Psychological Evaluation

Mini-Mental State Examination in Brazil: An Item Response Theory Analysis

Denise Mendonça de Melo[1]
Altemir José Gonçalves Barbosa[2]
Nelimar Ribeiro de Castro[3]
Anita Liberalesso Neri[4]

Abstract: The Mini-Mental State Examination (MMSE) is currently used to the track cognitive status of older adults in Brazil. Studies on its psychometric properties are lacking, especially ones that use Item Response Theory. The objective was to assess the difficulty of the items in a Brazilian version of MMSE using the Rasch model and to identify possible differential item functioning (DIF), considering schooling, age and sex of the sample of Brazilian older adults. This study used the answers of 2.734 older adults to the 30 items of MMSE. It was found that four items in the seven series were the most difficult, and items nine, 13, 22 and 23 were the easiest. The skill of respondents was higher than the items' difficulty level. DIF was observed for schooling, sex and age in, respectively, 27, 18 and 16 items. It is concluded that the use of MMSE in Brazilian older adults should be cautious due to the large number of biased items, mainly due to schooling.

Keywords: Minimental state examination, item response theory, psychometry, educational status

Miniexame do Estado Mental Brasileiro: Análise com Teoria de Resposta ao Item

Resumo: O Miniexame do Estado Mental (MEEM) tem sido utilizado para rastrear status cognitivo de idosos brasileiros. Há poucos estudos sobre suas propriedades psicométricas, especialmente com Teoria de Resposta ao Item. Este estudo teve por objetivo avaliar a dificuldade dos itens de uma versão brasileira do MEEM por meio do modelo de Rasch e identificar possível funcionamento diferencial dos itens (DIF), considerando escolaridade, idade e sexo dos integrantes de amostra de idosos da comunidade. Foram utilizadas as respostas de 2.734 idosos aos 30 itens do MEEM. Constatou-se que quatro itens do sete seriado foram os mais difíceis e os itens nove, 13, 22 e 23, os mais fáceis. O nível de habilidade das pessoas foi maior do que o nível de dificuldade dos itens. Observou-se DIF para escolaridade, sexo e idade, respectivamente 27, 18 e 16 itens. Conclui-se que o uso do MEEM em idosos da comunidade deve ser cauteloso devido ao grande número de itens enviesados, principalmente pela escolaridade.

Palavras-chave: Miniexame do estado mental, teoria de resposta ao item, psicometria, escolaridade

Miniexamen del Estado Mental Brasileño: Análisis con la Teoría de Respuesta al Ítem

Resumen: El Miniexamen del Estado Mental (MMSE) se ha utilizado para rastrear el estado cognitivo de los ancianos brasileños. Sin embargo, no hay estudios sobre sus propiedades psicométricas, especialmente utilizando la Teoría de Respuesta al Ítem. El objetivo de este estudio fue evaluar la dificultad de los ítems de una versión brasileña del MMSE utilizando el modelo de Rasch, así como identificar el posible funcionamiento diferencial de los ítems (DIF), considerando el nivel de estudios, la edad y el sexo de los ancianos participantes de la muestra. Se utilizaron las respuestas de 2.734 ancianos a los 30 ítems del MMSE. Se evidenció que cuatro ítems de siete series fueron los más difíciles, y los ítems 9, 13, 22 y 23 los más fáciles. El nivel de habilidad de las personas fue más alto que el nivel de dificultad de los ítems. Se observó que el DIF para nivel de estudios, sexo y edad fueron 27, 18 y 16 ítems, respectivamente. Se concluye que el uso del MMSE en los ancianos de la comunidad debe ser cauteloso debido a la gran cantidad de ítems sesgados, principalmente por el nivel de estudios.

Palabras clave: Miniexamen del estado mental, teoría de respuesta al ítem, psicometría, escolaridad

[1]*Centro Universitário UniAcademia, Juiz de Fora-MG, Brazil*
[2]*Universidade Federal de Juiz de Fora, Juiz de Fora-MG, Brazil*
[3]*Universidade de Viçosa, Viçosa-MG, Brazil*
[4]*Universidade Estadual de Campinas, Campinas-SP, Brazil*

Correspondence address: Denise Mendonça de Melo. Centro Universitário UniAcademia. Rua Halfeld, 1.179, Centro, Juiz de Fora-MG, Brazil. CEP 36.016-000. E-mail: denisemmelo@gmail.com

The Mini-Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975) – a measure of cognitive status, i.e., a group of basic information processing abilities (Zisberg, Zysberg, Young, & Schepp, 2009) – is the most widely used and extensively studied cognitive screening test worldwide (Carnero-Pardo, 2014). However, in Brazil, controversies remain over its psychometric properties, particularly its dimensions, and few studies analyzing these properties using Item Response Theory (IRT) have been conducted.

Some studies (Castro-Costa et al., 2014; Ideno, Takayama, Hayashi, Takagi, & Sugai, 2012; Jones & Gallo, 2000; Melo, Barbosa, & Neri, 2017) have identified a multidimensional structure for the MMSE using factor or principal component

analyses. However, the presence of multiple factors has been questioned when, for example, considering the value and factor loading of the items (Jones & Gallo, 2000) or the item difficulty levels analyzed using IRT (Ideno et al., 2012). According to Jones and Gallo (2000), the study by Melo et al. (2017) also detected a much higher value for one of the factors or dimensions of the MMSE, where all items loaded positively for one factor in the unrotated matrix.

Notably, one-dimensionality is a requisite for IRT. IRT is a theory of the latent trait which can be applied to tests of ability or performance to assess their psychometric properties (Pasquali, 2013). If applied to a single measure, IRT produces ways of representing the relationship between the likelihood of an individual providing a correct answer to an item and its latent traits. Latent trait or *theta* (θ) is defined as the characteristics of the individual which cannot be observed directly and need indirect assessments, i.e., items which detect them (Bond & Fox, 2015). This occurs with the MMSE, which makes use of a series of items to measure the latent trait of cognitive status.

