

Avaliação Psicológica

Análise de estrutura interna do Inventário de Depressão Maior (MDI)

 Makilim N. Baptista¹
<https://orcid.org/0000-0001-6519-254X>

 Gabriela S. Cremasco¹
<https://orcid.org/0000-0003-2075-8049>

 Felipe Augusto Cunha¹
<https://orcid.org/0000-0001-7977-9712>

 Samira R. Marcon²
<https://orcid.org/0000-0002-5191-3331>

 Hugo G. B. dos Santos²
<https://orcid.org/0000-0001-7713-1314>

Para citar este artigo: Baptista, M. N., Cremasco, G. S., Cunha, F. A., Marcon, S. R., & Santos, H. G. B. (2020). Análise de estrutura interna do Inventário de Depressão Maior (MDI). *Psicologia: Teoria e Prática*, 22(1), 22–40. doi:10.5935/1980–6906/psicologia.v22n1p22–40

Submissão: 08/02/2019

Aceite: 16/07/2019



Todo o conteúdo de *Psicologia: Teoria e Prática* está licenciado sob Licença Creative Commons CC – By 4.0

1 Universidade São Francisco (USF), Itatiba, SP, Brasil.

2 Universidade Federal do Mato Grosso (UFMT), Cuiabá, Brasil.

Resumo

O objetivo do estudo foi avaliar as qualidades psicométricas do Inventário de Depressão Maior (MDI). O MDI foi aplicado em 714 estudantes universitários e em 12 pessoas com diagnóstico de depressão. Realizaram-se análises de estrutura interna via TCT e TRI, bem como confiabilidade, sensibilidade e especificidade. A solução unidimensional foi a mais coerente teoricamente e os parâmetros via TCT adequados, com confiabilidade de 0,89. Já, via TRI, nem todos os parâmetros adequados foram respeitados, tais como índices de *infit* e *outfit*, além de categorias de respostas não discriminativas. Os índices de acurácia também demonstraram limitações, principalmente com relação aos falsos negativos (especificidade de 63,9). Assim, sugere-se a realização de novos estudos analisando-se os dados via TRI, bem como um grupo clínico maior, diagnosticados com entrevista estruturada padrão.

Palavras-chave: sintomatologia depressiva; estudantes universitários; propriedades psicométricas; TRI; depressão.

INTERNAL STRUCTURE ANALYSIS OF THE MAJOR DEPRESSION INVENTORY (MDI)

Abstract

The objective of the study was to evaluate the psychometric qualities of the Major Depression Inventory (MDI). The MDI was applied to 714 university students and 12 people diagnosed with depression. Internal structure analyses were performed using TCT and IRT, as well as reliability, sensitivity, and specificity. The one-dimensional solution was the most coherent theoretically and the parameters via suitable TCT, with a reliability of 0.89. Already, via IRT, not all appropriate parameters were respected, such as *infit* and *outfit* indexes, as well as categories of non-discriminative responses. Accuracy indexes also showed limitations, especially concerning false negatives (specificity of 63.9). Thus, it is suggested to perform new studies analyzing the data via IRT, as well as a larger clinical group, diagnosed with a standard structured interview.

Keywords: depressive symptomatology; college students; psychometric properties; IRT; depression.

ANÁLISIS DE ESTRUCTURA INTERNA DEL INVENTARIO DE DEPRESIÓN MAYOR (MDI)

Resumen

El objetivo del estudio fue evaluar las calidades psicométricas del Inventario de Depresión Mayor (MDI). El MDI fue aplicado en 714 estudiantes universitarios y en 12 personas con diagnóstico de depresión. Se realizaron análisis de estructura interna vía TCT y TRI, así como confiabilidad, sensibilidad y especificidad. La solución unidimensional fue la más coherente teóricamente y los parámetros vía TCT adecuados, con confiabilidad de 0,89. En cambio, a través de TRI, no se respetaron todos los parámetros adecuados, tales como índices de *infit* y *outfit*, además de categorías de respuestas no discriminatorias. Los índices de exactitud también demostraron limitaciones, principalmente con relación a los falsos negativos (especificidad de 63,9). Así, se sugiere la realización de nuevos estudios analizando los datos vía TRI, así como un grupo clínico mayor, diagnosticados con entrevista estructurada estándar.

Palabras clave: sintomatología depresiva; estudiantes universitarios; propiedades psicométricas; TRI; depresión.

1. Introdução

A depressão é um transtorno mental incapacitante, e estima-se que atualmente mais de 300 milhões de pessoas no mundo tenham a doença. É altamente relacionada ao suicídio que contabiliza aproximadamente 800 mil mortes por ano (World Health Organization, 2017). Identificar os transtornos mentais precocemente e de forma adequada possibilita a tomada de decisões e o planejamento de intervenções eficazes, sendo os testes psicológicos uma ferramenta importante para o processo de avaliação psicológica (Baptista, 2018).

