

Investigating the Measurement Precision of the Montreal Cognitive Assessment (MoCA) for Cognitive Screening in Parkinson's Disease Through Item Response Theory

[Pedro Renato de Paula Brandão](#) , Danilo Assis Pereira , Brenda Hanae Bentes Koshimoto , Vanderci Borges , Henrique Ballalai Ferraz , Artur Francisco Schumacher-Schuh , [Carlos Roberto de Mello Rieder](#) , Maira Rozenfeld Olchik , Ignacio F. Mata , [Vitor Tumas](#) , [Bruno Lopes Santos-Lobato](#) *

Posted Date: 20 December 2024

doi: 10.20944/preprints202412.1815.v1

Keywords: Montreal Cognitive Assessment (MoCA); cognition; cognitive screening; Parkinson's disease; item response theory; graded response model; differential item functioning; validity; reliability; factor analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Investigating the Measurement Precision of the Montreal Cognitive Assessment (MoCA) for Cognitive Screening in Parkinson's Disease Through Item Response Theory

Pedro Renato de Paula Brandão ¹, Danilo Assis Pereira ², Brenda Hanae Bentes Koshimoto ³, Vanderci Borges ⁴, Henrique Ballalai Ferraz ⁴, Artur F. Schumacher Schuh ⁵, Carlos Roberto de Mello Rieder ⁶, Maira Rozenfeld Olchik ⁵, Ignacio F. Mata ⁷, Vitor Tumas ⁸ and Bruno Lopes Santos-Lobato ^{9,*}

¹ Medical Department, Chamber of Deputies and Neurology Unit, Hospital Sírío Libanês (Brasília), Brasília, DF, BR

² Brazilian Institute of Neuropsychology and Cognitive Sciences (IBNeuro), Brasília, DF, BR

³ Laboratory of Human and Medical Genetics, Federal University of Pará, Belém, PA, BR

⁴ Universidade Federal de São Paulo, Escola Paulista de Medicina, Disciplina de Neurologia, São Paulo, SP, BR

⁵ Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, BR

⁶ Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, BR

⁷ Genomic Medicine Institute, Cleveland Clinic Foundation: OH, OH, US

⁸ Department of Neuroscience and Behavior Sciences, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil

⁹ Laboratory of Experimental Neuropathology, Federal University of Pará, Belém, PA, BR

* Correspondence: bruls4@ufpa.br

Abstract: Background: The Montreal Cognitive Assessment (MoCA) is widely used to evaluate global cognitive function in older Brazilian adults. However, concerns persist regarding its applicability in non-homogeneous socio-demographic groups. This study scrutinizes the Brazilian version of the MoCA, focusing on its measurement properties in a diverse cohort of patients with Parkinson's disease (PD). **Purpose:** This study examined the psychometric properties of the Brazilian Portuguese MoCA in a heterogeneous sample of PD patients using item response theory (IRT) methods. **Material and Methods:** In a multicenter cross-sectional study, 484 PD patients aged 26-90 years (mean \pm SD, 59.9 \pm 11.1 years), with disease durations ranging from 1 to 35 years (mean \pm SD, 8.7 \pm 5.4 years), underwent MoCA testing. IRT analyses, including the Graded Response Model, evaluated item parameters, such as difficulty (location) and discrimination. Differential item functioning was analyzed vis-à-vis age and education using multiple indicators multiple causes (MIMIC) modeling. **Results:** The MoCA exhibited essential unidimensionality and satisfactory model fit. Attention and naming demonstrated high discrimination. Orientation and naming items were less challenging. Multiple domains showed differential item functioning related to age and education, underscoring the necessity of considering background characteristics when interpreting total scores. **Conclusion:** This study enriches validity evidence for the MoCA in PD by providing a detailed analysis of its measurement properties and sources of score bias. Tailoring test content and norms based on education and establishing computerized scoring algorithms leveraging item parameters may optimize the tool's reliability and fairness. Refinements to mitigate differential item functioning could enable precise cognitive screening across diverse socio-demographic backgrounds.

Keywords: Montreal Cognitive Assessment (MoCA); cognition; cognitive screening; Parkinson's disease; item response theory; graded response model; differential item functioning; validity; reliability; factor analysis

1. Background

The Montreal Cognitive Assessment (MoCA) was designed as a brief cognitive screening test for identifying milder forms of cognitive impairment in the elderly population [1]. It has been extensively studied in Brazil and elsewhere, affirming its utility in assessing cognitive impairment in Parkinson's disease (PD) [2–4]. Despite its broad application, the MoCA faces challenges, including potential item biases and overall test structure. Originally developed in English, it has been adapted into several languages, including Brazilian Portuguese [5].

The effectiveness of MoCA in identifying cognitive impairment may vary, particularly in populations with lower education levels. This could lead to incorrect diagnoses or missed cases. Data from a normative study conducted in São Paulo, for example, has revealed that MoCA effectiveness in detecting cognitive impairment among Brazilian seniors varies significantly with age and education [6]. It was found that MoCA scores differ across cognitively normal individuals, those with “cognitive impairment no dementia” (CIND), and dementia patients, necessitating the adjustment of MoCA scores based on educational levels. The study notably pointed out the MoCA's limited effectiveness in detecting CIND in individuals with low educational levels despite its utility in diagnosing dementia in more educated groups.

The need to vary cutoff scores according to educational levels in different populations presents a significant challenge in validating cognitive assessment tools like MoCA. This variability complicates the efficiency of screening processes and increases the likelihood of human error. Research conducted in various countries highlights this issue, with recommendations for a range of cutoff scores tailored to different education and age groups [6–8]. This lack of universally applicable cutoff scores underscores the limitations of a one-size-fits-all approach and emphasizes the need for population-specific norms and scoring guidelines.

Central to this challenge is item bias, where certain test item characteristics, not directly related to the cognitive construct being measured, affect responses differently among groups [9]. Conventional scoring methods, which often treat all test items uniformly and may adjust scores for educational background, do not adequately tackle the influence of these factors on individual test items. This situation underscores the necessity for more refined methods to address such biases in cognitive assessments.

Item Response Theory (IRT) offers a valuable statistical framework for analyzing psychological and educational tests, playing a crucial role in evaluating test validity [10,11]. IRT allows for a detailed assessment of the quality of a test's items, offering insights into individual item difficulty, test-taker ability, and overall test performance. This information is instrumental in identifying the strengths and weaknesses of a test guiding decisions regarding which items should be included or excluded. By employing IRT, researchers can ensure that a test accurately measures the intended construct and yields reliable results, thus enhancing the overall effectiveness and precision of cognitive assessments [12]. Moreover, IRT provides a means to assess and mitigate item bias, allowing for developing of more equitable and culturally appropriate cognitive screening tools.

The MoCA has been extensively examined using the IRT framework in various contexts. For instance, a study conducted in Taiwan revealed that executive function tasks—including attention, working memory, and abstract thinking items—were more effective than language and visual-spatial subscales in distinguishing individuals with mild and moderate cognitive impairment [13]. Conversely, orientation items proved less effective in differentiating between cognitively unimpaired individuals and those with mild cognitive impairment (MCI) and dementia. Similarly, in Portugal, Rasch-based IRT models demonstrated a strong overall fit between MoCA items and individual values, underscoring the test's scalability and appropriateness [14]. These findings suggest that the discriminative power of MoCA items can vary significantly across different cognitive domains and levels of impairment, highlighting the necessity for a comprehensive understanding of the test's psychometric properties. However, despite these international insights, the analysis of MoCA using the IRT framework in Brazil remains limited.

