

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

DANILO ASSIS PEREIRA

**Adaptação e Evidências de Validade do Teste Neuropsicológico PBAC
para a População Brasileira**

Tese apresentada como requisito parcial para a obtenção do título de Doutor em Ciências da Saúde pelo Programa de Pós-Graduação em Ciências da Saúde da Universidade de Brasília.

Orientadora: Profa. Dra. Maria Clotilde Henriques Tavares

Co-orientador: Prof. Dr. Carlos Alberto Bezerra Tomaz

Brasília

2014



DANILO ASSIS PEREIRA

**Adaptação e Evidências de Validade do Teste Neuropsicológico PBAC
para a População Brasileira**

Tese apresentada como requisito parcial para a obtenção do título Doutor em Ciências da Saúde pelo Programa de Pós-Graduação em Ciências da Saúde da Universidade de Brasília.

BANCA EXAMINADORA

Profa. Dra. Maria Clotilde Henriques Tavares (Presidente)
Universidade de Brasília – UnB

Prof. Dr. Leandro Fernandes Malloy-Diniz (Examinador Externo)
Universidade Federal de Minas Gerais – UFMG

Prof. Dr. João Carlos Alchieri (Examinador Externo)
Universidade Federal do Rio Grande do Norte – UFRN

Profa. Dra. Corina Satler (Examinador Interno)
Universidade de Brasília, Campus Ceilândia – UnB

Prof. Dr. Harmut Günther (Examinador Interno)
Universidade de Brasília, Instituto de Psicologia – UnB

Prof. Dr. Joaquim Pereira Brasil-Neto (Examinador Suplente)
Universidade de Brasília, Instituto de Biologia – UnB

AGRADECIMENTOS

Agradeço à profa. Dra. Maria Clotilde Tavares por suas orientações e dedicação que tanto contribuíram para minha formação acadêmica. Muito obrigado pela prontidão!

Agradeço ao ilustre prof. Dr. Carlos Tomaz pela co-orientação deste trabalho, pelos ensinamentos e orientação em outras pesquisas e pela generosa amizade. Minha profunda admiração e respeito.

À Dra. Corina Satler pela ajuda na escolha do teste, no tema e incentivo neste trabalho.

Aos amigos e estudantes de medicina (UnB) Luciana Medeiros, Renan Pedroso e da estudante de especialização em Neuropsicologia Luciana Guimarães Vieira (IBNeuro) pela laboriosa coleta dos dados.

Aos queridos colegas de trabalho e professores-colaboradores do IBNeuro que tanto me incentivaram na Neuropsicologia, nas pesquisas científicas e em minha formação como profissional e como docente.

Aos meus pais, Adones e Denilde, e toda a minha família que tanto contribuíram para minha formação pessoal e incentivaram em minha jornada acadêmica.

*“Se um modelo é consistente com a realidade, então os dados deveriam ser consistentes com o modelo. Porém, se os dados são consistentes com o modelo, não significa que o modelo corresponda à realidade” (Bollen, 1989, p. 68 in *Structural Equations with Latent Variables*).*

RESUMO

Instrumentos de rastreio cognitivos são utilizados em todo o mundo por vários profissionais de saúde para detectar indivíduos com provável comprometimento cognitivo. O teste PBAC (*Philadelphia Brief Assessment of Cognition*) foi proposto a partir de uma série de instrumentos neuropsicológicos bem estabelecidos na literatura para avaliar comportamento social, memória, linguagem, percepção visoespacial e funções executivas. A literatura aponta que os subtestes do PBAC são correlacionados com perda de massa encefálica em pacientes com doenças neurodegenerativas. Este estudo tem como objetivo verificar evidências de validade da adaptação brasileira do teste de rastreio cognitivo PBAC usando amostras clínicas e populacionais, comparando-o a outros testes de rastreio cognitivo e funcionais. Participaram deste estudo 325 voluntários entre 18 e 94 anos de idade (mediana de 65), com escolaridade entre 0 e 19 anos (mediana de 13). Quatro amostras distintas foram utilizadas: 103 jovens universitários, 183 idosos hígidos, 24 pacientes portadores de demências (sendo mais comum a doença de Alzheimer) e 16 Kalungas moradores da região de Cavalcante-GO, descendentes dos quilombolas. Outros instrumentos de rastreio cognitivo (MEEM, Teste do Relógio, Trilhas A e B, Stroop) e funcionais (GDS e Pfeffer) foram utilizados numa subamostra para a validação convergente e discriminante. Análises especiais de modelagem de equação estrutural Bayesiana (BSEM) e teoria de resposta ao item multidimensionais (TRI_m) e multigrupo foram utilizadas. O teste PBAC possui boa confiabilidade na amostra geral ($\alpha = 0,86$), mas não na amostra de jovens (0,37). Quatro dimensões distintas foram encontradas: memória, linguagem, habilidade visoconstrutiva e funções executivas. O modelo BSEM mostrou para o grupo de idosos hígidos que a idade influencia negativamente a memória, mas não as demais habilidades. A escolaridade apresentou influência positiva em todos os constructos latentes. Com exceção dos itens Escrita e Trilhas, todos os demais apresentaram maior informação nas habilidades latentes dos idosos com demência, quando comparados aos hígidos nas funções de informações da TRI multigrupo. Tabelas normativas com a habilidade (θ), escore z e escala T foram calculadas para cada dimensão do teste, tendo como norma os dados dos idosos hígidos. O teste de rastreio cognitivo PBAC se mostrou tão sensível e específico quanto o MEEM para a detecção de demência do tipo Alzheimer. O teste é mais informativo em pessoas com habilidades latentes mais baixas. A retirada de alguns itens do teste pode aumentar sua acurácia para o diagnóstico de Alzheimer, no entanto mais estudos com outros grupos clínicos se tornam essenciais, uma vez que sua construção objetivou a avaliação multiclínica.

Palavras-chave: teste de rastreio cognitivo, demência, PBAC, neuropsicologia, psicometria.

ABSTRACT

Cognitive screening tests are used worldwide by many health professionals to detect individuals with probable cognitive impairment. The PBAC (*Philadelphia Brief Assessment of Cognition*) was proposed based on a series of well-established neuropsychological instruments to assess social behavior/comportment, memory, language, visuospatial perception, and executive functions. The literature indicates that the subtests of the PBAC are correlated with loss of brain tissue in patients with several neurodegenerative diseases. This study aims to find evidence of validation of the PBAC using clinical and population samples, and comparing it to other cognitive and functional screening tests. Participated in this study 325 individuals aged between 18 and 94 years-old (median 65), and schooling between 0 and 19 years (median 13). Four different samples were used: 103 university students, 183 healthy elderly, 24 patients with dementia (most common being Alzheimer's disease), and 16 afro-descendants (called Kalungas, living at Cavalcante-GO). All participants signed a consent form. Other cognitive screening instruments (MMSE, Clock Design Test, Trail Tests A and B, and Victoria Stroop Test) and functional tests (Pfeffer and GDS) were used in a subsample for the convergent and discriminant validity. Special analysis of Bayesian structural equation modeling (BSEM), multi-group and multidimensional response theory item (mIRT) were used. The PBAC test showed good reliability in the overall sample ($\alpha = 0.86$), but not in the youth sample (0.37). Four distinct dimensions were found: memory, language, executive functions, and visuospatial perception. The BSEM model showed that in healthy elderly people, age negatively influences memory, but not other skills. Education had a positive influence on all latent constructs. Except Writing and Oral Trails, all other items presented greater information on the latent abilities for the clinical group when compared to the information functions of the healthy elderly group. Normative tables with the ability (theta), z scale, and T scores were calculated for each dimension of the test, using the healthy elderly normative data. The PBAC cognitive screening test proved to be as sensitive and specific as the MMSE for detecting cognitive disorders associated to Alzheimer's dementia. The test is more informative on those individuals with lower latent abilities. The withdrawal of some items can increase PBAC's accuracy for the diagnosis of Alzheimer's, however further studies with other clinical groups are essential since its construction aimed to evaluate other kind of neurodegenerative disease.

Key words: cognitive screening test, dementia, PBAC, neuropsychology, psychometrics.

LISTA DE FIGURAS

Figura 1. Frequência de idade em faixas de cinco anos (N=325).....	30
Figura 2. Representação esquemática dos procedimentos estatísticos.....	35
Figura 3. Modelo de equações estruturais hipotéticos. Em (a), estão os índices das variáveis endógenas com duas variáveis latentes correlacionadas entre si. Em (b) os índices das variáveis exógenas, com apenas uma variável latente. Em (c) uma única variável observável representa uma medida perfeita para o construto latente.....	37
Figura 4. Modelo esquemático de uma análise Bayesiana.....	44
Figura 5. Exemplo da densidade da estimativa posterior Bayesiana dos valores da média, mediana e moda, com os intervalos de credibilidade de 95% (linhas azuis).	45
Figura 6. Normalidade Q-Q (qnorm) para cada grupo.....	50
Figura 7. Densidade de respostas nas categorias dos itens para todos os grupos (N = 325).	51
Figura 8. Densidade gaussiana de Kernel do escore total do PBAC para cada grupo avaliado.	53
Figura 9. Densidade kernel do escore total do PBAC amostra total (N=325, k=0,052).....	53
Figura 10. Dispersão com linhas de tendência e caixas laterais dos escores totais do PBAC e MEEM para cada grupo avaliado. Os pontos das caixas laterais representam os <i>outliers</i> de cada teste.	57
Figura 11. Análise <i>cluster</i> dos itens (N=325). O teste PBAC resulta um modelo com 11 <i>clusters</i> , com ajuste de 0,89; ajuste padrão de 0,98 e RMSR igual a 0,08. As intercorrelações corrigidas para atenuação diagonal da matriz de estrutura foi de 0,91.....	59
Figura 12. <i>Scree plots</i> das análises paralelas dos componentes principais (CP) e análises fatoriais (AF) dos dados atuais (N=325), dados simulados e dados reamostrados.....	61

Figura 13. Comparação entre os modelos Bayesianos (BSEM) dos participantes idosos (n=183). À esquerda (1º modelo), o item Leitura não possui carga significativa em F3 (DIC=4922,52; BIC 5293,74; $pD=68,76$; $PPp=0,446$). No centro-esquerdo (2º mod.), o item Leitura foi excluído (DIC=5145,88; BIC 5463,18; $pD=59,60$; $PPp=0,443$). No centro-direito (3º mod.), modelo de 4 fatores, onde Leitura não é significativo em F4 (DIC=4938,36; BIC 5309,92; $pD=71,86$; $PPp=0,215$). À direita (4º mod.), o modelo de 4 fatores (DIC=4939,35; BIC 5314,58; $pD=70,37$; $PPp=0,183$). 66

Figura 14. Modelo de equação estrutural Bayesiano para o grupo de idosos hígidos (n=181). Efeitos da idade e escolaridade nas variáveis latentes. Informação de ajuste ao modelo: $PPp=0,326$; DIC=5009,003; BIC=5353,062; $pD=62,105$. Algoritmo GIBBS(RW), com 100.000 iterações com 2 canais MCMC. 69

Figura 15. Curvas ROCs comparando o teste PBAC com o padrão-ouro MEEM (n=58; 24 pacientes demenciados e 34 controles)..... 71

Figura 16. Densidades dos escores totais do PBAC para os idosos controle (n=34; linha pontilhada) e pacientes demenciados (n=24; linha contínua). A linha vermelha na vertical sinaliza o ponto de corte de 58,8 pontos. 72

Figura 17. Curva característica do teste (esq.) e função de informação total (dir.) do teste PBAC. (N=325). 73

Figura 18. Regressão polinomial entre o escore do teste PBAC (abscissa) e a habilidade latente (média) dos indivíduos (N=325). $R^2=0,955$; $y=0,001x^2 - 0,002x - 2,23$ 74

Figura 19. Curvas características dos itens (em função da resposta na categoria). 75

Figura 20. Histograma empírico em função da frequência esperada para os grupos de idosos hígidos (n=183, rosa) e demência (n=24, azul)..... 78

Figura 21. Curvas de informações do teste para o grupo de idosos hígidos (azul) e demência (rosa)..... 79

Figura 22. Eficiência relativa do teste PBAC para o grupo Demência (linha rosa), considerando os idosos hígidos como grupo de comparação (linha de base, azul). 80

- Figura 23. Erros padrões do teste PBAC. A curva azul representa o grupo de idosos hígidos e a curva rosa os pacientes portadores de demência. ID = idosos hígidos; DM = pacientes com demência. 81
- Figura 24. Eficiência relativa dos itens. Na linha de base (rosa) os idosos hígidos tiveram o parâmetro de discriminação fixado em zero. Na linha azul, os pacientes demenciados com parâmetro de discriminação variando em diferentes habilidades latentes (Teta). 83
- Figura 25. Esquerda: Histograma e boxplot do Fator 1 (idosos hígidos). Oulier robusto ≤ 10 pontos. Direita: Curva de Informação Total do Fator 1. 85
- Figura 26. Esquerda: Histograma e boxplot do Fator 2 (idosos hígidos). Oulier robusto ≤ 13 pontos. Direita: Curva de Informação Total do Fator 2. 86
- Figura 27. Esquerda: histograma e boxplot do Fator 3 (grupo idosos hígidos). Oulier robusto ≤ 15 pontos. Direita: Curva de Informação Total do Fator 3. 87
- Figura 28. Esquerda: Histograma e boxplot do Fator 4 (Aprendizagem, Fluência e Trilhas), idosos hígidos. Oulier robusto ≤ 8 pontos. Direita: Curva de Informação Total do Fator 4. 88
- Figura 29. Esquerda: Histograma e boxplot do escore total do PBAC, idosos hígidos. Oulier robusto ≤ 52 pontos. Direita: Curva de Informação Total (azul) e erro padrão do teste (rosa). . 89

LISTA DE TABELAS

Tabela 1 - Estatísticas descritivas dos grupos.....	30
Tabela 2 - Estatística descritiva da distribuição do grupo Jovens.	47
Tabela 3 - Estatística descritiva da distribuição do grupo Idosos.	48
Tabela 4 - Estatística descritiva da distribuição do grupo Demência.....	48
Tabela 5 - Estatística descritiva da distribuição do grupo Kalunga.....	49
Tabela 6 - Estatística descritiva da distribuição da amostra total.	49
Tabela 7. Índices de confiabilidade para cada item do teste PBAC.	54
Tabela 8 - Coeficientes de confiabilidade do teste PBAC para cada grupo avaliado.	55
Tabela 9 – Estimativa posterior da correlação Bayesiana entre o teste PBAC e outras escalas de medidas cognitivas e funcionais.	56
Tabela 10 – Análises fatoriais exploratórias robustas (EFA MLR) e comparações entre os modelos (N=325).	60
Tabela 11 – Comparação entre os modelos exploratórios da teoria de resposta ao item multidimensional (TRIm).	63
Tabela 12 – Comparação entre os modelos EFA Bayesianos para o grupo de idosos hígidos (n=183).	65
Tabela 13 – Modelos TRIm confirmatórios de 4 dimensões para a amostra total e para o grupo de idosos hígidos.	67
Tabela 14 – Curvas ROC de 34 idosos controle e 24 pacientes demenciados.....	70
Tabela 15 – Estimativa dos parâmetros TRI bifatorial para a amostra total (N=325).	77
Tabela 16 – Escore do somatório dos itens do Fator 1, tabela normativa padrão e robusta para os idosos hígidos (n=183).....	85
Tabela 17 - Escore do somatório dos itens do Fator 2 (Cópia da Figura de Rey e Orientação de Linhas), tabela normativa padrão e robusta para os idosos hígidos (n=183).....	86

Tabela 18 - Escore do somatório dos itens do Fator 3 (Semântica, Nomeação, Leitura e Escrita), escores Teta, z e T para os idosos hígidos (n=183).	87
Tabela 19 - Escore do somatório dos itens do Fator 4 (Aprendizagem, Fluência e Trilhas), escores Teta, z e T para os idosos hígidos (n=183).	88
Tabela 20 - Escore total do PBAC, escores Teta, z e T para os idosos hígidos (n=183).	89

LISTA DE EQUAÇÕES

Equação 1	$\eta = B\eta + \Gamma\xi + \zeta$	37
Equação 2	$Y = \Lambda_y\eta + \varepsilon$	37
Equação 3	$X = \Lambda_x\xi + \delta$	37

LISTA DE ABREVIATURAS E SIGLAS

a – Parâmetro de Discriminação (*slope*)

DA – Doença de Alzheimer

AIC – Critério de Informação de Akaike

AN – Avaliação Neuropsicológica

AUC – Curva Sob a Área

b – Parâmetro de Dificuldade/Localização (*threshold*)

BIC – Critério de Informação Bayesiana

BSEM – Modelo de Equação Estrutural Bayesiana

bvFTD – Demência Frontotemporal com Alterações Comportamentais

CBS – Síndrome Corticobasal

CFA – Análise Fatorial Confirmatória

CFI – Índice Comparativo de Ajuste (*Comparative Fit Index*)

TCT – Teoria Clássica dos Testes

D.P. – Desvio Padrão

DIC – *Deviance Information Criterion*

DIF – Funcionamento Diferencial dos Itens

e.p. – Erro Padrão

EFA - Análise Fatorial Exploratória

EM - (*Expectation-Maximization*)

FIFA - (*Full-information Item Factor Analysis*)

FIML - (*Full Information Maximum Likelihood*)

g.l. – Grau de Liberdade

GDS – Escala de Depressão Geriátrica

GOF – Índice de Ajuste ao Modelo (*Goodness-Of-Fit*)

CCI – Curva Característica do Item

TRI – Teoria de Resposta ao Item

JOLO - Discernimento de Orientação de Linha de Benton

LCM – Modelo de Classe Latente (*Latent Class Model*)

MCMC – Markov *Chain* Monte Carlo

MHRM – Metropolis-Hastings Robbins-Monro

TRIm – Teoria de Resposta ao Item Multidimensional

ML – Máxima Verossimilhança (*Maximum Likelihood*)

MLR – Máxima Verossimilhança Robusta (*Maximum Likelihood Robust*)

MEEM – Mini Exame do Estado Mental

naPPA – Afasia Progressiva Primária com Agramatismo

OR-LCMs – *Ordered Latent Class Models*

PBAC – Avaliação Cognitiva Breve da Filadélfia (*Philadelphia Brief Assessment of Cognition*)

pD – Número de Parâmetros Posteriores Estimados no Modelo Bayesiano

PPp – Probabilidade Preditiva Posterior

RAVLT - Teste de Aprendizagem Auditivo-Verbal de Rey (*Rey Auditory Verbal Learning Test*)

RMSEA – *Root Mean Square Error of Approximation*

ROC – *Receiver Operating Characteristic*

SABIC – Critério de Informação Bayesiana Ajustada ao Tamanho da Amostra

SEM – Modelo de Equação Estrutural (*Structural Equation Modeling*)

SSCAN – Correlação Canônica Esparsada de Neuroimagem

svPPA – Afasia Progressiva Primária com Alterações Semânticas

T – Escala T

TCT – Curva Característica do Teste

Teta – Nível de Habilidade no Traço Latente

TMT – Teste das Trilhas

z – Escore z

SUMÁRIO

AGRADECIMENTOS.....	3
RESUMO.....	4
ABSTRACT	5
LISTA DE FIGURAS.....	6
LISTA DE TABELAS.....	9
LISTA DE EQUAÇÕES	11
LISTA DE ABREVIATURAS E SIGLAS.....	12
SUMÁRIO.....	14
1 INTRODUÇÃO	16
1.1 Medida e Avaliação Neuropsicológica.....	16
1.2 Objetivos da Avaliação Neuropsicológica.....	17
1.3 A Avaliação Neuropsicológica como uma Avaliação Diagnóstica	19
1.4 Fundamentos de Psicometria: TCT e TRI	20
1.5 PBAC – Proposta de um Instrumento Neurocognitivo.....	25
1.6 Objetivo	27
2 MATERIAIS E MÉTODOS	29
2.1 Sujeitos.....	29
2.2 Instrumentos.....	31
2.3 Adaptações do teste PBAC.....	31
2.4 Procedimentos	33

	15
2.5 Vantagens do Uso do Modelo de Equação Estrutural.....	36
3 RESULTADOS.....	46
3.1 Análise da Distribuição dos Dados.....	46
3.2 Confiabilidade do Teste PBAC.....	54
3.3 Validade Convergente e Validade Discriminante.....	55
3.4 Análise da Dimensionalidade do PBAC.....	58
3.5 Comparação entre Análise Fatorial e TRIm.....	64
3.6 Estimativa do Modelo Psicométrico.....	65
3.7 Sensibilidade e Especificidade do teste PBAC.....	69
3.8 Curvas Características e de Informação – TRI multigrupo.....	72
3.9 Tabelas Normativas e Parâmetro Teta de Habilidade Cognitiva.....	84
4 DISCUSSÃO.....	90
4.1 Limitações do Estudo.....	93
4.2 Próximos Passos: Calibração do PBAC e Análise de Invariância Multigrupo.....	95
5 CONSIDERAÇÕES FINAIS.....	96
6 REFERÊNCIAS BIBLIOGRÁFICAS.....	100
7 ANEXOS.....	113
Anexo A - Aprovação do Comitê de Ética em Pesquisa em Humanos CEP-UnB.....	114
Anexo B - Termo de Consentimento Livre e Esclarecido.....	115
Anexo C - Avaliação Cognitiva Breve de Filadélfia.....	116
Anexo D - Mini Exame do Estado Mental.....	119
Anexo E - PBAC original (em inglês).....	120

1 INTRODUÇÃO

1.1 Medida e Avaliação Neuropsicológica

Toda e qualquer medida (incluindo a medida da cognição humana em Psicologia) “consiste de regras para atribuir símbolos aos objetos de forma que 1) represente quantidades de atributos numéricos (escala), ou 2) define se o objeto incide em categorias iguais ou diferentes a respeito de um atributo (classificação)” (pg. 3) (1). Uma medida, no entanto, requer algumas regras: devem ser claras, práticas em sua aplicação, não demandar conhecimentos ou habilidades dos administradores (além do treinamento básico), e os resultados não podem depender do administrador. O ponto básico da padronização destas regras é que os usuários de um determinado instrumento possam obter resultados similares, e que sejam confiáveis. Na Psicologia, assim como em outras ciências, não se mensura o objeto em si, mas os atributos deste objeto, como a inteligência, peso, altura, etc. Desta forma, a medida requer um processo de abstração, uma vez que um atributo se refira às relações entre objetos de uma dimensão em particular (como uma habilidade cognitiva, por exemplo). Conhecer a natureza de um atributo antes de medi-lo, torna-se uma tarefa imprescindível, pois um atributo pode não existir na forma como é proposto (1).

Um teste psicológico, ou neuropsicológico, deve ser útil como medida, incluindo a extensão para qual o dado obtido deste método possa se encaixar num modelo matemático, na medida de um único atributo, que possa ser replicado ao longo do tempo (caso necessário), que seja válido em vários sentidos e que possa produzir relações de interesse com outras medidas científicas. Ainda, um método de medida pode contribuir para explicar fenômenos importantes (1).

A avaliação neuropsicológica (AN) é uma das principais práticas dos neuropsicólogos (2). Assim como a avaliação psicológica, a AN recebeu influência direta dos psicólogos educacionais Binet e Simon (1908), e Spearman (1904), que desenvolveram inicialmente as medidas de

inteligência. Algumas destas técnicas utilizadas na educação e no recrutamento de soldados para a guerra, como as Escalas de Inteligência de Wechsler e as Matrizes Progressivas de Raven, foram incorporadas na AN (3).

A AN se tornou, nos dias de hoje, uma atividade relevante para neurologistas e neurocirurgiões no auxílio do diagnóstico, da compreensão das capacidades e limitações cognitivas de pacientes e do planejamento de intervenções, como a estimulação cognitiva e a reabilitação (4). Em consultórios, a AN tem ganhado importante destaque na geriatria e gerontologia para a avaliação de idosos, como auxílio no diagnóstico diferencial das demências, no planejamento e na avaliação da eficácia de programas de reabilitação cognitiva e funcional (5, 6). Nas escolas, destacam-se as questões de desenvolvimento cognitivo das crianças, como o transtorno de déficit de atenção e hiperatividade (7), transtornos de aprendizagem (8) e outros tipos de deficiências intelectuais (9) que dificultem a aprendizagem.

1.2 Objetivos da Avaliação Neuropsicológica

Em muitos pacientes, o déficit cognitivo é tão evidente que pode ser observado pelos familiares durante seus afazeres. No entanto, outros défices não são tão evidentes e necessitam ser observados através de tarefas mais complexas (10). A avaliação do déficit psicológico é focada no prejuízo cognitivo, pois em algum grau, o prejuízo cognitivo quase sempre acompanha todas as disfunções cerebrais (2).

Historicamente, os neuropsicólogos adaptaram testes existentes nas avaliações psicológicas, usando-os em pacientes neurológicos ou neuropsiquiátricos, enquanto desenvolviam novos instrumentos para avaliar, especificamente, os funcionamentos cognitivos e comportamentais identificados durante pesquisas neuropsicológicas (2, 11, 12). Em 1996, o Conselho de Representantes da Associação Americana de Psicologia (APA) criou o Grupo de

Trabalho das Qualificações para o Usuário de Testes (TFTUQ, *Task Force on Test User Qualifications*) que elaborou as diretrizes gerais para a elaboração de instrumentos de avaliação, bem como as qualificações necessárias para os profissionais que fizerem usos dos testes (13). Estas diretrizes, obviamente, se estendem aos neuropsicólogos que buscam adaptar, validar e utilizar instrumentos neuropsicológicos.

A AN tem como objetivo avaliar problemas do funcionamento do indivíduo em seu cotidiano, como as capacidades cognitivas, qualidade de vida, emoção e integração social (11). A AN possui seis objetivos principais (2) que podem ser utilizados em diferentes contextos da sociedade: a) *complementação diagnóstico* – como os testes de rastreio usados para discriminar sintomas psiquiátricos e neurológicos (em demências, doenças autoimunes, encefalopatias tóxicas, traumas, etc); b) *planejamento de intervenção e cuidados com o paciente* – onde a avaliação informa sobre o estado cognitivo do paciente, alterações comportamentais e características personológicas (monitoramento do declínio cognitivo ao longo do tempo, capacidade de auto cuidado, lidar com questões financeiras, etc); c) *planejamento de tratamento e remediação* – para delinear a área problema e avaliar as potencialidades para a reabilitação; d) *avaliação do tratamento* – conhecer a eficácia dos programas e serviços de reabilitação prestados; e) *pesquisa* – para se determinar a precisão, sensibilidade, validade e confiabilidade de instrumentos e verificar as alterações cognitivas e comportamentais manifestadas após a patologia cerebral; e f) *Neuropsicologia forense* – que busca responder questões diagnósticas e comportamentais da área jurídica e forense (14).

Uma questão relevante sobre o uso da AN é a sua validade ecológica (6, 15). Entende-se por validade ecológica o quão bem os dados da avaliação refletem o funcionamento diário ou o quão bem podem prever comportamentos futuros. Estudos sobre a acurácia preditiva dos dados neuropsicológicos foram realizados para uma grande diversidade de condições comportamentais, normalmente focados no funcionamento de vida diária (15, 16).

Apesar de vários avanços tecnológicos e psicométricos, um dos principais desafios da AN é lidar com a medida do déficit cognitivo, já que este conceito pressupõe alguma norma, algum nível ideal do funcionamento cognitivo em níveis comportamental e funcional (17, 18).

1.3 A Avaliação Neuropsicológica como uma Avaliação Diagnóstica

A avaliação diagnóstica pode ser bastante útil em diversos contextos clínicos (19). Esta avaliação pode ser conceituada como a implementação de um processo que envolve os seguintes componentes (20): a) a descrição do objetivo da avaliação, b) a descrição de um modelo para as habilidades latentes (também chamadas de traços latentes) de interesse diagnóstico, c) o desenvolvimento e análise das tarefas de avaliação (como as tarefas ou os itens de um teste), d) a especificação de um modelo psicométrico necessário para associar o desempenho nas tarefas às habilidades latentes, e) a seleção de métodos estatísticos para a estimativa do modelo e avaliação dos resultados, e f) o desenvolvimento de sistemas para reportar os resultados da avaliação aos pacientes, médicos, psicólogos e outros.

O propósito da AN pode ser bastante amplo, como diagnosticar uma competência ou um conjunto de habilidades (2, 11, 12, 16, 20). Este propósito pode ter impacto se um determinado atributo, habilidade ou traço latente será observado a partir de uma ou mais variáveis observáveis (contínuas ou discretas), desde que considere dois tipos de habilidades práticas: a análise de dados de avaliações baseadas em modelos de habilidades mais complexas (que pode oferecer uma informação mais rica do que os modelos unidimensionais) (21), e o desenvolvimento de um instrumento que tenha como propósito a avaliação diagnóstica (7, 9). No entanto, para desenvolver um instrumento apropriado (no caso da AN) é necessário que esteja fortemente embasado na teoria cognitiva e bem fundamentado nas relações neurofisiológicas e neurofuncionais (22). Esta última fundamentação pode ser alcançada através de técnicas de imageamento cerebral, como a ressonância magnética funcional, ou através de comparação

entre instrumentos que sejam conhecidos por ativar uma determinada circuitaria cerebral, ou que pacientes com determinadas condições clínicas não sejam capazes de realizá-los (exemplo, pacientes com anomia não sejam capazes de executar tarefas de nomear objetos animados ou inanimados) (23).

Nos casos dos instrumentos neurocognitivos já existentes na literatura, uma questão importante é a quantidade de informação que uma tarefa ou um grupo de itens pode oferecer para uma determinada habilidade latente. Ainda, o número de habilidades avaliadas não pode ser muito grande, pois pode apresentar falta de suporte estatístico; e nem muito reduzido (como excluir ou combinar itens) para não prejudicar a interpretação e a utilidade diagnóstica do instrumento (24).

No desenvolvimento de um teste, é importante conhecer como certas habilidades se interagem entre si e como elas se correlacionam com as variáveis biológicas, sociais e culturais (25, 26). Uma distinção importante para as interações dentro da própria tarefa é a conjuntiva *versus* a compensatória (20). A interação conjunta de habilidades numa tarefa é tida quando todas as habilidades parecem ser necessárias para que o indivíduo seja bem sucedido na tarefa como um todo. Já a interação compensatória, é verificada quando um nível alto em uma habilidade pode compensar um nível mais baixo em outra. Ainda, pode-se pensar numa mistura destas duas interações, bem como na existência de interações entre as habilidades, que podem coexistir nas resoluções das tarefas (10, 20).

1.4 Fundamentos de Psicometria: TCT e TRI

De acordo com Cronbach, um teste psicológico inclui três componentes importantes: os testes envolvem algum tipo de amostra do comportamento, estas amostras precisam ser coletadas de forma sistemática, e o propósito do teste é comparar o comportamento de duas

peças ou mais. (27). Os testes psicológicos são frequentemente categorizados como referentes a um critério (a decisão é feita a partir da habilidade da pessoa, onde há um ponto de corte para definir um grupo que não atinja determinado desempenho) ou referentes a uma norma (onde os testes são usados para comparar pessoas a partir de uma amostra de referência) (28). Para testes baseados em normas, uma população normativa ou representativa bem definida é utilizada para que se possa inferir as classificações dos respondentes a serem comparados.

A Psicometria, por sua vez, é a ciência que estuda os atributos dos testes psicológicos. Três destes atributos são de interesse especial: tipo de informação gerado pelo teste psicológico (score do teste, por exemplo), a confiabilidade dos dados, e questões acerca da validade das informações obtidas a partir das informações do teste psicológico. Desta forma, pode-se dizer que a Psicometria é “o estudo das operações e procedimentos usados para medir a variabilidade no comportamento e conectar estas medidas ao fenômeno psicológico” (p. 21) (28)

A Teoria de Resposta ao Item (TRI), também conhecida como Psicometria Moderna (29), possui vários argumentos que favorecem o seu uso na psicometria, como: a) o erro padrão da medida difere entre os escores (ou padrões de resposta), mas é generalizado entre populações (25, 30); b) um teste mais curto pode ser tão fidedigno quanto uma versão mais extensa (31); c) a comparação entre os escores dos testes entre formas múltiplas é apropriada quando os níveis de dificuldade do teste variam entre indivíduos (32); d) estimativas não enviesadas das propriedades dos itens podem ser obtidas de amostras não representativas (33); e) escores dos testes adquirem significado quando são comparados com a distância (localização/dificuldade) dos itens (34); f) as propriedades de escalas intervalares são obtidas ao aplicar os modelos nas medidas (35); g) os itens com formatos mistos podem produzir resultados relevantes para o score do teste (35); h) as mudanças nos escores podem ser comparadas de forma mais significativa (25, 36); i) a análise fatorial de todos os itens fornece mais informação do que a análise fatorial do teste (21, 37); e j) as características do item pode ser diretamente relacionadas às propriedades psicométricas (34, 38).

Diferentemente da TRI, a Teoria Clássica dos Testes (TCT) é baseada no modelo do escore verdadeiro (no qual o escore observado da pessoa é tipicamente a soma não ponderada das respostas dos itens do instrumento) (25, 38). Ao avaliar a habilidade do indivíduo, esta soma reflete o número de respostas corretas. A TCT assume que os erros de um instrumento não são correlacionados com os escores do traço latente num instrumento diferente (34). O pressuposto de que os erros do instrumento não possuem correlação com os erros num instrumento diferente é considerado um pressuposto “fraco”, porque os instrumentos dependem do tipo de resposta que os indivíduos oferecem (35). O pressuposto da TRI é considerado “forte”, porque a estimativa da localização do indivíduo é invariante com respeito ao instrumento. Então, a acurácia desta estimativa é conhecida para o grupo e para o indivíduo, e a estimativa dos parâmetros avaliados vão além da amostra usada para sua estimação. Na TRI, itens e pessoas são colocadas dentro de um mesmo *continuum* (25). Embora os itens possam diferir em suas localizações, a capacidade de um item em diferenciar indivíduos permanece constante (35). Em geral, cada item oferece informação para estimar a localização da pessoa. Desta forma, o total de informação do instrumento é a soma das informações dos itens.

Schultz-Larsen e colaboradores (39) examinaram um instrumento de rastreio cognitivo bastante conhecido na Neuropsicologia, o Mini Exame do Estado Mental (MEEM) (40, 41), e usaram a TRI para reduzir o número de tarefas do teste. A TRI, assim como o modelo Rasch, passou a ser uma técnica psicométrica bastante utilizada por vários profissionais de saúde (42) para examinar a independência local dos itens, a dimensionalidade do teste e seu funcionamento diferencial dos itens (DIF) (43). Em outro artigo, resultados psicométricos obtidos através das análises da TRI foram avaliados em termos do impacto de versões mais curtas do MEEM (44), avaliando os índices de sensibilidade, especificidade e valores preditivos em relação ao diagnóstico de demência. Estes artigos levantaram diversas questões sobre as limitações, em termos de generalização, das análises psicométricas tradicionais que são realizadas em

amostras com características tão diferentes, como população normal *versus* população clínica (45).

Estes novos modelos psicométricos propõem uma alternativa para a classificação com base na pontuação de corte tradicional – a análise da habilidade ou traço latente (representada na TRI pela letra grega *Teta*) – segundo a qual, para ser classificada (localizada) dentro da faixa da habilidade latente prejudicada (ou preservada), é preciso ter uma série de itens ou respostas incorretas (ou corretas) de modo que a probabilidade de associação da classe seja alta (46). A análise da classe latente pode ser utilizada tanto para classificar os indivíduos quanto para desenvolver pontuações de corte para medidas (35). Schultz-Larsen e colaboradores usaram um modelo misto onde incorporaram as análises da classe latente para o desenvolvimento de pontos de corte para fins de diagnóstico, que foram examinados em relação ao diagnóstico dado aos pacientes (39, 44). Estes autores examinaram, ainda, as curvas ROC (*Receiver Operating Characteristic*), que é a plotagem da sensibilidade no eixo vertical e 1–especificidade no eixo horizontal para cada ponto de corte na escala, bastante usada na Teoria de Detecção do Sinal, a fim de avaliar o desempenho relativo das versões em relação ao MEEM original. Os autores verificaram que uma das duas versões mais curtas do MEEM teve resultados tão bons quanto o teste original. Em outros estudos, foi detectada a existência de itens do MEEM com DIF, *Differential Item Functioning*), que podem ser considerados como itens tendenciosos ou enviesados (39, 44, 46-48).