Na literatura internacional, existem mais de 280 medidas diferentes em avaliação de depressão, as quais podem variar em termos de concepção teórica, descritores, existência de itens com semântica positiva e negativa, tipos de respostas, qualidades psicométricas, pontos de corte e objetivos (Santor, Gregus, & Welch, 2006). No Brasil, são poucas as medidas utilizadas em pesquisas. Por exemplo, uma revisão integrativa da literatura científica brasileira a respeito de instrumentos de avaliação da depressão foi realizada nas bases de dados IndexPsi e BVS Psi (SciELO e PePSIC) entre 2005 e outubro de 2015, sendo identificadas escalas nacionais e estrangeiras. Foram encontrados apenas 24 instrumentos, dos quais 12 avaliam unicamente depressão e cinco são específicos para a população de crianças/adoles-

centes. As escalas mais citadas foram o Beck Depression Inventory (BDI), a Hospital Anxiety and Depression Scale (HADS), a Hamilton Depression Rating Scale for Depression (HAM-D), o Children Depression Inventory (CDI) e a Escala Baptista de Depressão Adulto (EBADEP-A) (Baptista & Borges, 2016).

Muitos testes de depressão parecem ser relevantes e mais estudados nos contextos internacionais, como o Inventário de Depressão Maior (Major Depression Inventory – MDI), que foi traduzido para vários países, como Turquia, Suécia, Alemanha, Inglaterra, Holanda, Dinamarca e Finlândia, além de ter sido desenvolvido em colaboração com a Organização Mundial de Saúde (OMS) (Bech, Rasmussen, Olsen, Noerholm, & Abildgaard, 2001). O MDI foi desenvolvido com base nos sintomas de depressão descritos no *Manual diagnóstico e estatístico de transtornos mentais* (DSM-IV) e na Classificação Internacional de Doenças (CID-10). Trata-se de um instrumento de autorrelato que contém nove descritores e mede a gravidade da depressão em leve, moderada e severa (Bech, Timmerby, Martiny, Lunde, & Soendergaard, 2015; Bech & Wermuch, 1998).

O MDI é considerado uma escala de rastreamento e aplicação rápida, objetivando levantar os principais sintomas da depressão (Behera et al., 2017). A escala talvez seja pouco divulgada em publicações internacionais por ter sido desenvolvida na Dinamarca e, comparada com outros instrumentos (por exemplo, Beck Depression Inventory ou Hamilton Depression Rating Scale), construídos na década de 1960, é ainda uma medida relativamente nova, já que foi desenvolvida em 2001.

Realizaram-se alguns estudos com o objetivo de verificar as propriedades psicométricas do MDI como a pesquisa de Bech et al. (2001), que teve o intuito de averiguar a sensibilidade e a especificidade do instrumento. Selecionaram-se 43 indivíduos de um hospital psiquiátrico, dos quais 21 apresentavam diagnóstico de depressão de moderada a grave, quatro foram diagnosticados com depressão leve e os demais tinham outros transtornos psiquiátricos, como ansiedade, esquizofrenia e alcoolismo. A amostra foi composta de 34 mulheres com média de idade de 48,9 anos. A sensibilidade e especificidade do instrumento a partir dos critérios do DSM-IV foram de 0,90 e 0,82 e a partir da CID-10 de 0,86 e 0,86, respectivamente, com ponto de corte 26 (pontuação de 1 a 6 na escala do tipo Likert). A consistência interna a partir do alfa de Cronbach foi de 0,94. Os autores concluíram que o instrumento consegue rastrear e mensurar os níveis da depressão adequadamente a partir dos critérios diagnósticos dos manuais.

O estudo de Fountoulakis et al. (2003) teve uma amostra composta por 30 sujeitos diagnosticados com depressão e 68 no grupo controle. Os resultados indicaram que, no ponto de corte 26/27, a sensibilidade apresentou valor de 0,86 e a especificidade 0,94, e a confiabilidade pelo alfa de Cronbach foi de 0,89. A Análise dos Componentes Principais, com rotação Varimax, indicou dois fatores para o instrumento, sendo o primeiro constituído por todos os itens, com exceção do 10b, explicativos de 54% da variância. Já o segundo fator incluiu apenas o item 10b com 9% da variância. Os autores concluíram que a solução unifatorial é a mais adequada.