Applying the Graded Response Model (GRM), one of the IRT models for polytomous items, to the MoCA could improve the estimation of individual cognitive abilities [15,16]. GRM can identify

the most informative items, potentially resulting in a more concise MoCA. Additionally, GRM may offer insights into the psychometric properties of test items, helping to refine the test by pinpointing poorly performing or redundant items. Utilizing GRM, researchers can develop a more efficient and precise cognitive screening tool tailored to the target population's specific needs.

This study aims to analyze the performance of specific test domains and the overall scores of the Brazilian version of the MoCA using IRT in varying degrees of PD-related cognitive dysfunction and conduct a Multiple Indicators Multiple Causes model approach to investigate the impact of educational level and age on test performance. By examining the MoCA's psychometric properties in a diverse PD sample, this study seeks to contribute to the growing body of evidence on the test's validity and reliability while also exploring potential sources of item bias. The findings could inform the development of more accurate and equitable cognitive screening strategies for PD patients in Brazil and beyond.

2. Materials and Methods

2.1. Study Design and Ethics

This multicentric observational cross-sectional study was approved by the local Research Ethics Committees at each study site, upholding the ethical principles of the Declaration of Helsinki (including subsequent amendments) regarding research involving human subjects. The study was conducted between May 2007 and July 2022 across five Brazilian centers: Porto Alegre (RS), Ribeirão Preto (SP), Belém (PA), Brasília (DF), and São Paulo (SP). The data originated from two sources: the LARGE study [17] conducted in Belém, Ribeirão Preto, São Paulo, and Porto Alegre, and the Brasília Parkinson Study [18].

2.2. Participant Selection and Inclusion Criteria

A non-probabilistic convenience sample of 484 patients with idiopathic PD was recruited from movement disorder clinics. Patients were enrolled through direct invitation by their neurologists. Inclusion criteria stipulated a minimum age of 18, Brazilian Portuguese as first language, and meeting PD diagnostic criteria established by the Queen Square Brain Bank [19]. Diagnosis involved medical history, physical examinations, diagnostic testing, and imaging (e.g. MRI, transcranial ultrasound, dopamine transporter scans). Comorbid neurological disorders did not warrant exclusion.

2.3. Instrument

The MoCA is a 10-15 minute screening tool that evaluates seven cognitive domains: visuospatial/executive, naming, attention, language, abstraction, memory, and orientation [1]. Performance in each domain is assessed using specified tasks: Visuospatial/Executive (0-5 points) - trail making, cube copying, clock drawing; Naming (0-3 points) - animal naming; Attention (0-6 points) - digit span, Go-No-Go, serial 7s subtraction; Language (0-3 points) - phrase repetition, phonemic fluency; Abstraction (0-2 points) - similarity identification; Memory (0-5 points) - delayed recall; Orientation (0-6 points) - place, city, date. Total scores range from 0-30, with lower scores indicating greater impairment. Individuals with <13 years of education receive a 1-point bonus.

2.4. Data Analysis

For descriptive statistics, categorical variables were summarized with frequency (n) and percentages (%). Continuous variables were assessed for normality using the Shapiro-Wilk test. Normally distributed variables were reported with means and standard deviations (SD), while non-normally distributed variables were described with medians and interquartile ranges (IQR). Statistical analyses were conducted in R (v4.3.2), using a significance level (α) of 0.05 for all two-sided tests.

2.4.1. Dimensionality and Model Fit

Unidimensionality is a fundamental assumption in IRT. Exploratory factor analysis (EFA) with weighted least squares (WLS) was used to examine dimensionality. Factors were rotated using Promax rotation, and parallel analysis guided factor retention decisions. Model fit was evaluated using the root mean square error of approximation (RMSEA; good fit ≤ 0.08 , 90% CI ≤ 0.10) and Tucker-Lewis index (TLI; good fit ≥ 0.90 , excellent fit ≥ 0.95) [20]. These fit indices provide a quantitative assessment of how well the model aligns with the observed data, with RMSEA values closer to 0 and TLI values closer to 1 indicating better fit.

2.4.2. Graded Response Model (GRM)

The GRM was used for IRT analyses. As originally proposed by Samejima (1968), the GRM posits that as an individual's latent ability increases, their likelihood of endorsing a higher item response category increases [16]. A key advantage is its ability to model polytomous items with multiple response options. The GRM characterizes items using difficulty and discrimination parameters. Difficulty parameters quantify item endorsability at varying ability levels, while discrimination indexes an item's ability to differentiate individuals across the latent trait. The GRM was selected given its flexibility in modeling local item dependencies within our polytomous item data. Analyses were conducted using the *mirt* and *ggmirt* packages in R [21,22], which provide a comprehensive suite of tools for estimating and visualizing IRT models.

2.4.3. Test Reliability

Two IRT-based indices were examined - conditional and marginal reliability (r_{xx}). Conditional reliability measures the reliability at precise levels along the latent trait, quantifying the consistency of estimated scores among individuals with equal standing on the latent variable. Marginal reliability integrates conditional reliability across the full distribution of latent trait scores, providing an overall reliability estimate. These reliability indices offer an understanding of the MoCA's precision across different cognitive ability levels, complementing traditional internal consistency measures.

2.4.4. Multiple Indicators Multiple Causes Modeling

A Multiple Indicators Multiple Causes (MIMIC) model was applied to examine the effects of age and education on cognitive performance across seven MoCA domains, using the *lavaan* package in R [23]. The model specified age and education as exogenous predictors and the MoCA domains as endogenous variables. Regression coefficients (β) quantified the relationships between predictors and domains, with positive values indicating positive associations and negative values suggesting inverse relationships. Statistical significance ($p < 0.05$) of these coefficients provided evidence for differential item functioning (DIF) related to age and education. Model fit was assessed using CFI, TLI, and RMSEA. This approach allowed for the examination of differential item/test functioning among age and education groups, helping to elucidate background influences on domain functioning and improve the test's fairness.

3. Results

3.1. Participants

The study sample consisted of 484 patients diagnosed with PD. Participants ranged from 26 to 90 years old (mean \pm SD: 59.9 ± 11.1 years). PD duration varied substantially, spanning 1 to 35 years (mean \pm SD: 9.1 ± 5.8 years). The average educational attainment was 9.1 ± 5.4 years. The majority of participants were male (57.6%), aligning with the known higher PD prevalence among males.

Table 1 summarizes the demographic and clinical characteristics, including education levels, PD duration categories, gender proportions, and MoCA scores. The mean total MoCA score for the full cohort was 20.61 ± 5.80 (SD).

Table 1. Participants' demographic features.