Em trabalhos realizados recentemente no Laboratório de Neurociências e Comportamento do Instituto de Biologia da Universidade de Brasília, em parceria com o Hospital Regional da Asa Norte (dados não publicados), verificou-se que certas tarefas do MEEM, como as de orientação temporal e espacial, são bidimensionais quando executadas pelos pacientes demenciados; isto é, são necessárias duas habilidades latentes para a execução dos itens, enquanto as mesmas são unidimensionais quando realizadas pelos idosos não demenciados (sendo necessária apenas uma habilidade latente). Estes resultados demonstram que o MEEM é variante,

dependendo da população avaliada. Este teste possui, portanto, viés de medida, não sendo confiável na avaliação das habilidades cognitivas, apesar de bastante útil como teste de rastreio cognitivo. Este viés (o DIF) não ocorreu em outra escala, como a Escala de Depressão Geriátrica de 15 itens (49), quando foram comparados os pacientes demenciados e os idosos cognitivamente preservados ou com declínio cognitivo leve (trabalho em parceria com os pesquisadores Jonas Jardim de Paula e Leandro Malloy-Diniz da UFMG, dados não publicados).

Nesse sentido, uma questão importante a ser ressaltada é o impacto desse viés de medida cognitiva como, por exemplo, o seu valor preditivo em relação ao diagnóstico clínico. Há dois tipos de DIF: uniforme e não uniforme (34, 35). Este último indica que o DIF favorece um determinado grupo em alguns pontos ao longo do contínuo da função cognitiva (35).

Podemos considerar uma situação comum na prática clínica onde uma bateria neuropsicológica é igualmente utilizada para avaliar dois grupos hipotéticos distintos, como um grupo composto por pacientes agnósticos (incapazes de reconhecer certas formas de imagens ou figuras, mas sem prejuízos aparente na memória verbal) e um grupo com pacientes demenciados (incapazes de realizar certas tarefas, em especial as que exijam memória). Os dois grupos podem possuir a mesma média no escore total na bateria, porém o grupo agnóstico visual pode obter escores apropriados nas tarefas de memória, mas não nas tarefas de nomeação ou reconhecimento de objetos, enquanto que o grupo com demência do tipo amnésica pode facilmente identificar e nomear objetos, mas não pontuam adequadamente nas tarefas mnemônicas. Neste caso, haveria uma grande variabilidade nos construtos ou traços latentes entre os dois grupos, porém não seriam detectáveis usando-se as estatísticas clássicas para, como o teste *t* ou análise de variância (ANOVA) (50-52). Uma alternativa para diferenciá-los seria dividir a bateria em várias provas e fazer as análises das variáveis observáveis, uma a uma para ambos os grupos. Porém, isto não garante a observação dos vieses de respostas entre os grupos (50). Uma alternativa mais confiável é avaliar estes vieses item a item, como se faz na análise de DIF na Teoria de Resposta ao Item (53).

Logo, para muitos pesquisadores em Neuropsicologia, cinco perguntas são fundamentais ao elaborar e validar um instrumento neurocognitivo de qualquer natureza: a) os itens que compõem o instrumento operam de forma equivalente entre diferentes populações (como gênero, idade, habilidade, cultura, escolaridade, patologias, etc.)?; b) a estrutura fatorial de um instrumento ou de seu construto teórico é equivalente entre as populações, e é medida pelos itens numa única dimensão da avaliação ou pelos escores de subescalas, usando instrumentos múltiplos? (54); c) as trajetórias de uma estrutura causal são equivalentes entre as populações (no caso dos modelos de equação estrutural ou regressões múltiplas entre grupos)?; d) as médias latentes de certos construtos num modelo diferem entre populações?; e) a estrutura fatorial de um instrumento de medida pode ser replicada entre amostras independentes obtidas da mesma população? (54, 55).

1.5 PBAC – Proposta de um Instrumento Neurocognitivo

Muitas vezes, os profissionais de saúde não possuem ferramentas devidamente calibradas para avaliar a perda cognitiva através de um instrumento de rastreio rápido. Neste contexto, o teste *Philadelphia Brief Assessment of Cognition* (PBAC) foi proposto com o objetivo de avaliar cinco destes domínios cognitivos, a saber: *memória de trabalho, habilidade visioespacial, linguagem, memória episódica e comportamento* (56). Este teste tem sido empregado por Libon e colaboradores em diferentes amostras populacionais e os resultados sugerem que este instrumento é sensível para detectar alterações cognitivas em pacientes portadores de doenças neurodegenerativas.

Técnicas de neuroimagem tem sido utilizadas conjuntamente com o PBAC para correlacionar as diferentes redes neurocognitivas comprometidas em pacientes com degeneração frontotemporal com déficit de fluência verbal (57) e no declínio longitudinal nas demências frontotemporais e Alzheimer (58). O principal estudo buscou relacionar as subescalas

do teste PBAC com os diagnósticos clínicos relatados através de atrofia da massa cinzenta, devidamente verificados através de imageamento de ressonância magnética. No estudo, os pesquisadores obtiveram dados de 108 participantes: 23 com doença de Alzheimer (AD), 31 com demência frontotemporal com alterações comportamentais (bvFTD), 14 com afasia progressiva primária com alterações semânticas (svPPA), 14 com afasia progressiva primária com agramatismo (naPPA), e 17 com síndrome corticobasal (CBS). Os resultados mostraram que alterações na tarefa de comportamento social esteve associada com a atrofia ventral bilateral e medial frontal e o diagnóstico de bvFTD, prejuízos na tarefa de memória estiveram relacionados com a atrofia do precúneo bilateral e hipocampo, junto com o diagnóstico de AD. O baixo desempenho na tarefa construtiva e visoespacial esteve relacionado com atrofia parieto-occipital e têmporo-occipital, além do diagnóstico de CBS. Escores baixos na tarefa de linguagem estiveram relacionados com atrofia temporal esquerda, que se estendeu até à região frontal inferior esquerda e relacionados com o diagnóstico de svPPA. Dificuldades nas funções executivas estiveram relacionadas à atrofia bilateral do córtex pré-frontal dorsal, mas não estiveram relacionadas com nenhum diagnóstico clínico em particular.

Num estudo mais recente (59), o teste PBAC foi utilizado de forma mais extensiva, sendo utilizado como variância cognitiva induzida pelas condições neurodegenerativas. Foram utilizadas técnicas de análise multivariada esparsada, no caso a análise de correlação canônica esparsada de neuroimagem (SSCAN, *sparse canonical correlation analysis for neuroimaging*). Os autores alegaram que estas técnicas multivariadas revelam de forma mais poderosa as relações entre cérebro e comportamento/cognição. Para testar a hipótese de que o teste PBAC mede a integridade de diferentes redes corticais, empregaram a SSCAN para associar as medidas de imagens cerebrais com as informações oferecidas pelas várias dimensões do teste. Tecnicamente, esta abordagem proposta pelos autores permitiu avaliar uma redistribuição ponderada das subescalas do PBAC buscando maximizar a medida psicométrica com a degeneração cerebral observada nas neuroimagens. Ao mesmo tempo, a SSCAN otimizou e

selecionou regiões da matéria cinzenta para maximizar a correlação com o dado psicométrico do teste. Isto resultou num grupo de regiões da matéria cinzenta que puderam ser interpretados com uma rede associada com um determinado domínio psicométrico. Resumidamente, o objetivo dos pesquisadores foi testar a hipótese de que a SCCAN pode empregar as subescalas individuais do PBAC para extrair as redes de matéria cinzenta que são associadas com a variação da habilidade latente/cognição.

O menor desempenho na subescala de *comportamento social* se correlaciona com perda de matéria cinzenta nos córtices pré-frontral medial e inferior. As *tarefas executivas* se relacionaram com a rede cortical que envolve áreas distribuídas nos lóbulos frontal e temporal. A sub-escala de *linguagem* se relaciona com a matéria cinzenta temporal esquerda e parieto-temporal esquerda. A sub-escala de *memória* se relaciona com a atrofia do precúneo, hipocampo e áreas bilaterais do lobo temporal superior, sendo mais proeminente no lado esquerdo que no direito. Prejuízos significativos na *tarefa construtiva/visioespacial* se relaciona com perda bilateral de massa nos lóbulos têmpero-occipital e parieto-occipital (59).

1.6 Objetivo

O objetivo deste estudo foi apresentar evidências de validade do teste de rastreio neurocognitivo PBAC para a população de idosos alfabetizados. Outras amostras foram utilizadas para se conhecer as características psicométricas do teste. Desta forma, quatro grupos foram avaliados: jovens universitários, idosos hígidos, moradores analfabetos de comunidade isolada de Cavalcante-GO (conhecidos como Kalungas pela sua descendência dos quilombolas) e pacientes portadores de demências, principalmente a doença de Alzheimer. A busca de evidências de validade consistiu na avaliação da dimensionalidade do instrumento através da análise de *cluster*, análise fatorial exploratória e confirmatória (validação de constructo), comparação do teste com outros instrumentos (validação convergente e discriminante), obtenção

dos parâmetros de discriminação e dificuldade dos itens, obtenção de modelos psicométricos que utilizem traços latentes, como as equações estruturais. Ainda, buscou-se avaliar sua sensibilidade e especificidade utilizando um grupo clínico e um grupo controle, e oferecer tabelas normativas para a população de idosos hígidos escolarizados. Por fim, discutiu-se como o modelo psicométrico proposto poderá contribuir para a avaliação clínica em idosos.

2 MATERIAIS E MÉTODOS

2.1 Sujeitos

Participaram deste estudo 325 voluntários entre 18 e 94 anos, com mediana de 65 anos de idade. Os dados foram coletados entre 2009 e 2014 por estudantes de psicologia e medicina devidamente treinados para a realização dos procedimentos de testagem. O nível educacional foi mensurado a partir dos anos em que o participante frequentou o ensino formal, variando entre 0 e 19 anos (mediana de 13 anos). Os participantes foram selecionados a partir de quatro critérios definidos: jovens (em sua maioria universitários, entre 18 e 29 anos), idosos hígidos (acima de 60 anos), pacientes portadores de demência (provenientes de dois hospitais públicos de Brasília-DF com probabilidade diagnóstica para a doença de Alzheimer, em sua maioria) e idosos provenientes da comunidade Kalunga, uma comunidade isolada que vive próxima da cidade de Cavalcante-GO. Estes últimos foram selecionados devido ao baixo nível instrucional e pela ingenuidade que os indivíduos apresentam acerca da avaliação psicológica. A Figura 1 mostra a frequência da idade da amostra utilizada.

Os jovens (n=103) foram em sua maioria universitários dos cursos de Saúde e de Humanas provenientes da Universidade de Brasília. Os idosos hígidos (n=183) foram selecionados na população ativa do Distrito Federal (ex.: SESC do Plano Piloto e Coral da 3ª Idade da Universidade de Brasília). O teste de rastreio cognitivo MEEM foi utilizado para excluir prováveis portadores de deficiências cognitivas, adotando como critério de inclusão os idosos com MEEM maior que 18 pontos. Os pacientes portadores de demência foram selecionados a partir de um grupo de pacientes que recebe tratamento psicoeducacional e medicamentoso do Centro de Medicina do Idoso, Hospital Universitário de Brasília (n=14) e do Hospital Regional da Asa Norte (n=10). Todos os pacientes tiveram como diagnóstico principal a doença de Alzheimer, seguindo os critérios do DSM-IV-TR. Estes pacientes foram avaliados por uma equipe que compõe, no mínimo de um geriatra e um neuropsicólogo. Todos os participantes assinaram o Termo de

Consentimento Livre e Esclarecido (ver *Anexo B*), de acordo com a resolução para pesquisa com sujeitos humanos 196/96 CNS/MS, e a pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de Brasília (CEP-FM/UnB) com registro no. 034/2009 (*Anexo A*).

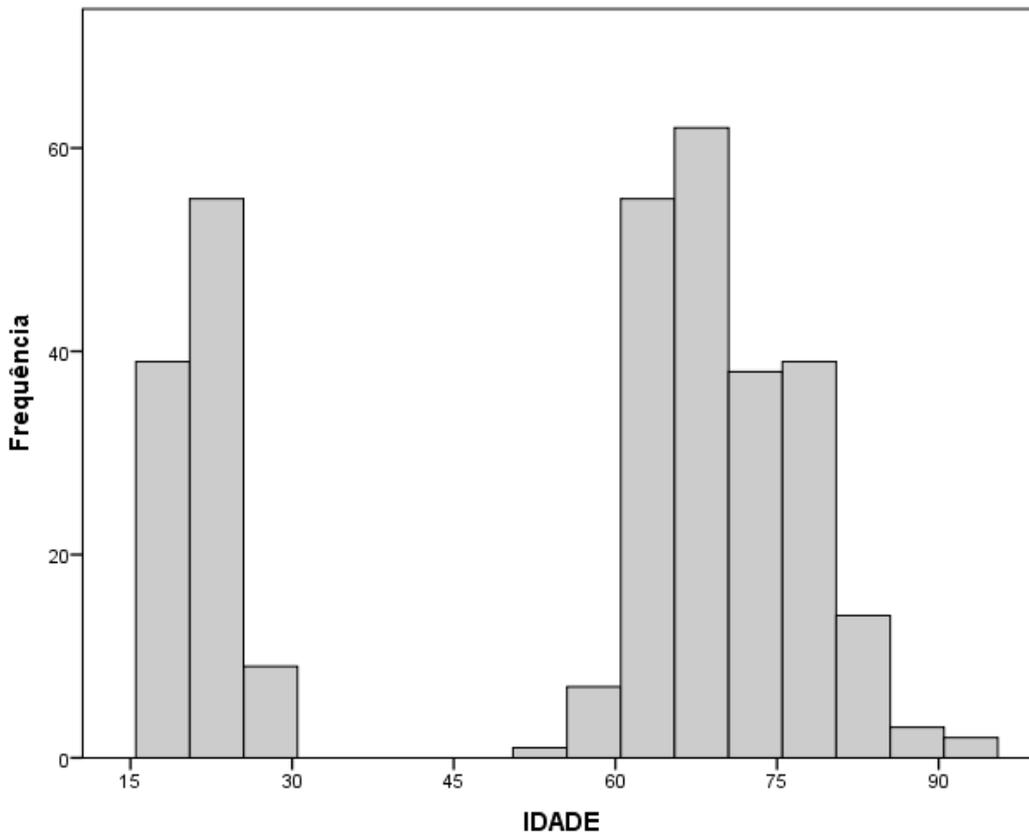


Figura 1. Frequência de idade em faixas de cinco anos (N=325).

Tabela 1 - Estatísticas descritivas dos grupos.

	Jovens (n=103)		Idosos (n=183)		Demência (n=24)		Kalungas (n=16)	
	Média (D.P.)	Med.(min-máx)	Média (D.P.)	Med.(min-máx)	Média (D.P.)	Med.(min-máx)	Média (D.P.)	Med.(min-máx)
Idade	21,7 (2,6)	21 (18-29)	70,3 (6,9)	69,1 (60-91)	72,3 (7,3)	72,5 (61-86)	68,4 (8,9)	66 (55-94)
Escolaridade	13,3 (1,3)	13 (9-16)	12,3 (4,5)	15 (0-18)	11,2 (4,8)	11 (1-16)	1,2 (1,9)	0 (0-5)
MEEM	29 (0,9)	29 (26-30)	27,5 (2,7)	28 (15-30)	20,3 (5,9)	23,5 (6-26)	18,6 (2,5)	19 (15-24)

2.2 Instrumentos

O teste de rastreio cognitivo chamado de Avaliação Cognitiva Breve da Filadélfia (PBAC, *Philadelphia Brief Assessment of Cognition*, 3ª versão) (56) foi criado em 2007 pelo Dr. David Libon e sua equipe do Departamento de Neurologia do Colégio de Medicina da Universidade de Drexel, na Filadélfia, Pensilvânia, e foi generosamente cedido para o Laboratório de Neurociências & Comportamento da Universidade de Brasília. O instrumento foi especificamente compilado para avaliação de rastreio de défices típicos de síndromes neurodegenerativas, tais como a doença de Alzheimer e a demência frontotemporal, e sua aplicação requer entre 15 e 20 minutos (60). A terceira versão do teste consiste de 15 variáveis, agrupadas em 5 sub escalas, variando entre 0 e 100 pontos, que avaliam o prejuízo dos seguintes domínios cognitivos: funções executivas (12 pontos), linguagem (20 pontos), habilidades visio perceptivas (18 pontos), memória episódica visual/verbal (26 pontos) e comportamento social (24 pontos).

Além do PBAC, outros testes foram utilizados para verificar as evidências de validade convergente e discriminante: Mini Exame do Estado Mental (MEEM), *Victoria Stroop Test* que mensura a efetividade da atenção focalizada e também função executiva (uma vez que requer controle inibitório), Teste do Relógio de Sunderland, Trilhas A e B, Escala de Depressão Geriátrica (de 15 pontos) e o Inventário de Vida Diária de Pfeffer.

2.3 Adaptações do teste PBAC

O teste PBAC, originalmente escrito em língua inglesa (ver Anexo E), foi traduzido para o português (ver Anexo C) por dois pesquisadores fluentes em ambos os idiomas. A tradução reversa foi utilizada para comparar as versões. Além da tradução, o teste sofreu algumas adaptações para a cultura brasileira, como modificações nos estímulos visuais da tarefa de nomeação.

No subteste *habilidade de memória de trabalho/execução* contém uma versão reduzida do Teste de Fluência Verbal Fonêmica, originalmente desenvolvido por Thurstone em 1938 (*Thurstone's Word Fluency Test*) (61), que consiste em medir a capacidade do indivíduo em gerar palavras que comecem com uma letra específica (ex.: FAS) e categoria semântica (ex: animais). Na Neuropsicologia, esta tarefa mostrou ser bastante sensível em lesões do lobo frontal, lobo temporal e núcleo caudado, na doença de Alzheimer, na doença de Huntington, amnésia e traumatismo crânio-encefálico (61). No PBAC, esta tarefa foi simplificada, consistindo em evocar apenas as palavras que se iniciam com a letra F, em 60 segundos. Outra tarefa consiste numa versão oral do Teste das Trilhas parte B (TMT, *Trail Making Test*) que busca avaliar o controle executivo, porém sem o componente visual e grafomotor do TMT e bastante utilizado clinicamente em pacientes com traumatismo (62, 63). A tarefa consiste em ligar oralmente números e letras em sequência, tais como 1-A, 2-B, 3-C, etc.

A *habilidade de linguagem* consiste em nomear imagens provenientes do Teste de Nomeação de Boston, criado em 1978 por Kaplan, Goodglass e Weintraub (64, 65). No PBAC, apenas seis imagens foram utilizadas, sendo 4 de objetos inanimados: serrote, cogumelo, aipo e pegador de gelo (que na versão brasileira foram substituídos pelo martelo, caju, alicate e abacaxi) e 2 de objetos animados: camelo e pelicano (que foram substituídos pelo macaco e arara/papagaio). As outras tarefas são: escrever uma sentença simples, leitura de palavras escritas e fala conversacional (observações da fala, como pausas, circunlocução, parafasias, agramatismos, disartria, esforço e discurso).

As *habilidades visoespaciais ou visoconstrutivas* são avaliadas através das tarefas de copiar uma versão simplificada da Figura Complexa de Rey (66) e do Discernimento de Orientação de Linha de Benton (JOLO, *Benton Judgment of Line Orientation*), onde o objetivo é combinar as linhas em um estímulo alvo, criado para avaliar pacientes neurológicos com défices visoespaciais (67).

A habilidade de *memória episódica visual/verbal* conta com a evocação tardia da versão simplificada da Figura Complexa de Rey e a aprendizagem de uma lista de seis palavras com três repetições e tentativas consecutivas, além da evocação livre e tardia desta lista de palavras. Esta tarefa foi baseada no Teste de Aprendizagem Auditivo-Verbal de Rey (RAVLT, *Rey Auditory Verbal Learning Test*), que foi desenvolvido em 1916 por Edouard Claparède, posteriormente adaptado por André Rey em 1958 e publicado em 1964 em seu livro *L'exame clinique en psychologie* (68). Originalmente, o RAVLT consiste numa lista de 30 palavras, sendo 15 na lista A e 15 na lista B. A lista A é lida em voz alta e é solicitado ao sujeito que se lembre do maior número de palavras possível, não importando a ordem. O processo é repetido cinco vezes, criando uma curva de aprendizagem entre as tentativas. Depois da quinta tentativa, a lista B é lida e pede-se que o sujeito a recorde. Esta lista serve como uma tarefa de interferência. Se o indivíduo evocar palavras da lista A quando lhe for solicitado para evocar da lista B, denomina-se este fenômeno de interferência proativa. Logo após a evocação da lista B, solicita-se ao sujeito que evoque as palavras da lista A, sem que ela seja apresentada neste momento. Caso o indivíduo evoque palavras da lista B, denominamos esta interferência de retroativa (68, 69).

A habilidade *comportamento social* é avaliada através da observação clínica de seis domínios comportamentais: apatia e iniciativa empobrecida, desinibição, comportamento social, agitação e irritabilidade, comportamento obsessivo ou ritualístico e falta de empatia. Devido à ausência de variabilidade de respostas para esta habilidade, os escores referentes à mesma não foram consideradas nas análises desta pesquisa.

2.4 Procedimentos

O procedimento psicométrico adotado neste trabalho é uma modificação dos modelos OR-LCMs (*Ordered Latent Class Models*) (70) para variáveis contínuas. Estes modelos são utilizados

para que se possa permitir ao avaliador o uso do escore total para decidir se um paciente apresenta alguma condição clínica detectável, como na triagem cognitiva.

As características psicométricas do teste PBAC foram analisadas observando primeiramente a distribuição e as frequências de respostas em cada categoria de cada item, buscando identificar o tipo de distribuição para cada grupo avaliado. Em seguida, o teste PBAC foi comparado com outras escalas de medidas cognitivas e funcionais (validade convergente e discriminante), e foram obtidos os coeficientes de confiabilidade dos itens e do teste para cada grupo, segundo os pressupostos da Teoria Clássica dos Testes.

Análises dimensionais foram realizadas para identificar a estrutura fatorial do PBAC. Para esta identificação, foram utilizadas análise de *cluster*, a análise paralela dos componentes principais e análises dos fatores para a obtenção dos autovalores, utilizando dados da amostra, dados simulados e reamostrados. Em seguida, as análises fatoriais exploratórias (EFA) e confirmatórias (CFA) foram realizadas e seus índices de ajuste ao modelo (GOF, *Goodness-Of-Fit*) foram comparados, buscando identificar a estrutura dimensional com menor erro residual. Após a identificação do modelo fatorial, foram avaliados os modelos de equação estrutural (SEM, *Structural Equation Modeling*) para a amostra total e para o grupo de idosos hígidos, seguindo duas abordagens: a frequentista e a Bayesiana. O objetivo é verificar se uma das abordagens é superior à outra para descrever o modelo estrutural do teste. Critérios de informações de Akaike (AIC) e Bayesianas (BIC), assim como o qui-quadrado do CFI (*Comparative Fit Index*) e valores do RMSEA (*Root Mean Square Error of Approximation*) serão utilizados como índices para a comparação entre os modelos empiricamente obtidos.

Utilizando as curvas ROC (*Receiver Operating Characteristic*) semiparamétricas e não-paramétricas, as áreas sob as curvas foram estimadas usando um subgrupo dos idosos hígidos e o grupo de idosos portadores de demência. Este subgrupo dos idosos hígidos foi selecionado

para que efeitos da idade e escolaridade não interfiram nas características de sensibilidade e especificidade e para a obtenção do ponto de corte clínico.

Procedimentos da Teoria de Resposta ao Item foram utilizados para a obtenção dos modelos de classes latentes (LCM, *latent class models*), das curvas características dos itens e curva característica do teste (estimativas das habilidades cognitivas dos indivíduos). Procedimentos especiais da função item-resposta multigrupo foram realizados para descrever a relação entre o atributo latente de interesse e a probabilidade em responder uma categoria particular do item, comparando simultaneamente o grupo de idosos hígidos e demenciados. Análise de DIF (*differential item functioning*) foi realizada afim de verificar possíveis vieses de resposta aos itens entre estes dois grupos.

Por fim, tabelas normativas informando o escore total de cada subteste, Teta (habilidade do indivíduo naquele escore), escore z e T foram obtidos para fins de classificação. Abaixo, uma representação esquemática dos procedimentos psicométricos adotados (Figura 2).

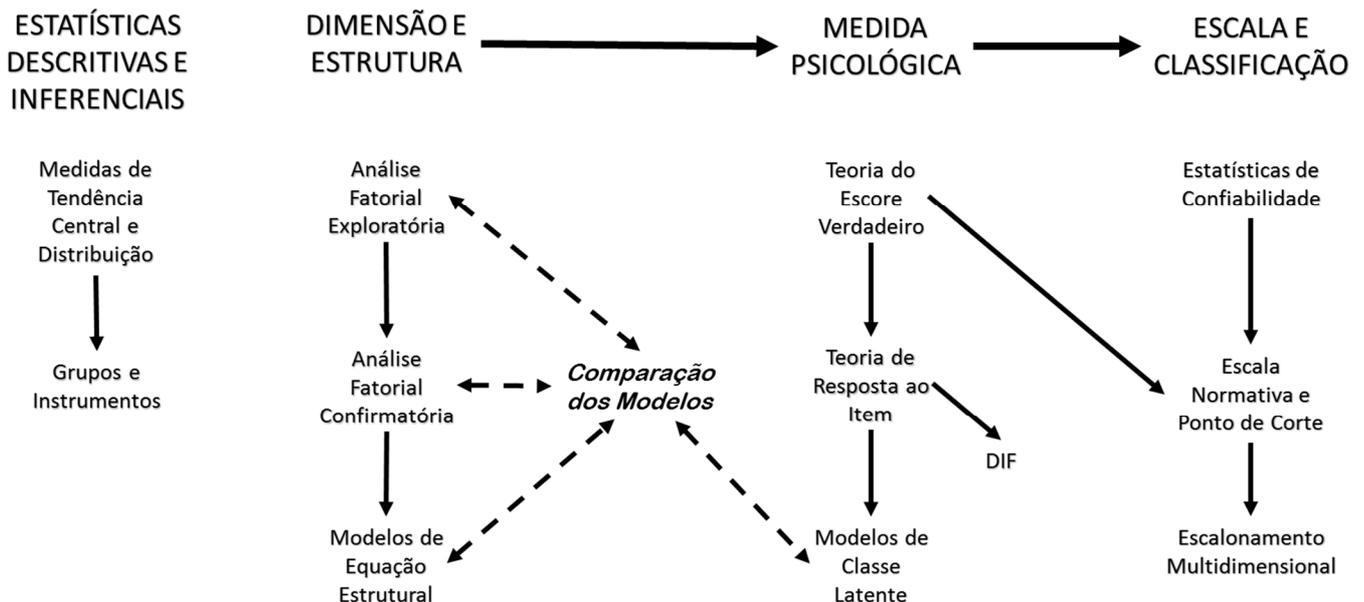


Figura 2. Representação esquemática dos procedimentos estatísticos.

2.5 Vantagens do Uso do Modelo de Equação Estrutural

Muitos construtos estudados na Neuropsicologia, como a inteligência, memória, atenção, velocidade de processamento, linguagem, etc., não podem ser observados diretamente (estes construtos hipotéticos são chamadas na psicometria de variáveis latentes) e tornando-se necessário a utilização de indicadores observáveis, como tarefas, provas ou perguntas (chamadas de itens) para mensurá-los (37). No entanto, analisar estas variáveis latentes requer análises estatísticas mais específicas, como a análise fatorial e modelos de equação estrutural. Pode-se dizer que o modelo de equação estrutural (SEM) consiste de duas partes: a medida do modelo que liga as variáveis observáveis às variáveis latentes não observáveis (fatores), e equações estruturais que ligam as variáveis latentes à cada outra via um sistema de equações simultâneas (71).

Nas duas últimas décadas, o SEM se espalhou rapidamente para várias áreas de conhecimento, como a psicologia, psiquiatria, biologia e ciências da saúde (72). Comparada com os métodos estatísticos tradicionais como a regressão múltipla, ANOVA, análise de trajetória (*path analysis*) e modelos multiníveis, as vantagens do SEM incluem os erros de medida (o que são ignorados nos métodos tradicionais), os modelos simultâneos de variáveis dependentes múltiplas, o teste geral de ajuste ao modelo, os efeitos de estimativa direta, indireta e total, a possibilidade de testar hipóteses complexas e específicas, de gerenciar dados complexos (como distribuição não normal, dados censurados (*censored*) ou categóricos), de testar a invariância dos parâmetros do modelo entre múltiplos grupos, de conduzir modelagem mista para lidar com a heterogeneidade da população, dentre outras (73).

Como acima mencionado, diferentemente dos métodos estatísticos tradicionais, o SEM tem como foco as variáveis latentes no lugar das variáveis observáveis (73). Normalmente, cinco etapas são realizadas para se obter um modelo de equação estrutural: formular, identificar, estimar, avaliar/comparar e modificar o modelo (51).

O diagrama SEM usa caixas para representar as variáveis observáveis (itens) e círculos para as variáveis latentes (fatores). As relações entre as variáveis são representadas por linhas, e a ausência delas indica que nenhuma relação foi hipotetizada entre as variáveis correspondentes. A linha com seta única representa a relação direta hipotetizada entre duas variáveis (com a cabeça da seta apontando para a variável que é influenciada pela outra variável). A seta bidirecional se refere às relações ou associações (e não efeitos) entre as variáveis (Figura 3).

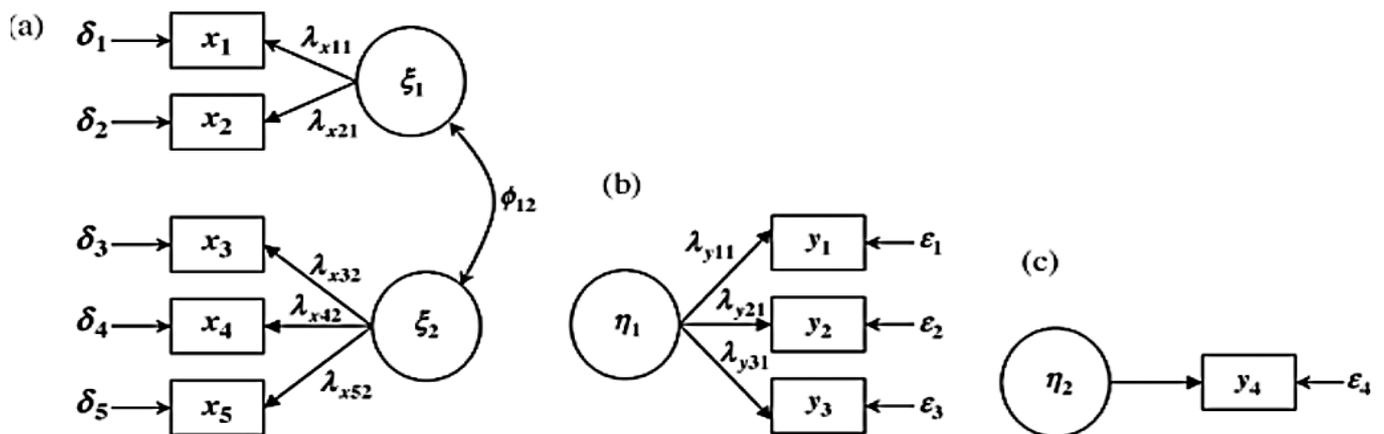


Figura 3. Modelo de equações estruturais hipotéticos. Em (a), estão os índices das variáveis endógenas com duas variáveis latentes correlacionadas entre si. Em (b) os índices das variáveis exógenas, com apenas uma variável latente. Em (c) uma única variável observável representa uma medida perfeita para o construto latente.

No modelo SEM, há p variáveis observáveis dependentes $y = (y_1, \dots, y_p)$, q variáveis observáveis independentes $x = (x_1, \dots, x_q)$, e m variáveis latentes endógenas $\eta = (\eta_1, \dots, \eta_m)$, ou exógenas $\xi = (\xi_1, \dots, \xi_m)$, formando o modelo geral de SEM:

Equação 1 $\eta = B\eta + \Gamma\xi + \zeta$

Equação 2 $Y = \Lambda_y\eta + \varepsilon$

Equação 3 $X = \Lambda_x\xi + \delta$

Estas três equações são expressas em formato de matrizes. A primeira equação representa o modelo estrutural que estabelece as relações ou equações estruturais entre as variáveis latentes. As variáveis latentes endógenas η e as exógenas ξ são conectadas por um sistema de equação linear com coeficientes das matrizes B (beta) e Γ (gama), assim como o vetor residual ζ (zeta), onde Γ representa os efeitos das variáveis latentes exógenas nas variáveis latentes endógenas, B representa os efeitos das variáveis latentes endógenas em outras variáveis latentes endógenas, e representa os termos residuais da regressão (73).

A segunda e a terceira equação representam os modelos das medidas que definem as variáveis latentes das variáveis observáveis. A segunda equação liga os indicadores endógenos (as variáveis observáveis y) às variáveis latentes endógenas (η), enquanto que a terceira equação liga os indicadores exógenos (as variáveis observáveis x) às variáveis latentes exógenas (ξ). As variáveis observáveis y e x são relacionadas às suas variáveis latentes correspondentes η e ξ pelas cargas fatoriais Λ_y (lambda y) e Λ_x (lambda x). O ε e δ são as medidas de erros associadas com as variáveis observáveis y e x , respectivamente (73).

No SEM, η pode incluir múltiplos conjuntos de análises fatoriais exploratórias (EFA) e confirmatórias (CFA) (21). Os fatores da CFA são identificados ao associar um fator a um conjunto de indicadores, como os itens. Itens específicos podem ter carga em um ou mais fatores. Na primeira etapa do processo, uma estrutura fatorial não restrita é estimada (realizado na EFA). Na segunda etapa, as restrições são impostas para confirmar a estrutura encontrada (o que corresponde a CFA) (74). As estatísticas de ajuste ao modelo são comparadas, podendo mudar as restrições impostas de modo a obter os melhores índices.

Enquanto a EFA é uma técnica exploratória aplicada para avaliar a estrutura fatorial de um instrumento (37, 75) a CFA é um modelo mais parcimonioso (73, 76), uma vez que calcula as relações apenas com itens específicos para os fatores (geralmente obtidas após a EFA ou pelos construtos teóricos pertinentes).

Os procedimentos de estimativa do SEM minimizam os resíduos que são as diferenças entre as variâncias/covariâncias da amostra e as variâncias/covariâncias estimadas a partir do modelo. Há muitas funções de ajuste ou procedimentos de estimativa disponíveis para a estimativa do modelo. O mais comum é a função de máxima verossimilhança (ML, *Maximum Likelihood*). No entanto, quando a não normalidade ameaça a validade deste teste de significância, alternativas de ajuste podem ser realizadas (77-79). Uma destas análises robustas, a MLR (*Maximum Likelihood Robust*), usa uma estimativa “sanduíche” com erros padrões robustos (80). Esta estimativa robusta é recomendada para amostras pequenas e médias e usa estratégias modernas para tratar os dados ausentes (*missing values*), como a estimativa FIML (*Full Information Maximum Likelihood*), que é mais eficiente e menos enviesada do que as abordagens tradicionais de *listwise* ou *pairwise* (81). Outra alternativa, acaba de ser proposta por Asparouhov e Muthén, que utilizam métodos de distribuição *t* (dentre outras soluções ainda não muito utilizadas, porém recém implementadas no software MPlus (versão 7.2)) para estimar SEM e modelos mistos com distribuições não-normais e assimétricas (82)

De forma combinada, índices de ajustes ao modelo (GOF) foram obtidos através da análise fatorial confirmatória (CFA), buscando conhecer a estrutura dimensional do instrumento e decidir qual modelo oferece melhor ajuste ao conjunto de itens (37, 83). Os índices, ou critérios, utilizados para medir o ajuste foram o critério de informação de Akaike (AIC), o índice de ajuste comparativo (CFI), $-2 \log\text{likelihood}$ e o qui-quadrado (χ^2). Estes resultados servem para a validação fatorial do instrumento (83).

Para a verificação de aspectos de validade do constructo, modelos de equações estruturais foram utilizados para testar e estimar as relações causais entre as variáveis, usando uma combinação de dados estatísticos e pressupostos qualitativos (32). Os modelos SEM permitem análises exploratórias (onde os parâmetros são livremente estimados) e confirmatórias (onde as hipóteses são testadas e uma é selecionada para representar um modelo causal).

A invariância fatorial de uma escala consiste de dois tipos diferentes de invariância: a invariância da medida (padrões ou valores das cargas fatoriais, variância do item e do erro) e a invariância da estrutura (que envolve a invariância das variâncias e covariâncias e médias dos fatores) (53, 84). Testar a invariância da medida garante que os itens do teste estão medindo os mesmos construtos teóricos (variáveis latentes ou fatores) em todos os grupos, o que permite a comparação entre os grupos (21, 50, 51, 53, 73). Se a invariância da medida não se mantém, a análise da medida não produz resultados com significado e as diferenças encontradas entre os grupos podem ser interpretados de forma ambígua. Nos testes estatísticos clássicos, como o teste *t* ou ANOVA, a invariância da medida entre os grupos analisados é um pressuposto que, muitas vezes, não se confirma (52). A invariância da medida envolve quatro níveis diferentes: invariância da configuração, invariância de mensuração fraca, invariância de mensuração forte e invariância de mensuração rigorosa (53, 84).