IRT presumes, among other premises, that a test needs to produce a true and identical score and that the result of testing cannot hinge on sample characteristics (Pasquali, 2013) such as education. This is fundamental, because in Brazil the results of cognitive screening using the MMSE are generally influenced by formal education (Melo & Barbosa, 2015).

One of the analyses performed by IRT is the widely known Rasch or one-parameter model, which assesses item difficulty, establishing a hierarchy of difficulty and the extent to which individuals are discriminated, i.e., separated by ability level (Fernandes, Pietro, & Delgado, 2015). This model presumes that item difficulty is a fundamental characteristic influencing responses and discriminates groups. In a systematic review on the use of IRT in cognitive tests, McGrory, Doherty, Austin, Starr, and Shenkin (2014) found that the most difficult items in the MMSE in a sample of demented individuals were the three items assessing memory (delayed recall), "what day is it today" (time orientation), and the "serial sevens" (attention/calculation). The subtraction items were also found to be more difficult by the study of Kim, Won, Kim, and Choi (2013). The least difficult items, according to McGrory et al. (2014), were "3-stage command", "pen naming" and immediate repetition of three names. The same study also found that the items with greatest discriminatory power were name the pencil, write a sentence, "what is the month", "wristwatch naming", "what is the day", "what is the year" and "close your eyes". The items which discriminated least were recall of two nouns and "3-stage command".

Differential item functioning (DIF) analysis is another function provided by IRT. DIF occurs when the parameters of a given item differ among different segments of the population, such as those based on education, resulting in some individuals being favored over others (Pasquali, 2013). Item-person interaction depends on two characteristics: item difficulty and person ability (Sisto, 2006). A test may be considered valid when it results in equal scores for persons with similar abilities. When this fails to occur for some items of a test, the item is said to exhibit DIF, i.e., the probability of being correct in this item is different in groups of individuals with the same level of competence (Sisto, 2006). Affirming that an item has DIF equates to stating the item has bias (Linacre, 2002). Some authors suggest excluding items exhibiting DIF (McGrory et al., 2014).

Thus, IRT is apparently able to clarify the complex relationship between variability in MMSE scores and education level, one of the major sources of controversy associated with the test (Melo & Barbosa, 2015). Apolinario, Mansur, Carthery-Goulart, Brucki, & Nitrini (2014), among others, found that education influences scores in the MMSE. The test contains several items directly associated with education level, where education constitutes a possible factor (Melo & Barbosa, 2015). A number of researchers take into account participants´ years of education to establish cut-off points for the MMSE (Melo & Barbosa, 2015), but this variable represents a relevant potential source of DIF. Jones and Gallo (2002) investigated the occurrence of DIF in the MMSE among groups stratified by years of education, in a sample of 8,556 community-dwelling individuals aged ≥ 50 years. Their results showed that a group of individuals with low education were more likely to err in the first item of serial sevens (100 - 7), spell "world" backwards, repeat the phrase, write a sentence, name the season of the year and copy a design. The researchers concluded that the items biased by education do not appear to be a source of observed differences in cognitive status.

Although they do not provide a totally satisfactory explanation for the variability in cognitive status in aging (Yassuda, Viel, Silva, & Albuquerque, 2013), the variables sex and age constitute a potential source of DIF for the measure in question, as demonstrated by Jones and Gallo (2002). In an analysis of DIF on the MMSE exploring age, the authors detected bias in one of the items of orientation, and in delayed recall, naming, repeat phrase, 3-stage command, write a sentence and copy design tasks in the older old in the sample (≥ 75 years). In the DIF test for sex they found that men were more likely to err in the spell word backwards, write a sentence and 3-stage command items but were more likely than women to give correct answers for the serial sevens and copy design items.

The objective of the present study was to assess item difficulty in the Brazilian version of the MMSE using the Rasch model and to detect possible DIF, taking into account education, age and sex in a sample of older adults from the community.

Method

Participants

This investigation drew on data from the Frailty in Brazilian Elderly study (2008-2009) conducted by the State University of Campinas (FIBRA UNICAMP),

a multi-center, multi-disciplinary study. The goal of FIBRA was to investigate associations among frailty indicators and demographic, health, psychosocial and cognitive variables in elderly aged ≥ 65 years from an urban zone including seven Brazilian sites selected by convenience criteria. The sample from each site was probabilistic and the sampling unit was census sector. The exclusion criteria were: having problems affecting memory, attention, spatial or temporal orientation and/or communication suggestive of dementia, being bedridden; presenting major stroke sequelae, with localized weakness and/or aphasia; having severe or unstable Parkinson´s Disease with severely impaired motricity, speech or affectivity; having severe visual or auditory deficits, hampering communication; and being at a terminal stage.

The sample comprised 2,734 older adults, 66.79% (n = 1,826) female and 33.21% (n = 908) male, stratified into four age groups, with 35.92% (n = 982) aged 65-69 years, 30.80% (n = 842) 70-74, 19.31% (n = 528) 75-79 and 13.97% (n = 382) ≥ 80 years of age. Mean age was 72.72 years (SD = 5.88). Regarding education, 16.25% (n = 444) were illiterate or had no formal education, 50.27% (n = 1,374) had 1-4 years of education, 19.03% (n = 520) 5-8 years and 14.45% (n = 395) ≥ 9 years of education. Mean years of study was 4.57 (SD = 4.01).

Instruments

Besides a demographic questionnaire (sex, age and education), the Brazilian version of the MMSE developed by Brucki, Nitrini, Caramelli, Bertolucci, and Okamoto (2003) was administered. This Brazilian version of the test and its corresponding cut-off points are recommended by the Brazilian Academy of Neurology for screening cognitive decline suggestive of dementia in older adults (Nitrini et al., 2005).