O MDI foi avaliado também por Cuijpers, Dekker, Noteboom, Smits e Peen (2007). Participaram desse estudo 258 pessoas, sendo 142 (55%) do sexo feminino, com média de idade de 36,45 anos com diagnósticos diversos, como distímia, transtorno bipolar e ansiedade. Além do MDI, aplicou-se a subescala de depressão do Symptom Checklist (SCL-90), e verificou-se uma correlação positiva e forte ($r = 0,79$; $p < 0,01$) entre os instrumentos. O alfa de Cronbach foi de 0,89 e a curva ROC indicou para sensibilidade de 0,66 e especificidade de 0,63 com 26 de ponto de corte (pontuação de 1 a 6). É importante citar que nesse estudo os próprios autores reconhecem as baixas taxas de acurácia diagnóstica, já que não se utilizou uma entrevista estruturada padronizada para os diagnósticos. Além disso, não foi possível calcular a confiabilidade entre psiquiatras no diagnóstico (*inter rater reliability*), o que provavelmente pode ter subestimado o número de participantes com depressão maior e a alta prevalência de comorbidades entre o antigo eixo I e eixo II do DSM nos pacientes.

Com o propósito de adaptar e verificar as propriedades psicométricas do MDI para a população brasileira, Parcias et al. (2011) realizaram um estudo com uma amostra de 120 pessoas. Desses participantes, 30 estavam diagnosticados com depressão, sendo 21 (70%) do sexo feminino com idade média de 50,23 anos e 90 sujeitos compuseram o grupo controle, com 63 (70%) do sexo feminino e idade média de 50,16 anos. O índice de Kaiser-Meyer-Olkin (KMO) foi de 0,50, sendo aplicado o método dos componentes principais para extrair os agrupamentos de itens e rotação Varimax. Indicaram-se dois componentes: o primeiro explicou 53,9% da variância e incluiu todos os itens com exceção do 10b, e o segundo explicou 13,6% da variância e incluiu os itens 4, 5 e 10b, e ambos justificaram 67,6% da variabilidade. A consistência interna por meio do alfa de Cronbach foi de 0,91. Os valores de sensibilidade e especificidade na curva ROC considerando a amostra

total foram de 0,86 e 0,75, respectivamente, para o ponto de corte 16 (pontuação de 0 a 5). A solução com dois componentes não pareceu adequada do ponto de vista dos descritores, já que o item 5 (culpa) teve cargas fatoriais similares nos dois agrupamentos, e o item de apetite aumentado (10b) apresentou maior carga no segundo componente.

A fim de verificar as propriedades psicométricas do instrumento em uma segunda amostra brasileira, o objetivo do presente estudo foi buscar evidências de validade baseada na estrutura interna para o MDI. Assim, foi conduzida uma análise fatorial exploratória (AFE), além de análises pela Teoria de Resposta ao Item (TRI), uma vez que a maior parte dos estudos tem sido conduzida a partir da Teoria Clássica dos Testes (TCT). Verificaram-se ainda os índices de sensibilidade e especificidade da escala.

2. Método

2.1 Participantes

Coletaram-se dados de 714 estudantes, dos 8.018 matriculados na Universidade Federal de Mato Grosso (UFMT), *campus* de Cuiabá. A seleção da amostra foi realizada por meio do método de amostragem probabilística por conglomerados (turma) e estratificada (grandes áreas). A seleção das salas em cada estrato se deu com probabilidade proporcional ao número de alunos matriculados. O tamanho da amostra foi determinado pela seguinte expressão: $n = N (p(1-p)) / ((N-1) (d^2/z^2) + (p(1-p)))$, em que N é o tamanho da população; d, o erro de amostragem (3,5%); z, a confiança estipulada (95%); e p, a proporção considerada (50%). Assim, a amostra mínima estipulada foi de 714 estudantes.

Para a definição do número de turmas a serem sorteadas em cada estrato, considerou-se o tamanho médio delas nos respectivos estratos, e, para a definição do número de alunos por estratos, realizou-se a multiplicação da fração de alunos matriculados pelo tamanho da amostra. Com o propósito de determinar o número de turmas que deveriam ser sorteadas, dividiu-se o número de alunos por estrato pelo número médio de turmas, o que totalizou 33 turmas.

Definido o número de turmas, foi considerado um acréscimo de, aproximadamente, 40% nelas, gerando um total de 46 turmas. O ajuste foi necessário, visto que o número de alunos por turmas não correspondeu aos dados disponibilizados

pelo Serviço de Tecnologia da Informação. Desse modo, o ajuste ocorreu com o intuito de manter o tamanho da amostra prevista para o estudo (714 sujeitos).