A invariância da configuração do teste (ou invariância do padrão) é definida como o mesmo número de fatores e o mesmo padrão de cargas fatoriais livres ou fixas entre grupos sem restrições de igualdade ou outro parâmetro do modelo (21, 37, 53, 75). Esta invariância é condição necessária para testar a invariância dos parâmetros de medida (carga fatorial, intercepto do item e variância do erro) que pertencem à característica da medida dos indicadores observados. Para se estabelecer este modelo, uma linha de base deve ser estabelecida para a CFA que melhor se ajusta para cada grupo avaliado, mas não precisa ser idêntico entre os grupos (51, 53).

A invariância de mensuração fraca é definida como a invariância das cargas fatoriais entre os grupos (84). As cargas fatoriais são, de fato, os coeficientes de inclinação da regressão dos indicadores ou itens observados em seus respectivos fatores e representa a força das relações lineares entre os itens e seus fatores (53). Se as cargas fatoriais são invariantes entre os grupos, então as medidas estão na mesma escala (invariância métrica). Quando se observa invariância da carga fatorial, as variáveis latentes são medidas da mesma forma em todos os grupos (85).

A invariância da mensuração forte é definida como a invariância das cargas fatoriais e interceptos dos itens entre os grupos. Um intercepto do item representa a origem ou a escala da medida. Se os interceptos dos itens são não-invariantes entre os grupos, isto indica que os participantes, em pelo menos um dos grupos, tendem a responder os itens sistematicamente de forma mais alta ou mais baixa na escala de estudo, mesmo que as cargas fatoriais sejam invariantes entre os grupos (85). Quando a invariância das cargas fatoriais e os interceptos dos itens permanecem, os escores dos itens de diferentes grupos deveriam ter a mesma métrica e a mesma escala, então as médias dos grupos podem ser comparadas (53).

Por fim, a invariância de mensuração rigorosa requer invariância métrica (das cargas fatoriais), escalar (dos interceptos dos itens) e da variância dos erros. No entanto, testar a invariância dos erros não tem muita utilidade prática (53). A invariância da mensuração é pré-requisito para se testar a invariância estrutural, testando a invariância dos parâmetros (tais como a variância, a covariância e a média do fator) entre os grupos (83). A não invariância dos parâmetros estruturais não significam problemas com o instrumento ou teste estudado, mas simplesmente a heterogeneidade da população. Os pesquisadores esperam que o instrumento opere de forma equivalente entre diferentes grupos ou populações (86). No entanto, não há nenhuma razão para acreditar que os níveis das variáveis latentes em estudo, ou que as relações entre as variáveis latentes permaneçam imutáveis entre diferentes populações. Torna-se, portanto, necessário garantir a invariância da medida entre os grupos e, então, testar a invariância da estrutura do instrumento ou comparar os parâmetros estruturais entre as populações (86).

Para se testar a invariância estrutural, os seguintes passos devem ser seguidos (53, 87):
1º - Teste de invariância da variância do fator (a variância do fator é uma medida de variação ou dispersão dos escores dos fatores); 2º - Teste de invariância da covariância dos fatores (útil para saber se as relações entre os fatores permanecem imutáveis em diferentes populações); 3º - Teste de invariância da média do fator (um dos grupos é tratado como grupo de referência e

estima-se a diferença deste com os outros grupos. No grupo de referência as médias dos fatores são fixadas em zero e nos grupos de comparação as médias dos fatores são parâmetros livres. Testar a invariância da média do fator ou comparar estas médias entre os grupos, as cargas fatoriais e os interceptos dos itens devem ser invariantes entre os grupos, ou seja, os escores dos itens observados nos diferentes grupos precisam ter a mesma métrica e a mesma escala de medida para poder comparar as médias dos fatores entre os grupos) (88).

Enfim, mudanças na CFI (ΔCFI) tem sido utilizadas para avaliar invariância em modelagem multi-grupo. Uma ΔCFI menor ou igual a 0,01 entre os modelos é considerada como um critério de invariância. Uma ΔCFI maior que 0,01 indica uma mudança significativa no ajuste do modelo para se testar invariância (88).

Teste de invariância da configuração – após o modelo de linha de base ser determinado para cada amostra, estes modelos são combinados num modelo multi-grupo para formar o modelo configuracional, no qual o mesmo número de fatores e o mesmo padrão são fixados e as cargas fatoriais ficam livres para cada um dos grupos, sem impor restrições de igualdade em nenhuma medida ou no parâmetro estrutural entre os grupos. Então, o modelo é estimado, simultaneamente, com todos os grupos. Apenas se a invariância configuracional for demonstrada entre os grupos, podem-se testar as invariâncias, tais como a invariância métrica, a invariância escalar, a variância de fator e a invariância da covariância e a invariância da média do fator (88).

Em suma, os parâmetros de interesse para responder questões relacionadas à invariância multi-grupo, no caso os idosos e os jovens, são: a) cargas fatoriais, b) covariâncias do fator, c) trajetórias das regressões estruturais, e d) médias dos fatores latentes.

2.6 BSEM: Modelo de Equação Estrutural Bayesiano

As análises frequentistas, que utilizam a máxima verossimilhança (ML) para as estimativas dos modelos de equação estrutural, e as análises Bayesianas diferem-se pela forma como tratam os parâmetros: como constantes ou como variáveis (89). As análises frequentistas encontram as estimativas através da maximização de uma probabilidade calculada para os dados, utiliza-se de valores intervalos de confiança e valores p para testar a significância da hipótese nula. As análises Bayesianas, por outro lado, combinam uma distribuição prévia para os parâmetros com a probabilidade dos dados para formar uma distribuição posterior para as estimativas dos parâmetros, oferecendo uma estimativa na forma de média, mediana ou moda da distribuição posterior. Utiliza-se do intervalo de credibilidade (ou graus de conhecimento), uma vez que os dados ou os resultados são indefinidos, carregando incertezas probabilísticas (89, 90).

As análises fatoriais Bayesianas, assim como os modelos de equações estruturais Bayesianas (BSEM, *Bayesian Structural Equation Modeling*) se tornaram mais populares nos últimos anos, em parte devido aos novos algoritmos computacionais, como o método Markov *chain* Monte Carlo (MCMC) (90). No entanto, as análises requerem conhecimento específico de softwares (como o WinBUGS, OpenBUGS, JAGS, SAS, MPlus ou pacotes específicos do R), e requerem especificações estatísticas mais complexas, como a utilização das estimativas prévias (91). Vários são os argumentos em favor do uso das análises Bayesianas, como a) as estimativas dos parâmetros não precisam possuir distribuição simétrica ou normal, pois o intervalo de credibilidade baseado nos percentis da distribuição posterior permite curvas fortemente assimétricas; b) não possui como pressuposto a teoria de grandes amostras, desempenhando análise superior mesmo com amostras menores que os demais modelos frequentistas, como o ML; c) lida melhor com modelos que possuam maior número de parâmetros, enquanto que modelos ML não oferecem uma solução tão “natural” (92). Os métodos Bayesianos permitem, no entanto, estimar modelos mais complexos, que muitas vezes não são possíveis através das

estimativas frequentistas tradicionais (92). Filósofos da ciência já assumiram que os métodos Bayesianos são ideias para inferência científica (93).

Em suma, o teorema Bayesiano é formado pela estimativa posterior, que é comparada com a estimativa prévia (ou anterior) dos parâmetros avaliados. Os dados informam sobre o parâmetro anterior, que é então considerado como uma nova variável e redistribuído, junto com a probabilidade da distribuição dos dados, de acordo com a informação prévia daquele parâmetro (ver Figura 4). Os valores posteriores são então obtidos dividindo os dados pelos parâmetros prévios (esta divisão é a probabilidade da distribuição), multiplicando esta divisão pelos parâmetros prévios e dividindo o resultado pelos dados, sendo expresso por parâmetros posteriores = α probabilidade x parâmetros prévios (94).

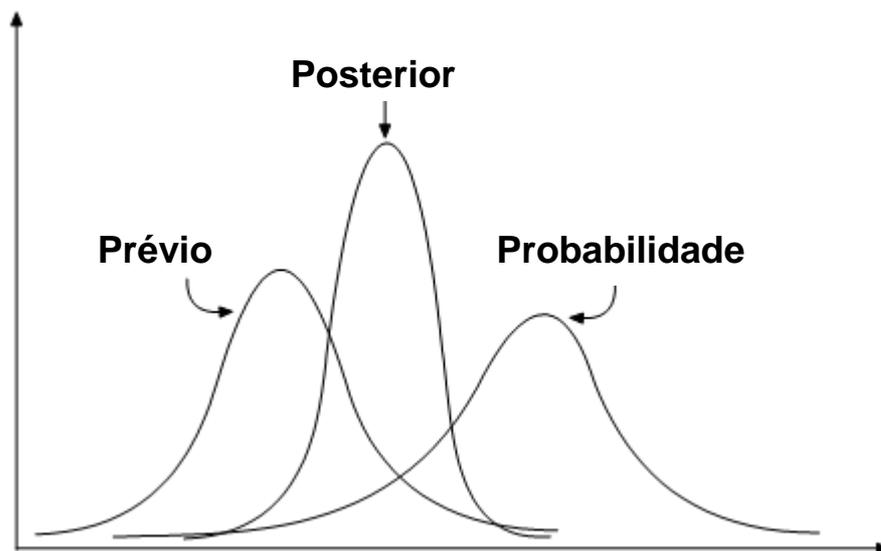


Figura 4. Modelo esquemático de uma análise Bayesiana.

As análises de modelo de equação estrutural Bayesiana deste trabalho seguirão as recomendações de Muthén e Asparouhov (95), usando o software MPlus (versão 7.2). É importante notar que o BSEM refere-se à abordagem Bayesiana usando parâmetros prévios com pequena variância (mas não fixados em zero, comumente realizado na análise frequentista).

Ainda, cargas fatoriais cruzadas (*cross-loadings*) serão analisadas na BSEM, uma vez que não pode ser avaliada nas análises fatoriais confirmatórias tradicionais (uma vez que a carga noutro fator é fixado em zero).

No caso do teste PBAC, os valores prévios dos parâmetros foram obtidos a partir da estimativa ML dos próprios dados em combinação com pequenas variâncias informadas, e a distribuição posterior obtida a partir do método MCMC, usando algoritmos GIBBS. A avaliação do ajuste ao modelo foi feita a partir da checagem preditiva posterior, proposta por Gelman (96, 97) e Scheines (98). Deve-se notar que o valor preditivo posterior p não se comporta como o valor p para o teste qui-quadrado de ajuste ao modelo no SEM frequentista. Por fim, as distribuições posteriores para as cargas cruzadas foram utilizadas como índices de modificação do modelo para os parâmetros livres em que o intervalo de credibilidade não seja igual a zero (ver Figura 5).

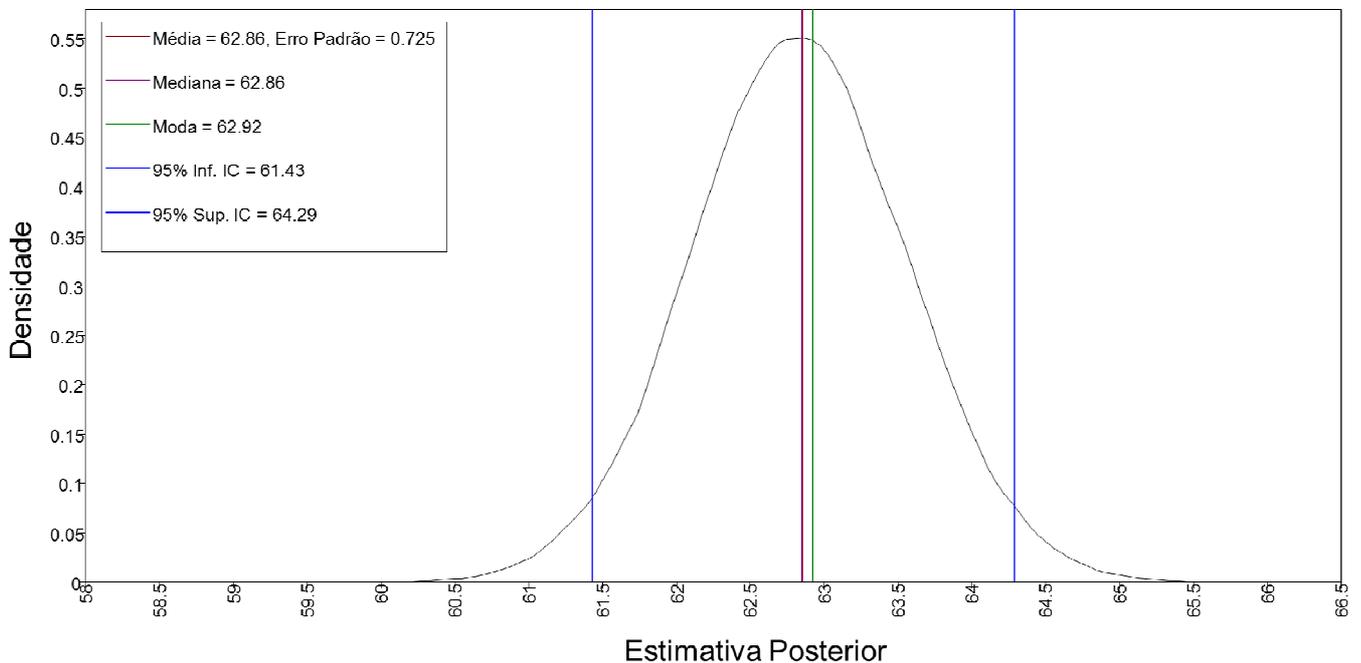


Figura 5. Exemplo da densidade da estimativa posterior Bayesiana dos valores da média, mediana e moda, com os intervalos de credibilidade de 95% (linhas azuis).

3 RESULTADOS

A estimativa posterior da média Bayesiana da amostra total no teste PBAC foi de 62,9 pontos com o intervalo de credibilidade (I.C. 95%) entre 61,4 e 64,3, com mínima de 7 e máxima de 76 pontos (excluindo os 24 pontos correspondentes à *habilidade de comportamento social*, devido à ausência de variância nas observações). A estimativa posterior da correlação Bayesiana entre o escore total do PBAC e a variável idade dos 325 indivíduos foi negativa ($r = -0,46$), (I.C. 95% entre -0,53 e -0,37). O valor da probabilidade preditiva posterior referente à diferença dos valores dos qui-quadrados observado e replicado (PPp) foi de 0,42 (considerado adequado, por ser próximo a 0,5); entre PBAC e a variável escolaridade foi positiva ($r = 0,68$; I.C. 95% = 0,62 a 0,73; PPp=0,49); e entre as variáveis idade e escolaridade foi negativa ($r = -0,23$; I.C. 95% = -0,33 a -0,12; PPp=0,49).

3.1 Análise da Distribuição dos Dados

O primeiro passo da análise psicométrica foi checar a distribuição da frequência de respostas para cada item em cada grupo, calculando a normalidade dos itens, através da assimetria e curtose (99). A assimetria é uma medida da distribuição da probabilidade de uma variável randômica/aleatória sobre a sua média obtida a partir de dados reais. Seu valor pode ser positivo ou negativo. Para uma distribuição unimodal, a assimetria negativa indica que a cauda esquerda da função de densidade/probabilidade é maior que a cauda direita. No entanto, este valor não representa a relação entre a média e a mediana. Já a curtose é uma medida de dispersão, que caracteriza o pico ou achatamento da curva da função da distribuição de probabilidade. A curtose da curva normal é 0. Seu limite inferior é -2, e não há limite superior. Se o valor da curtose é maior que 0, então a curva é mais afunilada e concentrada que a distribuição

normal (leptocúrtica), com valores se afastando da média. Se o valor for menor que 0, então a função é mais achatada que a distribuição normal (platicúrtica).

Para estas análises, foi utilizado o pacote estatístico R (*versão 3.0.3 for Windows*) (100). As Tabelas 2 a 5 mostram as estatísticas descritivas da tendência central e da dispersão para cada grupo: média, desvio padrão, mediana, média aparada em 20% (onde 20% dos menores e maiores valores são retirados da amostra), escore mínimo e máximo, assimetria e curtose. A Tabela 6 mostra as estatísticas da amostra total.

Tabela 2 - Estatística descritiva da distribuição do grupo Jovens.

Item	n	média	d,p,	mediana	média 20%	min	max	assimetria	curtose
Aprendizagem Verbal	103	6,0	0,1	6,0	6,0	5	6	-9,9*	96,1*
Fluência	103	5,2	1,0	6,0	5,4	1	6	-1,2	1,4
Evocação	103	5,2	1,0	6,0	5,4	2	6	-0,9	-0,3
Reconhecimento	103	5,6	0,8	6,0	6,0	4	6	-1,5	0,3
Nomeação	103	6,0	0,2	6,0	6,0	5	6	-4,1*	15,3*
Semântica	103	3,0	0,2	3,0	3,0	1	3	-9,9*	96,1*
Cópia F, Rey	103	12,0	0,2	12,0	12,0	11	12	-5,5*	28,7*
Orientação Linhas	98	5,9	0,7	6,0	6,0	0	6	-6,8*	51,6*
Evocação F, Rey	99	10,8	1,3	11,0	11,1	6	12	-1,1	0,9
Leitura	103	4,0	0,0	4,0	4,0	4	4	NaN	NaN
Escrita	103	3,0	0,0	3,0	3,0	3	3	NaN	NaN
Trilhas	103	4,9	1,2	5,0	5,2	0	6	-1,8	4,3
PBAC Total	103	70,7	3,9	71,0	71,4	55	76	-1,4	2,4
MEEM Total	103	29,1	1,0	29,0	29,2	26	30	-0,8	0,1

* *Efeito teto. NaN=sem variabilidade de resposta.*

Tabela 3 - Estatística descritiva da distribuição do grupo Idosos.

Item	n	média	d.p.	mediana	média 20%	min	max	assimetria	curtose
Aprendizagem Verbal	183	5,7	0,5	6,0	6,0	3	6	-2,1*	5,0*
Fluência	183	4,6	1,3	5,0	4,8	1	6	-0,6	-0,5
Evocação	183	4,1	1,6	4,0	4,4	0	6	-1,0	0,4
Reconhecimento	183	5,5	0,9	6,0	5,9	0	6	-2,5*	8,3*
Nomeação	183	6,0	0,2	6,0	6,0	4	6	-7,8*	65,5*
Semântica	183	2,8	0,6	3,0	3,0	0	3	-3,6*	12,4*
Cópia F. Rey	183	11,5	1,8	12,0	11,9	0	12	-5,2*	28,6*
Orientação Linhas	181	4,9	1,4	6,0	5,3	0	6	-1,2	0,5
Evocação F. Rey	183	7,9	2,9	9,0	8,3	0	12	-1,0	0,6
Leitura	183	4,0	0,2	4,0	4,0	2	4	-10,7*	118,2*
Escrita	183	2,9	0,4	3,0	3,0	0	3	-6,2*	40,9*
Trilhas	183	3,6	1,9	4,0	4,0	0	6	-0,7	-0,7
PBAC Total	183	63,5	8,6	66,0	65,3	23	74	-2,1	5,6
MEEM Total	183	27,4	2,7	28,0	28,0	13	30	-2,0	5,4

* *Efeito teto.*

Tabela 4 - Estatística descritiva da distribuição do grupo Demência.

Item	n	média	d.p.	mediana	média 20%	min	max	assimetria	curtose
Aprendizagem Verbal	24	4,9	1,3	5,0	5,2	1	6	-1,4	1,3
Fluência	24	3,5	1,8	3,0	3,5	0	6	0,0	-1,2
Evocação	24	1,8	1,9	2,0	1,5	0	5	0,4	-1,5
Reconhecimento	24	3,8	2,0	4,0	4,0	0	6	-0,4	-1,1
Nomeação	24	5,4	1,4	6,0	5,9	0	6	-2,7*	7,3*
Semântica	24	2,2	1,2	3,0	2,5	0	3	-0,9	-1,0
Cópia F. Rey	24	8,7	4,6	11,0	10,1	0	12	-1,0	-0,6
Orientação Linhas	24	3,4	1,9	4,0	3,4	0	6	-0,2	-1,2
Evocação F. Rey	24	3,6	3,8	2,5	3,0	0	10	0,4	-1,5
Leitura	24	3,8	0,5	4,0	4,0	2	4	-2,7*	6,5*
Escrita	24	2,6	0,9	3,0	2,9	0	3	-2,1*	3,2*
Trilhas	23	2,4	1,9	3,0	2,4	0	6	0,0	-1,3
PBAC Total	24	46,0	17,1	50,5	47,9	7	69	-0,5	-0,9
MEEM Total	24	20,3	5,9	23,5	21,5	6	26	-0,9	-0,5

* *Efeito teto.*

Tabela 5 - Estatística descritiva da distribuição do grupo Kalunga.

Item	n	média	d.p.	mediana	média 20%	min	max	assimetria	curtose
Aprendizagem Verbal	16	4,2	1,6	5,0	4,8	1	6	-1,1	-0,5
Fluência	16	0,9	1,4	0,0	0,4	0	4	1,0	-0,5
Evocação	16	2,4	1,8	3,0	2,3	0	5	0,1	-1,5
Reconhecimento	16	4,5	0,6	5,0	4,7	3	5	-0,9	-0,5
Nomeação	15	4,8	1,5	5,0	5,2	1	6	-1,1	0,1
Semântica	15	2,1	1,1	3,0	2,4	0	3	-0,8	-1,0
Cópia F. Rey	15	5,1	4,6	4,0	5,0	0	12	0,1	-1,8
Orientação Linhas	15	1,1	1,3	1,0	0,9	0	4	0,8	-0,5
Evocação F. Rey	15	1,0	1,8	0,0	0,2	0	5	1,4	0,3
Leitura	15	1,3	2,0	0,0	0,9	0	4	0,6	-1,7
Escrita	15	0,7	1,2	0,0	0,2	0	3	1,1	-0,6
Trilhas	15	0,2	0,6	0,0	0,0	0	2	2,4*	4,4*
PBAC Total	15	28,5	11,1	27,0	27,0	16	46	0,4	-1,6
MEEM Total	15	18,7	2,5	19,0	18,8	15	24	0,1	-0,6

* *Efeito chão.*

Tabela 6 - Estatística descritiva da distribuição da amostra total.

Item	n	média	d.p.	mediana	média 20%	min	max	assimetria	curtose
Aprendizagem Verbal	325	5,7	0,8	6,0	6,0	1	6	-3,6*	16,1*
Fluência	325	4,5	1,6	5,0	4,9	0	6	-1,1	0,6
Evocação	325	4,2	1,8	5,0	4,5	0	6	-1,0	0,2
Reconhecimento	325	5,4	1,1	6,0	5,8	0	6	-2,1	5,4
Nomeação	325	5,9	0,6	6,0	6,0	0	6	-6,4*	48,3*
Semântica	325	2,8	0,6	3,0	3,0	0	3	-3,3*	9,9*
Cópia F. Rey	325	11,1	2,6	12,0	12,0	0	12	-3,5*	11,5*
Orientação Linhas	318	4,9	1,7	6,0	5,5	0	6	-1,5	1,1
Evocação F. Rey	321	8,1	3,6	9,0	8,9	0	12	-1,0	0,1
Leitura	325	3,9	0,7	4,0	4,0	0	4	-5,0*	23,8*
Escrita	325	2,8	0,7	3,0	3,0	0	3	-3,8*	13,0*
Trilhas	324	3,8	2,0	4,0	4,2	0	6	-0,8	-0,6
PBAC Total	325	62,9	12,9	67,0	66,4	7	76	-2,0	3,9
MEEM Total	325	27,0	3,9	28,0	28,1	6	30	-2,2	5,5

* *Efeito teto.*

As tabelas acima mostram que alguns itens possuem efeito de teto (onde a probabilidade de acerto é maior que 95%) para a maioria dos grupos, como as tarefas Nomeação e Leitura. Na amostra total, metade dos 12 itens tiveram efeito de teto (curtoses acima de 5), mostrando uma

distribuição não-normal. O efeito chão (onde a probabilidade de acerto é menor que 5%) foi obtido no teste de trilhas para o grupo Kalungas.

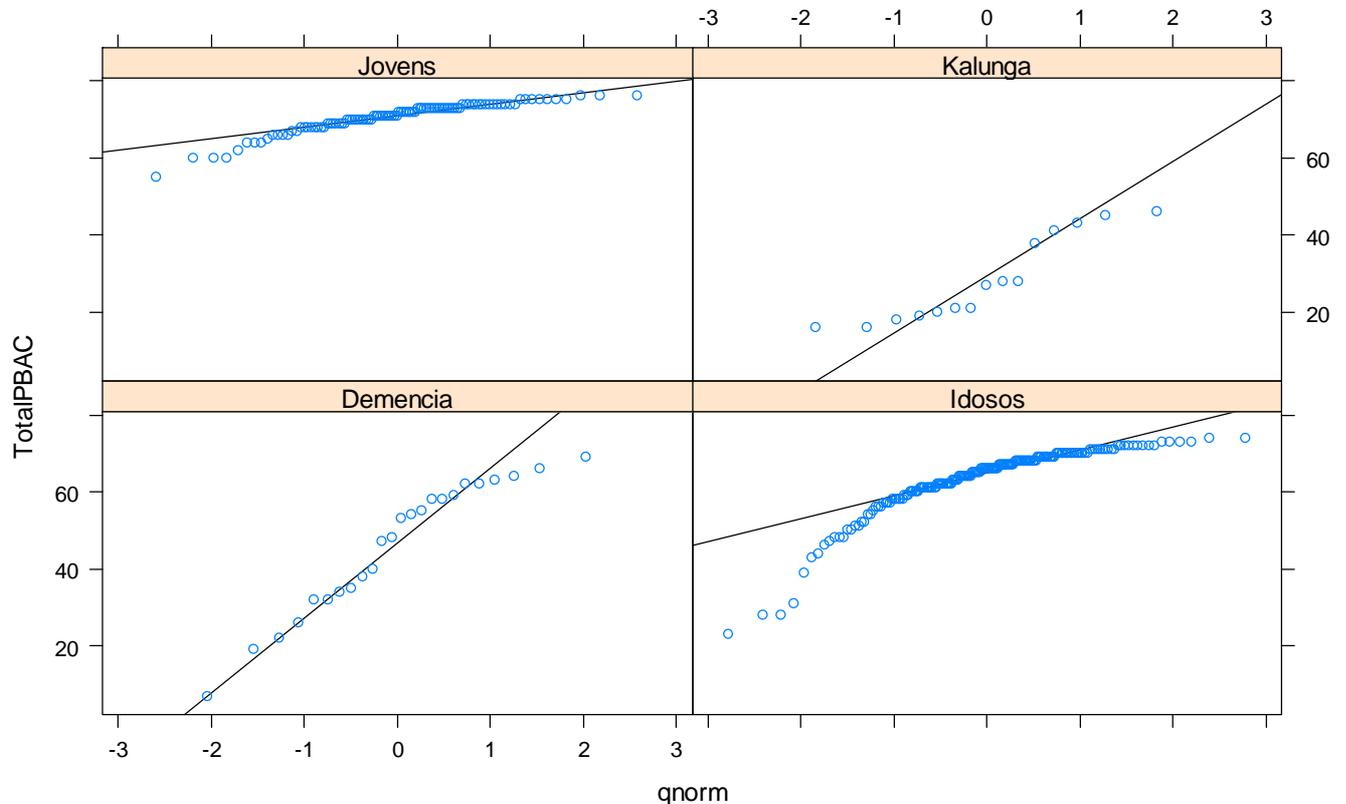


Figura 6. Normalidade Q-Q (qnorm) para cada grupo.

A técnica de transformação de Box-Cox (101) conhecida como Q-Q foi realizada para todos os grupos (Figura 6). Visualmente, é possível observar a distribuição não-normal de forma mais acentuada no grupo de Jovens e Idosos hígidos, uma vez que numa distribuição normal a linha de referência (linha contínua) teria ângulo aproximadamente de 45° , com os pontos uniformemente distribuídos sobre ela. Para os itens, a densidade de resposta em cada categoria foi mostrada na Figura 7. Nota-se que itens como Aprendizagem Verbal, Nomeação, Semântica, Leitura e Escrita tiveram frequência notavelmente maior nas categorias de respostas mais elevadas, mostrando pouca variabilidade de respostas nestes itens, mesmo para todos os sujeitos da amostra ($N = 325$).

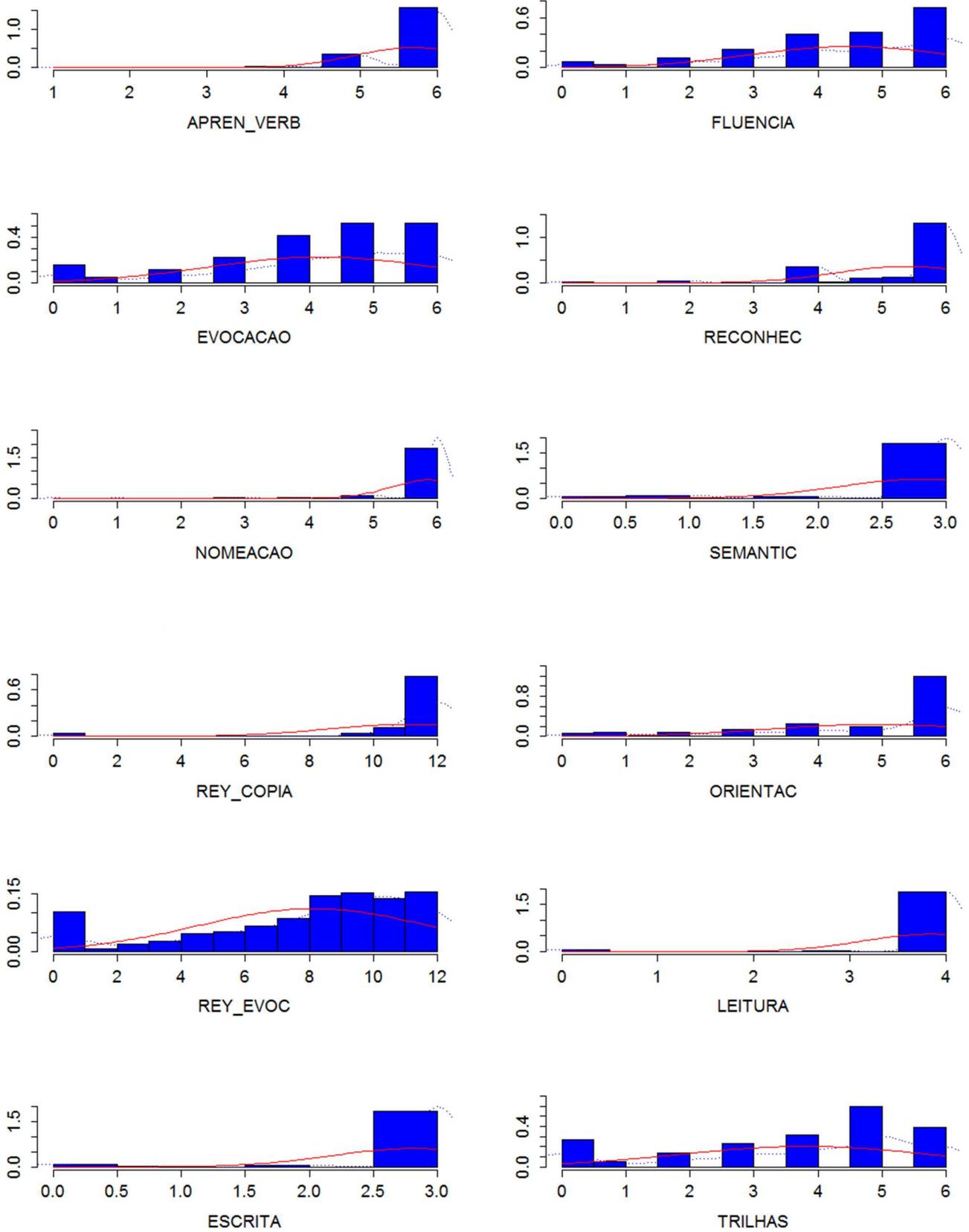


Figura 7. Densidade de respostas nas categorias dos itens para todos os grupos (N = 325).

Para o escore total do PBAC, a distribuição dos dados contínuos foi estimada através da densidade de Kernel, que é uma forma não paramétrica para estimar a probabilidade da função de uma determinada variável, sendo um dado fundamental para se fazer inferências sobre a população baseadas numa amostra finita (102), além de auxiliar na decisão sobre quais estatísticas utilizar. A Figura 8 mostra as densidades Kernel para cada grupo avaliado. O grupo de Jovens possui densidade maior à direita, com pico elevado e curva estreita, sinalizando que grande parte dos respondentes deste grupo obteve escore máximo, apresentando pouca variabilidade de resposta. Já o grupo de pacientes demenciados (linha contínua) apresenta larga base (achatada), mostrando que houve maior variabilidade de resposta.

A largura da banda de Kernel é um parâmetro livre obtido para informar quão ótima é a estimativa da amostra, tendo como referência a estimativa normal. Quanto mais distante de zero, mais distante é a distribuição dos dados da amostra de referência em relação à estimativa normal. Como era de se esperar de um instrumento de rastreio cognitivo, a distribuição da frequência dos escores obtidos pelos sujeitos da amostra não são normais (103, 104), concentrados na porção à direita, sugerindo efeito de teto no desempenho de grande parte das tarefas cognitivas apresentadas. A Figura 9 mostra a densidade *kerden*, uma transformação da densidade de Kernel com base na função gaussiana assimétrica, onde os limites da curva não ultrapassam os valores mínimos e máximos do escore total.

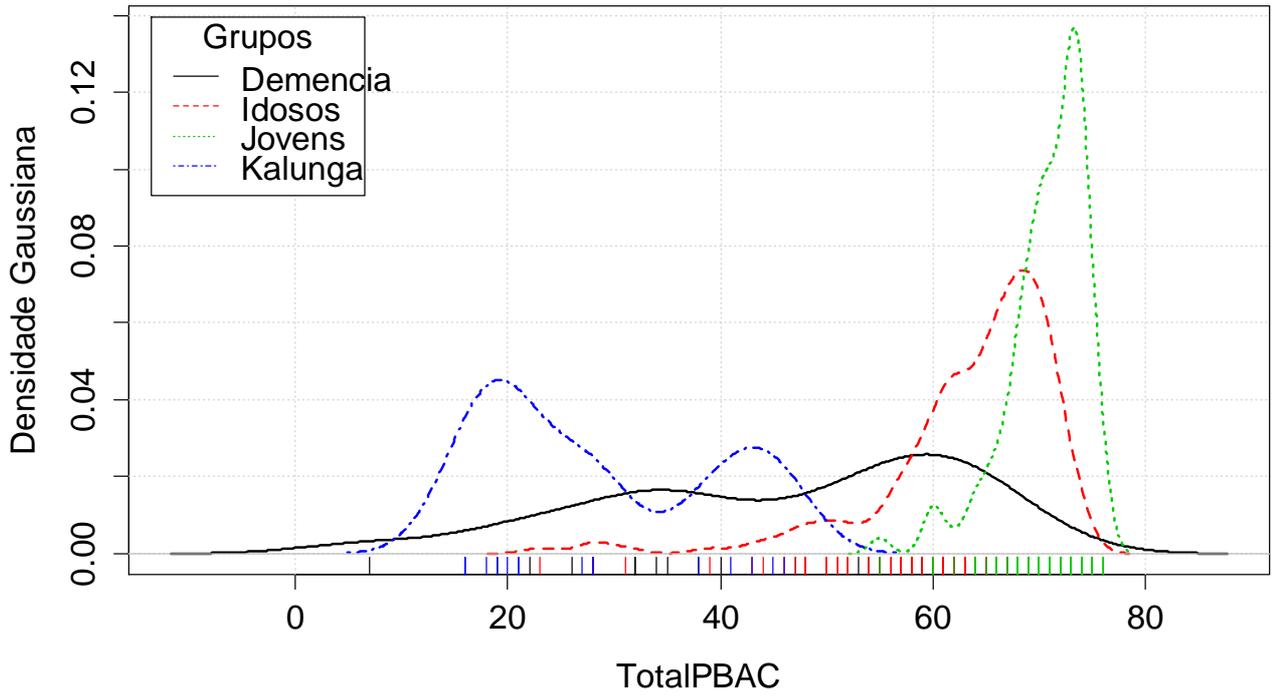


Figura 8. Densidade gaussiana de Kernel do escore total do PBAC para cada grupo avaliado.

