


ORIGINAL ARTICLE

Longitudinal measurement invariance of neuropsychological tests in a diverse sample from the ELSA-Brasil study

Laiss Bertola,¹  Isabela M. Benseñor,^{2,3} Alden L. Gross,⁴ Paulo Caramelli,⁵ Sandhi Maria Barreto,⁶ Arlinda B. Moreno,⁷ Rosane H. Griep,⁸ Maria Carmen Viana,⁹ Paulo A. Lotufo,³ Claudia K. Suemoto¹⁰

¹Faculdade de Medicina, Universidade de São Paulo (USP), São Paulo, SP, Brazil. ²Centro de Pesquisa Clínica e Epidemiológica, Hospital Universitário, USP, São Paulo, SP, Brazil. ³Departamento de Medicina Interna, Faculdade de Medicina, USP, São Paulo, SP, Brazil.

⁴Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA. ⁵Departamento de Clínica Médica, Faculdade de Medicina, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil. ⁶Hospital das Clínicas, Faculdade de Medicina, UFMG, Belo Horizonte, MG, Brazil. ⁷Departamento de Epidemiologia e Métodos Quantitativos em Saúde (DEMQS), Escola Nacional de Saúde Pública Sergio Arouca (ENSP), Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, RJ, Brazil. ⁸Laboratório de Educação em Ambiente e Saúde, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, RJ, Brazil. ⁹Departamento de Medicina Social, Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brazil. ¹⁰Divisão de Geriatria, Faculdade de Medicina, USP, São Paulo, SP, Brazil.

Objective: Longitudinal measurement invariance analyses are an important way to assess a test's ability to estimate the underlying construct over time, ensuring that cognitive scores across visits represent a similar underlying construct, and that changes in test performance are attributable to individual change in cognitive abilities. We aimed to evaluate longitudinal measurement invariance in a large, social and culturally diverse sample over time.

Methods: A total of 5,949 participants from the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) were included, whose cognition was reassessed after four years. Longitudinal measurement invariance analysis was performed by comparing a nested series of multiple-group confirmatory factor analysis models (for memory and executive function factors).

Results: Configural, metric, scalar and strict invariance were tested and supported over time.

Conclusion: Cognitive temporal changes in this sample are more likely to be due to normal and/or pathological aging. Testing longitudinal measurement invariance is essential for diverse samples at high risk of dementia, such as in low- and middle-income countries.

Keywords: Parallel tests; longitudinal change; cognition; low and middle income countries; memory; executive function

Introduction

Longitudinal studies can capture intra-individual cognitive trajectories over time¹ and can facilitate the identification of significant cognitive decline² and diagnostic accuracy.³ The repeated-measures design allows cognitive assessment over time with the same tests, which optimizes performance comparison.

Considering that neuropsychological assessments are mainly performed using tests that measure specific cognitive domains, as well as that the aging process is usually associated with increased cognitive interference during task performance,⁴ longitudinal studies should perform additional psychometric analysis.^{5,6} Measurement

invariance analysis aims to ensure reliable conclusions about real cognitive changes that can only be achieved with tests that can measure the same psychological trait over time.^{7,8}

Time-invariant neuropsychological batteries can guarantee that cognitive changes are attributable to normal and pathological brain aging and not to differences in the way tests measure the construct over time.⁹ An understanding of factors that contribute to real cognitive change is only possible when longitudinal cognitive assessment is invariant.

Despite the recommendation that measurement invariance should be verified in aging and dementia research,^{5,10} few cognitive studies of older adults have

Correspondence: Laiss Bertola, Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Arnaldo, 455, Cerqueira César, CEP 01246-903, São Paulo, SP, Brazil.

E-mail: laissbertola@gmail.com

Submitted Mar 24 2020, accepted Jun 03 2020.

How to cite this article: Bertola L, Benseñor IM, Gross AL, Caramelli P, Barreto SM, Moreno AB, et al. Longitudinal measurement invariance of neuropsychological tests in a diverse sample from the ELSA-Brasil study. *Braz J Psychiatry*. 2020;00:000-000. <http://dx.doi.org/10.1590/1516-4446-2020-0978>

assessed measurement invariance. Based on a highly educated sample ($n=2,265$) of 81% Caucasian and 11% African Americans, Salthouse employed a number of methodological features to maximize sensitivity for cognitive change. The results indicated significant loss of model fit with increased constraint, suggesting that, although not identical, the measurement profile was very similar over time.¹¹

At least partial strong longitudinal invariance in memory and executive function tests was found in a small, less-educated sample ($n=86$),⁹ and weak memory invariance was found over time in an ethnically diverse sample (13,308 Whites and 3,061 African Americans).¹² On the other hand, a cognitive battery showed strong measurement invariance over time in a sample of whites ($n=1,898$) and African Americans ($n=426$).¹³

Considering that 58% of the people with dementia live in low- and middle-income countries (LMIC)¹⁴ and that early diagnosis and intervention are important for addressing this health issue, it is imperative to assure the reliability of inferences about cognitive change in studies conducted in these countries. Even though some studies have investigated longitudinal invariance,¹⁵ there is still little literature on this topic, especially from a highly heterogeneous LMIC sample.