Variable	Total sample (n=484)	
Sex (male), % (n)	57.6% (279)	
Age, y, mean ± sd	59.9 ± 11.1	
Range (min-max)	26-90	
PD duration, y, mean ± sd	8.7 ± 5.4	
Range (min-max)	1-32	
Age at PD onset, % (n)		
	< 21 y	0.61% (3)
	21-49 y	42.7% (207)
	50+ y	56.6% (274)
Education level, % (n)		
	Elementary School (1-4 y)	26.2% (127)
	Middle School (5-9 y)	24.8% (120)
	High school (10-12y)	26.4% (128)
	Higher education (13+ y)	22.5% (109)
Study site, % (n)		
	São Paulo-SP	17.1% (83)
	Ribeirão Preto-SP	18.2% (88)
	Belém-PA	24.8% (120)
	Porto Alegre-RS	25.6% (124)
	Brasília-DF	14.3% (69)

Legend: PD - Parkinson's disease; y = years; sd = standard deviation; SP - São Paulo, PA - Pará, RS - Rio Grande do Sul; DF - Distrito Federal. Min - minimum; Max - Maximum.

Table 2 shows score distribution across MoCA's seven cognitive domains. Naming and Orientation exhibited ceiling effects, with 70% and 80.4% of participants scoring maximum points, respectively. Visuospatial/Executive scores were evenly distributed, peaking at 5 (21.5%). Attention centered around 6 (26%). Language and Abstraction had modal scores of 2 (36% and 39.3%). Memory showed wide distribution, with 0 most frequent (30%). These patterns indicate varying cognitive performance across domains, with Naming and Orientation showing strongest performance, while Visuospatial/Executive and Memory displayed more variability. This suggests differential sensitivities to PD-related cognitive impairment across MoCA domains, emphasizing the need for domain-specific score interpretation.

Table 2. Response frequencies for each MoCA Domain.

MoCA domain	0	1	2	3	4	5	6
<i>Visuospatial/executive (0-5)</i>	16.1%	14.0%	10.5%	18.2%	19.6%	21.5%	-
<i>Naming (0-3)</i>	2.7%	5.2%	22.1%	70.0%	-	-	-
<i>Attention (0-6)</i>	3.1%	4.8%	10.1%	15.1%	15.5%	25.4%	26.0%
<i>Language (0-3)</i>	14.9%	23.8%	36.0%	25.4%	-	-	-
<i>Abstraction (0-2)</i>	30.4%	30.4%	39.3%	-	-	-	-
<i>Memory (0-5)</i>	30.0%	14.7%	18.4%	16.3%	12.8%	7.9%	-
<i>Orientation (0-6)</i>	0.6%	1.4%	1.2%	1.2%	3.5%	11.6%	80.4%

3.2. IRT Graded Response Model

3.2.1. Dimensionality Analysis, Model Fit, and Reliability Assessment

The correlation matrix's factorability was confirmed by Bartlett's test ($\chi^2 = 808.03$, $df = 21.0$, $p < 0.001$) and KMO (0.867). Parallel analysis supported a unidimensional model (first eigenvalue: 3.205 vs. 1.176 in simulated data). High factor loadings across MoCA domains validated its use as a unidimensional cognitive assessment tool. EFA indicated satisfactory model fit ($\chi^2 = 26.106$, $df = 14$, $p = 0.025$; RMSEA = 0.042, 90% CI: 0.015-0.067; TLI = 0.977), reinforcing the model's robustness.

Figure 1 shows MoCA's conditional reliability peaks at moderate trait levels (θ between -1 and 1), indicating optimal precision for average cognitive abilities. Reliability decreases at cognitive spectrum extremes. The marginal reliability coefficient ($r_{xx} = 0.806$) demonstrates good overall precision. These findings support MoCA's suitability for assessing cognitive function in PD populations, especially for mild to moderate impairments.

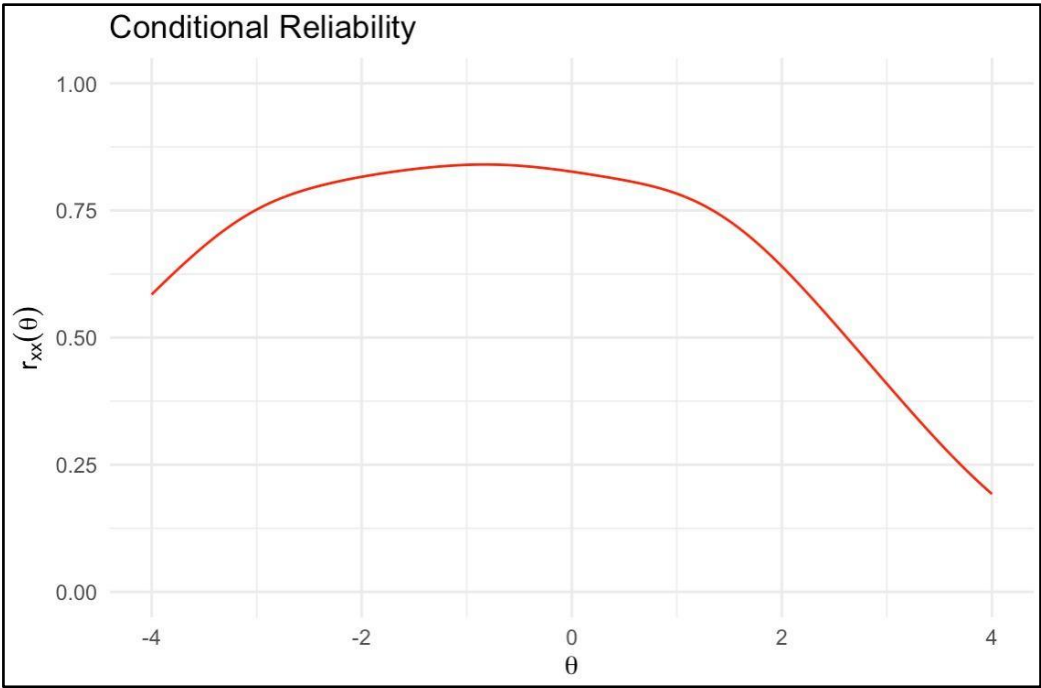


Figure 1. - Conditional reliability curve for the Montreal Cognitive Assessment (MoCA) based on item response theory analysis.

Legend. The x-axis shows the latent cognitive trait (θ) continuum. The y-axis displays reliability coefficients ($r_{xx}(\theta)$) at corresponding trait levels. The peak conditional reliability indicates the cognitive ability range at which the MoCA provides optimal measurement precision. As θ diverges from this peak, score reliability declines, reflecting variable consistency across the spectrum of cognitive faculties indexed by the MoCA.

3.2.2. Evaluation of Model and Item Fit

The GRM exhibited a robust overall fit to the MoCA data ($M2 = 12.94$, degrees of freedom (df) = 14, $p = 0.53$; RMSEA = 0.000 [90% CI: 0.000-0.041]; standardized root mean square residual [SRMSR] = 0.040), affirming its suitability for this analysis. Table 3 details the item-level fit statistics for each cognitive domain, all demonstrating good concordance with the model.

Specifically, the Visuospatial/Executive domain aligned well with the model's assumptions ($S-\chi^2 = 40.918$, $df = 50$, RMSEA = 0.000, $p = 0.817$). Similarly, acceptable fits were observed in the Naming ($S-\chi^2 = 34.498$, $df = 27$, RMSEA = 0.024, $p = 0.152$), Attention ($S-\chi^2 = 69.589$, $df = 50$, RMSEA = 0.028, $p = 0.035$), Language ($S-\chi^2 = 41.125$, $df = 41$, RMSEA = 0.003, $p = 0.465$), Abstraction ($S-\chi^2 = 21.000$, $df = 26$, RMSEA = 0.000, $p = 0.742$), Memory ($S-\chi^2 = 49.590$, $df = 49$, RMSEA = 0.005, $p = 0.450$), and Orientation domains ($S-\chi^2 = 34.239$, $df = 29$, RMSEA = 0.019, $p = 0.231$). These findings strongly support GRM-based analysis of the psychometric properties of the MoCA in a PD population.