Procedure

Data collection. Trained recruiters visited the domiciles of the elderly, inviting those who met the inclusion criteria to take part in a data collection session at a set time, date and place. All participants signed a Free and Informed Consent Form and the instruments analyzed in the study were then administered through an interview at the beginning of the data collection session. The detailed method of FIBRA is available in Neri et al. (2013).

Data analysis. As a prerequisite for subsequent analyses, the unidimensionality of the MMSE was tested using Modified Parallel Analysis with software R. For an empirical check of the unidimensionality assumption for the Rasch models, an analysis of latent dimensionality of dichotomous items was performed according to Drasgow and Lissak (1983), adopting the second eigenvalue of the tetrachoric correlation matrix of the items. The highest eigenvalue was considered an estimate of commonality. A type Monte Carlo method is used to approximate the distribution of this

statistic to test the null hypothesis. The Winsteps statistical program was employed. First, the analyses of indices of adequacy of the items and persons were performed, along with the item-person map using the Rasch model with the infit and outfit indices. Based on Linacre (2002), infit and outfit values of 0.5-1.5 were deemed acceptable, values of 1.5-2.0 were classified as moderate and values > 2.0 were considered unacceptable. The mean expected value for these indicators is one (Bond & Fox, 2015).

Subsequently, DIF analysis was carried out for the variables sex; age stratified into four groups (65-69 years, 70-74 years, 75-79 years and ≥ 80 years); and education also subdivided into four strata (illiterate subjects, those with no formal education, 1-4 years of education, 5-8 years, and ≥ 9 years of education). Three criteria were considered to determine whether the items had DIF: Criterion 1 – Difference between DIF values of the groups analyzed of ≥ 0.50 (Draba, 1977); Criterion 2 – Significant contrast according to the Mantel-Haenszel probability index (Linacre & Wright, 2009); and Criterion 3 – t-value ≥ 2.40 (Linacre & Wright, 2009). As three conditions were adopted, it was decided to mention in the table only if there was adhesion to the Mantel Haenszel's χ^2 and to the other two criteria.

In addition to IRT, tests of means (t-test and Analysis of Variance (ANOVA) with Tuke's *post hoc*) were carried out to compare performance in the MMSE considering the subgroups formed for the demographic variables sex, age and education. Multiple linear regression using the 'enter' method was also applied to determine the relationship between total MMSE score and the variables age, education and sex.

Ethical Considerations

The FIBRA study was approved by the Research Ethics Committee of the School of Medical Sciences of Unicamp and obtained authorization number 208/2007 (CAAE n. 0 151.1.146.000-07). All participants signed the free and informed consent term.

Results

The unidimensionality hypothesis of the MMSE was corroborated. The second eigenvalue was observed to be 2.2846 (p = 0.1881). The parameters of fit the items (Table 1) reveal a mean infit of 0.99 (SD = 0.14) and outfit of 1.04 (SD = 0.32). The adjustment ranges of infit were from 1.36 to 0.80 and of outfit from 1.91 to 0.58. All items showed good fit for infit whereas only four (13.00%) items (items 8, 11, 21 and 25) showed lack of fit indices for outfit. With regard to parameters of fit for persons, 8.23% lack of fitness was observed for infit and 19.57% for outfit.

The analysis of item difficulty (Table 2) revealed that four items of the serial sevens (15, 16, 17 and 18) proved to be the most difficult and items 22, 23, 09 and 13 were the easiest.

Table 1
Parameters of fit for the Rasch model with item difficulty and person ability indices (Theta) in the MMSE

Parameters	Items			Persons		
	Infit	Outfit	Errors	Infit	Outfit	Errors
Mean	0.99	1.04	0.08	1	1.01	0.66
SD	0.14	0.32	0.05	0.28	1.18	0.26
Maximum	1.36	1.91	0.22	2.27	9.90	1.85
Minimum	0.80	0.58	0.04	0.37	0.20	0.47
1.5 < > 2.0	0 (0.00%)	4 (13.00%)		132 (4.83%)	179 (6.55%)	
< 2.0	0	0		93 (3.40%)	356 (13.02%)	

Table 2
Difficulty indices of MMSE items

Item		Difficulty	Standard Error	Infit	Outfit
22	Recognition of Wristwatch	-3.33	0.22	0.99	0.78
23	Recognition of Pen	-3.28	0.22	0.97	0.96
09	City	-2.88	0.18	0.93	0.77
13	Brick (immediate repetition)	-1.78	0.11	0.99	0.96
05	Time	-1.63	0.11	0.99	1.21
04	Day of week	-1.56	0.10	0.99	1.04
10	State	-1.44	0.10	0.92	0.59
02	Month	-1.15	0.09	0.91	0.83
26	Fold correctly	-1.15	0.09	1.04	1.12
27	Put on floor	-1.12	0.09	1.04	1.25
11	Car (immediate repetition)	-1.07	0.09	1.13	1.91
12	Vase (immediate repetition)	-0.85	0.08	0.94	0.91
07	Broad or extended place (e.g., hospital)	-0.60	0.07	1.01	1.09
08	District or Street	-0.48	0.07	1.11	1.58
06	Specific or narrow place (e.g., nursery)	-0.47	0.07	1.01	1.18
24	Immediate repetition "NO IFS, ANDS, OR BUTS"	-0.39	0.07	1.00	1.05
25	Take the paper in your right hand	-0.17	0.07	1.22	1.53
03	Year	0.07	0.06	0.82	0.58
01	Day	0.46	0.06	1.08	1.25
14	100 – 7	0.84	0.05	0.80	0.61
28	Reading and execution "CLOSE YOUR EYES"	0.87	0.05	0.88	0.81
19	Car (recall)	1.50	0.05	1.16	1.29
29	Write a phrase	1.61	0.05	0.88	0.83
21	Brick (recall)	1.86	0.05	1.36	1.60
30	Design copy	2.33	0.04	0.98	0.97
20	Vase (recall)	2.56	0.04	1.28	1.41
17	79 – 7	2.60	0.04	0.83	0.75
18	72 – 7	2.79	0.05	0.81	0.73
16	86 – 7	2.79	0.05	0.86	0.80
15	93 – 7	3.08	0.05	0.86	0.79
MEAN		0.00	0.08	0.99	1.04
S.D.		1.83	0.05	0.14	0.32