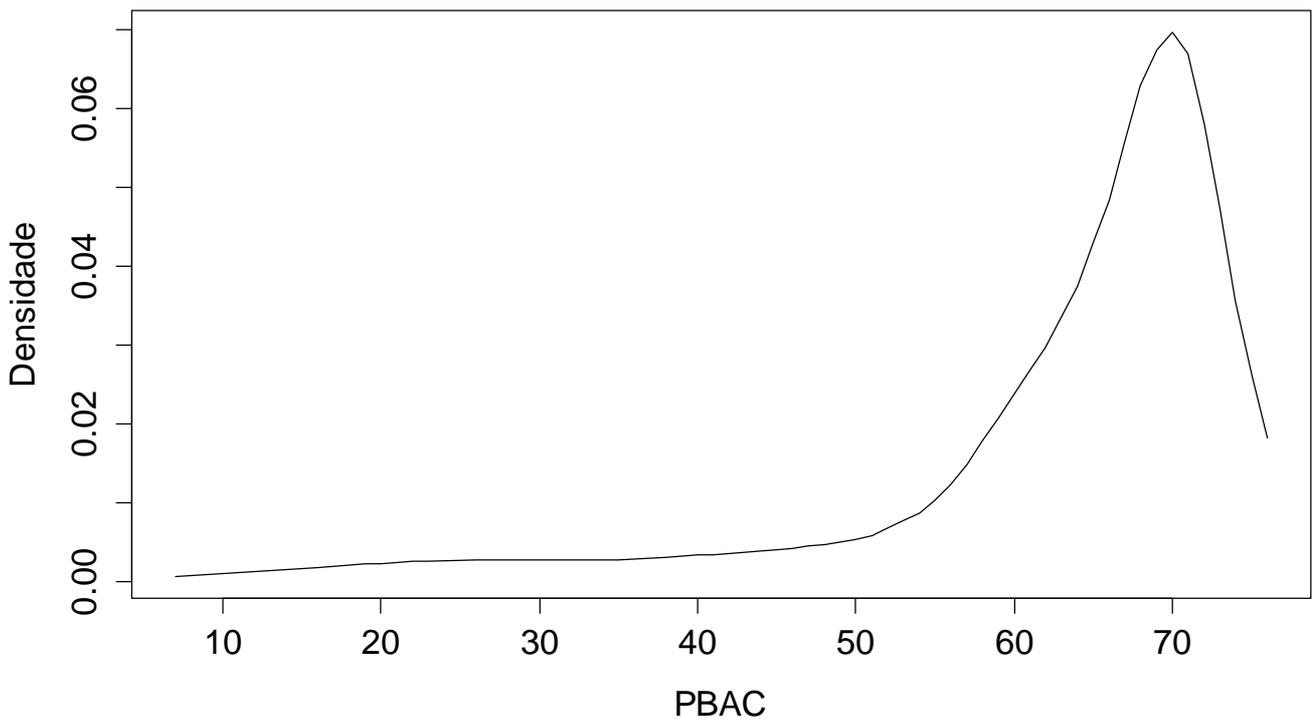


Figura 9. Densidade kerden do escore total do PBAC amostra total (N=325, k=0,052).

3.2 Confiabilidade do Teste PBAC

As análises de confiabilidade do teste foram realizadas através do pacote 'psych' (versão 1.4.5) (105) do software estatístico R. Além das estatísticas descritivas de médias e desvios padrões, foram calculadas as correlações de Pearson (r) do item com o teste, do item com o resto (que é o escore total do teste excluindo o item), a covariância inter-item e o alfa de Cronbach para cada item (ver Tabela 7).

Tabela 7. Índices de confiabilidade para cada item do teste PBAC.

Item	N	Alfa	G6 (smc)	r_{it} (item, total)	r_{ir} (item, resto)
Aprendizagem	325	0,86	0,91	0,73	0,65
Fluência	325	0,85	0,91	0,71	0,63
Evocação	325	0,85	0,91	0,63	0,59
Reconhecimento	325	0,86	0,92	0,62	0,56
Nomeação	325	0,86	0,91	0,69	0,57
Semântica	325	0,86	0,91	0,59	0,50
Cópia Rey	325	0,84	0,91	0,83	0,75
Orientação Linha	318	0,85	0,91	0,71	0,67
Evocação Rey	321	0,85	0,91	0,81	0,80
Leitura	325	0,86	0,91	0,68	0,58
Escrita	325	0,86	0,91	0,78	0,68
Trilhas Orais	324	0,85	0,92	0,67	0,63

O alfa de Cronbach (confiabilidade) total do teste foi de 0,86; o alfa padronizado total de 0,91; Guttman lambda 6 G6 (smc) total de 0,92; ômega total de 0,92.

Pode-se observar na Tabela 7 que o item Semântica possui a menor correlação com o escore total do teste ($r_{it} = 0,59$, quando o item está incluso no escore; e $r_{ir} = 0,50$ quando o item não está incluso no escore total do teste). O item Evocação da Figura de Rey, segundo a Teoria Clássica dos Testes (TCT), é o item mais discriminativo, uma vez que possui maior correlação com o escore total ($r_{it} = 0,81$ e $r_{ir} = 0,80$). De forma geral, o teste PBAC possui indicadores de confiabilidade (alfa = 0,86; ômega = 0,92), uma vez que apresentam valores superiores a 0,70, considerando-se o tamanho da amostra e o número de itens por sujeito (106).

Quando avaliados separadamente, os coeficientes de confiabilidade se tornam extremamente baixos para o grupo de Jovens ($\alpha = 0,37$ e $\lambda_6 = 0,46$), mas adequado para os demais grupos, conforme se pode observar na Tabela 8 abaixo. O coeficiente λ_6 de Guttman foi inserido, uma vez que considera o total da variância em cada item que pode ser considerado para a regressão linear de todos os outros itens (i.e., a variância dos erros).

Tabela 8 - Coeficientes de confiabilidade do teste PBAC para cada grupo avaliado.

Coeficientes	Jovens (n=103)	Idosos (n=183)	Demência (n=24)	Kalungas (n=15)	Total (n=325)
Alfa de Cronbach (I.C. 95%)	0,37(0,17–0,56)	0,75(0,67–0,82)	0,86(0,73–0,99)	0,82(0,7–0,96)	0,86(0,83–0,90)
Alfa Padronizado	0,41	0,81	0,90	0,82	0,91
Guttman Lambda 6 (G6)	0,46	0,83	0,95	0,96	0,93

3.3 Validade Convergente e Validade Discriminante

Validade convergente de um teste é a forte correlação deste com outros que medem teoricamente o mesmo traço latente, enquanto que validade discriminante é a fraca correlação com outros testes que não propõem mensurar o mesmo traço latente (107).

O teste PBAC possui estimativas posteriores das correlações Bayesianas positivas com os testes de rastreio cognitivos MEEM ($r=0,79$) e teste do Relógio de Sunderland ($r=0,54$), sendo o tamanho do efeito maior na correlação com o MEEM (o próprio valor de r). Já as correlações do PBAC foram negativas com os testes das Trilhas A e B ($-0,71$ e $-0,72$, respectivamente) e apresentaram efeitos mais evidentes com estes instrumentos do que com os testes Stroop, Escala de Depressão Geriátrica (GDS) e Pfeffer ($-0,30$; $-0,17$ e $-0,22$; respectivamente). Note na Tabela 9 que os valores do intervalo de credibilidade superior (2,5%), se aproximam de zero (0,05) para as estimativas da correlação do PBAC com os testes Stroop e GDS. Porém, estas estimativas são maiores que zero, mesmo no IC (5%), sendo 0,007 para a GDS e 0,091 para o

Stroop, o que na correlação frequentista seria o mesmo que dizer que a correlação é significativa considerando $p=0,001$.

Tabela 9 – Estimativa posterior da correlação Bayesiana entre o teste PBAC e outras escalas de medidas cognitivas e funcionais.

PBAC	Intervalo de Credibilidade 95%			Critérios de Informação			
	2,5% Inf.	r	2,5% Sup.	PPp	DIC	BIC	pD
MEEM (n=79)	0,70	0,79	0,87	0,486	914,9	927,1	4,7
Relógio (n=79)	0,54	0,67	0,79	0,486	919,8	923,0	4,7
TrilhasA (n=70)	-0,82	-0,71	-0,56	0,481	1268,0	1280,0	4,7
TrilhasB (n=67)	-0,85	-0,72	-0,56	0,485	1353,0	1365,0	4,6
Stroop (n=69)	-0,52	-0,30	-0,05	0,489	1158,0	1170,0	4,7
GDS (n=78)	-0,39	-0,17	-0,05	0,492	982,9	995,1	4,7
Pfeffer (n=77)	-0,60	-0,42	-0,22	0,491	910,8	923,2	4,6

Nota. MEEM = Mini-Exame do Estado Mental, GDS = Escala de Depressão Geriátrica. Especificações para as estatísticas Bayesianas: o ponto de estimativa utilizado foi a mediana, algoritmo do método MCMC (GIBBS-PX1 com 2 canais e amostragem de 50.000. As distribuições posteriores entre o canal 1 e canal 2 foram comparadas usando a estimativa de Kolmogorov-Smirnov), valores prévios (ML, probabilidade máxima). r = estimativa posterior da correlação Bayesiana entre as variáveis; PPp = valor da probabilidade preditiva posterior referente à diferença dos valores dos qui-quadrados observado e replicado; DIC = critério de informação do desvio, BIC = critério de informação Bayesiana, pD = número de parâmetros estimados no modelo.

A estimativa posterior da correlação Bayesiana entre os testes PBAC e MEEM para todos os sujeitos ($N=325$) foi ainda maior (0,88), com intervalo de credibilidade inferior de 0,85 e superior de 0,90 e PPp de 0,50. Analisando grupo a grupo, a estimativa posterior foi de 0,22 (PPp = 0,33; I.C. = 0,04 – 0,38) para o grupo Jovens ($n=103$); de 0,75 (PPp = 0,42; I.C. = 0,67 – 0,80) para o grupo Idosos hígidos ($n=183$); de 0,94 (PPp = 0,33; I.C. = 0,86 – 0,97) para o grupo Demência ($n=24$); e de 0,77 (PPp = 0,50; I.C. = 0,30 – 0,91) para o grupo Kalunga ($n=15$).

Na Figura 10 podemos observar a dispersão dos dados e linha de tendência de cada grupo entre os testes PBAC e MEEM, com os *box-plots* e *outliers* de cada teste. Esta figura demonstra a dispersão adequada para o grupo Demência (linha preta) e a pequena dispersão para o grupo de Jovens (linha verde).

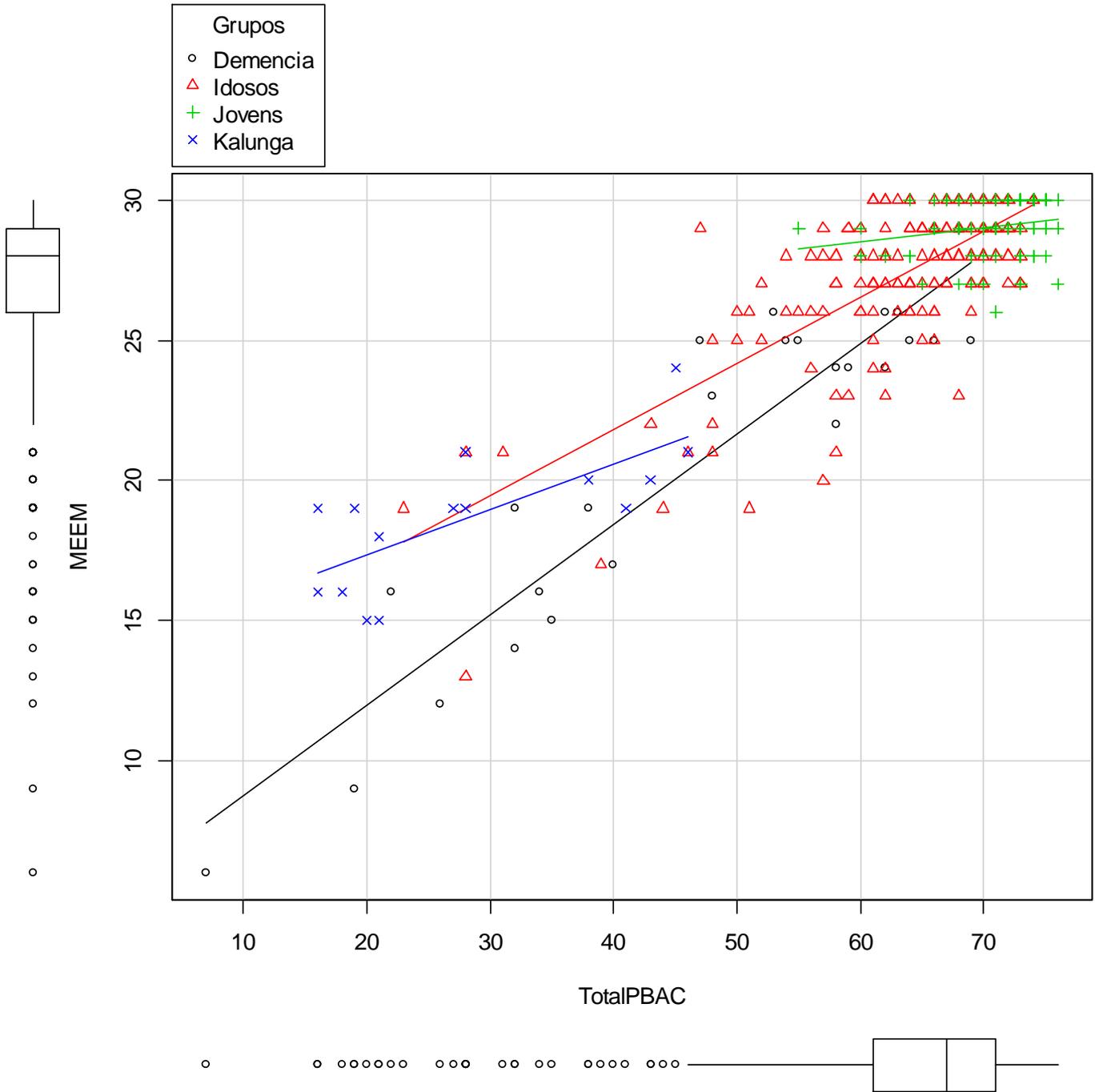


Figura 10. Dispersão com linhas de tendência e caixas laterais dos escores totais do PBAC e MEEM para cada grupo avaliado. Os pontos das caixas laterais representam os *outliers* de cada teste.

3.4 Análise da Dimensionalidade do PBAC

Uma técnica que está se tornando mais comum em pesquisas de Neuropsicologia para a redução dos dados e compreensão das intercorrelações é a análise do item por *clusters* (108). A análise *cluster* é uma alternativa mais simples que a análise fatorial, mas que permite observar quais variáveis correlacionam entre si, refletindo que os itens podem se correlacionar aos *clusters* que, por sua vez, podem se correlacionar com outros itens ou com outros *clusters* (*interclusters correlations*). O algoritmo utilizado para análise hierárquica dos *clusters* formam as escalas compostas. Os *clusters* são combinados se os coeficientes alfa e beta aumentam se estiverem conectados a outro *cluster* (105). O alfa é a média da correlação da divisão pela metade (*split-half*), e o beta é o valor da correlação *split-half* mais baixa. Estes coeficientes são uma estimativa da confiabilidade e da saturação dos fatores do teste. O teste PBAC resulta um modelo com 11 *clusters*, com ajuste de 0,89; ajuste padrão de 0,98 e RMSR igual a 0,08 (ver Figura 11). As intercorrelações corrigidas para atenuação diagonal da matriz de estrutura foi de 0,91. Estes resultados sugerem confiabilidade do teste e de que possui mais de uma dimensão.

Nesta análise *cluster*, podemos ver como os itens do PBAC se correlacionam. A figura mostra que Evocação e Reconhecimento se correlacionam intimamente, formando o *cluster* 2 (C2 0,78), assim como Escrita e Leitura (C3 0,86), Trilhas Orais e Orientação de Linhas (C4 0,76; tendo aí uma relação mais discreta com a tarefa de Fluência da Letra F, C6), Cópia e Evocação da Figura de Rey (C1 0,81), Nomeação e Aprendizagem Verbal (C5 0,77). Podemos notar que a tarefa de Semântica não está diretamente relacionada a nenhum outro item.

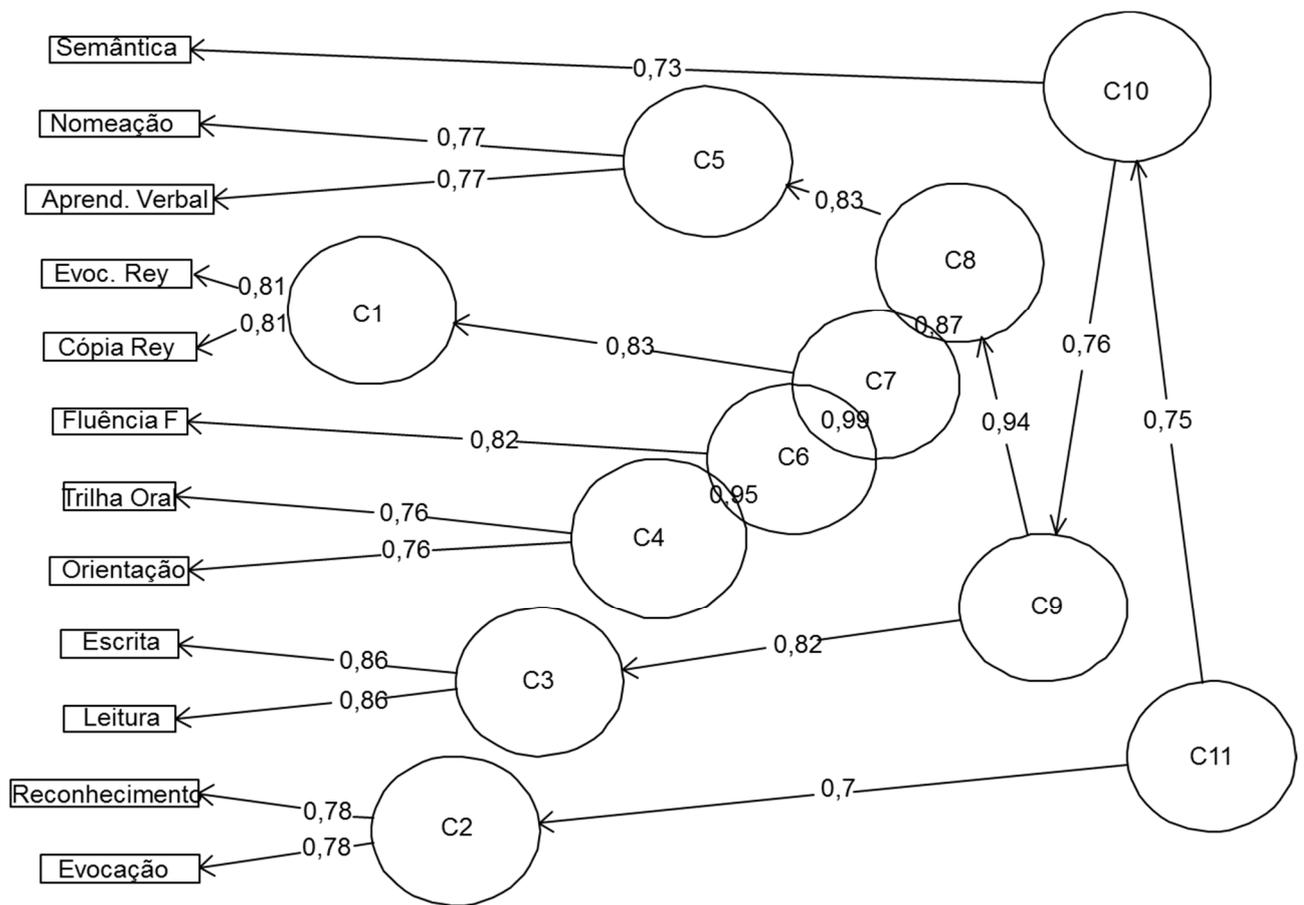


Figura 11. Análise *cluster* dos itens (N=325). O teste PBAC resulta um modelo com 11 *clusters*, com ajuste de 0,89; ajuste padrão de 0,98 e RMSR igual a 0,08. As intercorrelações corrigidas para atenuação diagonal da matriz de estrutura foi de 0,91.

Outra forma de averiguar o número de dimensões é o uso da análise fatorial exploratória (EFA) e confirmatória (CFA), utilizando-se o qui-quadrado como informação para averiguar qual modelo tem o melhor ajuste. Ambas as análises são empregadas para compreender a variância compartilhada das variáveis medidas que se pressupõe estar atribuída a um fator ou traço latente. O objetivo da EFA é identificar fatores buscando maximizar a quantidade de variância explicada (109), e o pesquisador não precisa ter nenhuma hipótese inicial sobre quantos fatores o teste irá apresentar. Já a CFA avalia a hipótese *a priori* do número de fatores, além de ser direcionado pela teoria (110). Diferentemente da EFA, a CFA é um modelo restritivo, pois as variáveis são impostas como pertinentes a um determinado fator.

Usando o software MPlus (versão 7.2 para Windows) (111), a EFA foi calculada usando a estimativa MLR, com rotação Geomin (do tipo oblíqua), obtendo-se uma solução de até 5 fatores. O modelo favorito é aquele que possui melhor ajuste dos dados e a decisão entre os modelos competidores é realizada entre a diferença dos valores dos testes de qui-quadrado, bem como a diferença entre os graus de liberdade. O modelo “maior” possui um maior número de parâmetros e menos graus de liberdade, enquanto que o modelo “menor” possui menor número de parâmetros e mais graus de liberdade. Se a diferença dos qui-quadrados (χ^2_{diff}) for significativa, então o modelo “maior” (que possui menos graus de liberdade, ver Tabela 10) é preferido (53, 73).

Tabela 10 – Análises fatoriais exploratórias robustas (EFA MLR) e comparações entre os modelos (N=325).

		N parâmetros	χ^2	<i>g.l.</i>	<i>p</i>
Modelos EFA	1 fator	36	205,87	54	0,000
	2 fatores	47	104,95	43	0,000
	3 fatores	57	51,71	33	0,020
	4 fatores	66	29,27	24	0,210
	5 fatores	74	9,63	16	0,885
	6 fatores	NA	NA	NA	NA
Comparação entre modelos EFA	1 fator x 2 fatores	-	70,91	11	0,000
	2 fatores x 3 fatores	-	45,52	10	0,000
	3 fatores x 4 fatores	-	18,54	9	0,029
	4 fatores x 5 fatores	-	16,30	8	0,038

Na Tabela 10, podemos observar que os dados de todos os sujeitos avaliados (N=325) possuem melhor ajuste ao modelo EFA de 5 fatores ($\chi^2= 9,63$; *g.l.*=16; $p=0,885$) de 74 parâmetros, e o modelo de 5 fatores é superior ao modelo de 4 fatores ($\Delta\chi^2= 16,30$; *g.l.*=8; $p=0,038$). Os autovalores (*eigenvalues*) foram obtidos para os componentes principais e análises fatoriais (Figura 12). O pacote ‘psych’ foi utilizado para os cálculos dos autovalores simulados e reamostrados (análise paralela). A análise paralela sugere a existência de 2 componentes principais e 3 fatores, porém o modelo de 4 fatores não pode ser descartado ($p=0,21$).

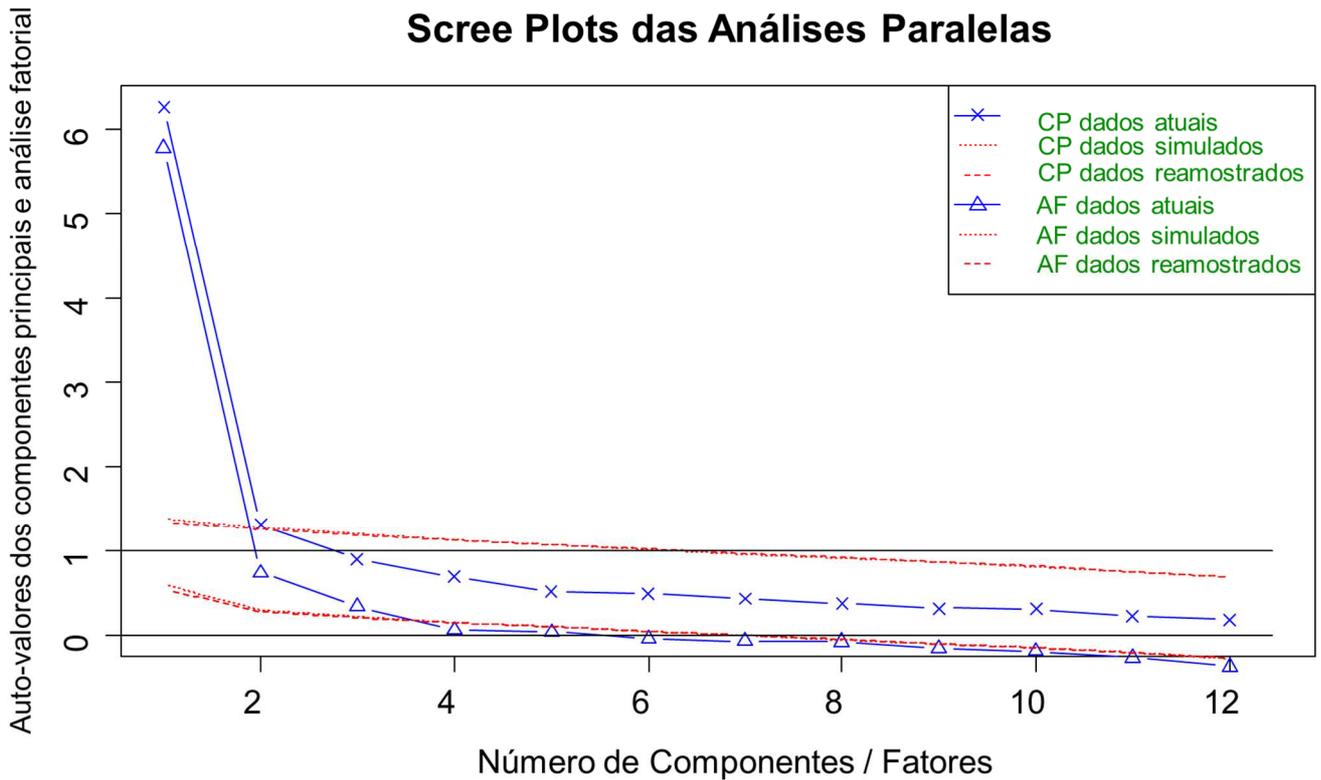


Figura 12. *Scree plots* das análises paralelas dos componentes principais (CP) e análises fatoriais (AF) dos dados atuais (N=325), dados simulados e dados reamostrados.

3.5 Teoria de Resposta ao Item Multidimensional (TRIm)

Comparada com a teoria clássica dos testes (TCT), a teoria de resposta ao item (TRI) é considerada o método preferido para conduzir avaliações sobre instrumentos psicológicos (29, 34, 54, 112). A TRI parte da premissa de que apenas dois elementos são responsáveis por descrever a resposta de uma pessoa numa determinada tarefa: a habilidade inerente ou traço latente desta pessoa, e as características do item (113). Os modelos da TRI estimam, no entanto, ambos os elementos com base no número total de indivíduos que responderam de forma correta um item, e o parâmetro da habilidade da pessoa é calculada com base no número total de itens que foram corretamente respondidos (29). Neste pressuposto, (a) uma pessoa com maior habilidade num determinado domínio cognitivo terá sempre maior probabilidade de endossar

corretamente um item do que uma pessoa com menor habilidade; (b) uma pessoa tem maior probabilidade em endossar itens que requeiram menor habilidade, do que os itens que requeiram maior habilidade.

A teoria de resposta ao item multidimensional (TRIm) é uma extensão da TRI e é usada para explorar a dimensionalidade do instrumento, segundo seus pressupostos, e para estimar simultaneamente os parâmetros de dificuldade e discriminação dos itens nestas diferentes dimensões obtidas (114, 115). Os índices de ajustes dos modelos, podem ser baseados nos métodos dos resíduos mínimos ou usando as estatísticas de razão da probabilidade do qui-quadrado (116). Estes índices são representações estatísticas que informam o quão bem os dados se ajustam ao modelo TRIm proposto. No entanto, os modelos não são razoavelmente acurados quando o número de respondentes é pequeno, sendo necessária a utilização de métodos que realizem a reamostragem dos dados, como o MCMC. Os autores divergem em relação ao tamanho da amostra nas estimativas dos modelos, alguns recomendam uma amostra de pelo menos 500 participantes, uma vez que as estimativas podem se tornar instáveis com menos do que 350 respondentes (34). Outros modelos mais simples, como a análise Rasch, onde somente contempla a estimativa do parâmetro de dificuldade é obtido, um número entre 50 a 100 respondentes pode ser considerado razoável (117).

Nesta pesquisa, diferentes modelos TRIm exploratórios foram analisados, usando uma generalização multidimensional do modelo de resposta gradual (GRM) de Samejima (118), proposto por Muraki e Carlson. Este modelo considera o item do teste requer um número de passos, mas que o sucesso num passo depende do sucesso de um passo anterior, ou seja, se o passo k foi alcançado, então significa que o passo $k-1$ também foi alcançado.

Para buscar um melhor ajuste para o modelo de análise fatorial com informação plena, baseada na máxima verossimilhança (FIFA, *Full-Information Item Factor Analysis*). Foram utilizados os algoritmos EM (*Expectation-Maximization*) e MHRM (*Metropolis-Hastings Robbins-*

Monro). Este último usa os mesmos modelos TRIm, mas testa os traços latentes empiricamente produzidos com uma estimativa de reamostragem de Monte Carlo (119). Para estas análises (Tabela 11), foi utilizado o pacote 'mirt' (versão 1.3) (120) do R. A Tabela 11 mostra as análises dos modelos TRIm exploratórios para a amostra total (superior) e a amostra de idosos hígidos (inferior).

Tabela 11 – Comparação entre os modelos exploratórios da teoria de resposta ao item multidimensional (TRIm).

	AIC	ACCI	SABIC	BIC	logLik	χ^2	<i>g.l.</i>	<i>p</i>
<i>Amostra total (N=325)</i>								
1 dimensão	7291,04	7348,90	7341,83	7605,10	-3562,52	-	-	-
2 dimensões	7192,84	7270,49	7250,36	7548,52	-3502,42	120,20	11	0
3 dimensões	7170,68	7269,96	7234,32	7564,20	-3481,34	42,15	10	0
4 dimensões	7155,40	7277,50	7224,54	7582,97	-3464,70	33,26	9	0
<i>Somente idosos (n=183)</i>								
1 dimensão	4244,95	4333,70	4247,91	4469,62	-2052,48	-	-	-
2 dimensões	4218,08	4349,60	4221,50	4478,04	-2028,04	48,88	11	0
3 dimensões	4206,55	4390,55	4210,40	4498,62	-2012,28	31,52	10	0
4 dimensões	4189,01	4435,36	4193,24	4509,96	-1994,51	35,54	9	0

Nota:

Os resultados mostrados na Tabela 11 mostram que o modelo com 4 dimensões é o preferido, uma vez que apresenta melhores índices de ajustes aos dados do PBAC quando comparado com os demais, tanto para a amostra total (AIC = 7155,40; $\chi^2 = 33,26$; $p < 0,001$) quanto para a amostra de idosos hígidos (AIC = 4189,01; $\chi^2 = 35,54$; $p < 0,001$). Note que os dados dos idosos hígidos possuem melhor ajuste para o modelo de 4 dimensões que os dados da amostra total, mesmo com menor número de sujeitos na amostra.

3.5 Comparação entre Análise Fatorial e TRI

Análise Fatorial (AF) e TRI possuem virtualmente formulações estatísticas idênticas quando são aplicados às matrizes de respostas dos itens: 1) Apesar do procedimento ser idêntico, a diferença reside no fato de que a AF é baseada na redução no número de fatores, buscando encontrar o menor número de fatores que reproduz a matriz de correlações observadas, enquanto que a TRI é uma técnica para modelar a interação entre pessoas e itens de um teste. Desta forma, TRI é uma ferramenta útil para modelar as contribuições mais significativas da interação entre pessoas e itens. 2) A AF praticamente ignora as características das variáveis de entrada (exceto quando se usa o modelo de equação estrutural, pois considera as médias, variâncias e covariâncias, mas não com o objetivo de entender melhor estas variáveis de entrada). Já a TRI, por outro lado, objetiva nas diferenças das médias e variâncias dos escores dos itens, pois elas estão diretamente relacionadas às características importantes dos itens do teste, tais como a dificuldade e a discriminação. 3) A análise TRI trabalha ativamente para encontrar soluções que usem o mesmo espaço latente entre os testes e as amostras. O objetivo é manter um sistema de coordenadas comum para todas as análises, de tal forma que os itens terão estimativas dos parâmetros numa métrica comum. Isto possibilita a criação de banco de itens (como em Testagens Adaptativas Computadorizadas, CAT) (121).

As análises TRI mostram claramente que os itens do teste podem ser sensíveis para examinar as diferenças nas várias dimensões. Se subgrupos da população examinada diferir em dimensões, o resultado relatado pode incluir informações sobre viés do teste, i.e., certos itens podem ser utilizados para medir um determinado traço latente numa população e não medir em outra, denominado funcionamento diferencial dos itens (DIF).

3.6 Estimativa do Modelo Psicométrico

Devido à característica de não-normalidade dos dados, a estimativa do modelo psicométrico para os idosos hígidos foi realizada utilizando a estatística Bayesiana, uma vez que não viola pressupostos importantes, como o tamanho da amostra. O software utilizado para estas estimativas foi o Mplus 7.2. Na figura dos modelos de equação estrutural Bayesiana (BSEM) (Figura 13), os retângulos representam as variáveis observáveis (os itens), enquanto que os círculos representam as variáveis latentes não observáveis (os fatores). A seta unidirecional indica o intercepto da regressão entre a variável latente e a variável observável. O valor de épsilon (ϵ) apresenta a variância das variáveis observáveis e as setas duplas apresentam as covariâncias entre as variáveis latentes.

A Tabela 12 mostra que o modelo EFA Bayesiano de 2 fatores não é diferente do modelo de 1 fator (comparação entre os modelos igual a 0,533), porém, o modelo de 4 fatores se mostra superior aos demais, uma vez que o valor de PPp é superior a 0,5. Caso o item Leitura seja excluído do teste, o modelo exploratório com 3 fatores é o candidato ideal, uma vez que apresenta ajuste PPp igual a 0,145.

Tabela 12 – Comparação entre os modelos EFA Bayesianos para o grupo de idosos hígidos (n=183).

	EFA Bayesiano	DIC	BIC	pD	PPp	Comparação	Log Bayes
Com Leitura	1 fator	5028,58	5142,52	36,72	0,000	-	-
	2 fatores	4989,31	5143,78	45,05	0,000	0,533	-0,63
	3 fatores	4964,85	5153,73	55,54	0,006	0,006	-4,98
	4 fatores	4945,06	5176,80	63,68	0,056	0,000	-11,53
Sem Leitura	1 fator	5186,98	5293,96	32,32	0,000	-	-
	2 fatores	5167,92	5310,05	39,04	0,010	0,000	-8,05
	3 fatores	5153,84	5331,98	52,29	0,145	0,000	-10,96

A partir dos modelos EFA Bayesianos, os modelos de equações estruturais BSEM foram computados, onde dezenas de possibilidades foram consideradas e analisadas, porém somente

alguns modelos tiveram ajustes adequados para fins de comparação ($PPp > 0,5$). A Figura 13 apresenta quatro modelos candidatos a representarem o modelo cognitivo definitivo para o teste PBAC.