3.2.3. Analysis of Item and Person Parameters

The GRM analysis was employed to evaluate how effectively each domain of the MoCA discriminates between levels of cognitive abilities and to assess item difficulty levels (as detailed in Table 3). Notably, the Attention and Naming domains exhibited the highest discrimination values, indicating their strong capability to differentiate individuals across various latent trait levels. The particularly high discrimination value in the Attention domain (1.985) highlights its sensitivity in identifying subtle cognitive impairment.

On the other hand, the Memory domain demonstrated the lowest discrimination (1.265), implying a relatively weaker capacity to differentiate between varying levels of cognitive abilities.

This lower discrimination in Memory could affect its effectiveness, particularly in identifying early stages of cognitive decline.

The analysis of location thresholds provided insights into the range of item difficulties. The Orientation domain was marked by the lowest difficulty levels (thresholds: -4.135 to -1.225), suggesting that these items are more easily endorsed across a range of cognitive abilities. Conversely, the Memory domain presented the highest difficulty (thresholds: -0.860 to 2.402), requiring a higher level of cognitive functioning for successful performance. These findings highlight MoCA's capabilities in capturing diverse cognitive performance profiles with varying levels of sensitivity and challenge across its different domains. Identifying highly discriminating domains (Attention and Naming) and those with a broader range of difficulty (Memory) can guide the refinement of the MoCA for optimal use in PD populations.

Table 3. Item parameters: discrimination and location (thresholds).

MoCA domain	Discrimination	Location					
	(a)	b1	b2	b3	b4	b5	b6
Visuospatial/executive (0-5)	1.473	-1.454	-0.677	-0.042	0.963	-	-
Naming (0-3)	1.740	-2.857	-2.013	-0.704	-	-	-
Attention (0-6)	1.985	-2.607	-1.922	-1.185	-0.538	- 0.018	0.854
Language (0-3)	1.331	-1.694	-0.441	1.064	-	-	-
Abstraction (0-2)	1.671	-0.722	0.397	-	-	-	-
Memory (0-5)	1.265	-0.860	-0.203	0.567	1.373	2.402	-
Orientation (0-6)	1.550	-4.135	-3.237	-2.851	-2.577	- 2.072	-1.225

Figure 2 presents Item Characteristic Curves (ICCs) for each MoCA domain, illustrating the relationship between latent cognitive abilities (θ) and response category endorsement probabilities. Steeper curves (e.g., Attention, Naming) indicate higher discrimination, while gradual slopes (e.g., Memory) suggest lower discrimination. Curve positioning along the x-axis reflects domain difficulty, with Orientation being the easiest and Memory the most challenging. These ICCs visually represent how difficulty and discrimination parameters influence response patterns across the cognitive ability spectrum.

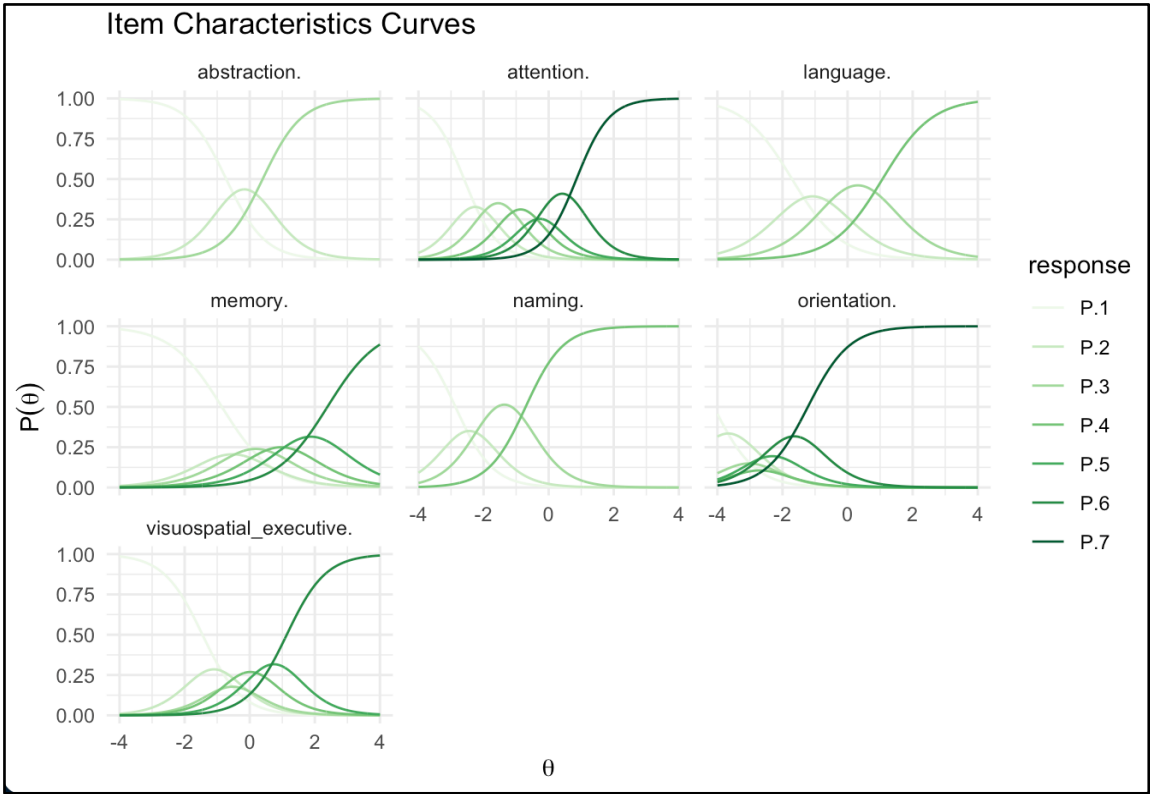


Figure 2. - Item Characteristic Curves for Montreal Cognitive Assessment (MoCA) domains.

Legend. The graph presents Item Characteristic Curves (ICCs) for polytomous item responses within the cognitive domains of the MoCA. The x-axis (θ) represents the latent trait of cognitive ability, ranging from -4 (indicating lower ability) to +4 (indicating higher ability). The y-axis depicts the probability $[P(\theta)]$ of a participant endorsing a particular response category at a given level of θ . Curves labeled P.1 through P.7 correspond to the ascending response categories for each domain, with P.1 representing the lowest category (e.g., no correct responses) and P.7 representing the highest category (e.g., all correct responses). A higher curve position on the y-axis reflects a higher probability of a participant with a corresponding θ level selecting that response category, providing a visual representation of the probability of achieving each score based on cognitive ability.

Figures 3 and 4 offer insights into item and test information functions. Figure 3's item information curves show psychometric information provided by each MoCA domain across the cognitive ability spectrum. Domains with higher peaks (e.g., Attention and Naming) offer more information, indicating greater reliability and discrimination at specific ability levels. Figure 4 displays the test information function, representing aggregate information from the entire MoCA. The curve's peak indicates the ability level where MoCA provides the most precise measurement. It also illustrates the inverse relationship between test information and standard error of measurement.