The map of items in the MMSE obtained by Rasch analysis is depicted in Figure 1. Participant ability level is shown to the left, ranging from -3 to +5, and item difficulty level is shown to the right, ranging from -3 to +3. The mean person difficulty was greater than the mean item difficulty, and thus person ability was higher than item difficulty. The easiest items were 22 (wristwatch recognition) and 23 (pen recognition) while the most difficult was 15 (93-7 subtraction).
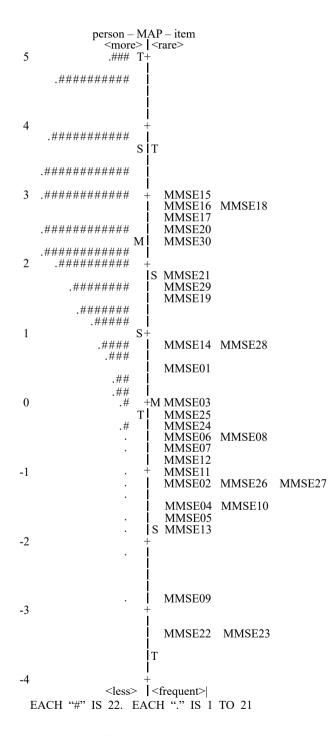
```
                person – MAP – item
                  <more> | <rare>
      5            .### T+
             .##########  |
                          |
                          |
      4                   +
             .##########  |
                         S|T
                          |
             .##########  |
                          |
      3     .###########  +    MMSE15
                          |    MMSE16  MMSE18
                          |    MMSE17
            .###########  |    MMSE20
            .###########  M|    MMSE30
            .###########   |
      2      .##########  +
                          |S MMSE21
              .########   |    MMSE29
                          |    MMSE19
               .#######   |
                .#####    |
      1                  S+
                 .####    |    MMSE14  MMSE28
                 .###     |
                          |    MMSE01
                  .##     |
                  .##     |
      0            .#    +M MMSE03
                  T|    MMSE25
                 .#     |    MMSE24
                  .     |    MMSE06  MMSE08
                  .     |    MMSE07
                       |    MMSE12
     -1            .    +    MMSE11
                  .     |    MMSE02  MMSE26  MMSE27
                  .     |
                       |    MMSE04  MMSE10
                  .     |    MMSE05
                  .    |S MMSE13
     -2                +
                  .     |
                       |
                       |
                       |
                  .     |    MMSE09
     -3                +
                       |
                       |    MMSE22  MMSE23
                      |T
                       |
     -4            <less>  | <frequent>|
       EACH "#" IS 22.  EACH "." IS 1 TO 21
```

Figure 1. Map of items and persons produced by the Rasch model for the Mini-Mental State Examination.

The DIF analysis was performed based on the variables sex, age and education (Table 3). Although this statistical treatment focuses on the items, the results for global performance of participants in the MMSE considering these demographic characteristics are presented first. Significant differences were found between the sexes ($t = 5,355$; $p < 0.001$), age groups ($F = 54,381$; $p < 0.001$) and groups stratified by educational level ($F = 359.083$; $p < 0.001$). In the first case, male participants ($M = 24.54$; $SD = 3.90$) attained better performance than women ($M = 23.67$; $SD = 4.17$). Tukey´s post hoc produced four homogenous subgroups for the variable age. MMSE scores of younger participants were higher than for older individuals: 65-69 years ($M = 24.92$; $SD = 3.48$); 70-74 years ($M = 24.07$; $SD = 3.94$); 75-79 years ($M = 23.40$; $SD = 4.15$); and ≥ 80 years ($M = 21.98$; $SD = 4.97$). With regards to education, four homogenous groups were produced *post hoc*. The elderly with more years of education had higher cognitive status: illiterate or with no formal education ($M = 19.60$; $SD = 4.32$); 1-4 years of education ($M = 23.97$; $SD = 3.54$); 5-8 years ($M = 25.34$; $SD = 3.14$); and ≥ 9 years ($M = 26.98$; $SD = 2.47$).

Male participants had more years of education ($t = 2,452$; $df = 2,730$; $p < 0.05$. Male – $M = 4.84$; $SD = 4.35$, Female – $M = 4.44$; $SD = 3.82$) and did not differ significantly ($t = 1.842$; $df = 2,732$; $p = 0.07$) from participants in the female group for age. Participants in the older age groups had fewer years of education than those in younger age groups ($F = 10.755$; $p < 0.001$). In this test of means, Tukey´s post hoc comprised three homogenous subgroups, where the group with the greatest mean years of education included individuals aged 65-69 years ($M = 5.09$) and 70-74 years (4.50). The latter age group also formed a subgroup with intermediate levels of education, together with elderly aged 75-79 years ($M = 4.23$). The third and final grouping comprised participants aged 75-79 years and ≥ 80 years and had the lowest mean ($M = 3.88$) years of education.

The regression analysis revealed that both age ($t = -12.388$; $p < 0.001$) and education ($t = 28.440$; $p < 0.001$) significantly influenced the total result in the MMSE ($F = 522.572$; $p < 0.001$). The value of $r2$ (0.277) showed that these variables explained only around 30% of the variability in the score of this measure, where the standardized coefficient beta for education (0.466) was greater than that for age (-0.202), indicating that the first variable influenced cognitive status more than the second.