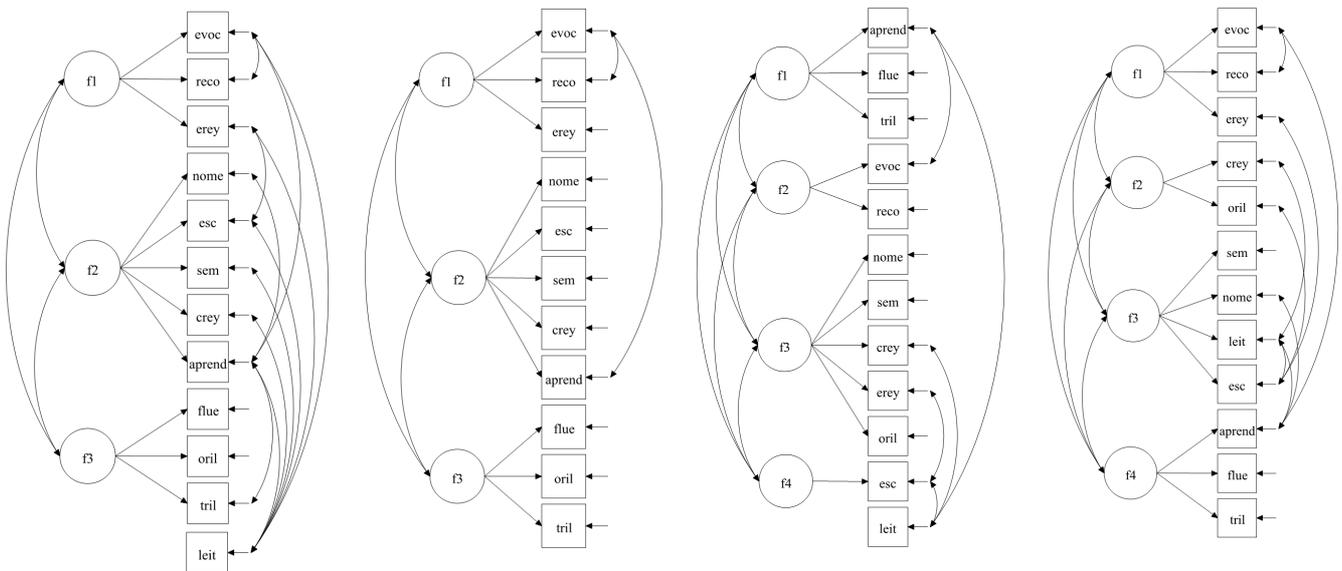


Figura 13. Comparação entre os modelos Bayesianos (BSEM) dos participantes idosos (n=183). À esquerda (1º modelo), o item Leitura não possui carga significativa em F3 (DIC=4922,52; BIC 5293,74; $pD=68,76$; $PPp=0,446$). No centro-esquerdo (2º mod.), o item Leitura foi excluído (DIC=5145,88; BIC 5463,18; $pD=59,60$; $PPp=0,443$). No centro-direito (3º mod.), modelo de 4 fatores, onde Leitura não é significativo em F4 (DIC=4938,36; BIC 5309,92; $pD=71,86$; $PPp=0,215$). À direita (4º mod.), o modelo de 4 fatores (DIC=4939,35; BIC 5314,58; $pD=70,37$; $PPp=0,183$).

O modelo BSEM com melhor ajuste para os dados dos idosos hígidos foi o 1º modelo de 3 fatores (DIC = 4922,52), porém o item Leitura não recebe carga em nenhum fator e tem covariância em quase todos os demais itens do teste. O 2º modelo de 3 fatores teve o item Leitura excluído, porém se torna um modelo menos preciso que o primeiro (DIC = 5145,88). O 4º modelo de 4 fatores, representa de forma mais fidedigna a proposta inicial do proponente do teste (56), onde F1 comporia o subtteste de *memória*, F2 *habilidades visoconstrutivas*, F3 *linguagem* e F4 *funções executivas*. O ajuste do teste é considerado inadequado ($PPp = 0,183$, pois não se

aproxima de 0,5) com DIC = 4939,35. Comparado com o 1º modelo, a diferença entre os DIC, Δ DIC, foi de 16,83, o que não chega a ser considerado um modelo inviável, devido à pequena diferença.

Análises TRIIm confirmatórias mostraram que este 4º modelo, que representa a proposta original do teste PBAC pelos seus idealizadores, tem melhor índice de ajustes quando utilizado somente os dados do grupo de idosos hígidos (AIC = 4341,63) do que quando considerados todos os grupos da amostra (AIC = 7759,84). Ainda, os modelos mostram que o uso do algoritmo MHRM melhora a estimativa do modelo quando considerados todos os grupos (AIC = 7214,45; $p < 0,001$), mas não há diferenças quando observado somente o grupo de idosos hígidos (AIC = 4336,04; $p = 0,237$).

Tabela 13 – Modelos TRIIm confirmatórios de 4 dimensões para a amostra total e para o grupo de idosos hígidos.

	AIC	ACCI	SABIC	BIC	logLik	χ^2	<i>g.l.</i>	<i>p</i>
<i>Amostra total (N=325)</i>								
4 dim (Libon) <i>nested</i>	7759,84	7817,70	7810,63	8073,90	-	-	-	-
4 dim (bifator)	7217,13	7296,78	7275,26	7576,60	-3513,57	-	-	-
4 dim (MHRM)	7214,45	7294,10	7573,91	7272,58	-3512,23	2,68	0	0
<i>Somente idosos (n=183)</i>								
4 dim (Libon) <i>nested</i>	4341,63	4430,38	4344,59	4566,29	-2100,81	-	-	-
4 dim (bifator)	4360,36	4445,85	4363,28	4581,82	-2111,18	14,69	1	0
4 dim (MHRM)	4336,04	4401,14	4338,67	4535,03	-2106,02	10,42	8	0,237

Nota:

O modelo teórico de Libon (*nested model*) foi fixado e concomitantemente comparado. O ajuste geral de um modelo é baseado na estatística do desvio, que segue uma distribuição do qui-quadrado. As estatísticas do desvio mudam quando parâmetros são adicionados ou excluídos do modelo (21). O qui-quadrado é calculado como sendo a diferença entre os modelos aninhados, usando os mesmos dados. A diferença entre os graus de liberdade (*g.l.*) entre os

modelos é utilizada no cálculo do qui-quadrado. Então, o qui-quadrado testa a hipótese de ajuste idênticos entre dois modelos. Uma falha em rejeitar a hipótese nula significa que os dois modelos se ajustam igualmente na população avaliada. Se dois modelos representam bem o conjunto de dados, então a escolha sempre se dá pelo modelo mais parcimonioso (122).

Ainda, considerando apenas os idosos hígidos, o 4º modelo foi calculado objetivando investigar as relações entre as variáveis independentes idade e escolaridade e as variáveis latentes do teste. A Figura 14 mostra que a idade influencia de forma negativa somente o fator F3 ($R^2 = -0,34$; I.C. = -0,47 a -0,22), que contém os itens de Evocação, Reconhecimento e Evocação da Figura de Rey, enquanto que a escolaridade influencia positivamente todos os quatro fatores do teste. O modelo sugere que quanto maior a escolaridade do idoso, maior a probabilidade do endosso nos fatores do teste, enquanto que quanto maior a idade, menor a probabilidade no endosso do terceiro fator. Neste modelo, o item Leitura foi excluído por não contribuir de forma significativa. O item Aprendizagem Verbal é covariante dos itens de Nomeação (0,21, I.C. = 0,03 a 0,36) e Evocação (0,23, I.C. = 0,09 a 0,35), e Evocação é covariante de Reconhecimento (0,30, I.C. = 0,14 a 0,42). Ainda, F1 e F4 não foram covariantes.

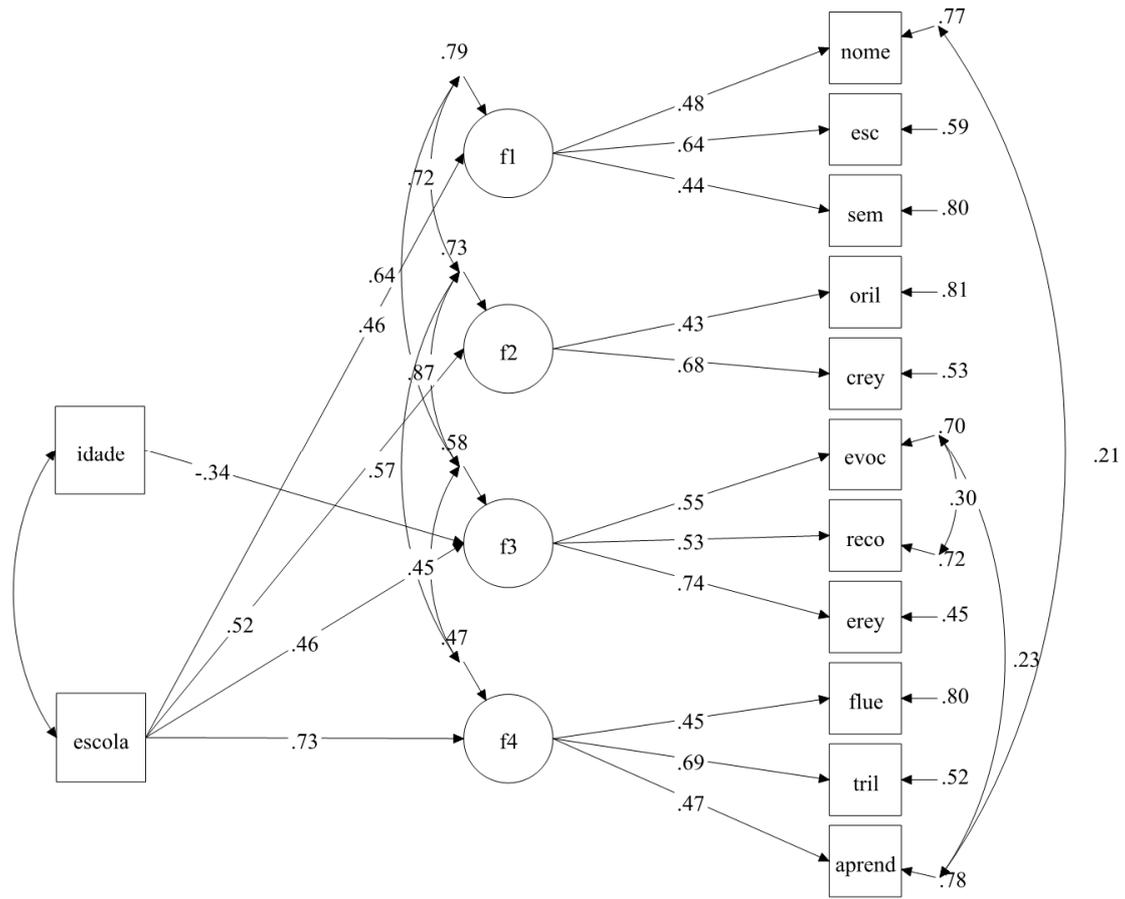


Figura 14. Modelo de equação estrutural Bayesiano para o grupo de idosos hígidos (n=181). Efeitos da idade e escolaridade nas variáveis latentes. Informação de ajuste ao modelo: Pp=0,326; DIC=5009,003; BIC=5353,062; pD=62,105. Algoritmo GIBBS(RW), com 100.000 iterações com 2 canais MCMC.

3.7 Sensibilidade e Especificidade do teste PBAC

Análises utilizando curvas ROC (*receiver operating characteristic*) não-paramétricas foram realizadas para comparar a acurácia do teste PBAC e seus subtestes em relação ao teste MEEM (considerado padrão ouro para detectar demências). Nesta análise, é estimada a sensibilidade (fração de pessoas com o diagnóstico de demência que é corretamente identificada como positivo pelo teste) e a especificidade (fração de pessoas sem a demência que o teste corretamente identifica como negativo). Se seleccionar um limiar alto (um ponto de corte, por exemplo), aumenta-se a especificidade do teste, mas perde-se em sensibilidade. Se seleccionar um limiar baixo, aumenta-se a sensibilidade do teste, mas perde-se em especificidade. A

acurácia do teste depende do quão bem ele separa o grupo clínico daquele considerado saudável. Esta acurácia é medida através da área abaixo da curva ROC. Uma área igual a 1 representa um teste perfeito, enquanto que uma área com valores próximos a 0,5 representa um teste medíocre. Podemos classificar a acurácia em: 0,90 a 1 (excelente), 0,80 a 0,90 (bom), 0,70 a 0,80 (razoável), 0,60 a 0,70 (fraco) e 0,50 a 0,60 (ruim).

Tabela 14 – Curvas ROC de 34 idosos controle e 24 pacientes demenciados.

	Área ROC	Erro Padrão	χ^2	g.l.	$p > \chi^2$	Bonferroni $p > \chi^2$
MEEM (padrão ouro)	0,927	0,035				
Aprendizagem	0,735	0,062	10,3	1	0,001	0,017
Fluência	0,718	0,073	8,3	1	0,004	0,051
Evocação	0,822	0,056	4,9	1	0,026	0,340
Reconhecimento	0,710	0,071	11,0	1	0,001	0,012
Nomeação	0,595	0,047	49,2	1	0,000	0,000
Semântica	0,663	0,053	19,6	1	0,000	0,000
Cópia Fig. Rey	0,683	0,067	13,5	1	0,000	0,003
Orientação Linha	0,757	0,064	7,3	1	0,007	0,090
Evocação Fig. Rey	0,788	0,063	5,6	1	0,018	0,229
Leitura	0,544	0,030	79,3	1	0,000	0,000
Escrita	0,577	0,049	40,7	1	0,000	0,000
Trilhas Oraís	0,712	0,068	10,2	1	0,001	0,019
Escore Total PBAC	0,856	0,050	2,6	1	0,104	1,000

Considerando isoladamente os subtestes do PBAC, apenas a tarefa de Evocação teve área superior a 0,80 (considerado bom). As tarefas de Nomeação, Leitura e Escrita são consideradas insuficientes para detectar demências, por terem áreas inferiores a 0,60. Observe, no entanto, que a diferença entre as áreas sob as curvas entre o MEEM (0,93, e.p.=0,035) e o PBAC (0,856, e.p.=0,048) não foi considerada significativa ($p=0,104$). Ou seja, não há evidências estatísticas de que o PBAC seja menos acurado que o MEEM na detecção de demências. Uma análise visual, mostrando uma cobertura ponto a ponto entre os testes avaliados, pode ser realizada na Figura 15 abaixo.

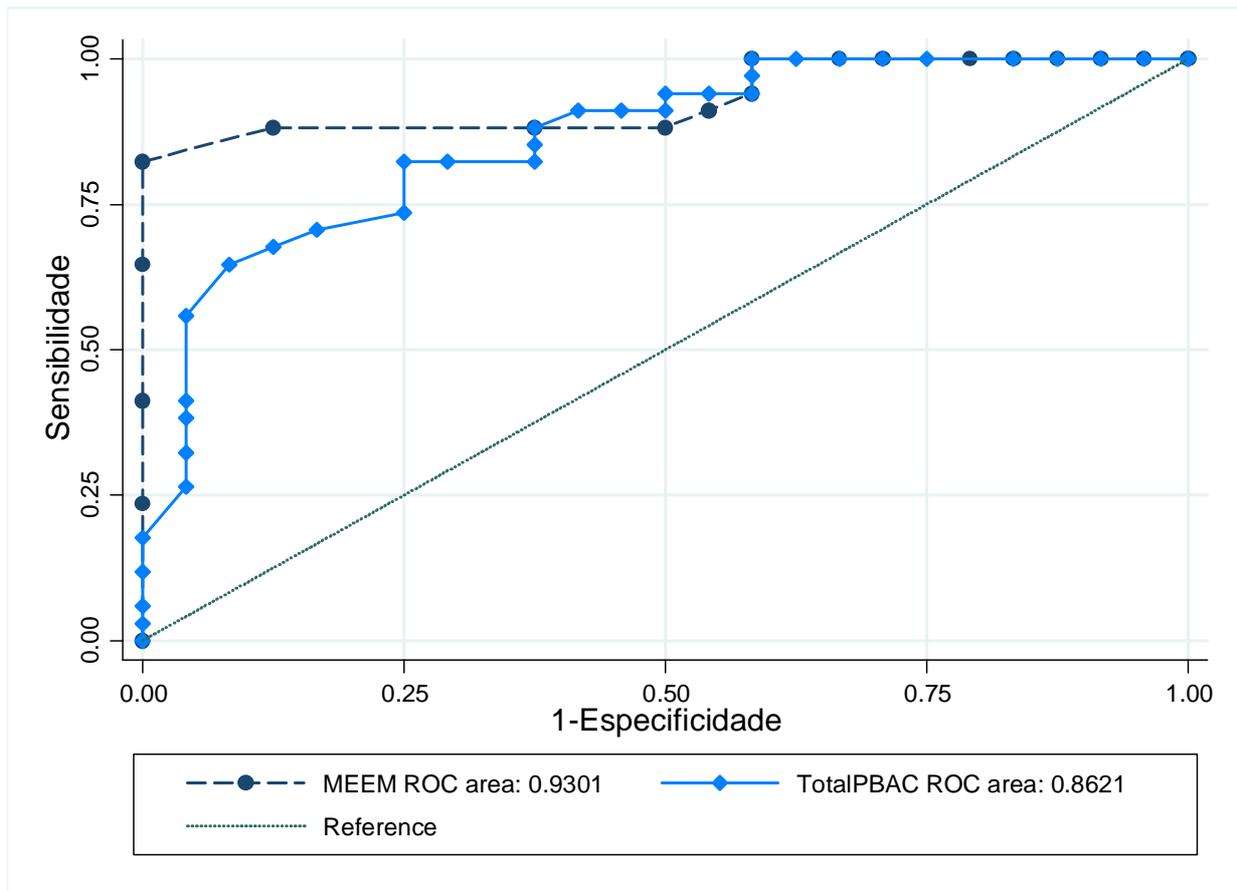


Figura 15. Curvas ROCs comparando o teste PBAC com o padrão-ouro MEEM (n=58; 24 pacientes demenciados e 34 controles).

Realizando uma análise Bayesiana semiparamétrica da estimativa da curva ROC para o teste PBAC (pacote Dppackage, versão 1.1.6) (123), um ponto de corte de 58,8 pontos (ver Figura 16) oferece 0,805 de sensibilidade, e 0,733 de especificidade (AUC = 0,843). Esta mesma estimativa foi realizada para o MEEM, obtendo um ponto de corte de 26 pontos, com 0,847 de sensibilidade e 0,887 de especificidade (AUC = 0,903).

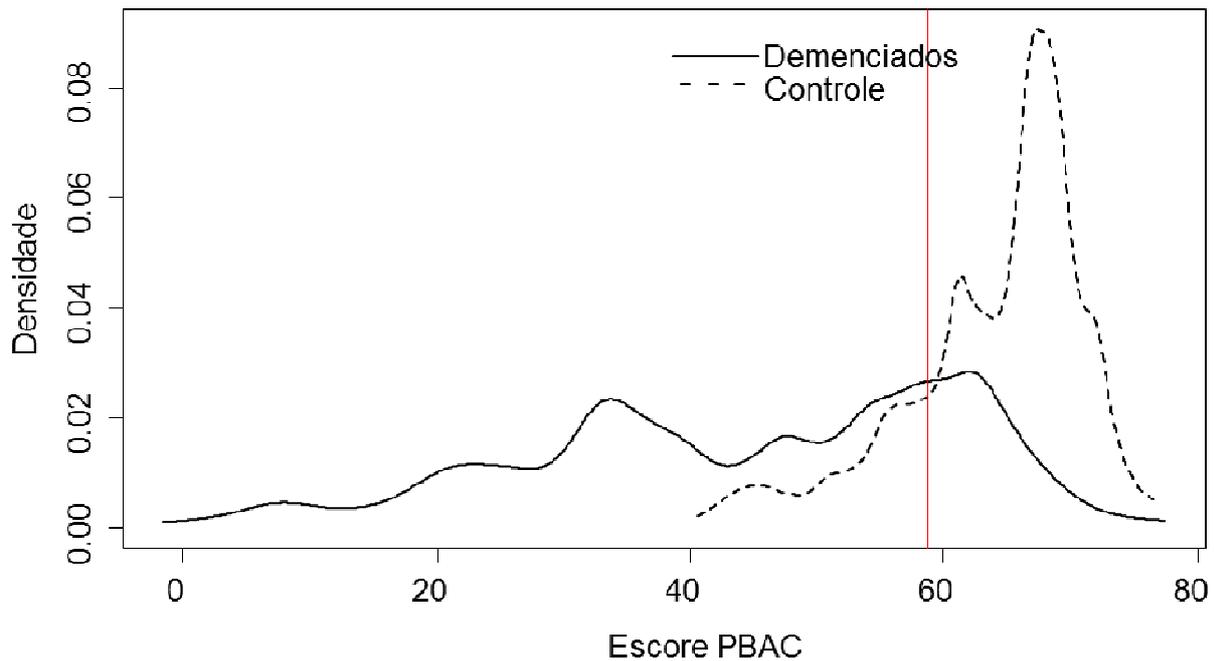


Figura 16. Densidades dos escores totais do PBAC para os idosos controle (n=34; linha pontilhada) e pacientes demenciados (n=24; linha contínua). A linha vermelha na vertical sinaliza o ponto de corte de 58,8 pontos.

No entanto, a obtenção de pontos de corte a partir das curvas ROC possui limitações que devem ser levadas em consideração (124). Alternativas, como a teoria de resposta ao item, serão consideradas como critério diagnóstico (125, 126).

3.8 Curvas Características e de Informação – TRI multigrupo

Análises da teoria de resposta ao item foram aplicadas ao teste PBAC para inferir a habilidade latente dos respondentes. Considerando que o PBAC seja um teste unidimensional, pode-se observar que indivíduos com habilidade cognitiva igual a -1, obtém 51 pontos no teste (Figura 17, esquerda). O teste se mostra bastante informativo para indivíduos com habilidade latente entre -2,5 e -0,5 (Figura 17, direita). Esta função de informação do teste demonstra que

o instrumento é pouco confiável quando avalia indivíduos de traços latentes mais elevados (acima de 0), como é esperado num instrumento de rastreio cognitivo.

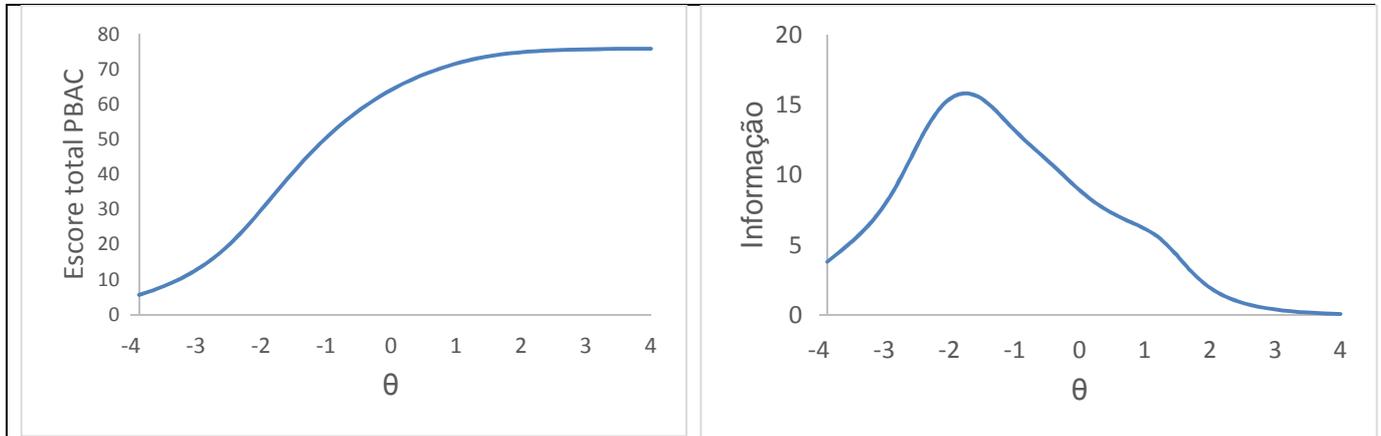


Figura 17. Curva característica do teste (esq.) e função de informação total (dir.) do teste PBAC. (N=325).

A Figura 18 mostra a plotagem dos escores do PBAC com a habilidade latente dos indivíduos numa regressão polinomial, demonstrando que o aumento do endosso no teste implica no aumento da habilidade cognitiva do indivíduo, semelhantemente à curva característica do teste, mostrada na Figura 17 (esq.).

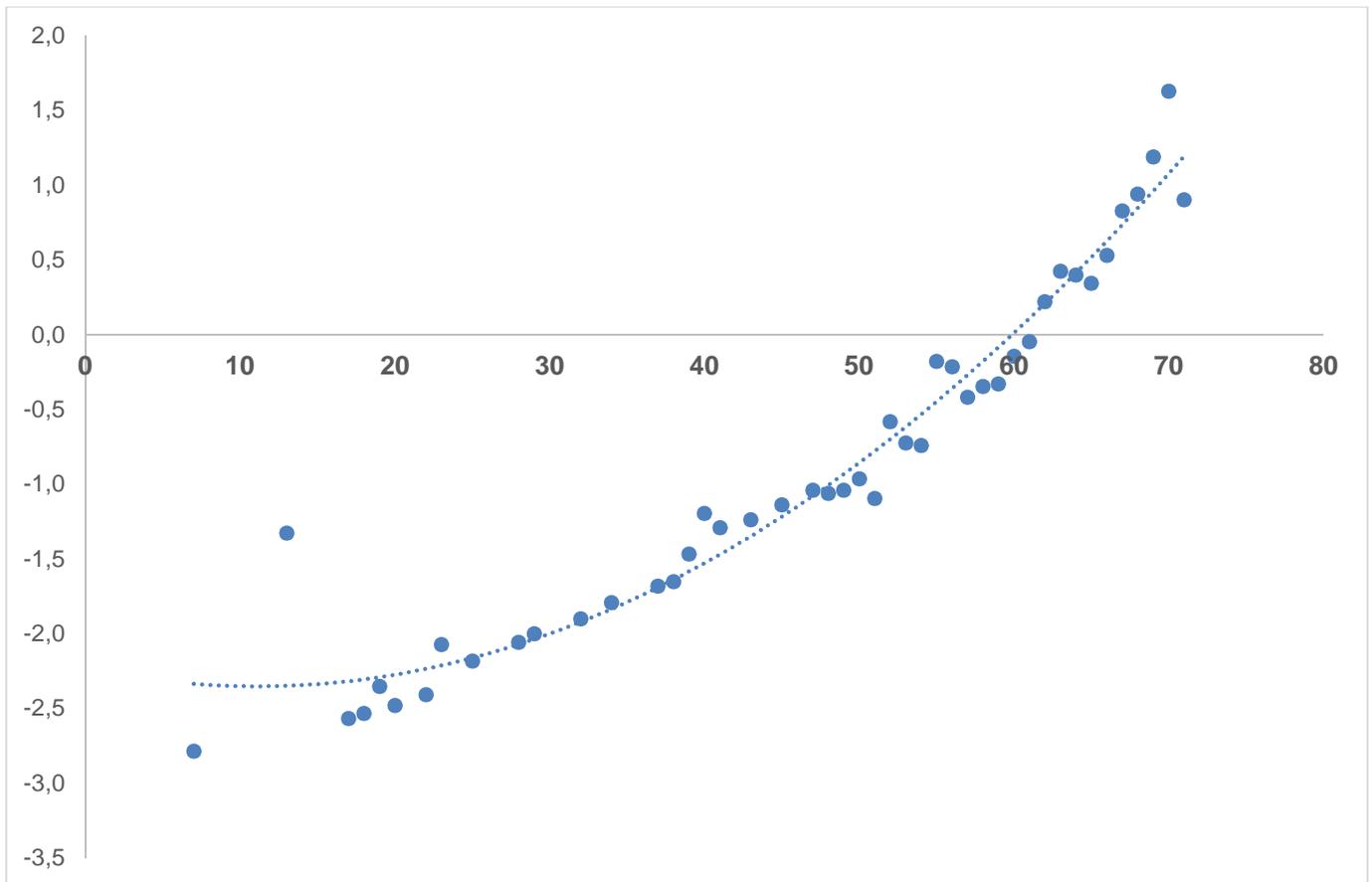


Figura 18. Regressão polinomial entre o escore do teste PBAC (abscissa) e a habilidade latente (média) dos indivíduos (N=325). $R^2=0,955$; $y=0,001x^2 - 0,002x - 2,23$.

Na Figura 19, pontos de corte podem ser obtidos para separar indivíduos com baixa habilidade latente (-1) dos demais. No teste de Trilhas Orais, por exemplo, um indivíduo com habilidade latente -1 endossaria 2 pontos, enquanto que um indivíduo com habilidade latente 0 endossaria 3,5 pontos. Na Evocação da Figura de Rey, um indivíduo com 4 pontos teria uma habilidade -1, enquanto que com 8 pontos, teria uma habilidade 0, e 11 pontos, habilidade 1. Observe que os itens Nomeação, Semântica, Leitura e Escrita, são pouco discriminativos para inferir, adequadamente, as habilidades latentes dos indivíduos, uma vez que os endossos nas categorias de resposta para a habilidade -1 são muito semelhantes para a habilidade 0 ou maior. Para que o indivíduo endosse o item Leitura numa categoria abaixo (3 pontos), sua habilidade latente deve ser próxima de -1,8. Curvas com pouca inclinação (*slope*) representam itens menos

discriminativos, ou seja, pouca capacidade para inferir a habilidade latente dos respondentes. A partir da habilidade 0, os itens se tornam pouco discriminativos, portanto, pouco informativos. Outra análise que pode ser feita a partir da Figura 19 é acerca da dificuldade relativa dos itens. Ao observar as curvas obtidas a partir das tarefas de Cópia e de Evocação da Figura Complexa de Rey, podemos notar que são praticamente paralelas. No entanto, o item Cópia está posicionado bem mais à esquerda do item Evocação da Figura de Rey, concluindo que Cópia é um item mais fácil que Evocação, uma vez que a probabilidade de endosso é maior, mesmo para indivíduos com baixa habilidade latente (10 pontos para habilidade -1).

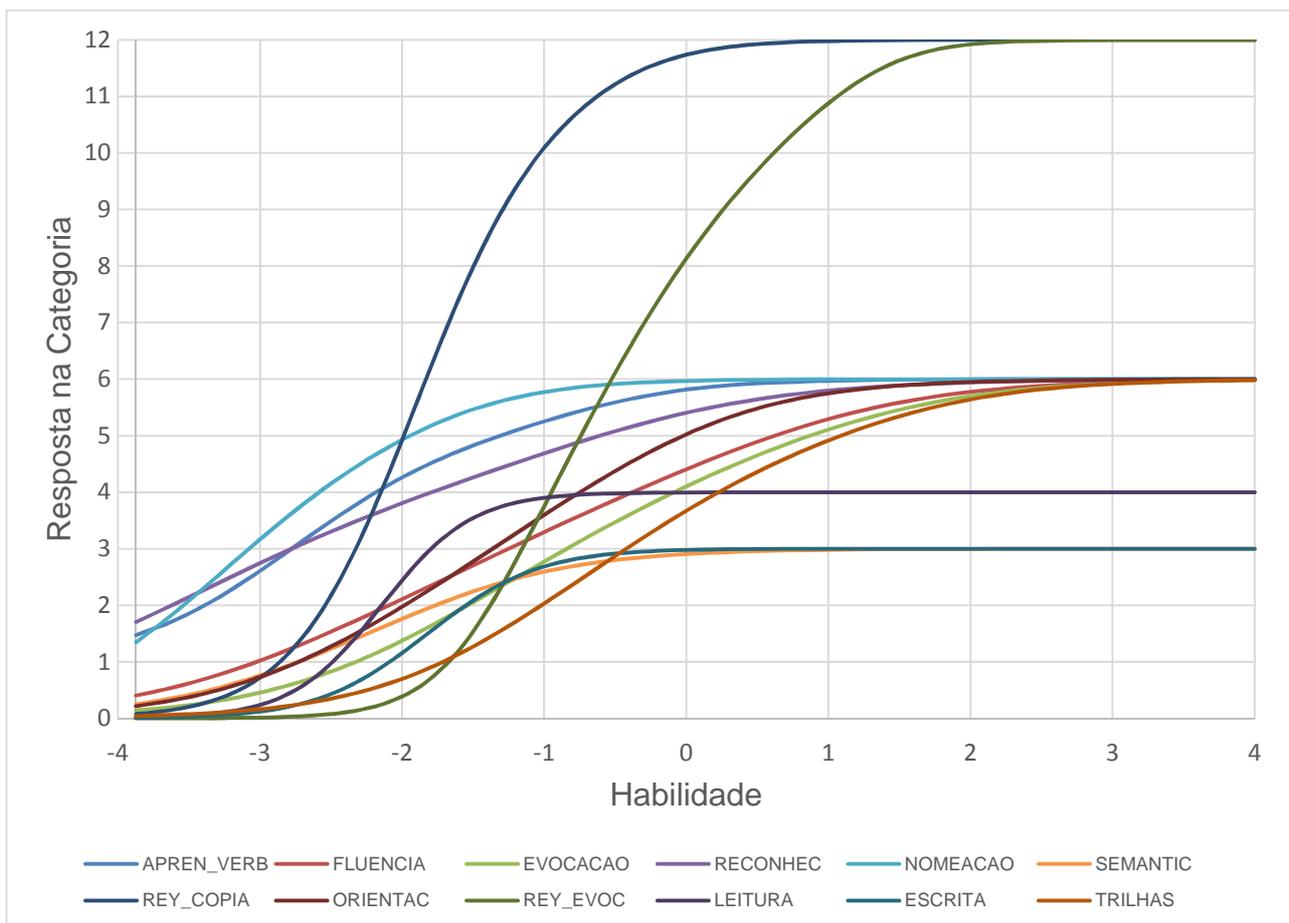


Figura 19. Curvas características dos itens (em função da resposta na categoria).

O termo 'informação' é usado na TRI como um indicador estatístico da qualidade da estimativa de um parâmetro, no caso o parâmetro Teta. O valor da informação está relacionado à extensão do intervalo de confiança do traço estimado para uma pessoa (127). No entanto, abordagens mais modernas mostram que a função de informação é um indicador de que o escore reportado de um teste ou de um item diferencia entre o traço latente de interesse (29). O termo 'dificuldade' do item refere-se à proporção da amostra de examinandos que oferece uma resposta correta para um item do teste. Este parâmetro indica o ponto na escala latente onde a curva característica do item (CCI) tem a probabilidade de acerto igual a 0,5. O termo 'discriminação' do item se refere à capacidade do item do teste em diferenciar os examinandos que possuam um nível mais alto no constructo avaliado daquele que possui pouco deste constructo. Na teoria clássica dos testes, este parâmetro é obtido a partir da correlação ponto-bisserial ou bisserial entre o escore do item e o escore total do teste.

Os parâmetros de discriminação (*slopes*) e dificuldade (*localização*) dos itens foram calculados através do modelo multidimensional da Teoria de Resposta ao Item, considerando um fator comum a todos os itens (a_1 , modelo bifatorial). Os parâmetros da TRI foram restritos aos resultados da análise fatorial confirmatória do modelo de 4 fatores compatível com a proposta de Libon, usando o algoritmo (MHRM) e considerando as tarefas de Evocação, Reconhecimento e Evocação da Figura de Rey numa primeira dimensão (a_2); Cópia da Figura de Rey e Orientação de Linhas numa segunda dimensão (a_3); Nomeação, Escrita, Semântica e Leitura numa terceira dimensão (a_4); e Aprendizagem Verbal, Fluência (letra F) e Trilhas numa quarta dimensão (a_5). A Tabela 15 mostra os parâmetros de discriminação no fator geral (a_1) e nos demais fatores/dimensões ($a_2 - a_5$). Desta forma, pode-se observar que o item mais discriminativo do teste foi Escrita ($a_1 = 5,69$), e o menos discriminativo foi o item Reconhecimento ($a_1 = 1,56$).

A localização multidimensional das categorias de respostas dos itens foram somadas para compor a dificuldade do item ($\sum d$). Para se tornar um valor comparativo, este parâmetro da

localização multidimensional foi dividido pelo número total de categorias de respostas dentro de cada item. Desta forma, pode-se observar que o item Trilhas é o mais difícil para a amostra (0,81), enquanto que o item Escrita é o mais fácil (11,85).

Tabela 15 – Estimativa dos parâmetros TRI bifatorial para a amostra total (N=325).

	a1	a2	a3	a4	a5	Σd	$\Sigma d/n \text{ cat.}$
Aprendizagem Verbal	1,80				0,26	24,52	4,90
Fluência	2,01				1,94	17,70	2,95
Evocação	2,07	2,34				12,96	2,16
Reconhecimento	1,56	1,96				22,93	4,59
Nomeação	1,98			0,37		31,69	6,34
Semântica	1,60			0,01		11,74	3,91
Cópia F. Rey	4,67		2,30			100,42	9,13
Orientação Linhas	2,16		1,07			19,31	3,22
Evocação F. Rey	3,33	0,54				23,70	1,97
Leitura	3,60			1,58		25,59	8,53
Escrita	5,69			3,78		35,56	11,85
Trilhas	1,65				0,47	4,87	0,81

As habilidades latentes dos indivíduos levam em consideração a dificuldade e a discriminação dos itens. No entanto, a presença de indivíduos não-alfabetizados na amostra pode levar à interpretações enviesadas do teste, como por exemplo o item Escrita, que obteve maior parâmetro de discriminação.

Uma vez que a proposta do teste PBAC é de rastreio cognitivo, apenas dois grupos serão considerados nas análises seguintes: idosos hígidos (n=183) e pacientes demenciados (n=24). O grupo de jovens tem apresentado, até o momento, pouca variabilidade de resposta. O teste PBAC se mostrou pouco informativo para este grupo, uma vez que apresenta alta habilidade latente (Figura 17, direita). Por outro lado, o grupo Kalunga apresentou variabilidade adequada, porém a ausência de escolaridade podem afetar significativamente a interpretação clínica do teste.