Dado que no método de seleção da amostra foi considerada uma estratificação, esta ocorreu a partir das quatro grandes áreas de ensino, formando quatro estratos com seus respectivos cursos: Centro de Letras e Ciências Humanas, com 17 cursos; Centro de Ciências Sociais, com 23 cursos; Centro de Ciências Exatas e Tecnológicas, com 18 cursos; e Centro de Ciências Biológicas e da Saúde, com oito cursos. Para o sorteio da amostra, os cursos e suas respectivas turmas foram listados e numerados por estratos. O sorteio foi realizado pelo programa Minitab, em que o primeiro número aleatório sorteado constituiu a primeira turma da listagem numerada, e assim por diante com as demais turmas. Após a coleta dos dados, 77 protocolos foram anulados por estarem incompletos, restando 637 questionários, o que corresponde a 90% da amostra estipulada inicialmente. A maior parte da amostra foi composta por sujeitos com idade entre 18 e 24 anos (70,0%), com 339 (53,2%) do sexo feminino. Já o grupo clínico foi composto por 12 pessoas com diagnóstico de episódio depressivo maior por um psiquiatra, sendo 11 do sexo feminino, com idade variando entre 19 e 61 anos ($M = 42,08$; $DP = 16,46$), provenientes de um serviço de neuropsicologia de uma instituição particular de uma cidade do Nordeste, e todos com sintomatologia de moderada a severa no BDI.

2.2 Instrumento

O MDI (Bech & Wermuch, 1998) foi desenvolvido de modo a abarcar os sintomas da depressão do DSM-IV e da CID-10. É composto por dez itens, e os itens 8 e 10 contêm duas opções (A e B), em que se considera o subitem com maior pontuação, totalizando 12 possibilidades de respostas e dez descritores. O inventário é apresentado em uma escala do tipo Likert de 6 pontos: 0 = nenhuma vez, 1 = algumas vezes, 2 = um pouquinho menos que a metade do tempo, 3 = um pouquinho mais do que a metade do tempo, 4 = maior parte do tempo e 5 = o tempo todo com pontuação mínima de zero e máxima de 50 pontos.

2.3 Procedimentos

O projeto foi submetido ao Comitê de Ética em Pesquisa da UFMT e aprovado sob o nº 1.021.217. Após a aprovação do Certificado de Apresentação para Apreciação Ética (CAAE), os participantes responderam ao instrumento mediante assi-

natura do Termo de Consentimento Livre e Esclarecido (TCLE), impresso em duas vias, contendo todas informações pertinentes ao estudo. As aplicações ocorreram de forma coletiva e em sala de aula com os universitários e de forma individual nas instituições de saúde com os participantes do grupo clínico, com duração aproximada de 10 e 20 minutos, respectivamente.

2.4 Análise dos dados

Os dados provenientes do estudo foram analisados com o programa Factor, versão 10.3.01 (Lorenzo-Seva & Ferrando, 2006), com o intuito de realizar a análise de dimensionalidade e a AFE. O método de estimação usado na AFE foi o Unweighted Least Squares (ULS), a partir de uma matriz de correlações policóricas, com método de retenção Minimum Average Partial (MAP). Como critério de inclusão, os itens teriam que apresentar carga fatorial mínima de 0,30. Os dados foram analisados também pela TRI, por meio do modelo Rating Scale, a partir do programa Winsteps 3.74. O ajuste dos itens foi avaliado segundo os índices de *infit* e *outfit*, usados para classificar se os itens são adequados, e o valor deveria estar entre 0,60 e 1,40 (Linacre, 2010). Por fim, as análises de sensibilidade e especificidade foram realizadas pelo SPSS, por meio da curva ROC, lembrando que apenas para essa análise utilizou-se a amostra clínica.

3. Resultados

Inicialmente, verificou-se a distribuição dos 12 itens (considerando os subitens A e B dos itens 8 e 10) com a análise de retenção dos dados no programa Factor, versão 10.5.01 (Lorenzo-Seva & Ferrando, 2006). Foram empregados os métodos HULL (Hull method) análise paralela e BIC (BIC method). A análise paralela e o HULL recomendaram a retenção de um fator, enquanto o BIC recomendou até três. As distribuições dos itens nos modelos de dois e três fatores não foram satisfatórias do ponto de vista dos descritores teóricos e da interpretabilidade. Por exemplo, no modelo de dois fatores, apenas o item 10 B (apetite aumentado) carregou no fator 2, e, na solução de três fatores, houve mescla de descritores afetivos, cognitivos e somáticos, além de novamente o item 10 B, por ser o único com aumento de sintomas, carregar em um fator isolado.

Em função disso, optou-se pelo modelo de um fator. Essa solução apresentou $\chi^2(54) = 133.080$, RMSEA= 0,083, CFI= 0,97 e GFI= 0,99. A interpretação da

solução unifatorial permitiu identificar que todos os itens carregaram significativamente no fator geral e o índice de confiabilidade foi de 0,89 pelo alfa de Cronbach. As cargas de cada item são apresentadas na Tabela 3.1. Foi estabelecida uma carga mínima de 0,30 para avaliar se os itens eram adequados a partir da AFE, e apenas o item 10 B (“Você esteve com o apetite aumentado?”) apresentou carga abaixo de 0,40.

Tabela 3.1. Matriz rodadas com as cargas fatoriais.