Separate regression analyses for the sexes revealed similar results across all participants, i.e., education plays a fundamental role in cognitive status. Among women, both age ($t = -10.324$; $p < 0.001$) and education ($t = 24.788$; $p < 0.001$) significantly influenced the total score in the MMSE ($F = 387.133$; $p < 0.001$) and together these variables explained around 30% of cognitive status ($r2 = 0.298$), where beta for education (0.489) was also higher than that for age (-0.204). Similarly in men, age ($t = -7.485$; $p < 0.001$) and education ($t = 14.426$; $p < 0.001$) significantly influenced the total score in the exam ($F = 148.128$; $p < 0.001$). The value of $r2$ (0.247) indicates that this model explained

25% of variability of the total in the instrument. Similarly, the beta coefficient for education (0.420) was greater than that for age (-0.218).

Notably, for education, only three items (10%) in the MMSE had no DIF, more specifically, items 2 "What is the month", 12 immediate repetition of the word "vase" and 24 immediate repetition of the phrase "No ifs, ands, or buts". In the case of the variable sex, 12 items exhibited no DIF, while for age, no DIF was detected for 14 items. Only items 12 and 24 exhibited no bias for sex, age and/or education.

Table 3
DIF analysis of MMSE based on sex, age and education

Items	Sex		Age[d]		Education	
	Criteria	Group	Criteria	Group	Criteria	Group
1	3	Male	2 and 3	Older old	1, 2 and 3	Greater
2	-	-	1 and 3	Older old	-	-
3	3	Female	1 and 3	Older old	1, 2 and 3	Lower
4	1, 2 and 3	Male	1, 2 and 3	Older old	1, 2 and 3	Greater
5	-	-	1	Older old	1	Greater
6	3	Male	-	-	1 and 3	Greater
7	1, 2 and 3	Male	-	-	1, 2 and 3	Greater
8	-	-	-	-	1 and 3	Greater
9	-	-	1	Mixed	1	Lower
10	1 and 3	Female	-	-	1 and 3	Lower
11	-	-	-	-	1 and 3	Greater
12	-	-	-	-	-	-
13	-	-	-	-	1, 2 and 3	Mixed
14	1, 2 and 3	Female	2	Younger old	1 and 3	Lower
15	1, 2 and 3	Female	2	Younger old	2 and 3	Mixed
16	1, 2 and 3	Female	2	Younger old	1 and 3	Lower
17	1, 2 and 3	Female	2	Younger old	1 and 3	Lower
18	1, 2 and 3	Female	2	Younger old	1, 2 and 3	Mixed
19	1, 2 and 3	Male	-	-	1, 2 and 3	Greater
20	1, 2 and 3	Male	-	-	1 and 3	Greater
21	3	Male	2	Older old	1, 2 and 3	Greater
22	-	-	1	Mixed	1	Mixed
23	-	-	1	Mixed	1, 2 and 3	Greater
24	-	-	-	-	-	-
25	-	-	1 and 3	Younger old	1, 2 and 3	Greater
26	-	-	-	-	1	Greater
27	1, 2 and 3	Male	-	-	1 and 3	Greater
28	1, 2 and 3	Male	2	Younger old	1, 2 and 3	Lower
29	1, 2 and 3	Male	-	-	1, 2 and 3	Lower
30	1, 2 and 3	Male	-	-	1, 2 and 3	Lower

Note. [a]Criterion 1: contrast ≥ 0.50; Criterion 2: Mantel-Haenszel probability $p < 0.01$; Criterion 3: $t \geq 2.40$.

Discussion

Even though four items presented outfit, that is, probability of erroneous response to an item that should have the correct answer, the Rasch analysis revealed satisfactory parameters of fit, according to the infit and outfit indices, for the unidimensional model, indicating that the MMSE indeed assesses the latent trait of cognitive status. No hypotheses were stated to explain why four very different items presented outfit. If the measure were under construction, they could be excluded. As it is not the case, further research is recommended. Notwithstanding this limitation, fundamental requisites for others analyses were observed, such as analyses of item difficulty and DIF (Fernandes et al., 2015).

The item difficulty analysis showed that the serial sevens, three-word recall, copy design and writing a phrase were the most complex. The difficulty of the serial sevens and delayed recall was also detected by McGrory et al. (2014) in a literature review. Kim et al. (2013) found that subtraction items of the MMSE – serial sevens – had a high level of difficulty, particularly for individuals with low education.

The easiest items were recognition of objects, immediate repetition of words and most orientation items. The finding that the first two groups of items are easy corroborates the results of the literature review by McGrory et al. (2014). Carnero-Pardo (2014) believed the use of 10 orientation items in the MMSE to be excessive, questioning their true utility for the envisaged screening. Perhaps the need to include all of them in the test should be reassessed, given that they were found to be very easy by this study. The map of items confirmed the greater difficulty of the serial sevens items, specifically "93 minus 7", and the easiness of the two object recognition items.

Knowledge of the quality of the items of the MMSE based on the analysis of their difficulty is a fundamental aspect for the tester, because it provides objective information on each of the items in the instrument. For example, according to the analysis of item difficulty level proposed by McIntire and Miller (2000), which defines indices of 0-0.2 as very difficult, and 0.8-1.00 as very easy, over half of the MMSE items are of very low difficulty. These results can be explained by the screening, as opposed to diagnostic, function of the instrument. In addition, the exam was administered in a sample of elderly selected on the basis of criteria that excluded those with dementia processes, which may have led to a greater number of correct answers.

Observation of the item difficulty and the standard administration sequence of the measure shows that the items are not arranged in an increasing scale of difficulty, i.e., from the easiest to the more complex (Kline, 2015). According to the result found, the MMSE starts with an item of medium difficulty and is followed by 12 items considered easy, and subsequently by eight items considered difficult, six easy items and finally by three difficult items. Although it is not a skills assessment, which needs to adhere to a progressive scale that trains the person undergoing the assessment, the failure to sequence the items in increasing order of difficulty may impact motivation to respond to them. In addition, without underestimating the analysis of item difficulty, it is important to assess the utility of the item for measuring the construct of the test, evidenced by the discrimination power of the item, i.e., its ability to separate persons with greater and lesser ability. But this should be investigated in future research.