Um histograma empírico (Figura 20) mostra a habilidade latente (Teta) em função da frequência esperada para o grupo de idosos hígidos e pacientes demenciados. Apesar de haver certa intersecção entre as habilidades latentes dos dois grupos, há uma separação nítida nas habilidades latentes, sendo que o grupo de pacientes demenciados estão, em sua maioria, num Teta menor ou igual a zero, enquanto que a frequência esperada para os idosos hígidos fica, em sua maioria, num Teta maior que zero.

As curvas de informação do teste PBAC foram calculadas para ambos os grupos e comparadas para verificar a possibilidade da análise TRI multigrupo (Figura 21).

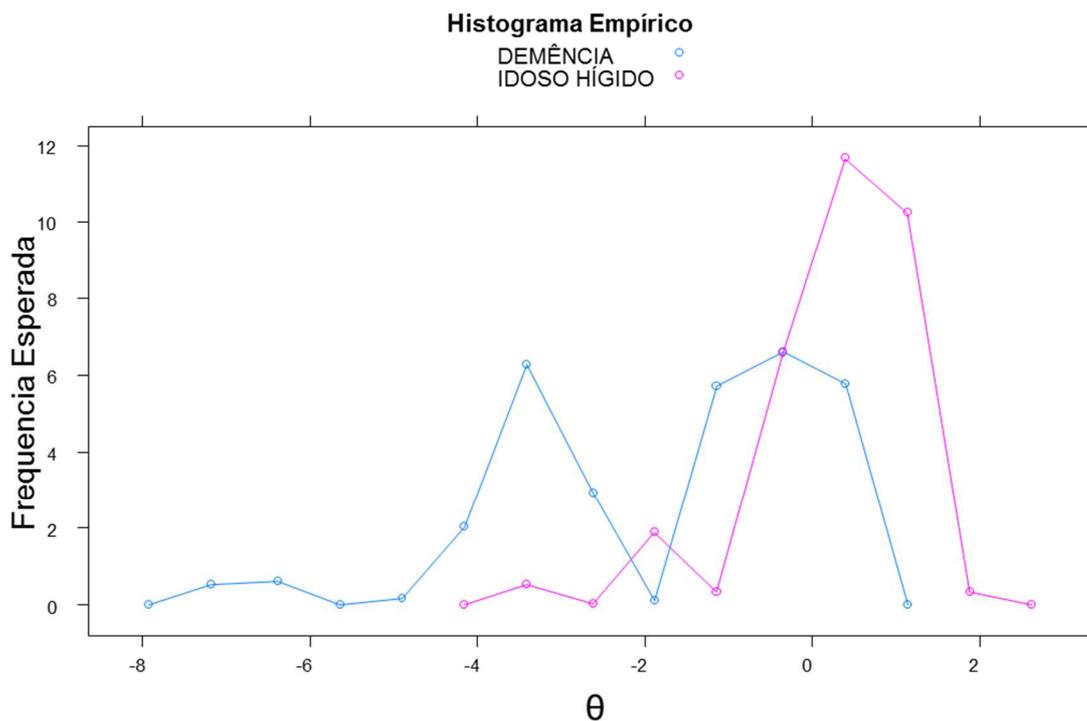


Figura 20. Histograma empírico em função da frequência esperada para os grupos de idosos hígidos (n=183, rosa) e demência (n=24, azul).

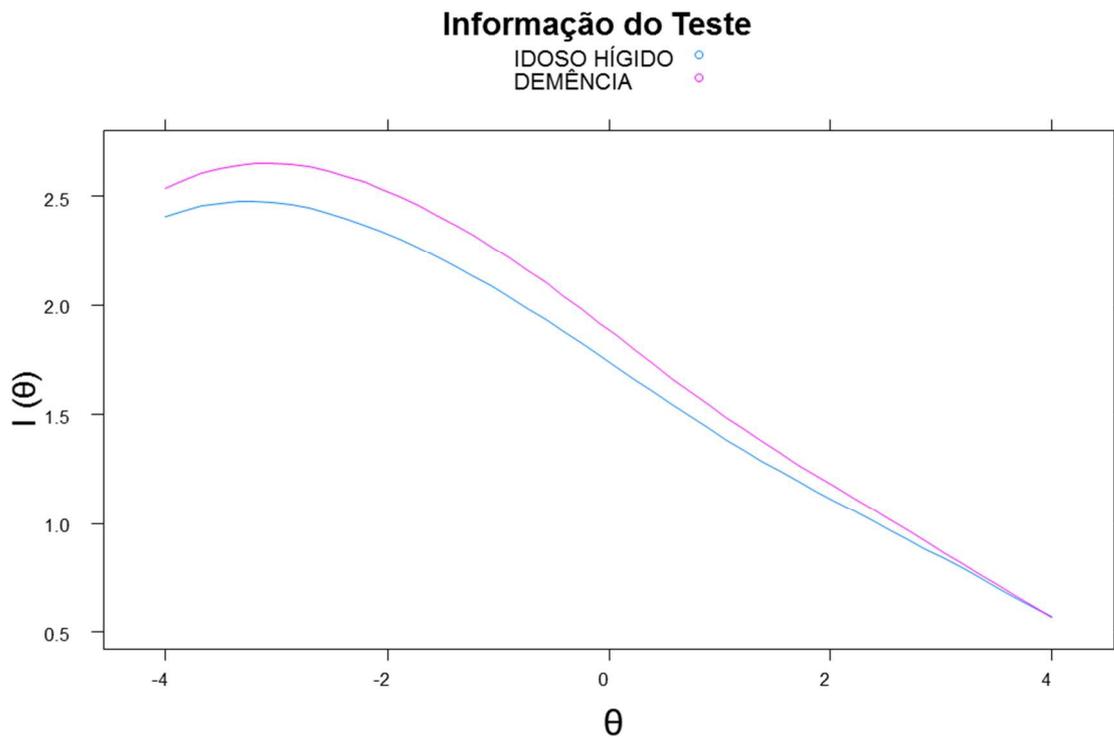


Figura 21. Curvas de informações do teste para o grupo de idosos hígidos (azul) e demência (rosa).

Foi realizada uma ANOVA comparando os índices de ajuste para as curvas obtidas na Figura 21, tendo como invariância a discriminação do teste, mostrando que não há diferença significativa entre as curvas ($\chi^2=3,25$; g.l.=12; $p=0,993$). Também não foi significativa a diferença entre as curvas quando o intercepto (dificuldade) foi igualmente fixado para os grupos, liberando as variâncias e as médias ($\chi^2=-174,22$; g.l.=64; $p=0,999$). Os resultados quando fixadas a média (*slopes*, interceptos e variâncias livres; $\chi^2=20,83$; g.l.=1; $p<0,001$) ou a variância (*slopes* e interceptos livres; $\chi^2=3,25$; g.l.=12; $p=0,993$) mostraram que a diferença entre as curvas é significativa (como esperado).

Estimativas multi-grupos (FIML) foram utilizadas para detectar o funcionamento diferencial do items (DIF). Para as análises, foi utilizado o algoritmo MHRM (119). Estas estimativas assumem que os modelos entre os grupos são independentes e que permitem, no entanto, impor limites ou restrições entre os grupos, tanto nos interceptos (fixando o parâmetro de dificuldade

entre os grupos) quanto nos slopes (fixando o parâmetro de discriminação para que seja igual entre os grupos), e liberando da mesma forma as variâncias e as médias.

Na Figura 22 pode-se observar a eficiência relativa do PBAC para o grupo de pacientes portadores de demência, tendo como linha de base (discriminação fixada em 1) o grupo de idosos hígidos. Esta análise revela que o teste possui maior eficiência entre os valores de Teta (habilidade) entre 1 e -4, atingindo seu pico de eficiência em -2. Resumidamente, o escore total do teste é capaz de discriminar indivíduos com declínio cognitivo. O erro padrão do teste (Figura 23) se torna maior nos Tetas acima de -1, tornando-se menos confiável para Tetas acima de 0.

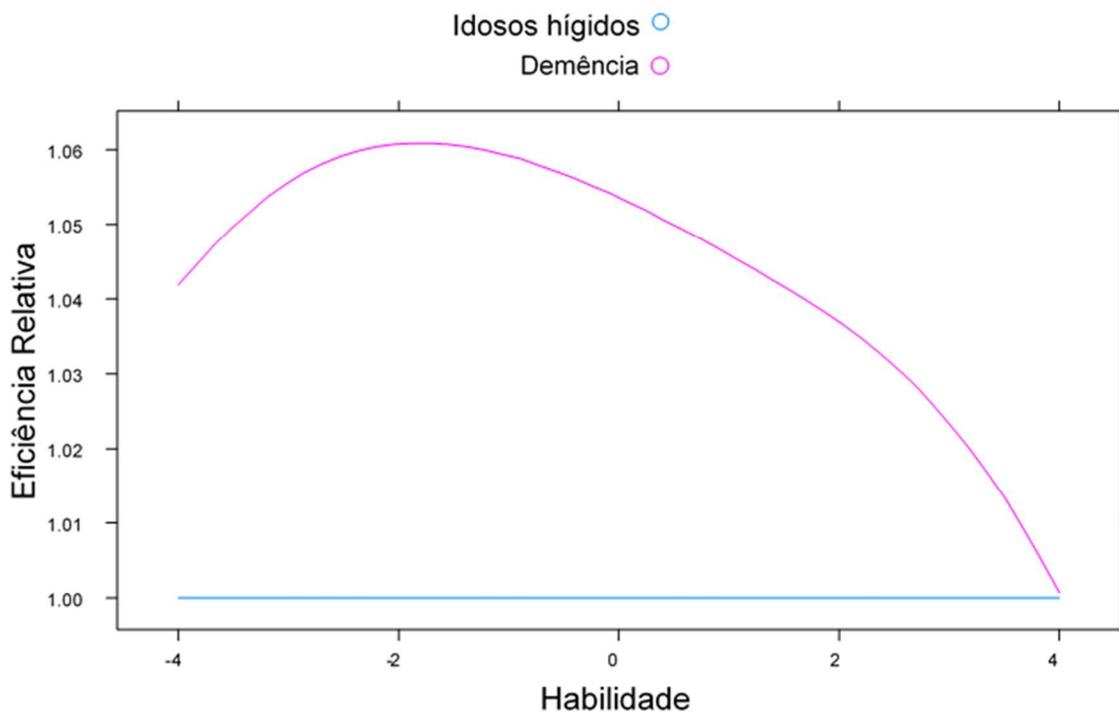


Figura 22. Eficiência relativa do teste PBAC para o grupo Demência (linha rosa), considerando os idosos hígidos como grupo de comparação (linha de base, azul).

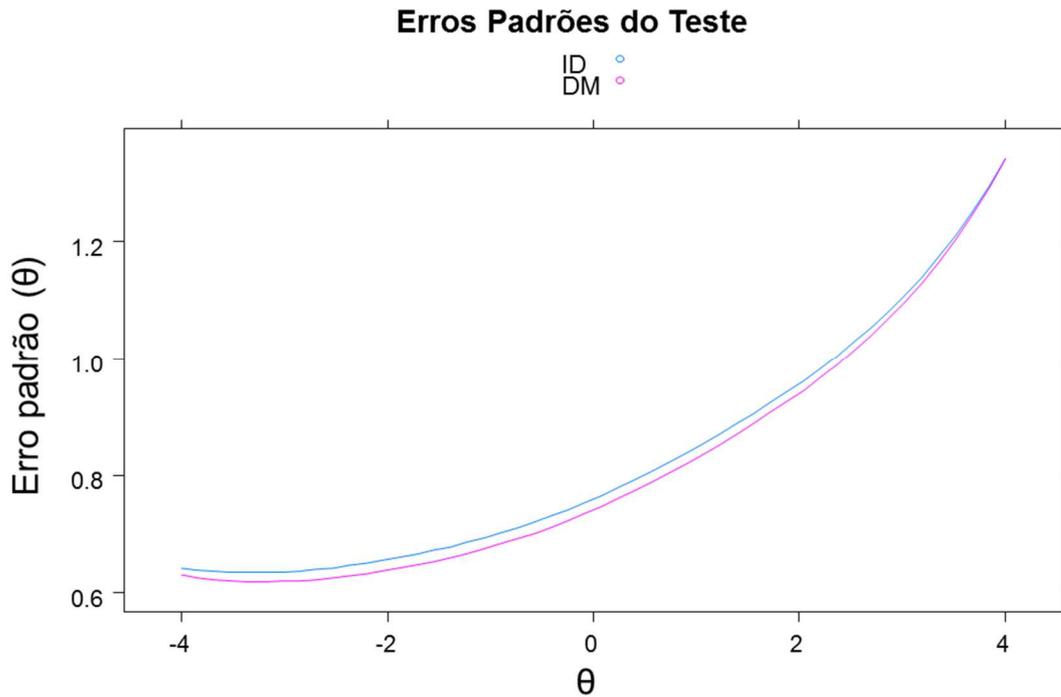
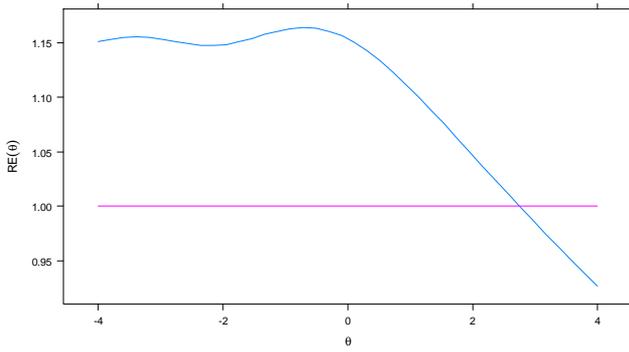


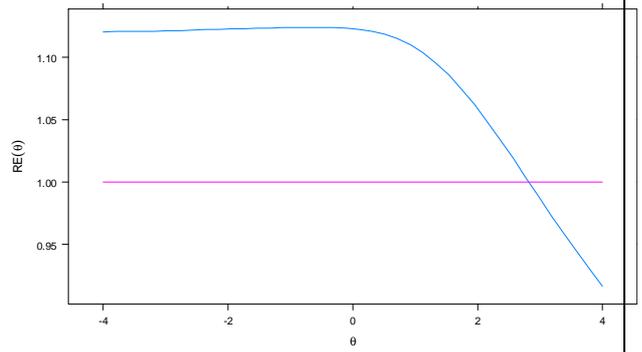
Figura 23. Erros padrões do teste PBAC. A curva azul representa o grupo de idosos hígidos e a curva rosa os pacientes portadores de demência. ID = idosos hígidos; DM = pacientes com demência.

Na Figura 24, os itens do teste foram analisados usando o mesmo princípio, tendo os idosos hígidos como grupo de comparação fixado em 1. Os itens Aprendizagem Verbal e Reconhecimento passam a discriminar pacientes com Teta igual ou menor que 0, enquanto o item Fluência da Letra F passa a discriminar pacientes com Teta igual ou menor que 1. Na prática, isto significa que a tarefa de Fluência é mais eficiente em discriminar indivíduos mais cognitivamente preservados do que os itens de Nomeação, Cópia da Figura de Rey e Leitura, que possuem picos em Tetas próximos a -2. Pode-se observar, ainda, que as tarefas de Evocação da lista de palavras, Orientação de Linhas e Evocação da Figura Complexa de Rey, a amplitude da área abaixo da curva é maior que nos demais itens. Isto demonstra que estes itens se tornam discriminativos em pacientes com habilidades mais elevadas. Os itens Escrita e Trilhas não são eficientes em detectar diferenças nos traços latentes.

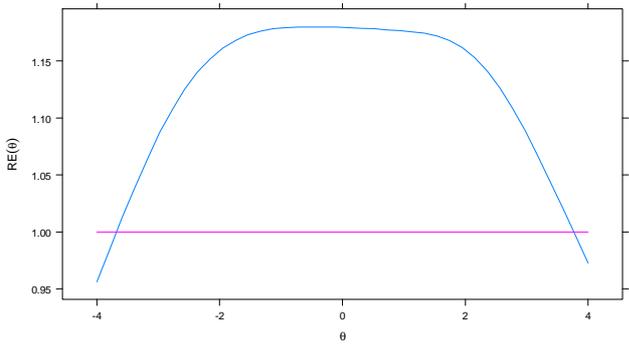
APRENDIZAGEM VERBAL



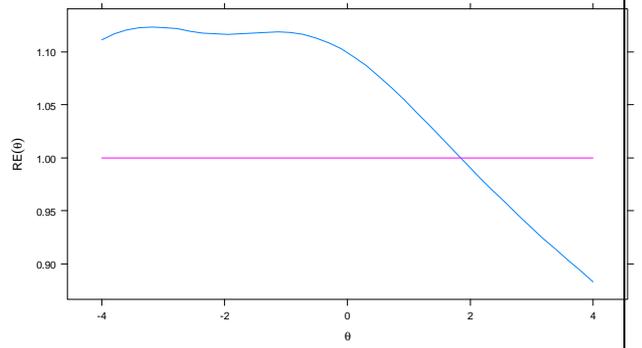
FLUÊNCIA



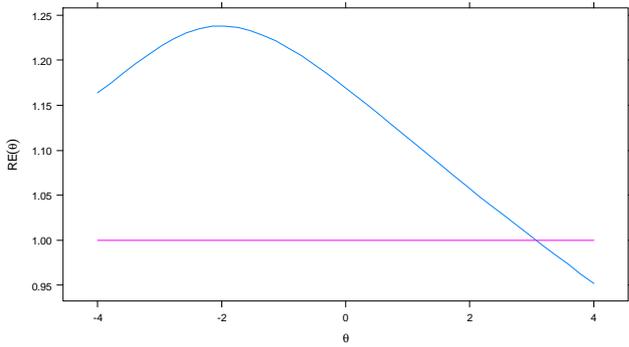
EVOCAÇÃO



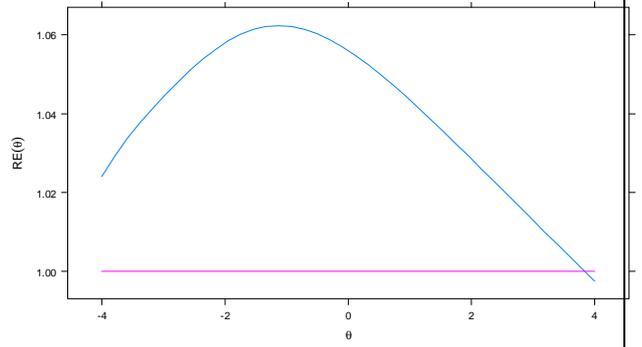
RECONHECIMENTO



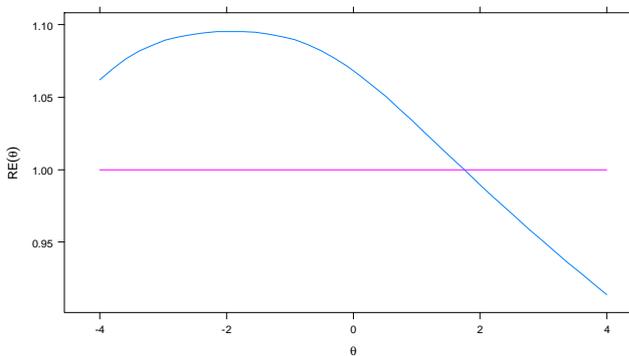
NOMEAÇÃO



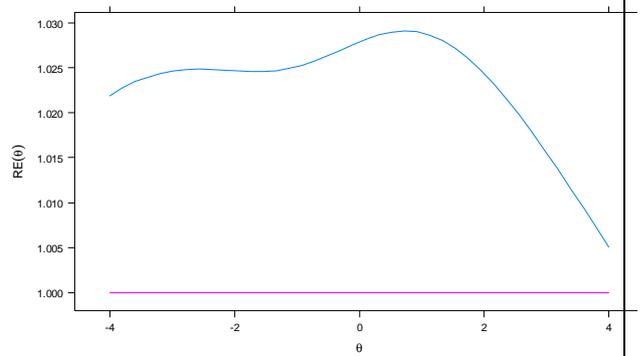
SEMÂNTICA



CÓPIA FIG. REY



ORIENTAÇÃO DE LINHAS



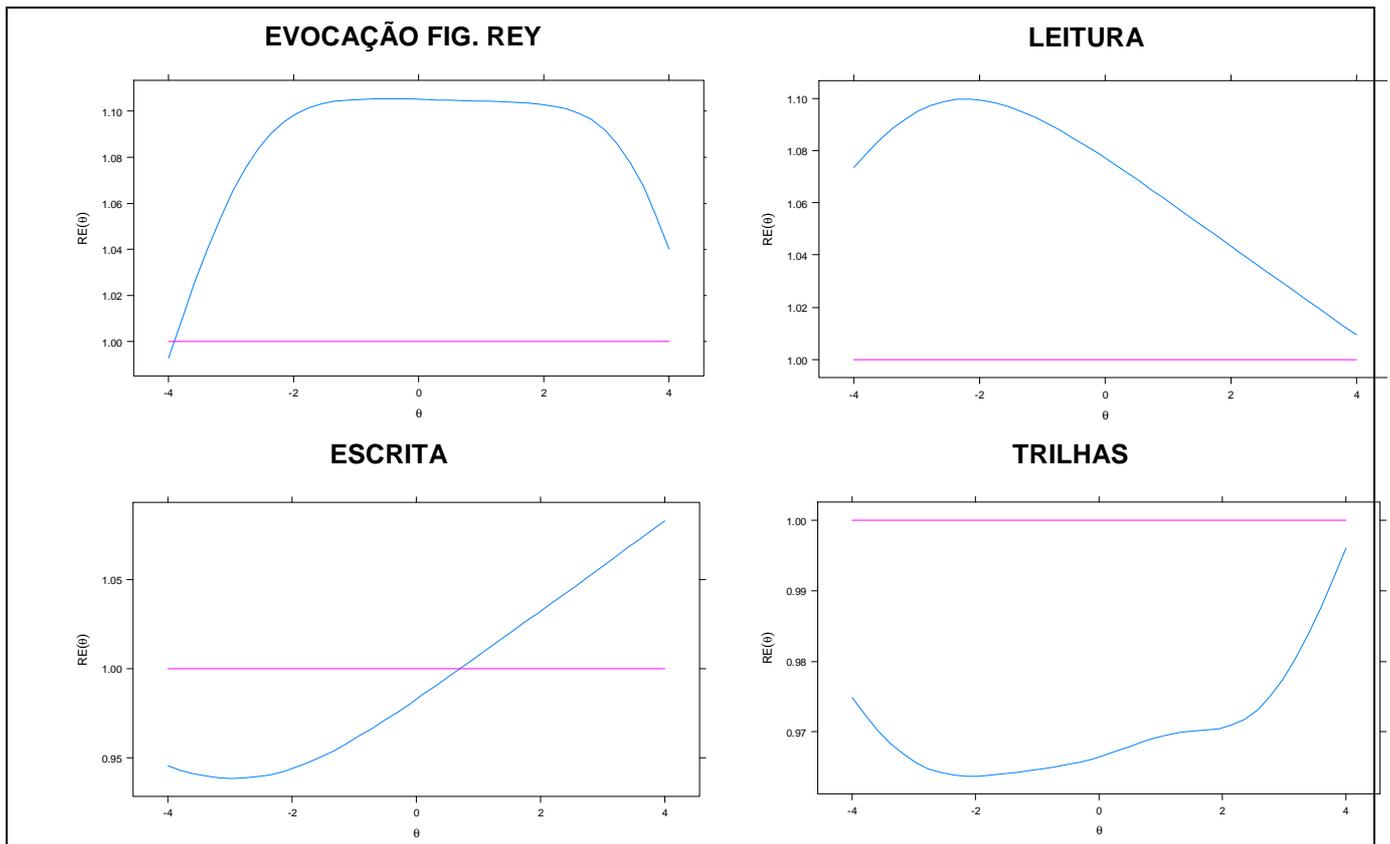


Figura 24. Eficiência relativa dos itens. Na linha de base (rosa) os idosos hígidos tiveram o parâmetro de discriminação fixado em zero. Na linha azul, os pacientes demenciados com parâmetro de discriminação variando em diferentes habilidades latentes (Teta).

Os resultados das análises de viés dos itens mostraram que o teste PBAC não possui itens com funcionamento diferencial (DIF). Para estas análises, várias alternativas foram isoladamente testadas: 1ª) sem ancoragem, todos os itens foram testados para DIF adicionando restrições aos itens, um de cada vez; 2ª) análises da matriz de informação com o método S-EM; 3ª) teste se a adição das restrições do slope ou do intercepto resultam em DIF; 4ª) o mesmo da alternativa anterior, mas usando os testes de Wald com ajustamento de Benjamini & Hochberg; 5ª) testar se a adição de restrições apenas no slope resulta em DIF para todos os itens; 6ª) usando alguns itens como âncoras; 7ª) uso de um modelo com restrição forte com a estimativa do parâmetro Teta (habilidade latente); e 8ª) busca sequencial usando o SABIC como critério de seleção.

Pode-se observar na Figura 24 que o subteste de Memória é mais informativo entre as habilidades latentes -1,7 a 0. O subteste de Memória possui os seguintes parâmetros: Evocação ($a=3,54$; $b=-0,64$), Reconhecimento ($a=2,33$; $b=-2,05$) e Evocação da Figura de Rey ($a=1,59$; $b=-0,73$), sendo a tarefa de Evocação da lista de palavras o item mais discriminativo ($a=3,54$) e o mais difícil ($b=-0,64$).

3.9 Tabelas Normativas e Parâmetro Teta de Habilidade Cognitiva

Mais informativo que um ponto de corte, as tabelas normativas se tornam extremamente eficientes para classificar os indivíduos dentro de uma gradação de medida, no caso a habilidade cognitiva. O escore z (que é uma normalização dos escores absolutos, considerando a média e o desvio-padrão do grupo) e escala T (que é uma transformação do escore z com média 50 e desvio-padrão de 10) são métricas bastante conhecidas na Psicometria. Na escala T , podemos classificar os indivíduos seguindo o seguinte critério: limítrofe $T < 35$, média inferior $35 < T < 42$, média $43 < T < 56$, média superior $57 < T < 62$, e superior $T > 62$. Nota-se que a escala $T < 35$ (limítrofe) refere-se a um escore $z < 1,5$. No entanto, estes escores são produzidos a partir da média, que naturalmente sofre influências de *outliers* e da não-normalidade dos dados. A TRI, propõe um sistema de classificação que considera a dificuldade e a discriminação dos itens. Um indivíduo, por exemplo que obteve 13 pontos no Fator 1 do PBAC (Tabela 16) tem habilidade cognitiva menor que um indivíduo que acertou apenas 10 pontos.

Nas Tabelas 16 a 20 estão disponíveis os Tetas, escore z e escala T para os idosos hígidos, correspondendo aos fatores 1 a 4 e ao escore total do teste. Ainda, histogramas dos escores totais de cada fator e do teste, bem como as curvas de informação estão demonstradas nas Figuras 25 a 29.

Tabela 16 – Escore do somatório dos itens do Fator 1, tabela normativa padrão e robusta para os idosos hígidos (n=183).

Escore	Teta	z	T
0	-2,79	-3,93	11
4	-2,17	-3,03	20
5	-1,55	-2,81	22
6	-1,49	-2,47	25
8	-1,44	-2,13	29
10	-0,79	-1,68	33
11	-1,25	-1,46	35
12	-0,91	-1,23	38
13	-1,00	-1,01	40
14	-0,70	-0,78	42
15	-0,42	-0,56	44
16	-0,32	-0,33	47
17	-0,15	-0,10	49
18	-0,12	0,12	51
19	0,06	0,35	53
20	0,21	0,57	56
21	0,34	0,80	58
22	0,97	1,02	60
23	1,11	1,25	62
24	1,62	1,47	65

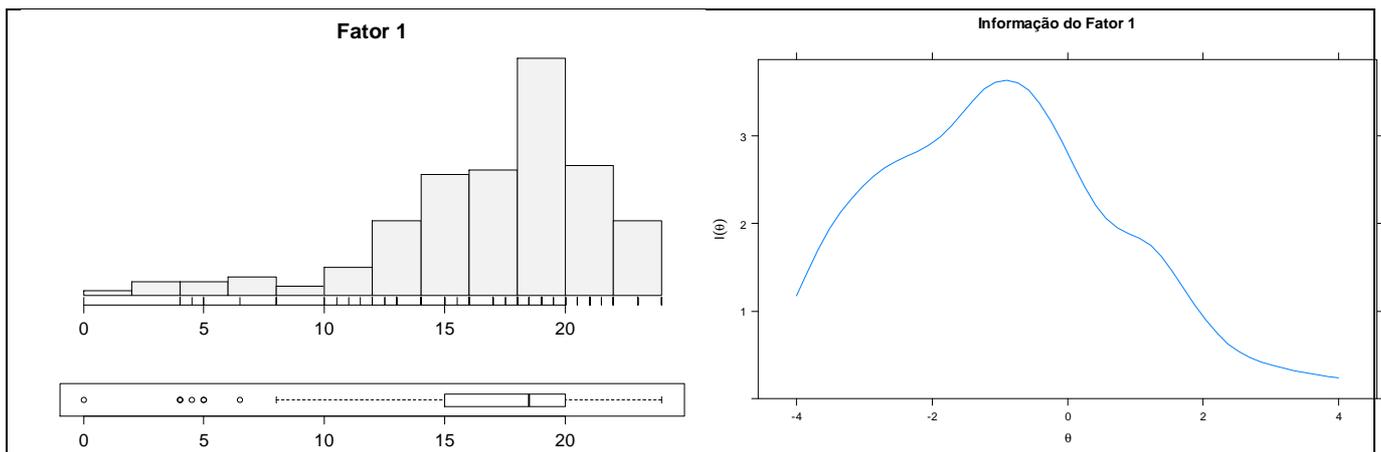


Figura 25. Esquerda: Histograma e boxplot do Fator 1 (idosos hígidos). Oulier robusto ≤ 10 pontos. Direita: Curva de Informação Total do Fator 1.

Tabela 17 - Escore do somatório dos itens do Fator 2 (Cópia da Figura de Rey e Orientação de Linhas), tabela normativa padrão e robusta para os idosos hígidos (n=183).

F2	Teta	z	T
1	-1,79	-12,62	-76
4	-0,65	-10,15	-51
5	-0,29	-9,32	-43
6	-1,60	-8,50	-35
11	-2,16	-4,37	6
12	0,07	-3,55	15
13	-1,16	-2,72	23
14	-1,19	-1,90	31
15	-0,80	-1,07	39
16	-0,36	-0,25	48
17	0,14	0,58	56
18	0,65	1,40	64

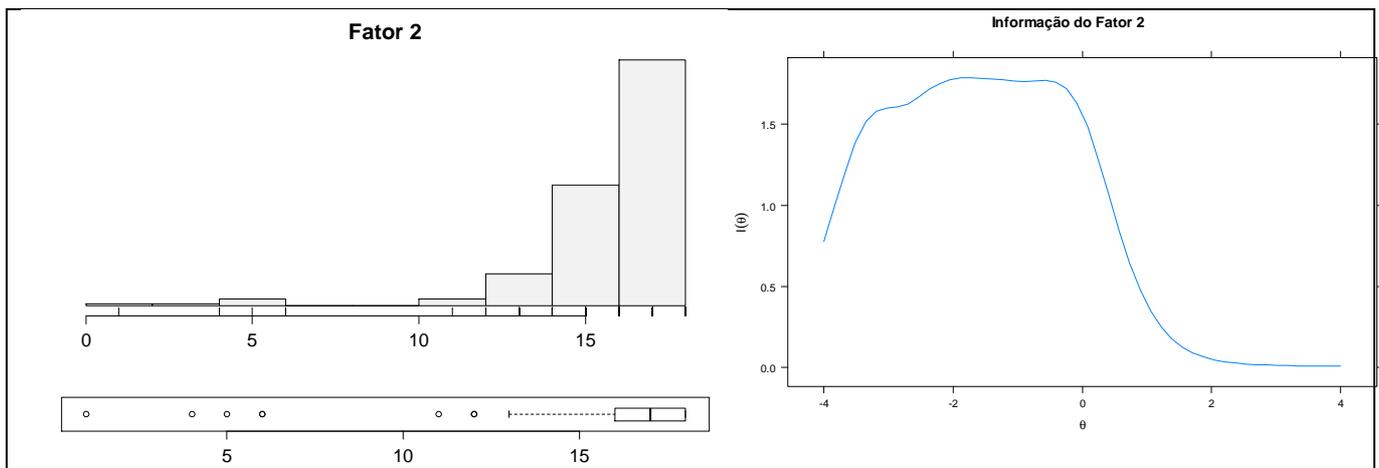


Figura 26. Esquerda: Histograma e boxplot do Fator 2 (idosos hígidos). Oulier robusto ≤ 13 pontos. Direita: Curva de Informação Total do Fator 2.

Tabela 18 - Escore do somatório dos itens do Fator 3 (Semântica, Nomeação, Leitura e Escrita), escores Teta, z e T para os idosos hígidos (n=183).

F2	Teta	z	T
10	-2,40	-6,30	-13
12	-1,77	-4,10	9
13	-1,19	-3,00	20
14	-1,01	-1,90	31
15	-1,03	-0,80	42
16	0,17	0,30	53

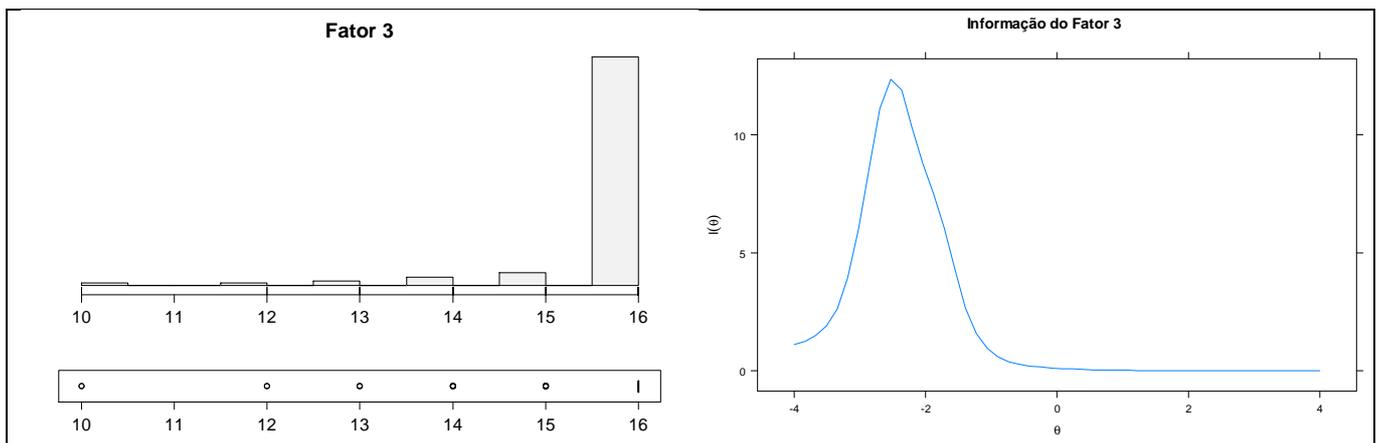


Figura 27. Esquerda: histograma e boxplot do Fator 3 (grupo idosos hígidos). Oulier robusto ≤ 15 pontos. Direita: Curva de Informação Total do Fator 3.

Tabela 19 - Escore do somatório dos itens do Fator 4 (Aprendizagem, Fluência e Trilhas), escores Teta, z e T para os idosos hígidos (n=183).