This study aimed to evaluate longitudinal measurement invariance of the neuropsychological battery of the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). We examined longitudinal measurement invariance in this sample across two cognitive assessments performed four years apart. We hypothesized that the measurement would be invariant over time in assessing the same constructs, and our goal was to assure that future studies with this sample will be able to correctly assess risk factors for cognitive decline.

Methods

Participants

The ELSA-Brasil sample consisted of active or retired employees from public institutions in six large cities (Belo Horizonte, Porto Alegre, Rio de Janeiro, Salvador, São Paulo, and Vitória).^{16,17} The total sample included 15,105 Brazilian Portuguese speakers, aged 35 to 74 years, who were free of dementia at enrollment (2008 to 2010). The baseline assessment included sociodemographic information, clinical history, a mental health evaluation, lifestyle factors, occupational exposure, and general health family history. Cognitive function was reassessed (2012 to 2014) only in participants who were 55 years or older ($n=7,066$) at the second visit.

We excluded participants with self-reported medical diagnosis of stroke and those who were using medications that indicate the presence of active neurologic or psychiatric diseases (i.e., neuroleptics, antiparkinsonian agents, and antiepileptic drugs) at baseline. We also excluded participants with missing cognitive scores at baseline, and those with incomplete assessment at follow-up. At baseline, 13,395 participants remained after application of the eligibility criteria. Among the 7,066 participants

who were 55 years old at the second visit, 5,949 were reassessed four years later and were considered the final sample (Figure 1).

Neuropsychological assessment

Baseline assessment used the Consortium's standardized memory tests to Establish a Registry for Alzheimer's Disease (CERAD)¹⁸ validated for the Brazilian population¹⁹ to assess learning, delayed word recall, and recognition (CERAD word list test).^{19,20} The baseline assessment also included the semantic (SVF) and phonemic (PVF) verbal fluency tests (animals and letter F, respectively),^{21,22} and the Trail Making Test B (TMT).²³ Follow-up assessment used the same cognitive measures, except for the verbal fluency tasks. The PVF of letter F was replaced by letter A, and the SVF of animals was replaced by vegetables to reduce learning effects. Trained examiners administered the tests in a fixed order during a single session, and all environmental requirements for psychometric testing were met.²⁴

We used the learning, delayed recall, and recognition scores from the CERAD word list test to determine the episodic memory factor. SVF and PVF scores were based on correct exemplars produced in one minute. The TMT score was based on time (in seconds) to complete the task. Verbal fluency tasks and the TMT are both executive function tests and were used to determine the executive function factor.²⁵ Since the tests were assessed with different measurement units, test scores were transformed into z-scores to be expressed on the same scale.

Test equating

Small alterations to the verbal fluency tests were made between the first and second waves. The test versions were parallel, but not equivalent. Parallel tests can assess and score the same domain using similar content; however, when there is a disparity in difficulty level, the same individual will score differently, even when no pathological process is present. To determine whether the ELSA-Brasil cognitive assessment is invariant over time, we first performed a test equating analysis. The purpose of this analysis is to guarantee that the distinct versions of the verbal fluency tests measure the construct at the same difficulty level by transforming one test score into the same metric and range of values the other test.^{26,27} In this study, we equated the SVF vegetables and the PVF letter A scores used in the second wave to the SFV animals and PVF letter F scores used in the first wave, respectively.

Various equating methods²⁶ differ according to the way the new score is assigned. For this study, we opted for equipercentile equating, which defines the new relative score position based on percentile ranks, because this method is more suitable when tests do not have normal distributions.^{26,28} This approach identifies scores on two measures that have the same percentile rank and transforms the score of one test to the corresponding score of the reference test. For example, a score of 20 words