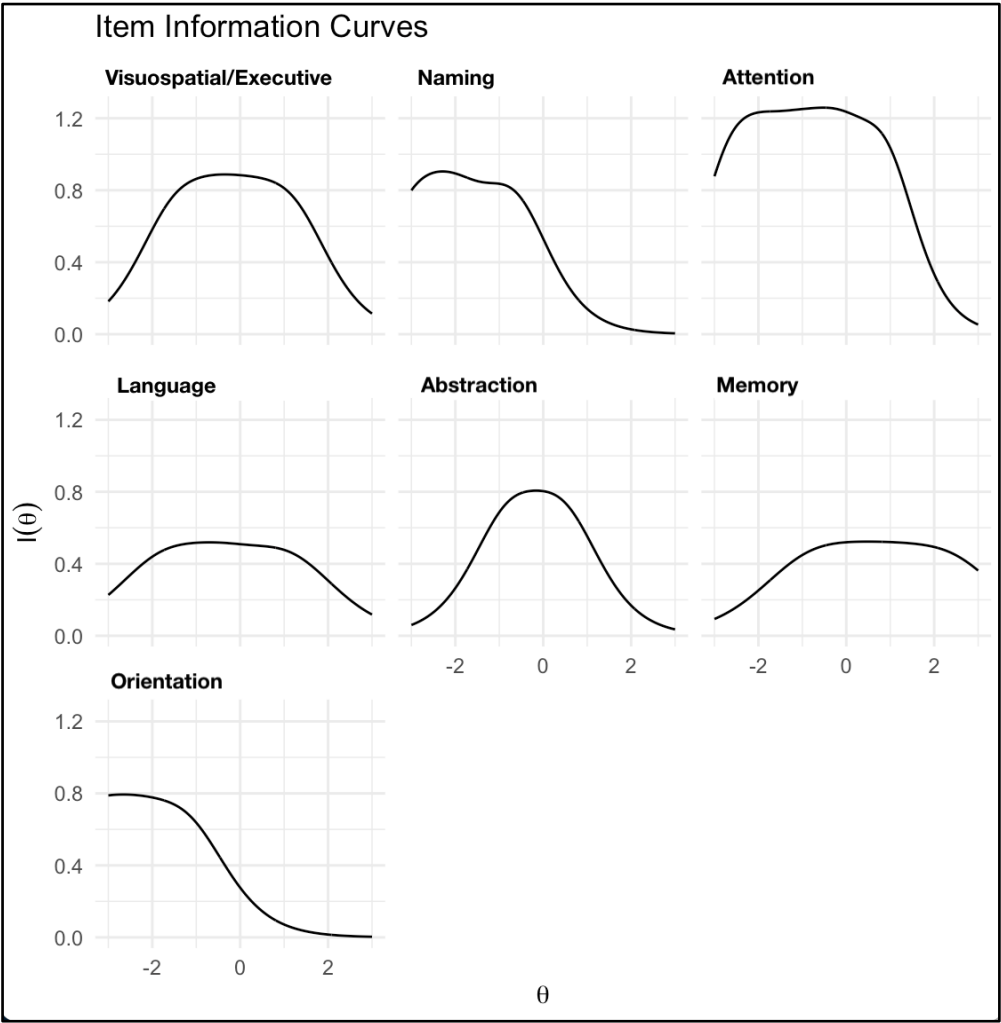


Figure 3. - Item Information curves for MoCA domains.

Legend. This figure illustrates the item information functions (IIFs) associated with the cognitive domains assessed by the Montreal Cognitive Assessment (MoCA). Each graph represents the sensitivity of the respective domain to detect changes in cognitive ability, denoted as theta (θ). The θ parameter is normatively scaled, ranging from -3 to +3, signifying low to high cognitive abilities, respectively. The acme of each curve symbolizes the juncture at which the domain is most informative, i.e., where it most accurately discerns between different levels of cognitive function. A steeper ascent and peak of the curve connote a higher information density, indicating a domain's heightened precision in measuring cognitive abilities around its peak θ value. Conversely, the flattening of the curve at the tails suggests diminishing informational yield for extreme ability levels.

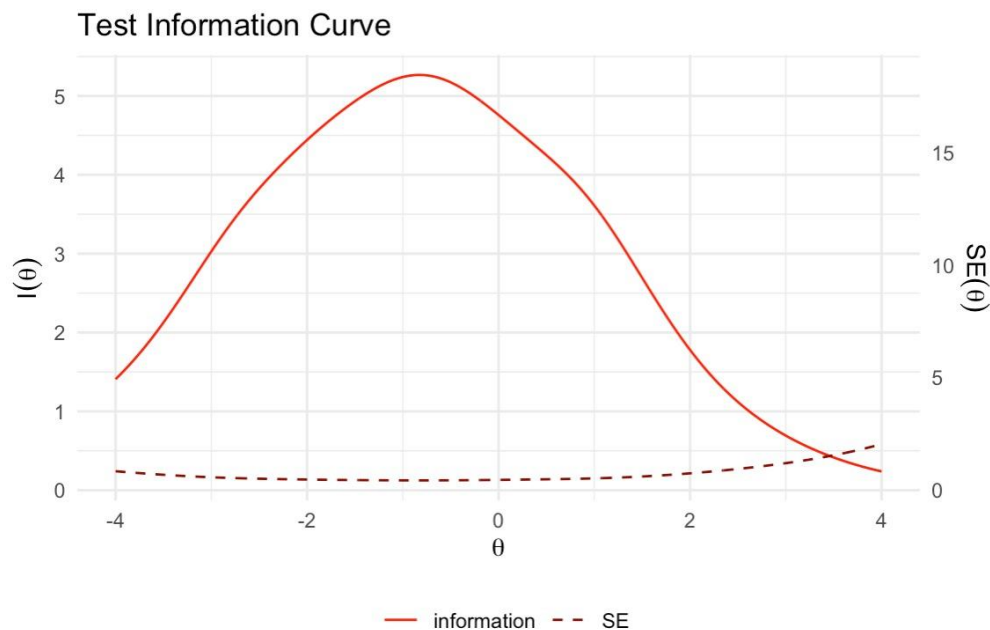


Figure 4. - Test information and standard error curves for the MoCA.

Legend. The solid red line shows the test information function $I(\theta)$ across cognitive ability levels (θ) . The peak indicates where the MoCA provides optimal measurement information. The dashed line represents the standard error of measurement $[SE(\theta)]$, which inversely relates to test information. Lower $SE(\theta)$ reflects greater precision in estimating θ . $SE(\theta)$ is minimized where test information is maximized, delineating the optimal range of measurement precision.

3.3. Multiple Indicators Multiple Causes (MIMIC) Modeling of Age and Education Effects on Latent Trait

The Multiple Indicators Multiple Causes (MIMIC) model was employed to examine differential item functioning (DIF) across demographic variables. The model demonstrated adequate fit ($\chi^2 p = 0.606$, CFI and TLI > 0.95 , RMSEA < 0.001 for standard model, 0.050 for scaled model). Regression analysis revealed significant associations between cognitive performance and demographics: age negatively impacted performance ($\beta = -0.149$, $p = 0.016$), while education showed a positive influence ($\beta = 1.001$, $p < 0.001$). Cognitive domains were differentially affected by age and education, with Memory Recall and Executive/Visuospatial tasks benefiting more from education, and Working Memory/Attention showing a moderate correlation with age but stronger association with education. A negative correlation between age and education ($\beta = -0.126$, $p = 0.008$) was observed. Educational attainment significantly affected the MoCA's latent trait, accounting for 70% of its variance. These findings underscore the complex interplay between age and education in cognitive performance, emphasizing the need for age- and education-adjusted norms in MoCA interpretation.