F4	Teta	z	T
5	-2,22	-3,11	19
6	-1,92	-2,76	22
7	-1,62	-2,41	26
8	-1,35	-2,06	29
9	-1,17	-1,72	33
10	-1,02	-1,37	36
11	-0,69	-1,02	40
12	-0,49	-0,67	43
13	-0,25	-0,32	47
14	-0,05	0,02	50
15	0,19	0,37	54
16	0,45	0,72	57
17	0,84	1,07	61
18	1,37	1,41	64

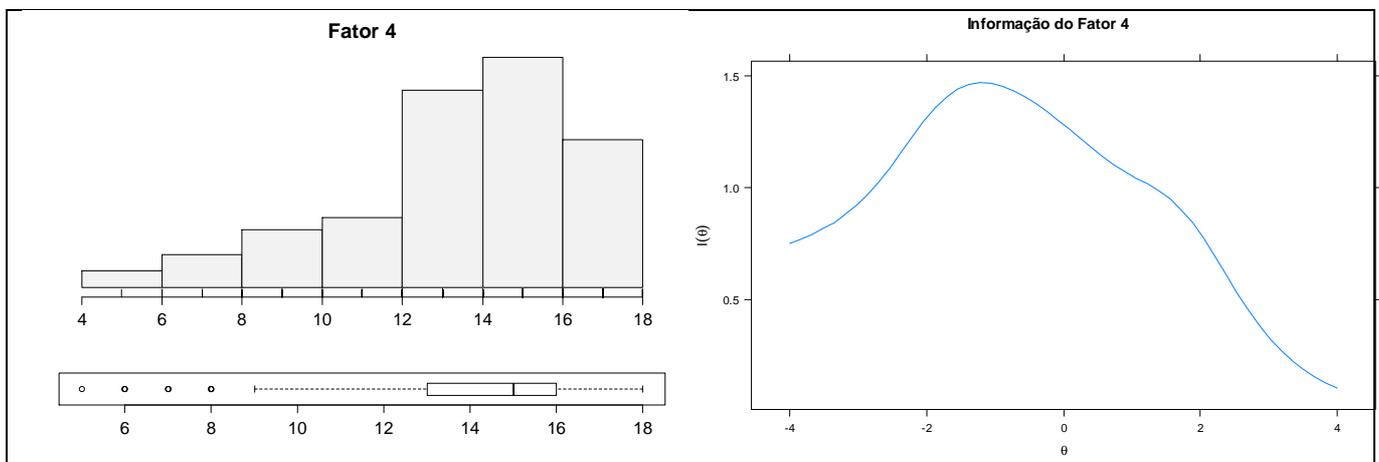
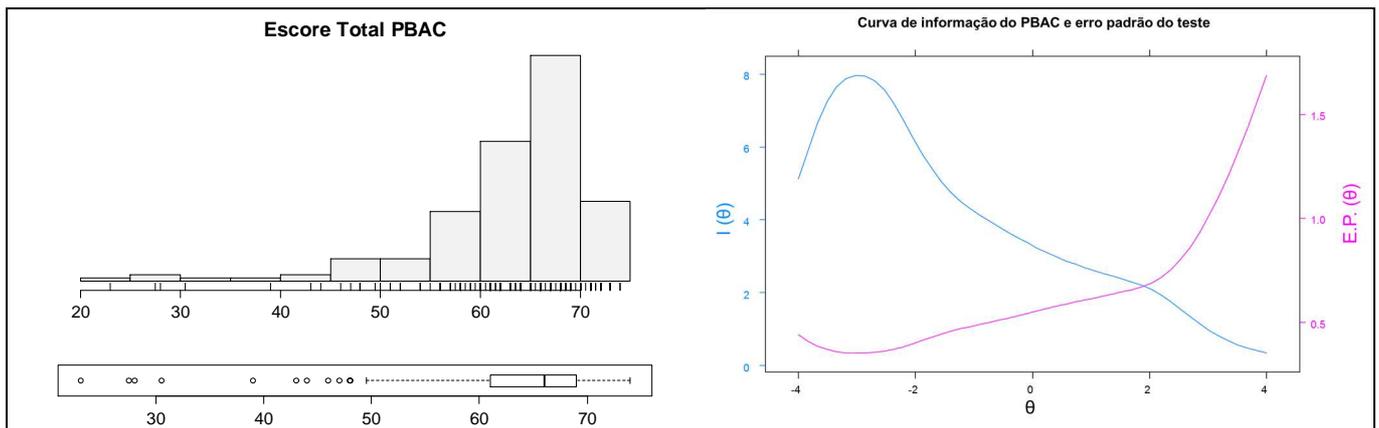


Figura 28. Esquerda: Histograma e boxplot do Fator 4 (Aprendizagem, Fluência e Trilhas), idosos hígidos. Oulier robusto ≤ 8 pontos. Direita: Curva de Informação Total do Fator 4.

Tabela 20 - Escore total do PBAC, escores Teta, z e T para os idosos hígidos (n=183).

PBAC	Teta	z	T	PBAC	Teta	z	T
23	-2,81	-4,72	3	60,5	-0,61	-0,34	47
27,5	-2,97	-4,19	8	61	-0,43	-0,28	47
28	-2,87	-4,14	9	61,5	-0,39	-0,22	48
30,5	-2,40	-3,84	12	62	-0,27	-0,17	48
39	-1,74	-2,85	21	63	-0,25	-0,05	50
43	-1,78	-2,38	26	63,5	-0,07	0,01	50
44	-2,15	-2,27	27	64	0,05	0,07	51
46	-2,12	-2,03	30	65	0,09	0,18	52
47	-1,70	-1,92	31	65,5	-0,09	0,24	52
48	-1,70	-1,80	32	66	0,26	0,30	53
49,5	-1,37	-1,63	34	66,5	0,20	0,36	54
50	-1,04	-1,57	34	67	0,22	0,42	54
51	-1,40	-1,45	35	67,5	0,15	0,48	55
52	-1,06	-1,33	37	68	0,47	0,53	55
54	-0,89	-1,10	39	68,5	0,28	0,59	56
55	-0,81	-0,98	40	69	0,56	0,65	57
56	-0,68	-0,87	41	69,5	0,52	0,71	57
57	-0,79	-0,75	42	70	0,78	0,77	58
57,5	-1,06	-0,69	43	70,5	0,62	0,83	58
58	-0,72	-0,63	44	71	0,88	0,88	59
58,5	-0,29	-0,58	44	71,5	0,91	0,94	59
59	-0,73	-0,52	45	72	1,28	1,00	60
59,5	-0,83	-0,46	45	73	1,31	1,12	61
60	-0,40	-0,40	46	74	1,50	1,23	62

Figura 29. Esquerda: Histograma e boxplot do escore total do PBAC, idosos hígidos. Oulier robusto ≤ 52 pontos.

Direita: Curva de Informação Total (azul) e erro padrão do teste (rosa).

4 DISCUSSÃO

O teste de rastreio cognitivo PBAC objetiva avaliar cinco domínios cognitivos: *funções executivas, linguagem, habilidades visoperceptivas, memória episódica visual/verbal e comportamento social*. Neste estudo, pudemos observar que o teste mede adequadamente quatro destes cinco domínios em idosos hígidos e demenciados, uma vez que o subteste *comportamento social* não pode ser devidamente avaliado por falta de variabilidade. Os itens que compõem este subteste tiveram forte correlação formando uma dimensão específica. As tarefas de Cópia e Evocação da Figura modificada de Rey e Orientação de Linha (JOLO) compõe uma segunda dimensão; e Escrita, Nomeação e Leitura, uma terceira dimensão. Prejuízos no subteste de *memória* estiveram relacionados com a atrofia do precúneo bilateral e hipocampo, junto com o diagnóstico de DA. Em nosso estudo, na curva ROC (Tabela 14), podemos ver que a tarefa de Evocação da lista de palavras é, de fato, mais sensível que as demais tarefas quando utilizada para detectar a presença de Alzheimer. Pode-se notar, ainda, que os itens de Leitura, Escrita e Nomeação possuem a menor área ROC, sendo considerados, portanto, menos indicativos para a detecção de demência.

A obtenção do modelo cognitivo de um teste, como proposto pela análise fatorial e a modelagem de equação estrutural, se torna extremamente relevante na Neuropsicologia moderna; uma vez que o conhecimento dos instrumentos de avaliação pode trazer à luz reformulações importantes acerca do atual conhecimento cognitivo. Os resultados obtidos na pesquisa de neuroimagem, usando como dado cognitivo o teste PBAC (59), podem corroborar com esta ideia de efeitos simultâneos de diferentes traços latentes na execução de uma tarefa. Os autores concluíram que o subteste de *funções executivas* ativou simultaneamente duas redes neurais distintas: o giro frontal superior (região comumente relacionada à memória) e o giro temporal superior.

Os resultados do modelo cognitivo demonstram que os quatro fatores obtidos podem ser assim classificados: F1 que contém os itens de Nomeação, Escrita e Semântica compõe o subteste de *linguagem*, F2 que contém os itens Orientação de Linhas e Cópia da Figura de Rey o subteste de *habilidade visoconstrutiva*, F3 formado pelos itens Evocação, Reconhecimento e Evocação da Figura de Rey compondo o subteste de *memória* e, F4 com os itens de Fluência, Trilhas Oraís e Aprendizagem Verbal, formando o subteste de *funções executivas*. No entanto, estes itens que compõe F4 possuem características típicas da habilidade *linguagem*, como seria o caso do item Fluência da letra F (tarefa de fluência fonêmica) e Aprendizagem Verbal que é covariante com o item de Nomeação, e da habilidade *memória*, como o item Aprendizagem Verbal, que é covariante do item de Evocação. No entanto, os fatores F4 e F1 não foram covariantes, sugerindo que ambos não são ativados por habilidades latentes comuns. Ainda, F4 (*funções executivas*) não sofre influência negativa da idade, ao passo que F3 (*memória*) sofre, o que podemos concluir que F4 é, de fato, um fator separado de F1 e de F3.

Devido à pouca variabilidade dos dados, distribuições não-normais foram observadas em praticamente todas as tarefas cognitivas apresentadas. Em decorrência disto, os modelos Bayesianos para as análises múltiníveis foram essenciais, uma vez que permitiram maior liberdade nas estimativas de covariância dos itens, além da reamostragem *Markov chain MonteCarlo*, que possibilitou a estimativa dos parâmetros mesmo em amostragens pequenas. Análises multiníveis e multigrupos podem ser extremamente complexas ou impossíveis nas estatísticas frequentistas (89, 94), uma vez que as mesmas requerem um número considerável de respondentes em cada grupo (nem sempre disponíveis quando se trata de amostras clínicas). A tarefa se torna mais difícil quando se busca explicar o fenômeno através dos conhecimentos em Neuropsicologia, uma vez que outros fatores intrinsicamente relevantes, como a idade, escolaridade, região geográfica, ou experiências prévias com testes clínicos ou psicológicos, possam interferir na estimativa dos parâmetros dos modelos.

A distribuição não-normal dos dados já era esperada para este tipo de instrumento de rastreio, uma vez que grande parte dos instrumentos de rastreio cognitivo são formados a partir de tarefas relativamente simples para a grande maioria da população. A inserção de idosos com baixa ou nenhuma escolaridade (no caso, o grupo Kalungas), aumentou a variabilidade de respostas para as tarefas de Leitura e Escrita, deslocando a média total do teste para valores mais baixos. Porém, a inserção desta amostra de idosos analfabetos pode não ter atingido o objetivo proposto, uma vez que muitos destes idosos nunca foram expostos à uma avaliação anterior. Ainda, hipóteses diagnósticas (como a doença de Alzheimer) não podem ser descartadas para este grupo, uma vez que nenhuma avaliação clínica foi realizada.

Como o propósito do estudo foi a adaptação cultural e a evidência de validade de um instrumento de rastreio cognitivo e de diagnóstico clínico, somente os dados dos idosos hígidos e dos pacientes demenciados foram considerados na segunda parte do estudo, onde foram obtidas as curvas de sensibilidade e especificidade, bem como na obtenção das curvas de informação do teste e dos itens, pela TRI multigrupo. Para a obtenção do modelo cognitivo para o teste PBAC, somente dados dos idosos hígidos foram utilizados, uma vez que o teste não se mostrou informativo para o grupo de jovens universitários. Uma dificuldade para se obter um modelo cognitivo do instrumento usando dados do grupo clínico é que certas habilidades latentes podem sofrer alterações, de modo que o indivíduo maneje de forma compensatória a forma de endossar a tarefa. Isto pode levar a efeitos indiretos de uma ou mais habilidades latentes preservadas numa outra que esteja clinicamente comprometida (35, 36, 114).

A magnitude da correlação entre o teste PBAC e outros testes de rastreio cognitivo, como o MEEM, teste do Relógio, Teste de Trilhas (A e B) e a fraca correlação com os testes de depressão (GDS) e de atividades funcionais (Pfeffer) mostraram evidências de sua validade convergente e discriminante. A correlação mais forte foi com o MEEM e com os testes de Trilhas, uma vez que estes compartilham características comuns à outras tarefas do PBAC, como Trilhas Orais, Nomeação, Aprendizagem e Evocação.

O alfa de Cronbach do teste foi considerado aceitável (0,86) para a amostra de pacientes demenciados e para a amostra total, adequado (0,75) para a amostra de idosos hígidos e a desejar (0,37) para a amostra de jovens universitários (106). Sabe-se, no entanto, que os valores de alfa são muito sensíveis, tornando-se menos específicos quando utilizados em dados não normais ou em instrumentos multidimensionais (128).

A TRI multigrupo mostrou que a eficiência relativa do teste PBAC é maior para os idosos portadores de demência que para os idosos hígidos, entre as habilidades -4 e 1. Com exceção de Escrita e Trilhas Orais, todos os itens são mais informativos para o grupo de idosos demenciados do que para o grupo de idosos hígidos. A retirada destes itens do teste PBAC (Escrita e Trilhas) pode deixar o teste mais informativo para o grupo clínico, permitindo melhor calibração para o mesmo. No entanto, estudos com outras amostras clínicas, bem como maior número de idosos hígidos de baixa escolaridade se tornam necessários antes da etapa de redução da escala.

Apesar do teste PBAC ser bastante promissor para o uso clínico, estudos com outras amostras, como pacientes portadores de demência fronto-temporal, demências subcorticais, dentre outras, precisam ser realizadas. Ainda, aumentar o tamanho da amostra de idosos hígidos para a obtenção de dados normativos para outros níveis educacionais se torna imprescindível para que o instrumento tenha maior aplicabilidade entre a população da terceira idade. Por fim, calibrar o teste PBAC, eliminando itens não informativos, pode aumentar seu potencial diagnóstico, mesmo que se torne um instrumento mais curto, com apenas 9 ou 10 itens.

4.1 Limitações do Estudo

Um problema apontado neste estudo está na elaboração da metodologia para a coleta de dados, uma vez que os sujeitos hígidos foram selecionados a partir do escore maior que 18

pontos. Este critério de inclusão afetou a variabilidade de respostas nas tarefas cognitivas, uma vez que delimitou o nível da habilidade latente dos participantes.

Os resultados obtidos mostraram que o MEEM teve maior acurácia que o PBAC nas curvas ROC (apesar de não serem significativamente diferentes) para detectar a presença de um quadro demencial. No entanto, este achado pode ser resultante do fato de que escores do MEEM foram utilizados para a realização do diagnóstico clínico nos pacientes demenciados (procedimento comumente realizado no Centro de Medicina do Idoso do Hospital Universitário de Brasília), enquanto que escores do PBAC não.

Os itens de Aprendizagem Verbal, Evocação e Reconhecimento de palavras do PBAC, diferentemente do RAVLT, não avaliam as interferências proativas e retroativas, uma vez que não utilizam a lista B de palavras. Tampouco avaliam a intrusão de palavras não contidas na lista, quando o sujeito diz uma palavra que não estava presente na lista apresentada, fenômeno conhecido como intrusão de memória. As palavras utilizadas no PBAC na tarefa de memória episódica verbal não foram escolhidas com base na frequência de seu uso na língua portuguesa e, ainda, tiveram extensões entre 4 e 7 letras, enquanto que variava apenas entre 3 a 5 letras no idioma original. Sabe-se que a extensão da palavra influencia na capacidade de armazenamento e na evocação das palavras (69), variável esta que deveria ter sido considerada na adaptação do teste. Outro problema é em relação às figuras utilizadas na tarefa de nomeação. A intenção dos autores na escolha dos estímulos do *Boston Naming Test* não é conhecida, pois as figuras podem ter sido propositalmente escolhidas por diferirem em termos da probabilidade de evocação correta (dificuldade do item) na população norte-americana, escolhendo, por exemplo, dois itens de grau de dificuldade fácil, dois itens médios e dois itens difíceis. Ao utilizar outras imagens não pertinentes ao *Boston Naming Test*, a adaptação brasileira pode ter desconsiderado esta possibilidade.

Estudos multigrupo, como o modelo de equação estrutural e a teoria de resposta ao item, não puderam ser adequadamente utilizados devido ao pequeno número de indivíduos na amostra Kalunga e idosos clínicos. Ainda, o grupo de jovens não pode ser adequadamente utilizado, por não apresentar variabilidade de respostas. Desta forma, os modelos cognitivos ficaram restritos à amostra total ou à amostra de idosos hígidos.

4.2 Próximos Passos: Calibração do PBAC e Análise de Invariância Multigrupo

Devido às características do teste de rastreio cognitivo PBAC, uma proposta de modificação do mesmo se justifica, uma vez que itens como Leitura, Escrita e Trilhas Orais podem contribuir pouco ou nada para a avaliação das capacidades cognitivas dos indivíduos. Com a exclusão dos itens, os modelos se tornam mais fidedignos, por possuírem melhor ajustes aos dados.

Uma vez que o objetivo é verificar se o teste de rastreio neuropsicológico PBAC apresenta evidências de validade para as diferentes amostras avaliadas, o mesmo deve ser invariante, isto é, medir a mesma habilidade latente para todos os grupos. Se algumas relações causais são testadas em mais de uma população, podemos perguntar se as mesmas relações causais se mantêm em outras populações. O modelo de equação estrutural multigrupo pode ser utilizado para examinar tal heterogeneidade da população (21, 50, 51, 53, 73). A modelagem multigrupo implementa, simultaneamente, análises de múltiplas populações e o modelo de análise fatorial confirmatória multigrupo é frequentemente utilizado para testar a invariância (usada aqui como sinônimo de equivalência) fatorial das escalas medidas (84, 129).

5 CONSIDERAÇÕES FINAIS

Um dos princípios da ciência é que, de alguma forma, se busca antecipar resultados. Alguns cientistas usam de teorias e metodologias rigorosas, enquanto outros são mais informais acerca do resultado obtido através de suas próprias observações. O mesmo vale para a Psicologia, em especial a Neuropsicologia. Uma das frases que mais tenho ouvido nos últimos anos sobre os resultados obtidos a partir de extensas análises estatísticas e psicométricas é: “E daí?”. Tal questionamento requer uma resposta que, muitas vezes, depende de um contexto relevante. Uma resposta que sempre me prontifico a oferecer é que os resultados obtidos podem servir para identificar um traço latente, uma habilidade do indivíduo, independentemente de sua condição clínica. Prever e generalizar comportamentos futuros com base nas respostas emitidas durante a tarefa, e que esta previsão não dependa única e exclusivamente de meus julgamentos pessoais, de minhas habilidades como clínico.

Há muitas teorias, mas podemos testar estas teorias uma vez que seus atributos possam ser adequadamente medidos. No entanto, muitas vezes, estamos medindo atributos diferentes quando usamos populações diferentes. Meu exagero no detalhamento estatístico se justifica pelo fato de que o uso de métodos mais poderosos é essencial na elaboração de teorias mais extensas. Conhecer que as demências podem causar vieses nas avaliações neuropsicológicas, até mesmo nos instrumentos de rastreio cognitivo, nos leva a elaborar teorias psicológicas importantes.

Muitos teóricos modernos tem discutido se o escore do teste é válido ou se o teste é válido; isto é, os valores de mensuração ou o instrumento de mensuração (130). A crítica é sobre os conceitos de validade e confiabilidade do escore de um teste, buscando enfatizar a necessidade de encontrar teorias mais substantivas dos atributos cognitivo/psicológicos que justifiquem a sua construção. Por exemplo, ao comparar dois instrumentos de avaliação (como o PBAC e MEEM),

assume-se que estes instrumentos possuem os mesmos atributos e que são representações válidas deste atributo. No entanto, a pergunta inicial sobre o que é de fato válido: o teste ou o escore do teste, leva os psicometristas a refletir sobre os conceitos de validade e também de fidedignidade ou confiabilidade. O foco é no teste, a questão é o que este teste mede. Se o foco é no escore do teste, a questão é para que propósitos o escore do teste poderá ser utilizado (130). O termo validade na Medicina e demais Ciências da Saúde foca, predominantemente, na utilidade prática do escore do teste e a pergunta sobre o que de fato o teste mede é amplamente suprimida.

A confiabilidade é um conceito mais técnico e restrito que a validade, uma vez que é menos problemático, apesar da estimativa da confiabilidade (como a característica do grupo) e da precisão de medida (como a característica do indivíduo) sejam passíveis de discussão (131).

Retomando a história da avaliação psicológica, já em meados de 1920 o atributo psicológico era considerado o agente que causava as respostas que uma pessoa oferecia aos itens, e o teste era considerado um instrumento para medir este agente. Medir era derivado desta relação causa-resposta. A validade de um teste era de acordo com um critério (como seleção de soldados, operários, estudantes, etc). Um mesmo teste poderia, por exemplo, ter diferentes validades, se tivesse vários critérios. Em 1989, Messick (pg. 13) propôs que “validade é um julgamento avaliativo integrado do grau de evidência empírica e razões teóricas que suportem a adequação e propriedade de inferências e ações baseadas nos escores dos testes e outros modelos de avaliação” (132). Desta forma, validade se torna um conceito que se refere à uma interpretação ou uso particular do escore de um teste. Como resultado, o que o teste mede se torna menos relevante do que a ênfase da utilidade prática do mesmo, como se vê claramente nos instrumentos de rastreio cognitivo, como o MEEM.

O problema básico em medida psicológica é a ausência, na maioria dos casos, de teorias bem fundamentadas e empiricamente testadas sobre os atributos psicológicos que fundamentem

a construção dos itens de um teste (133). Ainda, se há uma teoria disponível, o problema passa a ser a existência de várias outras teorias sobre o mesmo atributo, não havendo experimentos suficientes para decidir qual teoria favorecer. Em decorrência disto, se torna comum durante a construção de um teste selecionar itens (de um banco de itens, por exemplo) que definam o que o teste avalia, em vez de se basear na teoria que guiaria a operacionalização do atributo naquele item. O pesquisador fundamenta-se, então, na psicometria para achar o que o teste mede e, então, interpreta a estrutura da análise fatorial ou através das revelações da TRI para cada item e aceita estas interpretações e revelações como uma explicação de como os respondentes responderam os itens.

Neste estudo, utilizar grupos com habilidades notoriamente distintas como jovens e idosos hígidos, pacientes demenciados e uma amostra de pessoas de uma comunidade isolada, seria uma tentativa de comparar como diferentes atributos cognitivos poderiam contribuir para conhecer as diferentes características dos itens do teste. No entanto, estas diferenças entre os grupos (como a escolaridade ou mesmo a familiaridade que o grupo tem em ser avaliado) se torna uma armadilha para a pesquisa, uma vez que as variáveis intervenientes contaminam a interpretação teórica dos atributos cognitivos. De forma semelhante, buscar amostras com indivíduos de baixa habilidade latente no atributo para aumentar a variabilidade de resposta no escore do teste, pode ser útil apenas se algumas perguntas pudessem ser adequadamente respondidas: quais processos cognitivos foram ativados? Como a psicometria poderia contribuir nos modelos de processamento cognitivo ativos ou inativos durante a execução dos itens do teste? (134) Como a psicometria poderia auxiliar na elaboração de modelos de diagnósticos clínicos? (135). No entanto, é preciso refletir até que ponto uma determinada teoria pode ser inferida a partir dos dados coletados e analisados.

Em suma, as teorias definem os atributos como níveis elevados de abstração, mas os atributos apenas se tornam tangíveis e observáveis no comportamento. As estruturas teóricas dos atributos precisam ser traduzidos em comportamentos observáveis que sejam típicos

daquele atributo (processo conhecido como operacionalização) (130), ou seja, a especificação da operação precisa medir o atributo. Os comportamentos típicos são provocados por itens bem definidos e escolhidos de forma que os respondentes ofereçam respostas ou soluções que sejam informativas daquele atributo. Isto só funciona bem com uma teoria consistente e com evidências científicas ou psicométricas suficientes. Neste caso, o teste PBAC se torna forte candidato, uma vez que seus itens são bem definidos na literatura neuropsicológica, uma vez que possui forte correlação com estruturas cerebrais de pacientes com doenças neurodegenerativas, e uma vez que a maioria de seus itens informam sobre a alteração dos atributos em pacientes demenciados quando comparado com os idosos hígidos.

O teste PBAC segue, até certo ponto, o princípio de que escores mais altos reflitam níveis mais altos nos atributos avaliados, como proposto pelos teóricos clássicos. No entanto, esta hipótese nem sempre pode ser observada quando o instrumento é avaliado utilizando-se modelos modernos, como a teoria de resposta ao item. No PBAC, as análises psicométricas dos dados produziram informações relevantes sobre a estrutura do teste e sobre sua qualidade clínica, mas ainda é necessário investigar se os resultados obtidos podem contribuir com a teoria dos atributos avaliados, o que normalmente ocorre após repercussão do instrumento entre outros pesquisadores, e sua utilização em outros contextos preditivos. O uso de modelos estruturais, modelos de classe latente e as análises de item pela TRI podem ser bastante úteis para compreender os modelos dos processos cognitivos subjacentes. O objetivo do teste PBAC é medir os atributos propostos pela teoria neuropsicológica, utilizá-lo como instrumento preditivo de um determinado critério (como o declínio das funções cognitivas) e permitir uma discussão acerca das variáveis intervenientes dos atributos necessários para a execução de suas tarefas.

6 REFERÊNCIAS BIBLIOGRÁFICAS

1. Nunnally J, Bernstein I. *Psychometric Theory*. 3rd ed. New York, NY: McGraw-Hill; 1994. 752 p.
2. Lezak MD, Howieson DB, Bigler ED, Tranel D. *Neuropsychological Assessment*. 5th ed. New York, NY: Oxford University Press; 2012.
3. Barr WB. Historical development of the neuropsychological test battery. In: Morgan J, Ricker J, editors. *Textbook of clinical neuropsychological*. New York, NY: Psychology Press; 2008.
4. Malec JF. Ethical and evidence-based practice in brain injury rehabilitation. *Neuropsychological rehabilitation*. 2009;19(6):790-806.
5. da-Silva SL, Pereira DA, Veloso F, Satler CE, Arantes A, Guimarães RM. Programa de Reabilitação Neuropsicológica da memória aplicada à demência: um estudo não-controlado intra-sujeitos. *Estudos de Psicologia*. 2011;28(2):229-40.
6. da-Silva SL, Veloso FR, Pereira DA. Pressupostos e experiências de atendimentos de reabilitação neuropsicológica ecológica. In: Abrisqueta-Gomez J, editor. *Reabilitação Neuropsicológica - abordagem interdisciplinar e modelos conceituais na prática clínica*. Porto Alegre, RS: ArtMed GrupoA; 2012. p. 334-42.
7. Weyandt L, DuPaul GJ, Verdi G, Rossi JS, Swentosky AJ, Vilaro BS, et al. The Performance of College Students with and without ADHD: Neuropsychological, Academic, and Psychosocial Functioning. *J Psychopathol Behav Assess*. 2013:1-15.
8. Vakil E, Blachstein H, Wertman-Elad R, Greenstein Y. Verbal learning and memory as measured by the Rey-Auditory Verbal Learning Test: ADHD with and without learning disabilities. *Child neuropsychology : a journal on normal and abnormal development in childhood and adolescence*. 2012;18(5):449-66.

9. Yeates KO, Ris MD, Taylor HG, Pennington BF, editors. *Pediatric neuropsychology: Research, theory, and practice*. 2nd ed. New York, NY: Guilford Press; 2010.
10. Schmid R, Eschen A, Ruegger-Frey B, Martin M. Instruments for comprehensive needs assessment in individuals with cognitive complaints, mild cognitive impairment or dementia: a systematic review. *Int J Geriatr Psychiatry*. 2012;27(4):329-41.
11. Mitrushina M, Boone KB, Razani J, D'Elia LF. *Handbook of Normative Data for Neuropsychological Assessment*. 2nd ed. New York, NY: Oxford University Press; 2005.
12. Strauss E, Sherman EMS, Spreen O. *A Compendium of Neuropsychological Tests: administration, norms and commentary*. 3rd ed. New York, NY: Oxford University Press; 2006.
13. Turner SM, DeMers ST, Fox HR. APA's Guidelines for Test User Qualifications. *American Psychologist*. 2001;56(12):1099-113.
14. Larrabee GJ, editor. *Forensic Neuropsychology - a scientific approach*. 2nd ed. New York, NY: Oxford University Press; 2012.
15. Chaytor N, Schmitter-Edgecombe M. The Ecological Validity of Neuropsychological Tests: A Review of the Literature on Everyday Cognitive Skills. *Neuropsychol Rev*. 2003;13(4):181-97.
16. Marcotte TD, Grant I, editors. *Neuropsychology of everyday functioning*. New York, NY: Guilford Press; 2009.
17. Ashendorf L. A Compendium of Tests, Scales, and Questionnaires: The Practitioner's Guide to Measuring Outcomes After Acquired Brain Impairment. *Applied Neuropsychology*. 2012;19(1):78-9.
18. Brooks BL, Iverson GL, Lanting SC, Horton AM, Reynolds CR. Improving test interpretation for detecting executive dysfunction in adults and older adults: prevalence of low scores on the test of verbal conceptualization and fluency. *Applied neuropsychology Adult*. 2012;19(1):61-70.
19. Demakis GJ. Clinical practice of forensic neuropsychology: An evidence-based approach. *Archives of Clinical Neuropsychology*. 2013;28(3):304-5.

20. Dibello LV, Roussos LA, Stout W. Review of Cognitively Diagnostic Assessment and a Summary of Psychometric Models. In: Rao CR, Sinharay S, editors. *Handbook of Statistics on Psychometrics*. 26. Amsterdam: Elsevier; 2006. p. 979-1030.
21. Loehlin JC. *Latent Variable Models*. 4th ed. Mahwah, New Jersey: Lawrence Erlbaum Associates Inc; 2004.
22. Jones A, Prangnell SJ, Thomas C, Newby G. Neuropsychological assessment: the not-so-basic basics. In: Newby G, Coetzer R, Daisley A, Weatherhead S, editors. *The Handbook of Real Neuropsychological Rehabilitation in Acquired Brain Injury*. London: Karnac Books Ltd; 2013. p. 27-65.
23. Damasio H, Grabowski TJ, Tranel D, Ponto LL, Hichwa RD, Damasio AR. Neural correlates of naming actions and of naming spatial relations. *Neuroimage*. 2001;13(6 Pt 1):1053-64.
24. Roussos LA, Dibello LV, Henson RA, Jang EE, Templin JL. Skill diagnosis for education and psychology with IRT-based parametric latent class models. In: Embretson S, Roberts J, editors. *New Directions in Psychological Measurement with Model-based Approaches*: American Psychological Association; 2006.
25. Allen MJ, Yen WM. *Introduction to Measurement Theory*. Long Grove, Illinois: Waveland Press, Inc; 2002. 310 p.
26. Groth-Marnat G. *Handbook of Psychological Assessment*. 4th ed. New Jersey: John Wiley & Sons Inc; 2003. 797 p.
27. Cronbach LJ. *Essentials of Psychological Testing*. 2nd ed. New York: Harper; 1960. 650 p.
28. Furr MR, Bacharach VR. *Psychometrics: An Introduction*. 2nd ed. Thousand Oaks, CA: Sage; 2014. 447 p.
29. Lord FM. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1980.
30. Brennan RL. Generalizability Theory and Classical Test Theory. *Applied Measurement in Education*. 2010;24(1):1-21.

31. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res.* 2007;16(S1):5-18.
32. Harlow LL. Making Meaning of Multivariate Methods. In: Maydeu-Olivares A, McArdle JJ, editors. *Contemporary Psychometrics*. New Jersey: Lawrence Erlbaum Associates; 2005. p. 537-72.
33. Crane PK, Narasimhalu K, Gibbons LE, Mungas DM, Haneuse S, Larson EB, et al. Item response theory facilitated cocalibrating cognitive tests and reduced bias in estimated rates of decline. *Journal of Clinical Epidemiology.* 2008;61(10):1018-27.e9.
34. Embretson SE, Reise SP. *Item response theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates Inc. Publishers; 2000.
35. Ayala RJ. *The Theory and Practice of Item Response Theory*. New York: The Guilford Press; 2009.
36. Teresi JA, Ramirez M, Jones RN, Choi S, Crane PK. Modifying Measures Based on Differential Item Functioning (DIF) Impact Analyses. *Journal of Aging and Health.* 2012.
37. Mulaik SA. *The Foundations of Factor Analysis*. New York, NY: McGraw-Hill; 1972.
38. Chadha NK. *Applied Psychometry*. New Deli, India: Sage Publications Inc; 2009. 362 p.
39. Schultz-Larsen K, Kreiner S, Lomholt R. Mini-Mental Status Examination: Mixed Rasch model item analysis derived two different cognitive dimensions of the MMSE. *Journal of Clinical Epidemiology.* 2007;60(3):268-79.
40. Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician: Pergamon Press; 1975.
41. Mitchell A. The Mini-Mental State Examination (MMSE): An Update on Its Diagnostic Validity for Cognitive Disorders. In: Larner AJ, editor. *Cognitive Screening Instruments*: Springer London; 2013. p. 15-46.
42. Conrad KJ, Smith EV. International Conference on Objective Measurement: Applications of Rasch Analysis in Health Care. *Medical Care.* 2004;42(Supplement):I-1.

43. Bond TG, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences. 2nd ed. New York: Psychology Press; 2007.
44. Schultz-Larsen K, Lomholt R, Kreiner S. Mini-Mental Status Examination: A short form of MMSE was as accurate as the original MMSE in predicting dementia. *Journal of Clinical Epidemiology*. 2007;60(3):260-7.
45. Edelen MO, Thissen D, Teresi JA, Kleinman M, Ocepek-Welikson K. Identification of Differential Item Functioning Using Item Response Theory and Likelihood-Based Model Comparison Approach - application to the Mini-Mental State Examination. *Medical Care*. 2006;44(11):S134-S42.
46. Ramirez M, Teresi JA, Holmes D, Gurland B, Lantigua R. Differential Item Functioning (DIF) and the Mini-Mental State Examination (MMSE) - overview, sample, and issues of translation. *Medical Care*. 2006;44(11):S95-S106.
47. Morales LS, Flowers C, Gutierrez P, Kleinman M, Teresi JA. Item and Scale Differential Functioning of the Mini-Mental State Exam Assessed Using the Differential Item and Test Functioning (DFIT) Framework. *Medical Care*. 2006;44(11):S143-S51.
48. Teresi J. Mini-Mental State Examination (MMSE): Scaling the MMSE using item response theory (IRT). *Journal of Clinical Epidemiology*. 2007;60(3):256-9.
49. Yesavage JA, Sheikh JI. 9/ Geriatric Depression Scale (GDS). *Clinical Gerontologist*. 1986;5(1-2):165-73.
50. Bartholomew D, Knott M, Moustaki I. Latent Variable Models and Factor Analysis - a unified approach. Balding DJ, Cressie NAC, Fitzmaurice GM, Goldstein H, Molenberghs G, Scott DW, et al., editors. London, UK: Wiley, John Wiley & Sons, Ltd; 2011.
51. Wang J, Wang X. Structural Equation Modeling - applications using MPlus. Balding DJ, Cressie NAC, Fitzmaurice GM, Goldstein H, Johnstone IM, Molenberghs G, et al., editors. West Sussex, UK: John Wiley & Sons Ltd; 2012.
52. Seltzer M, Choi K. Sensitivity Analysis for Hierarchical Models: downweighting and identifying extreme cases using the t distribution. In: Reise SP, Duan N, editors. *Multilevel*

Modeling - methodological advances, issues, and applications. 3rd ed: Lawrence Erlbaum Associates; 2003. p. 25-52.

53. Byrne BM. Structural Equation Modeling with MPlus - basic concepts, applications, and programming. New York: Routledge, Taylor & Francis Group; 2012.

54. McDonald RP. Factor Analysis and Related Methods. New Jersey: Lawrence Erlbaum Associates; 1985.

55. Wilson M. Constructing Measures: an item response modeling approach. Mahwah, NJ: Lawrence Erlbaum Associates; 2005.

56. Libon DJ, Rascovsky K, Gross RG, White MT, Xie SX, Dreyfuss M, et al. The Philadelphia Brief Assessment of Cognition (PBAC): a validated screening measure for dementia. Clin Neuropsychol. 2011;25(8):1314-30.

57. Libon DJ, McMillan C, Gunawardena D, Powers BS, Massimo L, Khan B, et al. Neurocognitive contributions to verbal fluency deficits in frontotemporal lobar degeneration. Neurology. 2009;73(7):535-42.

58. Tan KS, Libon DJ, Rascovsky K, Grossman M, Xie SX. Differential Longitudinal Decline on the Mini-Mental State Examination in Frontotemporal Lobar Degeneration and Alzheimer Disease. Alzheimer disease and associated disorders. 2013; Publish Ahead of Print:10.1097/WAD.0b013e31827bdc6f.

59. Avants BB, Libon DJ, Rascovsky K, Boller A, McMillan CT, Massimo L, et al. Sparse canonical correlation analysis relates network-level atrophy to multivariate cognitive measures in a neurodegenerative population. NeuroImage. 2014;84(0):698-711.

60. Libon DJ, Massimo L, Moore P, Coslett HB, Chatterjee A, Aguirre GK, et al. Screening for frontotemporal dementias and Alzheimer's disease with the Philadelphia Brief Assessment of Cognition: a preliminary analysis. Dementia and geriatric cognitive disorders. 2007;24(6):441-7.