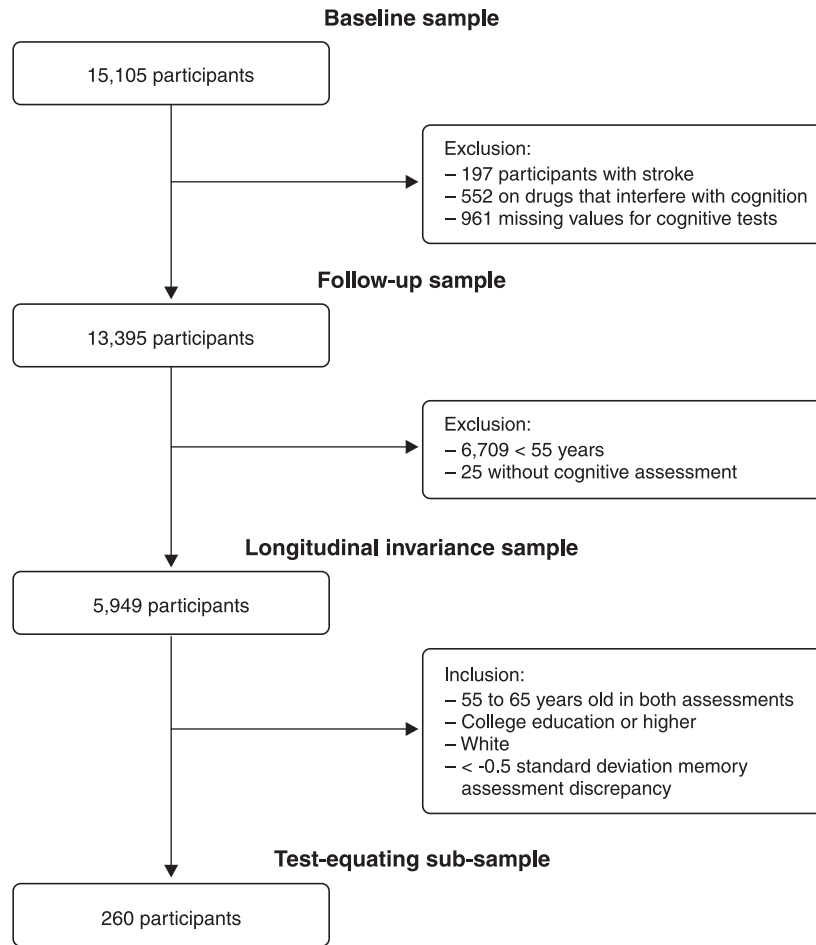


Figure 1 Flowchart of the study participants and the test equating sample.

in the SVF animals test had a percentile rank of 50. To obtain a similar percentile in SVF vegetables it was necessary to produce 18 words. This difference demonstrates that the second version was more difficult than the first at the median.

To perform this analysis, we selected a strictly homogenous sample that was not expected to have measurable cognitive change over the four-year follow-up to guarantee that the differences in test scores in the two waves were due to differences in the test versions and not to cognitive change.²⁸ The homogenous sample was selected considering the following characteristics related to a more stable cognitive trajectory: a) being 55 to 65 years old in both assessments; b) having a college education or higher; c) being white; d) not having more than half of a standard deviation discrepancy between the baseline and follow-up episodic memory (CERAD word list); and e) having an equal proportion of male and female participants. The final homogenous sample included 260 participants.

The R package *equate*²⁹ was used to extract the equipercenile algorithm based on an equating sample that was subsequently applied to the entire sample. The equipercenile algorithm used a log-linear smoothing method to reduce irregularities due to sampling error in the score distribution.²⁹ The equated SVF vegetables and

PVF letter A scores were used to assess longitudinal measurement invariance.

Baseline and follow-up cognitive comparison

The baseline and follow-up assessments were compared using a paired sample *t*-test to assess performance stability over time. We also computed the Pearson *r* effect size of the difference between baseline and follow-up performance.

Longitudinal measurement invariance

A two-factor confirmatory factor analysis (CFA) model revealed previous measurement invariance across socio-demographic characteristics for the ELSA-Brasil baseline data.²⁵ CERAD word list test learning, delay recall, and recognition scores were used as an episodic memory factor, while SVF, PVF, and TMT were used as an executive function factor. The model included covariance between the two factors (Figure 2).

We used CFA to investigate the two-factor measurement structure over time. Considering that the TMT and the CERAD word list recognition test did not meet the normality assumption, we used the maximum likelihood