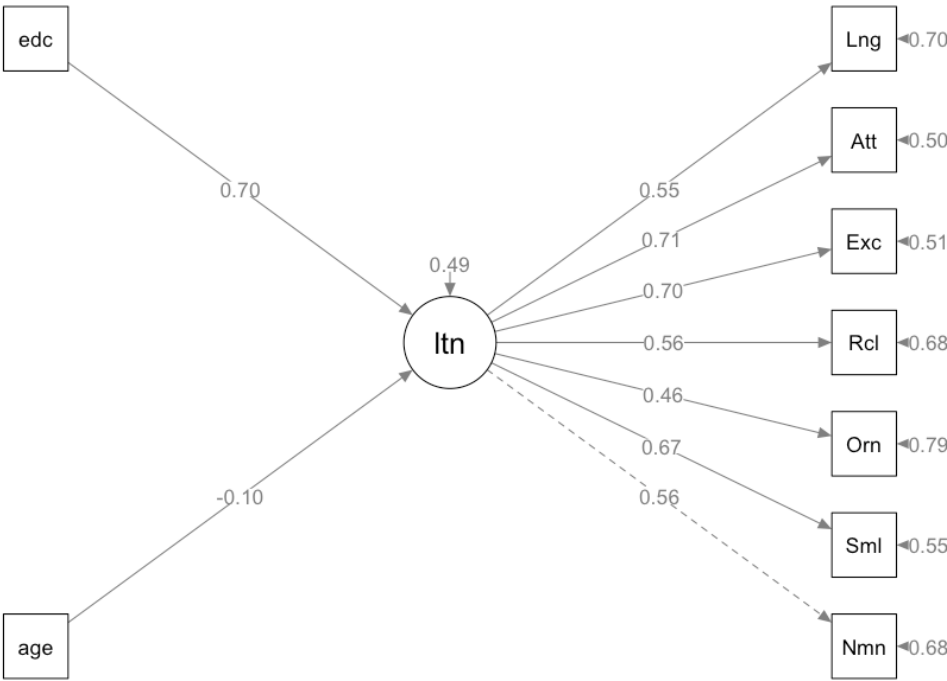


Figure 5. - Influence of Age and Education on Latent Cognitive Traits: A MIMIC Model Representation.

Legend. In the presented model, 'age' and 'education' (edc) are depicted as influencing factors, represented by the rectangles labeled 'age' and 'edc'. These factors impact the latent trait (ltn) of cognitive performance, which is represented by the central circle. The latent trait, in turn, influences various cognitive domains assessed by the Montreal Cognitive Assessment (MoCA), including Language (Lng), Attention (Att), Executive Functions (Exc), Recall (Rcl), Orientation (Orm), Similarities (Sml), and Naming (Nmnm). These domains are represented by the squares connected to the latent trait by arrows, indicating the strength and direction of their relationship. Solid arrows reflect significant positive associations, while dashed arrows indicate non-significant or weaker associations. The numbers on the arrows represent the standardized path coefficients, reflecting the strength of the relationships. This model aims to illustrate how age and education affect overall cognitive abilities and their subsequent impact on specific cognitive domains within the MoCA framework.

A second MIMIC model explored the influences of age and education on MoCA domain performance. The model demonstrated strong fit (CFI and TLI ~0.99, RMSEA ≤0.035). Educational attainment significantly impacted Language, Attention, Executive Functions, and Similarities (Abstraction) domains (β s from 0.76 to 2.677, $p < 0.05$), while age had a smaller, negative effect on Naming, Similarities, Orientation, and Executive Functions (β s from -0.175 to -0.283, $p < 0.05$). Age's impact on Memory, Attention, and Language was marginal ($p > 0.05$). These findings highlight the differential impact of age and education across cognitive domains, emphasizing the need for demographic-specific norms in MoCA interpretation to enhance assessment accuracy and fairness in diverse populations.

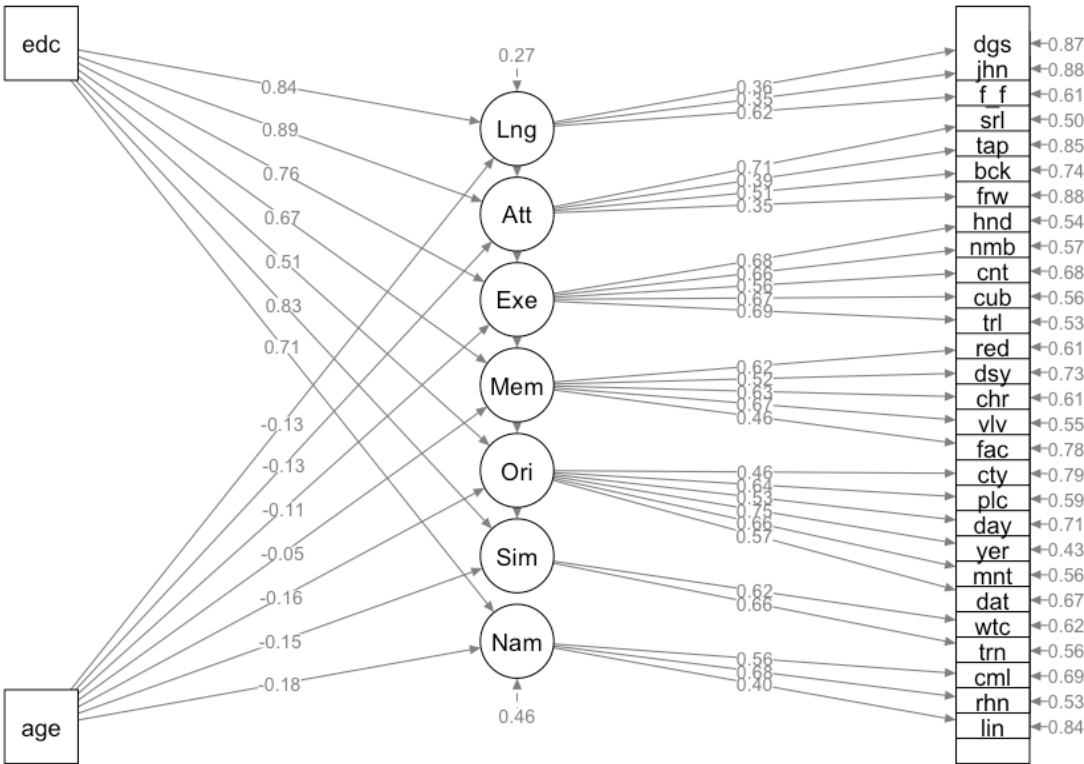


Figure 6. - Multiple Indicators Multiple Causes (MIMIC) model quantifies age and education's effects on MoCA domains rather than on a latent trait.

Legend. The model provides specific insights into how these variables impact different cognitive domains within the Montreal Cognitive Assessment (MoCA). Rectangles indicate the predictor variables ('age' and 'edc'), and the circles represent the cognitive domains: Language (Lng), Attention (Att), Executive/Visuospatial (Exe), Memory Recall (Mem), Orientation (Ori), Abstraction/Similarities (Sim) and Naming (Nam). The numerical values next to the single-headed arrows from 'age' and 'edc' denote the strength and direction of the effects on each domain, with negative values indicating an inverse relationship.