1. Você se sentiu de baixo astral ou triste?	0,76
2. Você perdeu interesse em suas atividades diárias?	0,75
3. Você sentiu falta de energia ou força?	0,72
4. Você se sentiu menos autoconfiante?	0,76
5. Você sentiu peso na consciência ou sentimento de culpa?	0,63
6. Você sentiu que viver não vale a pena?	0,68
7. Você teve dificuldade de concentração? Por exemplo, ao ler jornal ou assistir à TV?	0,66
8a. Você se sentiu agitado?	0,44
8b. Você se sentiu desanimado ou mais lento?	0,80
9. Você teve problemas para dormir à noite?	0,53
10a. Você esteve com o apetite diminuído?	0,47
10b. Você esteve com o apetite aumentado?	0,34

Os dados também foram analisados a partir da TRI pelo *software* Winsteps, versão 3.70.0 (Linacre, 2010), e considerou-se a correlação item *theta* com valores acima de 0,30, e o número menor encontrado foi 0,44. As medidas de *infit* e *outfit* foram consideradas adequadas quando no intervalo entre 0,60 e 1,40 (Linacre, 2010). As estatísticas descritivas sumarizadas das pessoas e dos itens referentes ao traço latente (*theta*), a dificuldade dos itens (b) os índices de *infit* e *outfit* podem ser observados na Tabela 3.2.

Tabela 3.2. Estatísticas descritivas sumarizadas das pessoas e dos itens.

Medidas	Pessoas			b	Itens	
	Theta	Infit	Outfit		Infit	Outfit
Média	-0,90	1,00	1,04	0,00	1,12	1,04
DP	0,91	0,64	0,72	0,43	0,53	0,46
Máximo	2,00	4,92	6,12	1,09	2,21	2,16
Mínimo	-3,83	0,06	0,10	-0,38	0,50	0,53

A média de *theta* dos respondentes ficou quase 1 logit abaixo da média dos itens, demonstrando que os itens foram difíceis para a amostra. Os itens do instrumento foram analisados quanto aos índices de *infit* e *outfit*, à dificuldade (b) e à correlação item-*theta*, cujos resultados são apresentados na Tabela 3.3.

Tabela 3.3. Estatística dos itens do MDI pelo modelo Rating Scale.

Item	Escore bruto	N	Dificuldade	Infit	Outfit	Correlação item- <i>theta</i>
1	1089	637	-0,34	0,51	0,54	0,72
2	1036	637	-0,27	0,65	0,64	0,71
3	1094	637	-0,35	0,74	0,74	0,70
4	990	637	-0,21	0,71	0,69	0,71
5	906	637	-0,08	0,96	0,97	0,62
6	339	637	1,12	1,74	1,13	0,54
7	782	637	0,11	0,94	0,93	0,93
8a	1127	637	-0,39	1,32	1,33	0,54
8b	1091	637	-0,35	0,53	0,54	0,74
9	819	637	0,05	1,44	1,29	0,56
10a	477	637	0,72	1,79	1,53	0,44
10b	853	637	0,00	2,14	2,09	0,43
Média	883,6	637	0,00	1,12	1,04	-
DP	241,7	0,0	0,45	0,53	0,45	-



Figura 3.1. Mapa de pessoas e itens da escala.

Como indicado na Tabela 3.3, os valores de *infit* para os itens 1, 6, 8b, 9, 10a e 10b ficaram fora do intervalo. Já em relação ao *outfit*, os itens 1, 10a e 10b também estiveram fora do intervalo de 0,60 e 1,40. Assim, das 12 possibilidades de resposta, seis ficaram fora dos parâmetros considerados como adequados no *infit*. Verificaram-se as correlações com magnitudes de moderadas a fortes, que variaram entre 0,43 e 0,93, e os valores de dificuldade dos itens variaram entre -0,39 e 1,12. A Figura 3.1 mostra o mapa de pessoas e itens.

Como mostra o mapa de itens da Figura 3.1, a média dos itens foi superior à média das pessoas, indicando que os itens foram difíceis para a amostra em questão, denotando uma lacuna entre os itens e as pessoas. A maior parte das pessoas ficou agrupada entre -2 e 0, e cada # equivale a cinco pessoas, enquanto os itens ficaram alocados entre -1 e 1. Os itens mais difíceis foram o 6 (“Você sentiu que viver não vale a pena?”) e o 10a (“Você esteve com o apetite diminuído?”), com valores de dificuldade de 1,12 e 0,72, respectivamente. Os itens mais fáceis foram o 8a (“Você se sentiu agitado?”) e o 8b (“Você se sentiu desanimado ou mais lento?”), com dificuldades de -0,39 e -0,35. Realizou-se uma análise das categorias de pontuação na escala Likert (de 0 a 5) do MDI, como indica a Tabela 3.4.

Tabela 3.4. Categorias de respostas do MDI.

