORIENTATION
           .      |
                  |
                  |
-4                +
     (EACH "#" IS 7. EACH "." IS 1 TO 6)
```

Figure 1. *Variable map: Conjoint measurement (persons and items)*

Note. **M** = Person mean value; *M* = Item mean value.