61. Tombaugh TN, Kozak J, Rees L. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists. 1999;14(2):167-77.

62. Ricker JH, Axelrod BN, Houtler BD. Clinical Validation of the Oral Trail Making Test. *Cognitive and Behavioral Neurology*. 1996;9(1):50-3.
63. Abraham E, Axelrod B, Ricker J. Application of the oral trail making test to a mixed clinical sample. *Archives of Clinical Neuropsychology*. 1996;11(8):697-701.
64. Whitehouse P, Caramazza A, Zurif E. Naming in aphasia: interacting effects of form and function. *Brain Lang*. 1978;6(1):63-74.
65. Borod JC, Goodglass H, Kaplan E. Normative data on the boston diagnostic aphasia examination, parietal lobe battery, and the boston naming Test. *Journal of Clinical Neuropsychology*. 1980;2(3):209-15.
66. Osterrieth PA. Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. *Archives de psychologie*. 1944;30:206-356.
67. Lindgren SD, Benton AL. Developmental patterns of visuospatial judgment. *J Pediatr Psychol*. 1980;5(2):217-25.
68. Schmidt M. *Rey auditory verbal learning test: a handbook*. Los Angeles: Western Psychological Services; 1996.
69. Malloy-Diniz LF, Lasmar VA, Gazinelli Lde S, Fuentes D, Salgado JV. The Rey Auditory-Verbal Learning Test: applicability for the Brazilian elderly population. *Rev Bras Psiquiatr*. 2007;29(4):324-9.
70. Emons WHM, Flore PC. Evaluating Scales for Ordinal Assessment in Clinical and Medical Psychology. 2013. In: *New Developments in Quantitative Psychology - Presentations from the 77th Annual Psychometric Society Meeting [Internet]*. New York, NY: SpringerSpringer Proceedings in Mathematics & Statistics.
71. Jöreskog KG. A general method for estimating a linear structural equation system. In: Goldberger AS, Duncan OD, editors. *Structural Equation Models in the Social Sciences*. New York, NY: Seminar Press; 1973.
72. Wang J, Wang X. *Structural Equation Modeling - applications using Mplus [E-Reader]*. Chichester, West Sussex, UK: John Wiley & Sons Ltd; Higher Education Press; 2012.

73. Bollen KA, editor. *Structural Equations with Latent Variables*. New York, NY: John Wiley & Sons, Inc; 1989.
74. Byrne BM. *Structural Equation Modeling with Eqs: Basic Concepts, Applications, and Programming (Multivariate Applications)(Multivariate Applications Series)*. 2006.
75. Comrey AL, Lee HB. *A First Course in Factor Analysis*. Hillsdale, NJ: Erlbaum; 1992.
76. Brown TA. *Confirmatory Factor Analysis for Applied Research*. New York, NY: The Guildford Press; 2006.
77. Bentler PM. Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika*. 1983(48):493-517.
78. Bentler PM. On tests and indices for evaluating structural models. *Personality and Individual Differences*. 2007(42):825-9.
79. Santorra A, Bentler PM, editors. *Scaling Corrections for Statistics in Covariance Structural Analysis*. Los Angeles, CA: UCLA; 1988.
80. Yuan KH, Bentler PM, editors. *Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data*. Washington, DC: The American Sociological Association; 2000.
81. Muthén B. Beyond SEM: General latent variable modeling. *Behaviormetrika*. 2002(29):81-117.
82. Asparouhov T, Muthén B. *Structural Equation Models And Mixture Models With Continuous Non-Normal Skewed Distributions*. Mplus Web Notes 2014.
83. Lubke GH, Muthén BO. Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*. 2004;11(4):514-34.
84. Marsh HW, Hocevar D. Application of confirmatory factor analysis to the study of self-concept: First-and higher-order factor models and their invariance across groups. *Psychological bulletin*. 1985;97(3):562-82.

85. French BF, Finch WH. Multigroup Confirmatory Factor Analysis: Locating the Invariant Referent Sets. *Structural Equation Modeling: A Multidisciplinary Journal*. 2008;15(1):96-113.
86. MH01259 NG, Lubke GH. Investigating population heterogeneity with factor mixture models. *Psychological methods*. 2005;10(1):21-39.
87. Schmitt N, Kuljanin G. Measurement invariance: Review of practice and implications. *Human Resource Management Review*. 2008;18(4):210-22.
88. Cheung GW, Rensvold RB. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*. 2002;9(2):233-55.
89. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. Third Edition ed. Dominici F, Faraway JJ, Tanner M, Zidek J, editors. Boca Raton, FL: Chapman & Hall / CRC Press; 2014. 640 p.
90. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. *The BUGS Book - A Practical Introduction to Bayesian Analysis*. Dominici F, Faraway JJ, Tanner M, Zidek J, editors. Boca Raton, FL: Taylor & Francis / CRC Press; 2012. 373 p.
91. Marin J-M, Robert CP. *Bayesian Essentials with R*. Second Edition ed. Casella G, DeVeaux R, Fienberg SE, Olkin I, editors. New York, NY: Springer Inc.; 2014.
92. Lesaffre E, Lawson AB. *Bayesian Biostatistics*. Senn S, Scott M, Jank W, editors. West Sussex, UK: John Wiley & Sons Ltd.; 2012.
93. Orlitzky M. How can significance tests be deinstitutionalized? *Organizational Research Methods*. 2012;15(2):199-228.
94. Lee MD, Wagenmakers E-J. *Bayesian Cognitive Modeling - A Practical Course*. New York, NY: Cambridge University Press; 2013.
95. Muthén B, Asparouhov T. Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological methods*. 2012;17(3):313.
96. Gelman A, Meng X-L, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*. 1996;6(4):733-60.

97. Gelman A. Parameterization and Bayesian modeling. *Journal of the American Statistical Association*. 2004;99(466).
98. Scheines R, Hoijtink H, Boomsma A. Bayesian estimation and testing of structural equation models. *Psychometrika*. 1999;64(1):37-52.
99. Royston JP. Some techniques for assessing multivariate normality based on the shapiro-wilk W. *Applied Statistics*. 1983;32(2):121-33.
100. R CTR. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
101. Sakia RM. The Box-Cox transformation technique: a review. *The Statistician*. 1992;41:169-78.
102. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society Series B (Methodological)*. 1991;53(3):683-90.
103. Miller LS, Rohling ML. A statistical interpretive method for neuropsychological test data. *Neuropsychol Rev*. 2001;11(3):143-69.
104. Wilcox RR, Keselman HJ. Modern robust data analysis methods: measures of central tendency. *Psychol Methods*. 2003;8(3):254-74.
105. Revelle W. *psych: Procedures for Psychological, Psychometric, and Personality Research* Evanston, Illinois: Northwestern University; 2014. Available from: <http://CRAN.R-project.org/package=psych>.
106. Ponterotto JG, Ruckdeschel DE. An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Percept Mot Skills*. 2007;105(3 Pt 1):997-1014.
107. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959(56):81-105.

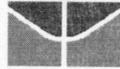
108. Cross CL. Statistical and Methodological Considerations when Using Cluster Analysis in Neuropsychologic Research. In: Allen DN, Goldstein G, editors. Cluster Analysis in Neuropsychological Research - recent applications. New York, NY: Springer; 2013.
109. Suhr DD. Exploratory or confirmatory factor analysis? Statistics and Data Analysis [Internet]. 2006 16/06/2013; 31:[17 p.]. Available from: <http://www2.sas.com/proceedings/sugi31/200-31.pdf>.
110. Tompson B. Exploratory and confirmatory factor analysis: understanding concepts and applications. Washington, DC: American Psychological Association; 2004.
111. Muthén LK, Muthén BO. Mplus User's Guide. Seventh ed. Los Angeles, CA: Muthén & Muthén; 1998-2012.
112. Urbina S. Essentials of psychological testing. Hoboken, NJ: Wiley; 2004.
113. Wainer H, Thissen D. True Score Theory: the traditional method. In: Thissen D, Wainer H, editors. Test Scoring. Mahwah, NJ: Lawrence Erlbaum Associates; 2001. p. 23-71.
114. Embretson S, Yang X. A Multicomponent Latent Trait Model for Diagnosis. Psychometrika. 2013;78(1):14-36.
115. Fletcher TD. Applied Psychometric Theory 2012 January, 13th 2013. Available from: <http://cran.r-project.org/web/packages/psychometric/psychometric.pdf>.
116. Masters G, Wright B. The Partial Credit Model. In: Linden W, Hambleton R, editors. Handbook of Modern Item Response Theory: Springer New York; 1997. p. 101-21.
117. Linacre J. Winsteps Rasch measurement computer program. 3.73 ed. Beaverton, Oregon 2011.
118. Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometric Monograph Supplement. 1969;34(17).
119. Cai L. A Two-Tier Full-Information Item Factor Analysis Model with Applications. Psychometrika. 2010;75(4):581-612.
120. Chalmers RP. mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software. 2012;48(6):1-29.

121. Reckase MD. Multidimensional Item Response Theory. Fienberg SE, van der Linden WJ, editors. New York, NY: Springer; 2009. 353 p.
122. Kline RB. Principles and practices of structural equation modeling. 3rd ed. New York, NY: The Guilford Press; 2011.
123. Jara A, Hanson T, Quintana F, Mueller P, Rosner G. DPpackage: Bayesian Semi- and Nonparametric Modeling in R. *Journal of Statistical Software*. 2011;40(5):1-30.
124. Szmukler G, Everitt B, Leese M. Risk assessment and receiver operating characteristic curves. *Psychological medicine*. 2012;42(5):895.
125. Tsai C-F, Lee W-J, Wang S-J, Shia B-C, Nasreddine Z, Fuh J-L. Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: validation of the Taiwanese version of the MoCA and an item response theory analysis. *International Psychogeriatrics*. 2012;24(04):651-8.
126. Cooper-Vince CE, Emmert-Aronson BO, Pincus DB, Comer JS. The Diagnostic Utility of Separation Anxiety Disorder Symptoms: An Item Response Theory Analysis. *Journal of abnormal child psychology*. 2013:1-12.
127. Birnbaum A. Some latent trait models and their use in inferring an examinee's ability. In: Lord FM, Novick MR, editors. *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley; 1968.
128. Cortina JM. What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*. 1993;78(1):98-104.
129. Steinmetz H, Schmidt P, Tina-Booh A, Wieczorek S, Schwartz SH. Testing measurement invariance using multigroup CFA: differences between educational groups in human values measurement. *Qual Quant*. 2008;43(4):599-616.
130. Sijtsma K. Theory Development as a Precursor for Test Validity. In: Millsap RE, Ark LA, Bolt DM, Woods CM, editors. *New Developments in Quantitative Psychology - Presentations from the 77th Annual Psychometric Society Meeting*. 66. New York, NY: Springer; 2013.

131. Zumbo BD. Validity: Foundational issues and statistical methodology. In: Rao CR, Sinharay S, editors. Handbook of statistics. Psychometrics. 26. Amsterdam: Elsevier; 2007. p. 45-79.
132. Messick S. Validity. In: Linn RL, editor. Educational measurement. 3rd ed. Washington, DC: American Council on Education and National Council on Measurement in Education; 1989. p. 13-103.
133. Michel J. Measurement in psychology. A critical history of a methodological concept. Cambridge: Cambridge University Press; 1999.
134. De Boeck P, Wilson M. Explanatory item response models: A generalized linear and nonlinear approach. New York: Springer; 2004.
135. Rupp AA, Templin JL, Henson RA. Diagnostic measurement: Theory, methods, and applications. New York: Guilford Press; 2010.

7 ANEXOS

Anexo A - Aprovação do Comitê de Ética em Pesquisa em Humanos CEP-UnB



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE MEDICINA
Comitê de Ética em Pesquisa em Seres Humanos

ANÁLISE DE PROJETO DE PESQUISA

Registro de Projeto: CEP-FM 034/2009.

Título: "Validação da prova de avaliação cognitiva breve de Filadélfia (PBAC) em uma amostra populacional do Distrito Federal."

Pesquisador Responsável: Carlos Alberto Bezerra Tomaz.

Documentos analisados: Folha de rosto, carta de encaminhamento, declaração de responsabilidade, protocolo de pesquisa, termo de consentimento livre e esclarecido, cronograma, bibliografia pertinente e currículo (s) de pesquisador (es).

Data de entrega: 13/05/2009.

Proposição do (a) relato (a)

Aprovação

Não aprovação.

Data da primeira análise pelo CEP-FM/UNB: 24/06/2009.

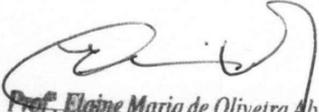
Data do parecer final do projeto pelo CEP-FM/UNB: 30/09/2009.

PARECER

Com base na Resolução CNS/MS nº 196/96 e resoluções posteriores, que regulamentam a matéria, o Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de Brasília decidiu **APROVAR**, conforme parecer do (a) relator (a) o projeto de pesquisa acima especificado, quanto aos seus aspectos éticos.

1. Modificações no protocolo devem ser submetidas ao CEP, assim como a notificação imediata de eventos adversos graves;
2. O (s) pesquisador (es) deve (m) apresentar relatórios periódicos do andamento da pesquisa ao CEP-FM.

Brasília, 01 de Outubro de 2009.


Prof. Elaine Maria de Oliveira Alves
Coordenadora do Comitê de Ética em Pesquisa
Faculdade de Medicina-UnB

Anexo B - Termo de Consentimento Livre e Esclarecido

Validação psicométrica da Escala de Avaliação Cognitiva Breve da Filadélfia (PBAC)

Você foi convidado a participar voluntariamente como sujeito em uma pesquisa sobre rastreamento de alterações das funções cognitivas conduzida pela equipe do Laboratório de Neurociências da Universidade de Brasília. O objetivo do trabalho é validar o instrumento de rastreamento cognitivo (PBAC), de forma a obter na nossa população tabelas normativas.

Esta participação é voluntária, o que significa que não haverá qualquer tipo de benefício financeiro, em troca de sua participação. Você poderá desistir de participar desta pesquisa, em qualquer momento que desejar, sem precisar justificar sobre os seus motivos aos pesquisadores. Da mesma forma, você poderá recusar-se a responder qualquer pergunta ou questionário, caso se sinta desconfortável com alguma pergunta.

Durante a avaliação, você será solicitado a desempenhar várias tarefas experimentais incluindo palavras, figuras ou sentenças ou realizar cópias de desenhos, ou símbolos. O estudo levará cerca de 25 minutos.

Não haverá desconfortos ou riscos envolvidos ao participar neste estudo. Durante as sessões, não será pedido para fazer qualquer coisa que o deixe constrangido.

Você tem o direito de perguntar o que desejar ao experimentador e de recusar-se a responder qualquer pergunta feita neste estudo. Perguntas que possam influenciar de certa forma o estudo serão respondidas somente ao final da sessão.

Suas respostas nesta pesquisa são confidenciais, ou seja, sigilosas. As informações pessoais sobre o que você responderá neste estudo serão mantidas somente pelo investigador e seus colaboradores e ninguém mais.

Se o resultado desta pesquisa possibilitar uma publicação ou outro uso educacional, como estudos de caso, nenhum participante será identificado em nenhum destes tipos de materiais publicados ou não, e informações que possibilitariam identificações serão substancialmente modificadas.

Não haverá nenhum custo para você nesta participação, a não ser o tempo envolvido durante os procedimentos.

Por favor, assine no local indicado abaixo indicando que você leu ou que tenham lido para você e está ciente deste "Termo de Consentimento". A natureza, riscos e benefícios do projeto foram explicados para você, e você concorda em participar no estudo como descrito.

Este termo de consentimento livre e esclarecido foi elaborado em duas vias, uma das vias sendo entregue ao paciente ou ao seu representante legal.

Eu, _____
 declaro concordar em participar dessa pesquisa de livre e espontânea vontade.

_____, ____/____/____

Anexo C - Avaliação Cognitiva Breve de Filadélfia

PACIENTE: _____ Escolaridade: _____

AVALIADOR: _____ DATA: ___ / ___ / ___

DIAGNÓSTICO: _____ tempo de diagn: _____

(1) **Aprendizagem Verbal** (1 até t3): Eu vou ler uma lista de palavras. Por favor, ouça com atenção e, quando eu acabar, repita na ordem que o Sr quiser. (aqui ele repetia isso para cada tentativa “agora eu vou ler novamente...”, “uma última vez”, a gente já dizia no começo que seria 3 vezes e depois seguia como eles.)

Palavra	Tentativa 1	Tentativa 2	Tentativa 3
Sapato	_____	_____	_____
Livro	_____	_____	_____
Cadeira	_____	_____	_____
Maçã	_____	_____	_____
Xícara	_____	_____	_____
Lago	_____	_____	_____

Pontuação (Máx. 6) =
Avalie apenas a 3ª tentativa. Atribuir 1 pt para cada palavra evocada corretamente.

(2) **Fluência de Letras** (letra “F”): Por favor, diga todas as palavras que o Sr conseguir lembrar que comecem com a letra “F”. Mas por favor não diga numeros (a gente não dizia esse, já que não tem número que comece com F), ou nomes de pessoa, como Francisca, ou nome de lugares como Fortaleza.

15s	30s	45s	60s

Pontuação (Max. 6) = _____ (0=0; 1-3 = 1; 4-6 = 2; 7-9 = 3; 10-12 = 4; 13-15 = 5; >16 = 6)

(3) **Aprendizagem Verbal (Evocação)**

A um tempinho atrás eu li uma lista de palavras para o Sr. Por favor tente lembrar quantas palavras o Sr conseguir. O Sr pode falar em qualquer ordem.

Evocar a lista de palavras de (1)

Atribuir 1 pt para palavra evocada corretamente.
Pontuação (Máx. 6) =

(4) **Aprendizagem Verbal (Reconhecimento)**

(corretas e distratoras=1/2 pt)

Eu vou ler uma palavra para você. Se essa palavra estava na lista que eu li para o Sra um tempinho atrás, por favor diga SIM. Se não estava na lista que eu li a um tempinho atrás, diga NÃO.

1 sapato _____	3 chapéu _____
4 livro _____	2 papel _____
6 cadeira _____	5 sofá _____
7 maçã _____	8 manga _____
9 xícara _____	11 prato _____
12 lago _____	10 montanha _____

Pontuação =
Atribuir ½ pt para o seguinte: SIM para a palavra correta + NÃO para a palavra enganadora. Senão, pontuar um 0.

(5) **Nomeação** (1pt)

Mostre cada figura para o paciente e peça para ele nomear, dizendo: por favor me diga o nome dessa figura
1.martelo 2.caju 3.macaco 4.alicate 5.abacaxi 6.papagaio/arara
Pontuação (Max. 6) =

(6) **Questão Semântica de (5)**

Peça ao paciente para agrupar os desenhos de (5), dizendo: “por favor aponte duas figuras que têm algo em comum”

animais _____
frutas _____
ferramentas _____

Pontuação (3pts):

(7) **Figura de Rey (Cópia)**: Peça ao paciente para copiar a figura no espaço abaixo

1 pt atribuído para cada um dos 12 itens representados na figura (ver folha de pontuação)

Retângulo grande ____ Cruz esquerda ____ Triângulo superior ____ Diagonais internas ____ Círculo ____ Linha triangular superior ____ Linha vertical interna ____ 3 linhas ____ Triângulo direito ____ Linha horizontal direita ____ Diamante ____ Linha horizontal interna ____ Linha triangular superior ____

(8) **Discernimento de Orientação de Linha**

Peça ao paciente para combinar cada uma das 6 linhas oblíquas abaixo da linha horizontal com as linhas acima da linha horizontal. (A gente fazia assim: “Com qual dessas linhas aqui de baixo essa aqui se parece? – e apontava para cada uma das linhas de cima”). 1 pt atribuído para cada combinação correta.

(9) **Figura de Rey (Evocação)** : Peça ao sujeito para desenhar a figura que foi mostrada a ele um tempinho atrás. Que o paciente a desenhe abaixo dos estímulos do Discernimento de Orientação de Linha. (“Agora eu quero que o Sr desenhe de cabeça aquela figura que acabou de copiar”) 1 pt atribuído para cada um dos 12 itens representados na figura.

(10) **Leitura**

Peça ao paciente para ler em voz alta cada uma das 4 palavras irregulares. 1 pt para cada resposta correta

TÁXI	CASA	CAMA	BRUXA
------	------	------	-------

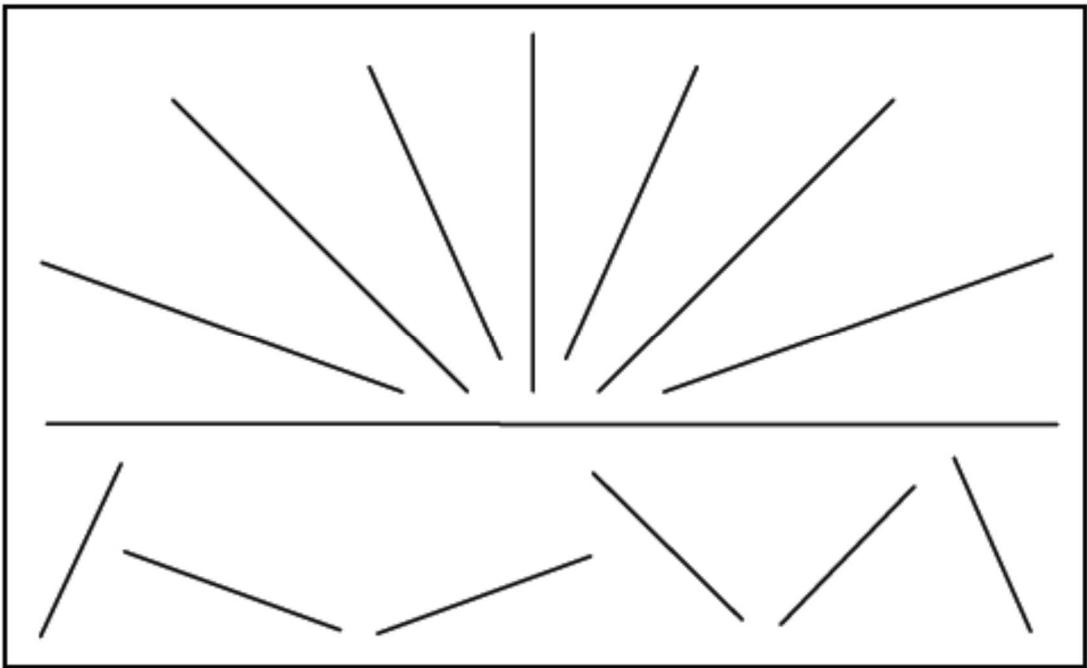
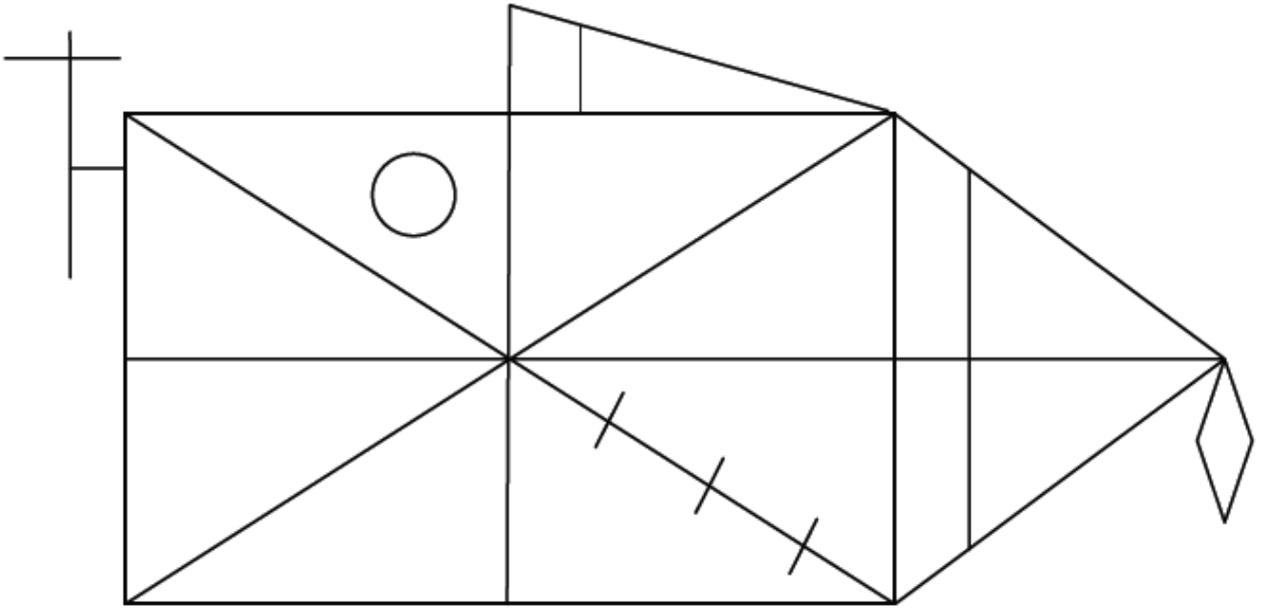
(11) **Escrita**: pontuar a frase do MEEM

1 pt atribuído para conteúdo, 1 pt atribuído para gramática, 1 pt atribuído para pronúncia

(12) **Teste de Trilhas Oral** (6pt) - limite de 60s. 1-a, 2-b, etc.

O Sr sabe contar e o Sr conhece o alfabeto. Eu quero que o Sr alterne entre um número e uma letra. Por exemplo, por favor diga 1-a, 2-b, 3-c e assim por diante.

PONTUAÇÃO DE TEMPO Até o “z” = 3 pts Até o “r” = 2 pts Até o “l” = 1 pt < “l” = 0 pts	PONTUAÇÃO DE ERRO O paciente recebe 3 pts no começo da tarefa, mas perde 1 pt para cada erro, até um máximo de 3 perdas.
--	--



Anexo D - Mini Exame do Estado Mental

		Escore
Orientação Temporal (5 pontos) Dê um ponto para cada item	Ano	
	Mês	
	Dia do mês	
	Dia da semana	
	Hora aproximada (variação de 1h aproximadamente)	
Orientação Espacial (5 pontos) Dê um ponto para cada item	Estado	
	Cidade	
	Bairro ou nome da rua próxima	
	Local geral: que local é este aqui (apontando ao redor num sentido mais amplo: hospital, casa de repouso, casa própria)	
Andar ou local específico: em que local nós estamos (consultório, dormitório, sala, apontando para o chão)		
Registro (3 pontos)	Carro	
	vaso	
	tijolo	
Cálculo (5 pontos)	Subtrair: $100 - 7 = 93$	
	$93 - 7 = 86$	
	$86 - 7 = 79$	
	$79 - 7 = 72$	
	$72 - 7 = 65$	
Evocação das palavras (3 pontos)	Quais os três objetos que você aprendeu anteriormente?	
	carro	
	vaso	
Nomeação (2 pontos)	tijolo	
	Relógio	
Repetição (1 ponto)	Caneta	
	“NEM AQUI, NEM ALI, NEM LÁ”	
Comando (3 pontos) Dê 1 ponto para cada ação correta	“Apanhe esta folha de papel com a mão direita, dobre-a ao meio e coloque-a no chão”	
Leitura (1 ponto)	FECHE SEUS OLHOS	
Frase (1 ponto)	“Escreva alguma frase que tenha começo, meio e fim”	
Copia do desenho (1 ponto)	Copiar dois pentágonos com interseção	
		
pontos		Escore Total = _____ / 30

Anexo E - PBAC original (em inglês)

Philadelphia Brief Assessment of Cognition v3.0brief Nov 1, 2009

Name: _____ Educ: _____ Age: _____ Dx: _____ Date: _____ MMSE _____/30

VERBAL LEARNING (1pt for t3)

Word	Trial 1	Trial 2	Trial 3
shoe	_____	_____	_____
book	_____	_____	_____
chair	_____	_____	_____
apple	_____	_____	_____
cup	_____	_____	_____
lake	_____	_____	_____

LETTER FLUENCY (letter 'F') (0=0; 1-3 = 1; 4-6 = 2; 7-9 = 3; 10-12 = 4; 13-15 = 5; >16 = 6)

15s 30s 45s 60s

RECALL (1pt)

RECOG (correct target and foil=1/2pt)

1 shoe _____ 3 hat _____
4 book _____ 2 paper _____
6 chair _____ 5 couch _____
7 apple _____ 8 peach _____
9 cup _____ 11 plate _____
12 lake _____ 10 mountain _____

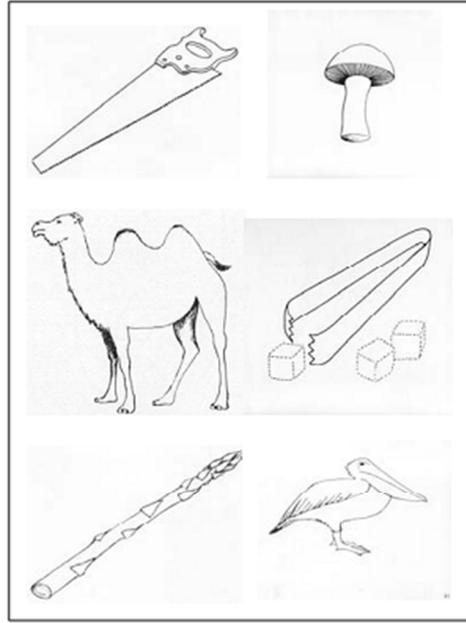
NAMING (1pt)

1.) saw _____
2.) mushroom _____
3.) camel _____
4.) tongs _____
5.) asparagus _____
6.) pelican _____

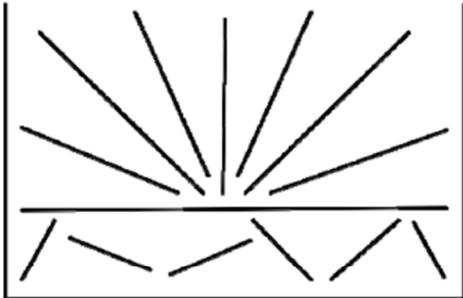
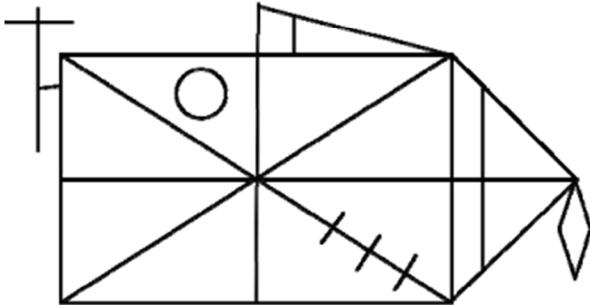
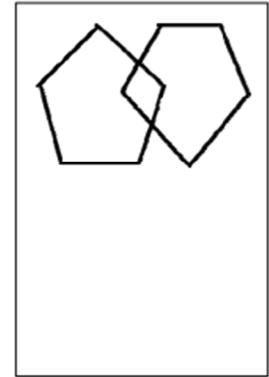
SPEECH (4pt total; -1/2pt)

word-finding pauses _____
circumlocutory _____
paraphasic-sem _____
paraphasic-phon _____
agrammatic _____
effortful _____
dysarthria _____
discourse _____

SEMANTIC (3pt)



PLACE
TIME
REGIS
WORLD/DLROW
RECALL
NAMING pen, watch
REPET no ifs, ands, or buts
COMP right hand, fork, lap
READ
WRITE
DRAW



SENTENCE ABOUT WEATHER Content=1, gram=1, spell=1

ORAL TRAILS (45 sec=6pt): -1 complete @ 60 sec; -1 to R; -1 to L; -1 each error up to 3

BEHAVIOR/PERSONALITY

Apathy/Poor Initiative

4 not observed
3 mild
2 moderate
1 severe

Disinhibition

not observed
mild
moderate
severe

Social Compartment

not observed
mild
moderate
severe

Agitation/irritability

not observed
mild
moderate
severe

Ritual/OCD

not observed
mild
moderate
severe

Empathy/Insight

not observed
mild
moderate
severe

REY FIG COPY__ RECALL__ Lg rect__ Lt cross__ Upper tri__ In diags__ Circle__ Up tri line__ In vert line__ 3 lines__ Rt tri__ In horiz line__ Diamond__ Rt tri line__

CHOIR
YACHT
PINT
COUGH

ORAL READ (1 pt)

PBAC Administration and Scoring Instructions.
(Nov 1, 2009)

Note: The score derived from the MEEM is not part of the total PBAC score

1.) Verbal List Learning

a.) Verbal List Learning Trial 1 – I am going to read a list of words. Listen carefully, when I am done please say them back in any order you wish. Please record the order of recall and intrusions responses.

b.) Verbal List Learning Trial 2 – Now I will read the list again, just as before, when I am done say back as many words as you can. Please record the order of recall and intrusions responses.

b.) Verbal List Learning Trial 3 – One last time, listen careful and after I read the words please say back as many as you can in any order you wish.

Scoring – Record verbatim the order of all responses including intrusion errors. For trial 3 award 1 point for each correct response, range 0 - 6

2.) Letter Fluency (letter 'F') – Please give me all of the words you can think of that begin with the letter “F”. But please do not give numbers such as four or five or the names of people such as Bob or Barbara or places such as Boston or Burlington. Also, please give only one form of a word. For example, if you were to say “buy”, please do not say “buying”.

Scoring – Record verbatim in 15s epochs. The scoring key is as follows: 0= 0; 1-3= 1; 4-6= 2; 7-9= 3; 10-12= 4; 13-15= 5; >16= 6, range 0 - 6.

3.) Delay Verbal Free Recall – A little while ago I read a list of words to you. Please try to remember as many of these words as possible. You can say them in any order.

Scoring – Record verbatim all responses. Score 1 point for each correct response, range 0 - 6

4.) Delay Verbal Recognition – I am going to read a word to you. If the word was on the list I read to you a little while ago please say yes. If it was not on the list I read a little while ago please say no.

Scoring – Read each word in the order indicated. Score ½ point for the correct identification of all targets along with its respective foil, range 0 - 3.

5.) Visual Confrontational Naming – Show each picture to the patient and ask for the name saying, please tell me the name of this picture.

Scoring – Score 1 point for each correct response, range 0 - 6.

6.) Semantic Knowledge – Using the six pictures from the Boston naming Test ask the patient to “please point to two pictures that are alike”. Record patient’s response verbatim.

Scoring – One point is awarded for each object pair

7.) Rey Copy – In the space provided below ask the patient to copy the Rey Complex Figure

Scoring – Award 1 point for each of the 12 figure elements, range 0 - 12.

8.) Line Orientation – Ask the patient to match each of the 6 oblique lines below the horizontal line to the lines above the horizontal line.

Scoring – Award 1 point for each correct answer, range 0 - 6.

9.) Delay Rey Recall – In the space provide ask the patient to drawing the figure they were shown a little while ago.

Scoring – Award 1 point for each of the 12 figure elements, range 0 - 12.

10.) Word Reading – Ask the patient to read each of the following four words: choir, yacht, pint, cough.

Scoring – Score 1 point for each correct response, range 0 - 4

11.) Sentence Writing – Ask the patient to write a sentence about the weather. This is obtained from the MEEM.

Scoring – One point for grammar, one point for spelling, one point for content; range 0 - 3

12.) Conversational Speech – Deduct ½ point for any of the problems or errors listed: word finding pauses; circumlocutory speech; semantic paraphasic errors; literal paraphasic errors; agrammatical speech; effortful speech, discourse; range 0 - 4.

Scoring – score ranges from 0 - 4.

13.) Behavior/ Personality – On the basis of your interaction with the patient and your interview with the family rating each of the following six behavioral characteristics: apathy/ poor initiative; disinhibition; social comportment; agitation/ irritability; ritualistic OCD behavior; capacity of empathy.

Scoring – Score each six behaviors using a 4-point score, range 0 - 24.

14.) Oral Trails – You know how to count and you know the alphabet. I want you to alternate between a letter and number. For example, please say a-1-b-2-c-3 and so on. Start timing when patient starts.

Scoring - Continue up to 60s, mark the 45 sec time-point.

Oral trails scoring: 6 points for accurate performance completed by 45s; 1 pt for accurate completion up to 60s; another 1 point for completion accurate to letter 'R' by 60s; another 1 point for accurate completion to letter 'I' by 60s; 1 pt for each error

15.) MEEM – The MEEM can be interspersed throughout the PBAC administration. Please give the MEEM registration, recall, and figure copy test items toward the end of the PBAC so as to not cause any interference with other PBAC test items.

executive index	ordinal letter F, oral trails	range 0 – 12
visuospatial index	copy, line orientation	range 0 – 18
language index	naming, word reading, conversational speech, word reading, semantic knowledge	range 0 – 20
memory	immediate free recall trial 3, delay free recall, delay recognition, rey recall	range 0 – 27
behavior	total behavior index	range 0 – 24
<hr/>		
total PBAC score		range 0 – 101
<hr/>		

Mean (standard deviation) total PBAC scores (range 0 – 96; digit span and MEEM not included)

AD	60.22 (10.22)
SOC	71.33 (12.53)
PNFA	63.89 (13.83)
SemD	65.82 (10.73)
CBD	55.38 (14.20)
DLD	50.05 (17.47)
<hr/>	
NC	86.15 (5.19)
<hr/>	