The item map analysis also showed that person ability level was greater than item difficulty, supporting the criticism by Spencer et al. (2013) regarding the ceiling effect of the MMSE, particularly for screening mild cognitive impairment. Nonetheless, this may have occurred because the sample contained community-dwelling elderly, supposedly without dementia. Thus, this result was somewhat expected, since there is a much higher proportion of normal, active elderly in the community than in institutions or clinics, for example. Further studies involving other samples, more specifically with different levels of cognitive decline, are needed.

Since this research, doesn't aim to modify items or the structure of the MMSE analyzed here, it is recommended that, in clinical practice, the ceiling effect be the object of attention. One possibility is to re-evaluate items that presented a lower difficulty index with other instruments that measure the same functions. Thus, Teste de Aprendizagem Auditivo-Verbal de Rey (RAVLT) (Paula & Malloy-Diniz, 2018) can, for example, be used as a measure of immediate memory, as this is an easy item in the MMSE. However, it is noted that the results of this study relate to a community sample. The ceiling effect may not be necessarily repeated in outpatient samples.

The performance of the elderly from the sample in the MMSE confirmed the results of Xie et al. (2015), showing higher scores among participants that were male, younger and with higher education. Thus, education appears to be the key variable for understanding performance in this measure, given that younger men were also those with a higher educational level. Very elderly women were part of a generation that traditionally studied less than men of the same age and were responsible for full-time care of the children and home (Almeida, Mafra, Silva, & Kanso, 2015). These individuals currently represent a high risk group for developing cognitive decline suggestive of dementia.

Given that the instrument had a good fit for the Rasch model, despite the predominance of easy items, DIF analysis of the MMSE was feasible, where variation of the items was tested according to sex, age and education. Twelve items analyzed according to sex and 14 to age exhibited no DIF. When applying only the Mantel-Haenszel criteria – the best model for considering an item as biased according to Holland and Thayer (1986) – the absence of DIF rose to 17 items for sex and 21 items for age. The items with DIF for sex included five items from the serial sevens, showing bias according to three analysis criteria. In this case, the favored group was women, in contrast with the results of Jones and Gallo (2002).

Among the items exhibiting DIF for age, five orientation items favored the older old, who, in the present study sample, had less education. The orientation domain is less dependent on education than other cognitive screening tasks (Xavier, D'Orsi, Sigulem, & Ramos, 2010). Decline in this ability represents a marker of cognitive problems (Xavier et al., 2010) and is more frequent in the older old.

The younger old, and thus more educated, were favored in items dependent on education, such as serial sevens. However, this same group of items, when analyzed according to education, favored the less educated or mixed extracts. In fact, four items (immediate repetition of the word brick, 93 minus seven, 72 minus seven and identifying a watch) favored both low and high education groups concomitantly. These results differ from those of Jones and Gallo (2002), who found that participants with low education were most likely to err in the first item of the serial sevens, repeat phrase, write a sentence, name season of the year (specific command from the version analyzed by the authors) and copy design.

Therefore, DIF of the MMSE based on education, owing to its inconsistent pattern, warrants further studies. Among other aspects, it is important to consider limitations in the use of years of education as a sole criterion for forming education groups, in view of the heterogeneity of Brazilian schools (Melo & Barbosa, 2015). This disparity occurs, for example, between systems (public or private) and regions (South, Southeast, North, Northeast or Center-West) in Brazil. In addition, changes in the education system from one decade to another marked by events in history must also be considered (Malloy-Diniz, Fuentes, & Cosenza, 2013), such as the Brazilian military dictatorship, which negatively impacted the formal education of cohorts of different ages.

The DIF exam considering education warrants special attention, given that this is a key domain for cognitive tests (Apolinario et al., 2014) and the fact that, in regression analysis, the variable predicts performance in the MMSE more than age, both for women and men. The only three items not exhibiting DIF apparently had no direct relationship with education, in that they only require immediate repetition of a word, phrase or awareness of the month when the assessment takes place. All other items, i.e., 27 items of the MMSE, were found to be biased according to at least one analysis criterion. This finding corroborates the international literature (Ramirez, Teresi, Holmes, Gurland, & Lantigua, 2006) which has detected substantial DIF in the MMSE. This is a cause for concern because the biased items should ideally be removed from the measure (McGrory et al., 2014), although that would render the MMSE version investigated in the present study unusable. DIF has been attributed by some authors to differences in translation of the instrument into different languages (Ramirez et al., 2006). However, this hypothesis needs to be corroborated, since DIF has also been confirmed in studies involving the original instrument (Jones & Gallo, 2002).

The number of items with DIF varied depending on the criteria adopted. When applying all three criteria at once, the number of items with DIF due to education was 13, whereas applying only the Mantel-Haenszel criterion (Holland & Thayer, 1986) the number was 14 items. Nevertheless, this is a high figure, accounting for almost 50% of the MMSE items, which would disfigure the test were they to be excluded.

It is important to point out that DIF based on education indicates that a given group has a higher or lower likelihood of providing a correct answer to the item and not that one group or another has a higher rate of correct answers. In the present study, groups with high education had a greater likelihood of providing correct answers in half of the items of the MMSE and, in regression analysis, education strongly predicted cognitive status. According to Yassuda et al. (2013), education is part of the group of dimensions that explains a large proportion of variability in the cognitive performance of the elderly. The authors emphasized that functions such as calculation, reaction time, long-term verbal memory and also motor tasks are highly influenced by education.

The analysis of the MMSE with IRT provided some results that would not be obtained with other strategies. They are summarized below. It was identified that people's ability was greater than the difficulty of the items, evidencing the ceiling effect. Inspection of DIF detected that more items were biased by education than by sex or age, where individuals with high and low education can be favored. Therefore, the MMSE for screening cognitive decline in community-dwelling elderly should be used with caution in light of evidence that the measure has a ceiling effect and contains a large number of biased items, particularly by education. It is noteworthy that this measure is widely used and extensively studied. In view of the controversies surrounding its psychometric properties, further similarly rigorous analyses should be carried out. Future analyses should employ not only IRT but also Classical Test Theory, given that these have proven to be complementary.