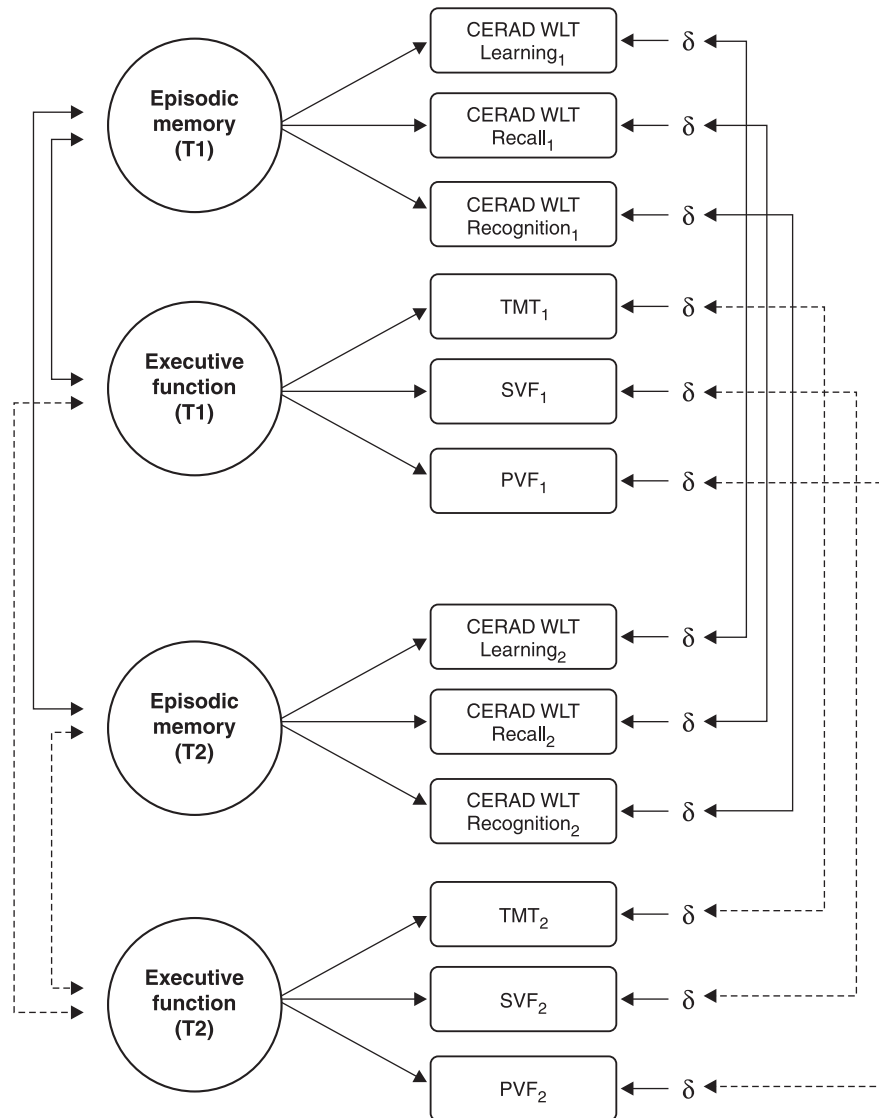


Figure 2 Two-factor measurement invariance CFA model specification. CERAD WLT = Consortium to Establish a Registry for Alzheimer's Disease word list test; PVF = phonemic verbal fluency; SVF = semantic verbal fluency; T1 = baseline assessment; T2 = follow-up assessment; TMT = Trail Making Test.

estimator with robust standard errors and χ^2 (Satorra-Bentler scaled).

Measurement invariance analysis was conducted in four steps.⁷ The first step was to evaluate longitudinal evidence of equal form in the CFA measurement model by evaluating whether the fits were comparable at each time point. Provided evidence of equal form, the second step is to test equivalence of factor loadings, assessing whether the tests have evidence of equivalent relationships to the latent variables (episodic memory and executive function factors) over time. The third step (equal indicator intercepts) tests whether the test response levels match the levels of latent traits (memory and executive function) are the same across testing time points to assure that changes in the factor are due to a change in the construct and not to the measurement of the construct at different times.

The fourth step is the equality of item residuals or unique variances. This step tests whether the sum of the specific variance (not shared with the factor) and the error variance are similar over time. Given that this model is highly constrained and unnecessary for measurement invariance (since the residuals are not part of the latent factor), this step is usually not met in most studies.⁷ We hypothesized the same would occur with our data. If the four steps reveal invariance, score changes over time can be attributed to a real change in cognitive performance and not to measurement error.

To test for measurement invariance at each level, the goodness of fit values for each step were compared to the previous model. We prioritized the root mean square error of approximation (RMSEA < 0.05) and the comparative fit index (CFI > 0.95) to evaluate overall model fit. Lower RMSEA and higher CFI results indicate better

fit. A change ≥ -0.010 in CFI and ≥ 0.015 in RMSEA indicates non-invariance.³⁰ Of the two indices, the CFI was selected as the primary criterion, since the RMSEA is sensitive to sample size and model complexity.

To determine whether missing data would impact the invariance results, we also performed a sensitivity analysis using multiple imputation for eligible participants with missing cognitive data (online-only supplementary material).