Categoria de resposta	%	Média observada	Limiar
0	33	-1,66	-2,56
1	33	-0,97	-0,96
2	10	-0,54	-0,23
3	10	-0,17	0,29
4	9	0,18	0,99
5	4	0,59	2,41

Constatou-se que, quando as opções de resposta do instrumento aumentavam de valor (0 a 5), as médias observadas dos níveis de *theta* também se elevavam, ou seja, ampliava-se a probabilidade do sujeito em endossar uma categoria de resposta conforme o nível de sintomatologia depressiva. A Figura 3.2 apresenta as categorias de pontuação.

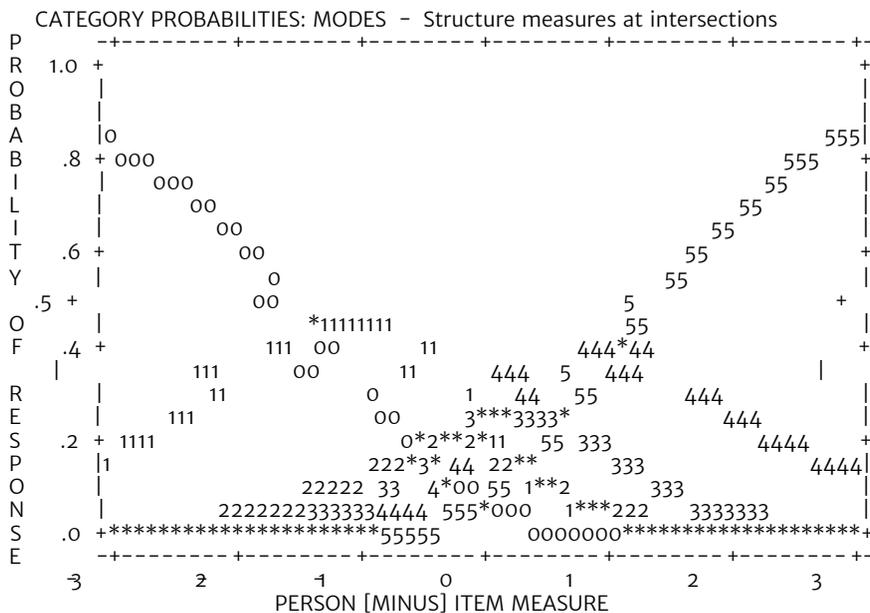


Figura 3.2. Mapa de categorias de respostas do MDI.

A Figura 3.2 indica a probabilidade da pessoa em endossar uma categoria de resposta. A intersecção entre duas categorias de resposta indica o valor do limiar (*threshold*) estimado delas. O limiar de transição da categoria 0 para a categoria 1 equivale ao valor de $-1,33$ de *theta*; da 1 para a 2, ao valor $0,39$; da 2 para a 3, ao valor de $-0,31$; da 3 para a 4, ao valor de $0,18$; e da 4 para a 5, ao valor de $1,07$. Observou-se que as categorias 2 e 3 foram pouco representativas, não se distinguindo, ao contrário das demais categorias, o que indica que não é necessária uma escala Likert de 6 pontos para o instrumento.

Também se calcularam os índices de sensibilidade e especificidade a partir da curva ROC. Para tanto, adotaram-se duas maneiras de pontuar a escala: com a pontuação dos itens entre 0 e 5 pontos (Tabela 3.5) e entre 1 e 6 pontos. Os valores de sensibilidade e especificidade para os pontos de corte 17 (quando pontuado de 0 a 5) e 29 (quando pontuado de 1 a 6) foram de 83,3% e 63,9%, respectivamente, sendo a *area under the curve* (AUC) de 76,3. Esse ponto foi o que apresentou a maior sensibilidade e especificidade, preservando a sensibilidade como maior em relação à especificidade.

Tabela 3.5. Sensibilidade e especificidade do MDI.

Ponto de corte	Sensibilidade (%)	Especificidade (%)
8,5000	1,000	0,727
9,5000	1,000	0,673
10,5000	1,000	0,634
11,5000	0,833	0,595
12,5000	0,833	0,551
13,5000	0,833	0,510
14,5000	0,833	0,463
15,5000	0,833	0,419
16,5000	0,833	0,389
17,5000	0,833	0,361
18,5000	0,750	0,350
19,5000	0,667	0,322
20,5000	0,667	0,295
21,5000	0,583	0,276
22,5000	0,500	0,254
23,5000	0,500	0,239
24,5000	0,500	0,226
8,5000	1,000	0,727
9,5000	1,000	0,673
10,5000	1,000	0,634

4. Discussão

O presente estudo teve como principal objetivo avaliar as qualidades psicométricas do MDI, com base na TCT e TRI, já que o único estudo encontrado no Brasil (Parcias et al., 2011) gerou resultados baseados apenas em TCT e encontrou uma solução componencial que poderia ser questionada do ponto de vista psicométrico e teórico (distribuição das cargas fatoriais e itens nos dois componentes), justificando uma nova análise, agora fatorial, e por métodos mais robustos.