Furthermore, several possible limitations of the present investigation should be mentioned, especially those related to sampling (external validity) and educational status (internal validity). The FIBRA sample did not include elderly people from rural areas and with indicators of significant cognitive decline, which limits the external validity of this research. Internal validity was limited by the fact that years of education constituted the only information about educational status collected. In Brazil, quality of education is a variable equally important to the amount of time people spend in school, and that was not measured.

Among other implications, this study suggests that the MMSE and its cut off points should be used carefully. For prudence, this measure needs to be administered in conjunction with other instruments that evaluate the same construct or at least related constructs, such the Clock Drawing Test (Malloy-Diniz, Fuentes, Mattos, & Abreu, 2018), even if the goal is only to screen cognitive status.

Another implication of this research concerns the need to make the new standard versions of the MMSE - MMSE-2: Brief Version (MMSE-2:BV) and MMSE-2: Expanded Version (MMSE-2:EV) - (Folstein, Folstein, White, & Messer, 2018) - available to the Brazilian scientific community and health professionals. It is valid that these new Brazilian versions will also be submitted, shortly, to a review of their psychometric properties through IRT and the classical theory of tests.

References

Almeida, A. V., Mafra, S. C. T., Silva, E. P., & Kanso, S. (2015). A feminização da velhice: Em foco as características socioeconômicas, pessoais e familiares das idosas e o risco social [The feminization of old age: A focus on the socioeconomic, personal and family characteristics of the elderly and the social risk]. *Textos & Contextos, 14*(1), 115-131. doi:10.15448/1677-9509.2015.1.19830

Apolinario, D., Mansur, L. L., Carthery-Goulart, M. T., Brucki, S. M., & Nitrini, R. (2014). Detecting limited health literacy in Brazil: Development of a multidimensional screening tool. *Health Promotion International* 29(1), 5-14. doi:10.1093/heapro/dat074

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.

Brucki, S. M. D., Nitrini, R., Caramelli, P., Bertolucci, P. H. F., & Okamoto, I. H. (2003). Sugestões para o uso do mini-exame do estado mental no Brasil [Suggestions for utilization of the mini-mental state examination in Brazil]. *Arquivos de Neuro-Psiquiatria, 61*(3B), 777-781. doi:10.1590/S0004-282X2003000500014

Carnero-Pardo, C. (2014). Should the mini-mental state examination be retired? *Neurologia, 29*(8), 473-481. doi:10.1016/j.nrl.2013.07.003

Castro Costa, E., Dewey, M. E., Uchôa, E., Firmo, J. O., Lima-Costa, M. F., & Stewart, R. (2014). Construct validity of the mini mental state examination across time in a sample with low-education levels: 10-year follow-up of the Bambui Cohort Study of Ageing. *International Journal of Geriatric Psychiatry, 29*(12), 1294-1303. doi:10.1002/gps.4113

Draba, R. E. (1977). *The identification and interpretation of item Bias (MESA Memorandum No. 25)*. Chicago, IL: The University of Chicago. Retrieved from http://www.rasch.org/memo25.htm

Drasgow, F., & Lissak, R. I. (1983). Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. *Journal of Applied Psychology, 68*(3), 363-373. doi:10.1037/0021-9010.68.3.363

Fernandes, D. C., Prieto, G., & Delgado, A. R. (2015). Construction and analysis by the Rasch model of two computerized recognition memory tests. *Psicologia: Reflexão e Crítica, 28*(1), 49-60. doi:10.1590/1678-7153.201528106

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (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. doi:10.1016/0022-3956(75)90026-6

Folstein, M. F., Folstein, S. E., White, T., & Messer, M. A. (2018). *MMSE-2®: Mini-Mental State Examination* (2nd ed.). Lutz, FL: PAR.

Holland, P. W., & Thayer, D. T. (1986). Differential item functioning and the Mantel-Haenszel procedure. *ETS Research Report Series, 1986*(2), i-24. doi:10.1002/j.2330-8516.1986.tb00186.x

Ideno, Y., Takayama, M., Hayashi, K., Takagi, H., & Sugai, Y. (2012). Evaluation of a Japanese version of the mini-mental state examination in elderly persons. *Geriatrics & Gerontology International, 12*(2), 310-316. doi:10.1111/j.1447-0594.2011.00772.x

Jones, R. N., & Gallo, J. J. (2000). Dimensions of the mini-mental state examination among community dwelling older adults. *Psychological Medicine, 30*(03), 605-618. doi:10.1017/S0033291799001853

Jones, R. N., & Gallo, J. J. (2002). Education and sex differences in the mini-mental state examination effects of differential item functioning. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 57*(6), P548-P558. doi:10.1093/geronb/57.6.P548

Kim, J. S., Won, C. W., Kim, B. S., & Choi, H. R. (2013). Predictability of various serial subtractions on global deterioration scale according to education level. *Korean Journal of Family Medicine, 34*(5), 327-333. doi:10.4082/kjfm.2013.34.5.327

Kline, P. (2015). *A handbook of test construction: Introduction to psychometric design*. New York, NY: Routledge.

Linacre, J. M. (2002). What do infit and outfit, mean-squared and standardized mean? *Rasch Measurement Transactions, 16*(2), 878. Retrieved from http://www.rasch.org/rmt/rmt162f.htm

Linacre, J. M., & Wright, B. D. (2009). *WINSTEPS: Multiple-choice, rating scale, and partial credit Rasch analysis* [Computer Software]. Chicago, IL: MESA Press.

Malloy-Diniz, L. F., Fuentes, D., Mattos, P., & Abreu, N. (2018). *Avaliação neuropsicológica [Neuropsychological assessment]*. Porto Alegre, RS: Artmed.

Malloy-Diniz, L. F., Fuentes, D. & Cosenza, R. M. (2013). *Neuropsicologia do Envelhecimento: Uma abordagem multidimensional* [Neuropsychology of Aging: A multidimensional approach]. Porto Alegre, RS: Artmed.