4. Discussion

Our investigation into the MoCA using the GRM within a Brazilian PD patient cohort offers detailed insights into the test's psychometric attributes. The findings on dimensionality align with the MoCA's design as a tool for globally assessing cognitive function, reinforcing its suitability as a comprehensive cognitive evaluation instrument. This aligns with previous research, such as a study from Portugal, which also upheld the MoCA's unidimensional nature [14]. However, our study sets itself apart by employing the GRM to address the polytomous characteristics of the MoCA items, offering a distinct analytical perspective.

Contrasting with the findings of Smith et al., who were unable to definitively determine a factor structure for the MoCA in a large cohort of recent-onset PD patients (n=1738) with normal to mildly impaired cognition [24], our results demonstrate a unidimensional factor structure for the Brazilian Portuguese MoCA in a PD context. The IRT analysis underpins the reliability of the MoCA, as evidenced by its adequate model fit and robust factor loadings. This suggests that within our study's framework, the MoCA effectively captures the cognitive dimensions it intends to measure in PD patients, offering a reliable assessment tool in this population.

The MIMIC model revealed significant effects of age and education on MoCA performance. Age negatively correlated with Naming domain scores ($\beta = -0.175$ to -0.283 , $p < 0.05$), suggesting decreased semantic memory and language fluency with aging. Conversely, higher education positively

influenced various cognitive domains, including Executive Functions and Naming, indicating enhanced cognitive reserve. The pronounced impact of education on Similarities (Abstraction) domain highlights its role in strengthening abstract reasoning. These findings emphasize the importance of considering both age and educational background when interpreting MoCA results, as higher education may mask early cognitive decline. Clinicians should adjust their interpretations accordingly to ensure accurate cognitive assessments.

Our findings partially align with those of Luo et al. (2020) in Hong Kong, where an IRT analysis of the MoCA was conducted among older Cantonese-speaking individuals. Using a sample of 1873 participants and the Chinese version of MoCA, the study implemented exploratory and confirmatory factor analyses along with the GRM. It revealed significant variation in MoCA item performance by educational background, particularly in visuospatial items, and found greater reliability of MoCA among uneducated participants. This advocated for the broader use of IRT to improve the measurement precision of clinical scales.

However, this study also contrasts with Sala et al.'s research in Japan, which thoroughly examined MoCA's psychometric properties in a large elderly cohort [25]. With 2408 participants across three age groups, their study identified multidimensionality in initial tests, followed by a hierarchical EFA revealing a general factor and seven first-order factors. A confirmatory factor analysis confirmed this structure, and importantly, measurement invariance was established across diverse demographics, including age, education, sex, and economic status, indicating MoCA's robustness across these variables. This finding differs from our study's emphasis on the impact of age and education. These divergent findings could be due to several factors, including differences in sample characteristics (elderly Japanese vs. PD patients in Brazil), analytical approaches (hierarchical EFA vs. GRM), and the handling of item scores (dichotomous vs. polytomous). Further research is needed to clarify these discrepancies and establish the generalizability of MoCA's psychometric properties across different populations and contexts.

In Portugal, MoCA's analysis by Freitas et al. (2014) using the Rasch model for dichotomous items and DIF analysis also provides a relevant comparison [14]. Assessing MoCA in 897 participants, including healthy individuals and a clinical group with various dementia forms, the study evaluated the test's fit and reliability. It explored DIF related to pathology, gender, age, and education. The results showed good item and person fits with no severe misfit among items and strong discriminant validity, especially between control and clinical groups. Although some items exhibited DIF, it was balanced across variables, mitigating score bias concerns. This balanced DIF contrasts with our findings, where educational background significantly influenced domain performance. These differences might stem from the use of distinct IRT models (Rasch vs. GRM), the handling of item scores (dichotomous vs. polytomous), and the focus on different clinical populations (dementia vs. PD). Future studies could directly compare these analytical approaches and populations to highlight these discrepancies.

Our study provides novel insights into the MoCA's application in PD patients. The Attention and Naming domains demonstrated high discrimination ability, underscoring their importance in detecting cognitive impairments. However, the Orientation and Naming domains showed limitations in differentiating individuals with higher cognitive abilities, suggesting a potential ceiling effect. Clinicians should interpret perfect or near-perfect scores in these domains cautiously and consider supplementing the MoCA with more challenging tasks when assessing individuals with previously high cognitive functioning. The significant effect of educational attainment on MoCA scores necessitates careful consideration in clinical practice. Adjusting cutoff scores based on patients' educational backgrounds could mitigate the risk of misdiagnosing cognitive impairment, while developing education-specific norms for PD populations could enhance diagnostic accuracy.

The MIMIC analysis revealed the differential impact of age and education on MoCA domains. Age exhibited a modest negative effect on cognitive performance, while education demonstrated a strong positive influence, particularly on Executive Functions (96%) and Naming (19%). These findings emphasize the importance of considering patients' educational backgrounds when interpreting MoCA performance and suggest the potential for education-based interventions to

promote cognitive resilience in PD. The analysis also indicated that age and education independently influenced MoCA performance, with no significant covariance between them. This suggests that the cognitive benefits of education may persist across the lifespan, potentially buffering against age-related decline. Clinically, these results highlight the importance of lifelong learning and cognitive stimulation in maintaining brain health, especially in neurodegenerative disorders like PD.

To address the psychometric limitations of the MoCA identified in our study, we propose a refined approach for its application in PD patients. We advocate for the calibration of test items using IRT to account for educational variability and the development of tailored MoCA versions with reduced educational bias. Additionally, we recommend supplemental assessments to better evaluate high cognitive performers and the establishment of large sample demographic-specific norms for more accurate score interpretation. The integration of technology, such as a web-based application utilizing the GRM for automated MoCA scoring, could significantly enhance diagnostic precision by individualizing item weighting. These refinements could substantially improve the MoCA's utility in diagnosing and tracking cognitive impairment in PD patients across diverse educational backgrounds.

We acknowledge the limitations of the study. The relatively small sample size and non-random participant selection may challenge the generalizability of our findings. Our sample, drawn from specialized clinics, might not fully represent the broader PD population, potentially excluding patients with more severe cognitive or motor impairments. Future research should aim to recruit larger, more diverse PD samples using random sampling methods. Additionally, while our analysis focused on the impact of education and age on Differential Item Functioning (DIF), it did not explore other potential contributing factors such as gender or cultural influences. Future studies should investigate a broader range of sociodemographic and cultural variables on MoCA performance in PD and develop strategies to mitigate their influence. Furthermore, the cross-sectional design of our study limits causal inferences about the relationships between demographic factors and cognitive performance. Longitudinal studies are necessary to elucidate the temporal dynamics of these associations and investigate the protective role of education against cognitive decline in PD over time.

Furthermore, our study focused on the MoCA as a global cognitive screening tool and did not include a comprehensive neuropsychological assessment battery. While the MoCA has demonstrated utility in detecting cognitive impairment in PD, it may not capture the full spectrum of cognitive deficits associated with the disease. Future research should explore the relationship between MoCA performance and more comprehensive cognitive assessments in PD and its predictive validity for clinical outcomes such as dementia risk and functional decline.