All analyses were performed in R,²⁹ Stata 13³¹ and Mplus 7.0.³² R (with the equate package²⁹) was used to perform the test equating analysis for the verbal fluency tests, allowing the longitudinal invariance analysis and future studies. Mplus 7.0 was used to perform the longitudinal invariance models, given that this program allows a better selection of estimators for a structural equation analysis and model specification according to a theoretically driven hypothesis. Stata 13 was used to perform descriptive and imputation analysis.

Ethics statement

The local institutional review board approved this study, and all participants provided informed consent.

Results

Descriptive information about the sample at baseline and after four years is shown in Table 1. Most of the baseline sociodemographic distribution was retained in the follow-up sample except age, as expected. The sample included 38.9% blacks, and the education of 41.5% of the

participants was less than college level; approximately 12% had not graduated from high-school.

Equating results

The equating results are plotted in Figure 3. The SVF vegetable scores at visit 2 were equated with the SVF animal scores at visit 1 to account for test version differences. The same process was performed to equate the raw PVF letter A scores with letter F scores. The mean and standard deviation for baseline verbal fluency scores and follow-up raw and equated scores suggested a successful score transformation (Table 2).

Longitudinal invariance results

Despite significant differences between baseline and follow-up mean cognitive scores, cognitive parameters at follow-up showed major stability over time according to the small effect sizes for repeated measurement comparison (Table 3). The cross-sectional CFA model revealed appropriate fit indices for the baseline model ($\chi^2_{(3)} = 21.334$, $p < 0.001$, CFI = 1.000, RMSEA = 0.021, confidence interval_(RMSEA) = 0.013-0.030), $p_{(RMSEA)} = 1.000$). The longitudinal measurement invariance results are presented in Table 4. The configural invariance (step 1) of the unconstrained model showed an adequate fit ($\chi^2 = 265.258$, CFI = 1.000, RMSEA = 0.035). The comparison of this invariance with the second step of equal factor loadings (weak or metric), despite a significant χ^2 difference, showed no significant decrease in fit ($\Delta CFI = -0.006$, $\Delta RMSEA = 0.006$), suggesting that the tests have

Table 1 Demographic and cognitive characteristics (n=5,949)

Variables	Baseline		Follow-up	
	Mean	SD	Mean	SD
CERAD WLT Learning z-score	-0.07	1.01	0.04	1.02
CERAD WLT Recall	-0.07	1.01	0.07	1.01
CERAD WLT Recognition	-0.06	1.04	0.09	0.89
Trail Making Test	-0.08	1.12	0.00	1.05
Semantic verbal fluency	-0.07	0.98	0.14	1.06
Phonemic verbal fluency	-0.05	1.02	-0.18	1.08
Age (years)	58.49	5.83	62.50	5.82
Baseline and follow-up, n (%)				
Education				
Primary		276 (4.64)		
Elementary		415 (6.98)		
High School		1,774 (29.82)		
College (or higher)		3,484 (58.56)		
Sex				
Female		3,307 (55.59)		
Male		2,642 (44.41)		
Race*				
White		3,347 (57.06)		
Black		2,283 (38.92)		
Other		236 (4.02)		

Neuropsychological tests are z-scored.

SD = standard deviation; CERAD WLT = word list test from the Consortium to Establish a Registry for Alzheimer's Disease.

* Race: 83 missing data from participants who refused to self-report their race.

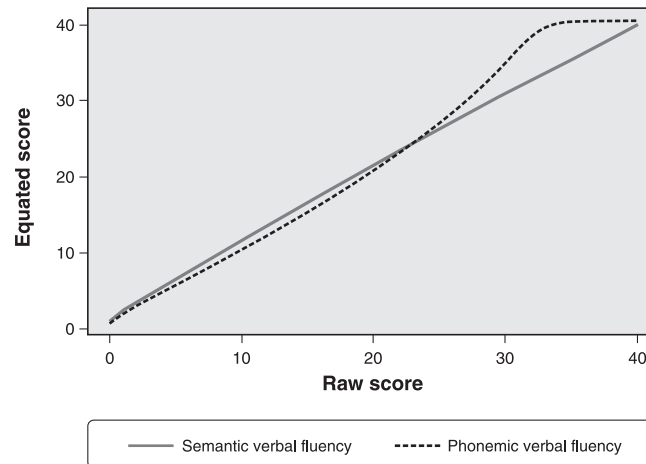


Figure 3 Plot of the raw (x axis) and equated (y axis) scores for semantic (vegetables) and phonemic (letter A) verbal fluency. The raw score was converted into the equated scores to account for different test versions.

Table 2 Mean and standard deviation for baseline scores, and raw and equated follow-up scores.

Scores	Semantic		Phonemic	
	Mean	SD	Mean	SD
Baseline raw score	18.25	5.05	12.51	4.41
Follow-up raw score	17.01	5.30	11.63	4.49
Follow-up equated score	19.01	5.31	11.80	4.80

SD = standard deviation.