Tanto na análise de Parcias et al. (2011) como na atual, pode-se verificar a possibilidade de itens antagônicos rebuscarem as soluções fatoriais. De acordo com

Podsakoff, MacKenzie, Lee e Podsakoff (2003), o efeito de método representa um tipo de variação sistemática introduzida como um resultado de determinada abordagem de pesquisa ou método usado para medir qualquer característica que esteja em investigação, que, no caso da pesquisa, seriam as ideias antagonistas de determinados itens do teste psicológico, como os itens 10a e 10b, que captam a mesma informação referente ao apetite do respondente (aumentada e diminuída), podendo representar as variações entre os resultados do instrumento por causa de respostas para características de medição ao conteúdo do traço latente do item. Além disso, cabe considerar que alterações no apetite podem ocorrer, mas não são sintomas unicamente típicos de depressão.

Ainda, o estudo de avaliação das propriedades psicométricas do MDI de Parcias et al. (2011) demonstrou uma solução com dois fatores, a partir um índice de KMO de 0,50, valor não considerado adequado para fatorabilidade de acordo com Friel (2009) que sugere valores entre 0,50 e 0,59 como sendo ruins. Da mesma forma, o estudo de Fountoulakis et al. (2003) utilizou a TCT com o objetivo de buscar evidências de validade para o MDI. Nesse estudo, identificaram-se as mesmas fragilidades na análise, como o método de rotação Varimax que revelou dois fatores para o instrumento: um composto por todos os itens, com exceção do 10b, e o outro com apenas o item excluído.

A solução via TCT demonstrou adequados valores de cargas fatoriais dos itens e ajustes, apesar de o RMSEA se encontrar no limiar do esperado, como aponta também o estudo de Olsen, Jensen, Noerholm, Martiny e Bech (2003). A consistência encontrada também se aproxima dos estudos de Fountoulakis et al. (2003) e Cuijpers et al. (2007). No entanto, as análises via TRI demonstraram pontos críticos, alguns coerentes com estudos muito recentes realizados com amostras de atenção primária e com hipótese diagnóstica de depressão, diferentes da atual amostra de universitários (Christensen, Oernboel, Nielsen, & Bech, 2019; Nielsen, Ørnboel, Vestergaard, Bech, & Christensen, 2017).

Foi possível observar que vários itens apresentaram inadequação quando comparados aos valores de referência do *infit* e *outfit*. Esses resultados indicaram que, tanto em relação aos itens quanto no que concerne às pessoas, foi encontrado um padrão de resposta esperado, porém a escala tem alguns erros de ajuste segundo a classificação de Linacre (2010). Analisou-se também se as categorias de respostas para o instrumento foram consideradas adequadas. No geral, as categorias

demonstraram uma representatividade de fácil compreensão dos respondentes, no entanto a atual divisão do MDI compõe uma escala de seis pontos (de 0 a 5), e as pontuações 2 e 3 não foram bem discriminadas pelos respondentes.

É interessante notar que os resultados encontrados na amostra brasileira vão ao encontro daqueles também desenvolvidos no exterior. Por exemplo, Christensen et al. (2019) também sugerem que alguns itens, como o 9 e 10, demonstraram desajuste no modelo Rasch, além de as categorias de resposta 2 e 3 também não serem discriminativas, apontando para uma escala de resposta Likert de 5 pontos. Além disso, os índices adequados de unidimensionalidade da escala só foram possíveis quando se retirou o item 9 da escala. Em relação aos itens desajustados, os autores comentam que esse fenômeno pode ocorrer, na amostra de pacientes de atenção primária, em decorrência de serem itens que se sobrepõem a sintomas de doenças somáticas, ansiedade ou mesmo com sintomas que são decorrentes do uso de medicamentos.

Nielsen et al. (2017) também realizaram uma análise de Rasch em uma amostra de 245 pessoas com suspeita de depressão, indicadas por profissionais de saúde. Os resultados também corroboraram os nossos achados a respeito de itens mais fáceis (por exemplo, o item 2) e do mais difícil (o item 6), além de indicarem os itens 1, 3, 8, 9 e 10 como problemáticos em termos de ajustes e resíduos. As categorias de resposta da escala Likert também se mostraram de difícil discriminação, sugerindo o colapsamento das categorias 2, 3 e 4, resultando em uma escala com quatro possibilidades de resposta, de modo a ajustar a escala para bons índices e boa unidimensionalidade.