McGrory, S., Doherty, J. M., Austin, E. J., Starr, J. M., & Shenkin, S. D. (2014). Item response theory analysis of cognitive tests in people with dementia: A systematic review. *BMC Psychiatry, 14*, 47. doi:10.1186/1471-244X-14-47

McIntire, S. A., & Miller, L. A. (2000). *Foundations of psychological testing: A practical approach*. Boston, MA: McGraw-Hill.

Melo, D. M., & Barbosa, A. J. G. (2015). O uso do mini-exame do estado mental em pesquisas com idosos no Brasil: Uma revisão sistemática [Use of the mini-mental state examination in research on the elderly in Brazil: A systematic review]. *Ciência & Saúde Coletiva, 20*(12), 3865-3876. doi:10.1590/1413-812320152012.06032015

Melo, D. M., Barbosa, A. J. G., & Neri, A. L. (2017). Miniexame do estado mental: Evidências de validade baseadas na estrutura interna [Minimental state examination: Validity evidence based on internal structure]. *Avaliação Psicológica, 16*(2), 161-168. doi:10.15689/AP.2017.1602.06

Neri, A. L., Yassuda, M. S., Araújo, L. F., Eulálio, M. C., Cabral, B. E., Siqueira, M. E. C., ... Moura, J. G. A. (2013). Metodologia e perfil sociodemográfico, cognitivo e de fragilidade de idosos comunitários de sete cidades brasileiras: Estudo FIBRA [Methodology and social, demographic, cognitive, and frailty profiles of community-dwelling elderly from seven Brazilian cities: The FIBRA Study]. *Cadernos de Saúde Pública, 29*(4), 778-792. doi:10.1590/S0102-311X2013000800015

Nitrini, R., Caramelli, P., Bottino, C. M. C., Damasceno, B. P., Brucki, S. M. D., & Anghinah, R. (2005). Diagnóstico de doença de Alzheimer no Brasil: Avaliação cognitiva e funcional. Recomendações do Departamento Científico de Neurologia Cognitiva e do Envelhecimento da Academia Brasileira de Neurologia [Diagnosis of Alzheimer's disease in Brazil: Cognitive and functional evaluation. Recommendations of the Scientific Department of Cognitive Neurology and Aging of the Brazilian Academy of Neurology]. *Arquivos de Neuro-Psiquiatria, 63*(3A), 720-727. doi:10.1590/S0004-282X2005000400034

Pasquali, L. (2013). *Psicometria: Teoria dos testes na psicologia e na educação* [Psychometrics: Test theory in psychology and education] (5th ed.). Petrópolis, RJ: Vozes.

Paula, J. J., & Malloy-Diniz, L. F. (2018). Teste de Aprendizagem Auditivo-Verbal de Rey (RAVLT) [The Rey Auditory-Verbal Learning Test (RAVLT)]. São Paulo, SP: Vetor.

Ramirez, M., Teresi, J. A., Holmes, D., Gurland, B., & Lantigua, R. (2006). Differential Item Functioning (DIF) and the Mini-Mental State Examination (MMSE): Overview, sample, and issues of translation. *Medical Care, 44*(11 Suppl. 3), S95-S106. doi:10.1097/01.mlr.0000245181.96133.db

Sisto, F. F. (2006). O funcionamento diferencial dos itens [Differential item functioning]. *Psico USF, 11*(1), 35-43. doi:10.1590/S1413-82712006000100005

Spencer, R. J., Wendell, C. R., Giggey, P. P., Katzel, L. I., Lefkowitz, D. M., Siegel, E. L., & Waldstein, S. R. (2013). Psychometric limitations of the mini-mental state examination among nondemented older adults: An evaluation of neurocognitive and magnetic resonance imaging correlates. *Experimental Aging Research, 39*(4), 382-397. doi:10.1080/0361073X.2013.808109

Xavier, A. J., D´Orsi, E., Sigulem, D., & Ramos, L. R. (2010). Time orientation and executive functions in the prediction of mortality in the elderly: Epidoso study. *Revista de Saúde Pública, 44*(1), 148-158. doi:10.1590/S0034-89102010000100016

Xie, H., Zhang, C., Wang, Y., Huang, S., Cui, W., Yang, W., ... Huo, Y. (2015). Distinct patterns of cognitive aging modified by education level and gender among adults with limited or no formal education: A normative study of the mini-mental state examination. *Journal of Alzheimer's Disease, 49*(4), 961-969. doi:10.3233/JAD-143066

Yassuda, M. S., Viel, T. A., Silva, T. B. L., & Albuquerque, M. S. (2013). Memória e envelhecimento: Aspectos cognitivos e biológicos [Memory and aging: Cognitive and biological aspects]. In E. V. Freitas & L. Py (Eds.), *Tratado de geriatria e gerontologia* (3rd ed., pp. 1477-1485). Rio de Janeiro, RJ: Guanabara Koogan.

Zisberg, A., Zysberg, L., Young, H. M., & Schepp, K. G. (2009). Trait routinization, functional and cognitive status in older adults. *The International Journal of Aging and Human Development, 69*(1), 17-29. doi:10.2190/AG.69.1.b

Denise Mendonça de Melo is a professor at Centro Universitário UniAcademia, Juiz de Fora-MG, Brazil.

Altemir José Gonçalves Barbosa is a professor at Universidade Federal de Juiz de Fora, Juiz de Fora-MG, Brazil.

Nelimar Ribeiro de Castro is a professor at Universidade de Viçosa, Viçosa-MG, Brazil.

Anita Liberalesso Neri is a professor at Universidade Estadual de Campinas, Campinas-SP, Brazil.

Authors' Contribution:
All authors made substantial contributions to the conception and design of this study, to data analysis and interpretation, and to the manuscript revision and approval of the final version. All the authors assume public responsibility for the content of the manuscript.