Table 3 Paired-sample *t*-test for cognitive assessment (baseline and follow-up) (n=5,949)

Cognitive assessment	<i>t</i>	p-value	<i>r</i>
CERAD WLT Learning	-3.46	< 0.001	0.04
CERAD WLT Recall	-5.57	< 0.001	0.07
CERAD WLT Recognition	-7.78	< 0.001	0.10
Trail Making Test	-3.98	< 0.001	0.05
Semantic verbal fluency	-11.12	< 0.001	0.14
Phonemic verbal fluency	12.86	< 0.001	0.16

Neuropsychological tests are standardized in z-scores.

CERAD WLT = word list from the Consortium to Establish a Registry for Alzheimer's Disease.

The equated follow-up verbal fluency scores were used for the comparison.

Table 4 Measurement invariance for the two-factor model over time (n=5,949)

Models	χ^2	df	RMSEA (90%CI)	Cfit	Δ RMSEA	CFI	Δ CFI
Equal form (step 1)	265.258	32	0.035 (0.031-0.039)	1.000		0.990	
Equal factor loadings (step 2)	421.093	38	0.041 (0.038-0.045)	1.000	0.006	0.984	-0.006
Equal indicator intercepts (step 3)	604.241	42	0.047 (0.044-0.051)	0.892	0.006	0.977	-0.007
Equal indicator error variances (step 4)	595.097	48	0.044 (0.041-0.047)	0.999	0.003	0.977	0.000

90%CI = 90% confidence interval; χ^2 = chi-square (Satorra-Bentler); CFI = Comparative Fit Index; Cfit = p-value for RMSEA; df = degrees of freedom; RMSEA = root mean square error of approximation.

equivalent relationships to the latent constructs (Episodic Memory and Executive Function factors) over time. The third step, which verified equal indicator intercepts (strong or scalar), also suggested that the properties of the tests were invariant across testing occasions (Δ CFI = -0.007, Δ RMSEA = 0.006). The fourth step, which assesses equal indicator error variances (strict), indicated that the test error variances might be stable over time, suggesting that no meaningful change in score variations occurred over time (Δ CFI = 0.000, Δ RMSEA = 0.003). In a

sensitivity analysis, we found configural, metric, scalar, and strict invariance over time when we used multiple imputed data (Table S1, available as online-only supplementary material).

Discussion

This study verified the longitudinal measurement invariance of a cognitive battery over time in a diverse sample from a LMIC. The findings revealed that the two-factor

model, episodic memory and executive function were characterized by strict longitudinal measurement invariance. The longitudinal invariance results show that this cognitive battery can assess true change in cognitive performance that is not due to psychometric variability over time. This approach guarantees that further studies in this sample that use cognitive change as an outcome will have unbiased results that are not attributable to changes in test properties.

Our results are in line with those of Barnes et al.,¹³ who reported strong invariance over time in a cognitive battery applied to a racially diverse sample. Additionally, our results demonstrated that we probably achieved strict invariance because our sample was significantly younger than that of Barnes et al.¹³ (i.e., a significant change in residual variance was not expected). However, although the sample of Blankson & McArdle¹² was similar to ours in age, education, and race distribution (whites and blacks), our results differed. They reported only weak measurement invariance over time, while we found strict invariance. Salthouse¹¹ and Moreira et al.⁹ also failed to reach the strictest invariance step, finding a significant change in residual variance over time. These differences might be attributable to sample characteristics or psychometric properties for assessing a specific cognitive domain. Both our study and that of Barnes et al. used the CERAD word list memory test, for example.¹³

Although longitudinal studies from developed countries have assessed the association between risk factors and cognitive decline, these risk factors might have distinct profiles and distributions in LMIC.³³⁻³⁶ It is important to consider socioeconomic and sociocultural aspects when designing studies to investigate risk factors for cognitive decline to promote more effective local public policy. Our results indicate that the longitudinal ELSA-Brasil findings can contribute reliable results, as shown in other studies from developed countries.

Cognition is directly and indirectly related to functional and social independence, and the assessment of reliable cognitive trajectories is important for epidemiological studies analyzing brain disorders. Our study expands the literature on methodological issues in aging research,⁵ reducing bias and increasing the reliability of the results. Achieving strong and strict invariance should also motivate studies from developed countries with diverse samples to seek better assessment and reduce bias in research results.

Despite hypothesizing that strict invariance would not be achieved, no significant difference in residual variance was found for scores across the two assessment points. One possible explanation is that the follow-up period was short for the mainly healthy and young sample of the ELSA-Brasil study. The preserved cognitive ability in this sample might have contributed to an absence of changes in the residual variability of scores.