In conclusion, our study demonstrates the utility of the IRT GRM approach in assessing the MoCA test's psychometric properties, revealing unidimensionality with variable domain discrimination and significant age and education influences. The findings highlight the necessity of considering these factors in MoCA score interpretation and advocate for refined, education-adjusted norms in clinical practice. MIMIC analysis indicates a strong positive impact of education and a modest negative effect of age on MoCA performance, suggesting education is a protective factor against cognitive decline in PD. Future research should expand DIF investigations, develop strategies to address measurement invariance and conduct longitudinal studies to explore the protective role of education over time. Integrating IRT and MIMIC modeling in cognitive screening tool validation promises to enhance our understanding of socio-demographic influences on cognitive health, leading to more personalized and equitable cognitive assessments in PD and other neurological conditions.

Funding: This work was partially supported by a grant from the Michael J Fox Foundation for Parkinson's Disease Research. The funding sources had no role in the study design, data collection and analysis, manuscript preparation, or the decision to submit the article for publication.

Data Availability: The anonymized datasets generated and analyzed during the current study are available from the corresponding author on reasonable request and subject to approval on a case-by-case basis. Data

access will only be granted to qualified researchers for legitimate research purposes, contingent on institutional ethical policies and applicable laws.

Acknowledgments: The authors gratefully acknowledge the Michael J. Fox Foundation and FAP-DF (Fundação de Apoio à Pesquisa do Distrito Federal) for funding the study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* 2005;53(4):695-699. doi:10.1111/J.1532-5415.2005.53221.X
2. Tumas V, Borges V, Ballalai-Ferraz H, et al. Some aspects of the validity of the Montreal Cognitive Assessment (MoCA) for evaluating cognitive impairment in Brazilian patients with Parkinson's disease. *Dement Neuropsychol.* 2016;10:333-338. doi:10.1590/S1980-5764-2016DN1004013
3. Almeida KJ, Carvalho LCL de S, Monteiro THO de H, Gonçalves PC de J, Campos-Sousa RN. Cut-off points of the Portuguese version of the Montreal Cognitive Assessment for cognitive evaluation in Parkinson's disease. *Dement Neuropsychol.* 2019;13(2):210-215. doi:10.1590/1980-57642018dn13-020010
4. Camargo CHF, Tolentino E de S, Bronzini A, et al. Comparison of the use of screening tools for evaluating cognitive impairment in patients with Parkinson's disease. *Dement Neuropsychol.* 2016;10:344-350. doi:10.1590/S1980-5764-2016DN1004015
5. Sarmento ALR [UNIFESP. Apresentação e aplicabilidade da versão brasileira da MoCA (Montreal Cognitive Assessment) para rastreio de Comprometimento Cognitivo Leve. *Presentation and applicability of the Brazilian version the MoCA (Montreal Cognitive Assessment) for Screening of Mild Cognitive Impairment.* Published online November 25, 2009. Accessed July 12, 2023. <https://repositorio.unifesp.br/handle/11600/8967>
6. Cesar KG, Yassuda MS, Porto FHG, Brucki SMD, Nitrini R. MoCA Test: normative and diagnostic accuracy data for seniors with heterogeneous educational levels in Brazil. *Arq Neuropsiquiatr.* 2019;77(11):775-781. doi:10.1590/0004-282X20190130
7. Wong A, Law LSN, Liu W, et al. Montreal Cognitive Assessment: One Cutoff Never Fits All. *Stroke.* 2015;46(12):3547-3550. doi:10.1161/STROKEAHA.115.011226
8. Lu J, Li D, Li F, et al. Montreal cognitive assessment in detecting cognitive impairment in Chinese elderly individuals: a population-based study. *J Geriatr Psychiatry Neurol.* 2011;24(4):184-190. doi:10.1177/0891988711422528
9. Balsis S, Choudhury TK, Geraci L, Bengtson JF, Patrick CJ. Alzheimer's Disease Assessment: A Review and Illustrations Focusing on Item Response Theory Techniques. *Assessment.* 2018;25(3):360-373. doi:10.1177/1073191117745125
10. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care.* 2000;38(9 Suppl):II28-42. doi:10.1097/00005650-200009002-00007
11. Reise SP, Waller NG. Item response theory and clinical measurement. *Annu Rev Clin Psychol.* 2009;5:27-48. doi:10.1146/annurev.clinpsy.032408.153553
12. Reise SEE Steven P. *Item Response Theory.* Psychology Press; 2000. doi:10.4324/9781410605269
13. Tsai CF, Lee WJ, Wang SJ, Shia BC, Nasreddine Z, Fuh JL. Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: validation of the Taiwanese version of the MoCA and an item response theory analysis. *Int Psychogeriatr.* 2012;24(4):651-658. doi:10.1017/S1041610211002298
14. Freitas S, Prieto G, Simões MR, Santana I. Psychometric properties of the Montreal Cognitive Assessment (MoCA): an analysis using the Rasch model. *Clin Neuropsychol.* 2014;28(1):65-83. doi:10.1080/13854046.2013.870231
15. Luo H, Andersson B, Tang JYM, Wong GHY. Applying Item Response Theory Analysis to the Montreal Cognitive Assessment in a Low-Education Older Population. *Assessment.* 2020;27(7):1416-1428. doi:10.1177/1073191118821733

16. Samejima F. Estimation of Latent Ability Using a Response Pattern of Graded Scores¹. *ETS Res Bull Ser.* 1968;1968(1):i-169. doi:10.1002/j.2333-8504.1968.tb00153.x
17. Sarihan EI, Pérez-Palma E, Niestroj LM, et al. Genome-Wide Analysis of Copy Number Variation in Latin American Parkinson's Disease Patients. *Mov Disord.* 2021;36(2):434-441. doi:10.1002/mds.28353
18. Brandão PR de P, Pereira DA, Grippe TC, et al. Parkinson's Disease-Cognitive Rating Scale (PD-CRS): Normative Data and Mild Cognitive Impairment Assessment in Brazil. *Mov Disord Clin Pract.* 2023;10(3):452-465. doi:10.1002/mdc3.13657
19. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry.* 1992;55(3):181-184. doi:10.1136/jnnp.55.3.181
20. Brown TA. *Confirmatory Factor Analysis for Applied Research*. The Guilford Press; 2006:xiii, 475.
21. Chalmers P, Pritikin J, Robitzsch A, et al. mirt: Multidimensional Item Response Theory. Published online May 30, 2023. Accessed June 5, 2023. <https://cran.r-project.org/web/packages/mirt/index.html>
22. Masur PK. ggmirt. Published online October 25, 2023. Accessed November 22, 2023. <https://github.com/masurp/ggmirt>
23. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *J Stat Softw.* 2012;48:1-36. doi:10.18637/JSS.V048.I02
24. Smith CR, Cavanagh J, Sheridan M, Grosset KA, Cullen B, Grosset DG. Factor structure of the Montreal Cognitive Assessment in Parkinson disease. *Int J Geriatr Psychiatry.* 2020;35(2):188-194. doi:10.1002/gps.5234
25. Sala G, Inagaki H, Ishioka Y, et al. The Psychometric Properties of the Montreal Cognitive Assessment (MoCA). *Swiss J Psychol.* 2020;79(3-4):155-161. doi:10.1024/1421-0185/a000242

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.