Our longitudinal analysis primarily consisted of equating verbal fluency scores. Longitudinal studies are known to be susceptible to practice effects. Although using parallel test versions has been suggested,¹ this approach increases comparison errors due to version discrepancies.²⁸ Equipercentile equation revealed adequate score

transformations, allowing direct comparison across the verbal fluency tasks. The differences in score more likely reflected an actual cognitive performance difference, and they are probably not due to version discrepancies across time points.

Our study has some limitations. The design of the ELSA-Brasil study foresees that only participants who were 55 years or older at the second visit underwent the cognitive assessment. Therefore, we could not assess measurement invariance over time for adults younger than 55. This limitation might be minimized by other longitudinal studies that have revealed stability or subtle decline in this young age group.^{37,38} In addition, despite presenting a complete case analysis for not considering participants with missing data in the second visit, we also performed a sensitivity analysis that revealed similar results.

The strengths of the current study include data from a longitudinally followed sociodemographically diverse sample with a wide age and education range. We also corrected the parallel verbal fluency test versions with a robust harmonization analysis, which can guarantee reliable longitudinal analysis across visits.

In conclusion, we showed that the ELSA-Brasil sample had valid and invariant cognitive measurements over time. These results from a large, diverse sample in a LMIC will help point out similarities and discrepancies in the field of normal cognitive aging and dementia compared to the massive data produced by developed countries.

Acknowledgements

The ELSA-Brasil baseline study was supported by the Brazilian Ministry of Health (Science and Technology Department) and the Brazilian Ministry of Science and Technology (Financiadora de Estudos e Projetos [FINEP] and the Conselho Nacional de Desenvolvimento Científico e Tecnológico [CNPq]; grants 01 06 0010.00 RS, 01 06 0212.00 BA, 01 06 0300.00 ES, 01 06 0278.00 MG, 01 06 0115.00 SP, 01 06 0071.00 RJ). SMB and RHG are CNPq research fellows (grants 300159/99-4 and 301807/2016-7 respectively).

Disclosure

The authors report no conflicts of interest.

References

- 1 Salthouse TA. Robust cognitive change. *J Int Neuropsychol Soc.* 2012;18:749-56.
- 2 Yaffe K, Fiocco AJ, Lindquist K, Vittinghoff E, Simonsick EM, Newman AB, et al. Predictors of maintaining cognitive function in older adults: the Health ABC study. *Neurology.* 2009;72:2029-35.
- 3 Zahodne LB, Wall MM, Schupf N, Mayeux R, Manly JJ, Stern Y, et al. Late-life memory trajectories in relation to incident dementia and regional brain atrophy. *J Neurol.* 2015;262:2484-90.
- 4 Pettigrew C, Martin RC. Cognitive declines in healthy aging: evidence from multiple aspects of interference resolution. *Psychol Aging.* 2014;29:187-204.
- 5 Weuve J, Proust-Lima C, Power MC, Gross AL, Hofer SM, Thiébaud R, et al. Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. *Alzheimers Dement.* 2015;11:1098-109.

- 6 Hofer SM, Sliwinski MJ. Two - design and analysis of longitudinal studies on aging. In: Birren JE, Schaie KW, Abeles RP, Gatz M, Salthouse TA, editors. *Handbook of the psychology of aging*. 6th ed. Burlington: Academic Press; 2006. p. 15-37.
- 7 Brown TA. *Confirmatory factor analysis for applied research*. 2nd ed. New York: Guilford Press; 2015.
- 8 Widaman KF, Ferrer E, Conger RD. Factorial invariance within longitudinal structural equation models: measuring the same construct across Time. *Child Dev Perspect*. 2010;4:10-8.
- 9 Moreira PS, Santos N, Castanho T, Amorim L, Portugal-Nunes C, Sousa N, et al. Longitudinal measurement invariance of memory performance and executive functioning in healthy aging. *PLoS One*. 2018;13:e0204012.
- 10 Salthouse T. *Major issues in cognitive aging*. Oxford: Oxford University; 2010.
- 11 Salthouse TA. Correlates of cognitive change. *J Exp Psychol Gen*. 2014;143:1026-48.
- 12 Blankson AN, McArdle JJ. Measurement invariance of cognitive abilities across ethnicity, gender, and time among older Americans. *J Gerontol B Psychol Sci Soc Sci*. 2015;70:386-97.
- 13 Barnes LL, Yumoto F, Capuano A, Wilson RS, Bennett DA, Tractenberg RE. Examination of the factor structure of a global cognitive function battery across race and time. *J Int Neuropsychol Soc*. 2016;22:66-75.
- 14 Alzheimer's Disease International. *World Alzheimer report 2015: the global impact of dementia* [Internet]. 2015 [cited 2020 Jul 22]. www.alz.co.uk/research/world-report-2015
- 15 Hayden KM, Jones RN, Zimmer C, Plassman BL, Browndyke JN, Pieper C, et al. Factor structure of the National Alzheimer's Coordinating Centers uniform dataset neuropsychological battery: an evaluation of invariance between and within groups over time. *Alzheimer Dis Assoc Disord*. 2011;25:128-37.
- 16 Aquino EM, Barreto SM, Bensenor IM, Carvalho MS, Chor D, Duncan BB, et al. Brazilian Longitudinal Study of Adult health (ELSA-Brasil): objectives and design. *Am J Epidemiol*. 2012;175:315-24.
- 17 Schmidt MI, Duncan BB, Mill JG, Lotufo PA, Chor D, Barreto SM, et al. Cohort profile: longitudinal Study of Adult Health (ELSA-Brasil). *Int J Epidemiol*. 2014;44:68-75.
- 18 Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989;39:1159-65.
- 19 Bertolucci PHF, Okamoto IH, Brucki SM, Siviero MO, Toniolo Neto J, Ramos LR. Applicability of the CERAD neuropsychological battery to Brazilian elderly. *Arq Neuropsiquiatr*. 2001;59:532-6.
- 20 Atkinson RC, Shiffrin RM. The control of short-term memory. *Sci Am*. 1971;225:82-90.
- 21 Fichman HC, Fernandes CS, Nitri R, Lourenço RA, Paradela EM, Carthery-Goulart MT, et al. Age and educational level effects on the performance of normal elderly on category verbal fluency tasks. *Dement Neuropsychol*. 2009;3:49-54.
- 22 Machado TH, Fichman HC, Santos EL, Carvalho VA, Fialho PP, Koenig AM, et al. Normative data for healthy elderly on the phonemic verbal fluency task -- FAS. *Dement Neuropsychol*. 2009;3:55-60.
- 23 Hamdan AC, Hamdan EMLR. Effects of age and education level on the trail making test in a healthy Brazilian sample. *Psychol Neurosci*. 2009;2:199-203.
- 24 Passos VM, Caramelli P, Benseñor I, Giatti L, Barreto SM. Methods of cognitive function investigation in the Longitudinal Study on Adult Health (ELSA-Brasil). *Sao Paulo Med J*. 2014;132:170-7.
- 25 Bertola L, Bensenor IM, Barreto SM, Moreno AB, Griep RH, Viana MC, et al. Measurement invariance of neuropsychological tests across different sociodemographic backgrounds in the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *Neuropsychology*. 2020;34:227-34.
- 26 Kolen MJ, Brennan RL. *Test equating, scaling, and linking: methods and practices*. 3th ed. New York: Springer; 2014.
- 27 Gonzalez J, Wiberg M. *Applying test equating methods: using R*. New York: Springer International Publishing; 2017.
- 28 Gross AL, Inouye SK, Rebok GW, Brandt J, Crane PK, Parisi JM, et al. Parallel but not equivalent: challenges and solutions for repeated assessment of cognition over time. *J Clin Exp Neuropsychol*. 2012;34:758-72.
- 29 Albano AD. *equate: an R package for observed-score linking and equating*. *J Stat Softw*. 2016;74:1-36.
- 30 Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Modeling*. 2007;14:464-504.
- 31 StataCorp. *Stata statistical software: release 13*. College Station: StataCorp LP; 2013.
- 32 Muthén LK, Muthén BO. *Mplus user's guide*. 7th ed. Los Angeles: Muthén & Muthén; 1998-2012.
- 33 Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. *Lancet*. 2017;390:2673-734.
- 34 Mukadam N, Sommerlad A, Huntley J, Livingston G. Population attributable fractions for risk factors for dementia in low-income and middle-income countries: an analysis using cross-sectional survey data. *Lancet Glob Health*. 2019;7:e596-603.
- 35 Suemoto CK, Ueda P, Beltrán-Sánchez H, Lebrão ML, Duarte YA, Wong R, et al. Development and validation of a 10-year mortality prediction model: meta-analysis of individual participant data from five cohorts of older adults in developed and developing countries. *J Gerontol A Biol Sci Med Sci*. 2017;72:410-6.
- 36 Suemoto CK, Ferretti-Rebustini RE, Rodriguez RD, Leite RE, Soterio L, Brucki SM, et al. Neuropathological diagnoses and clinical correlates in older adults in Brazil: a cross-sectional study. *PLoS Med*. 2017;14:e1002267.
- 37 Salthouse TA. When does age-related cognitive decline begin? *Neurobiol Aging*. 2009;30:507-14.
- 38 Hartshorne JK, Germine LT. When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychol Sci*. 2015;26:433-43.