Por fim, para analisar o ponto de corte da escala utilizada, foi realizado uma curva ROC para averiguar o padrão classificatório segundo o instrumento. Nas pesquisas que utilizaram o MDI como instrumento para avaliação (Bech & Wermuch, 1998; Fountoulakis et al., 2003), obtiveram-se pontos de corte de 26 e 27 (pontuação de 1 a 6 na escala do tipo Likert), ao passo que o presente estudo apresentou ponto de corte de 17 (pontuação de 0 a 5) e 29 na pontuação de 1 a 6. Quando se utilizou o ponto de corte do presente estudo, os valores de sensibilidade e especificidade foram de 83,3 e 63,9, respectivamente. O ponto de corte adotado levou em consideração a maior sensibilidade possível, além da maior especificidade, mas tendo como parâmetro a sensibilidade sempre maior do que a especificidade, característica importante para escalas de rastreamento.

Como apontam Mihura, Bombel, Dumitrascu, Roy e Meadows (2018), a acurácia de um teste pode ter maior credibilidade quando há revisões sistemáticas sobre ele, o que não foi encontrado na literatura. De qualquer forma, a sensibilidade, especificidade e AUC poderiam apresentar valores mais adequados, já que, como indicam Pettersson, Bostrom, Gustavsson e Ekselius (2015), valores de sensibilidade acima de 80% e de especificidade acima de 70% seriam mais próximos do esperado em escalas de rastreamento de sintomas, o que não foi alcançado no atual estudo, talvez por conta do tipo de amostra.

Por fim, o MDI é um instrumento de rastreio de sintomas da depressão, com tradução para diversos idiomas, no entanto ainda são poucos os estudos acerca de suas propriedades psicométricas. Sugere-se a realização de novos estudos para avaliar se o entendimento da resposta na escala Likert continua sendo pouco discriminativa, considerando reagrupar os dois pontos que não se diferenciaram no Brasil. Também se aconselha acrescentar uma amostra maior de pacientes depressivos no intuito de fazer um comparativo entre amostras clínicas e não clínicas, bem como a utilização de entrevista estruturada na amostra clínica, uma das maiores limitações deste estudo.

Referências

- Baptista, M. N. (2018). Avaliando “depressões”: Dos critérios diagnósticos às escalas psicométricas. *Avaliação Psicológica*, 17(3), 301–310. doi:10.15689/ap.2018.1703.14265.03
- Baptista, M. N., & Borges, L. (2016). Revisão integrativa de instrumentos de depressão em crianças/adolescentes e adultos na população brasileira. *Avaliação Psicológica*, 15(esp.), 19–32. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712016000300004&lng=pt&tlng=pt
- Bech, S., Rasmussen, N. A., Olsen, L. R, Noerholm, V., & Abildgaard, W. (2001). The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. *Journal of Affective Disorders*, 66, 159–164. doi:10.1016/S0165-0327(00)00309-8
- Bech, P., Timmerby, N., Martiny, K., Lunde, M., & Soendergaard, S. (2015). Psychometric evaluation of the Major Depression Inventory (MDI) as depression severity scale using the LEAD (Longitudinal Expert Assessment of All Data) as index of validity. *BMC Psychiatry*, 15(1), 1–7. doi:10.1186/s12888-015-0529-3

- Bech, P., & Wermuch L. (1998). Applicability and validity of the MDI in patients with Parkinson's disease. *Nordic Journal of Psychiatry*, 52(4), 305–310. doi:10.1080/08039489850149741
- Behera, P., Gupta, S. K., Nongkynrih, B., Kant, S., Mishra, A. K., & Sharan, P. (2017). Screening instruments for assessment of depression. *Indian Journal of Medical Specialities*, 8(1), 31–37. doi:10.1016/j.injms.2016.11.003
- Christensen, S., Oernboel, E., Nielsen, M. G., & Bech, P. (2019). Diagnosing depression in primary care: A Rasch analysis of the major depression inventory. *Scandinavian Journal of Primary Health Care*, 37(1), 105–112. doi:10.1080/02813432.2019.1568703
- Cuijpers, P., Dekker, J., Noteboom, A., Smits, N., & Peen, J. (2007). Sensitivity and specificity of the Major Depression Inventory in outpatients. *BMC Psychiatry*, 7, 39–44. doi:10.1186/1471-244X-7-39
- Fountoulakis, K. N., Iacovides, A., Kleanthous, S., Samolis, S., Gougoulas, K., Tsiptios, I., Kaprinis, G. S., & Bech, P. (2003). Reliability, validity and psychometric properties of the Greek translation of the Major Depression Inventory. *BMC Psychiatric*, 3(1), 1–8. doi:10.1186/1471-244X-3-2
- Friel, C. M. (2009). *Notes on factor analysis*. Huntsville, Texas: Criminal Justice Centre, Sam Houston State University.
- Linacre, J. M. (2010). *Auser's guide to Winsteps, Program Manual 3.70.0*. Chicago: Winsteps.com.
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods*, 38(1), 88–91. doi:10.3758/BF03192753
- Mihura, J. L., Bombel, G., Dumitrascu, M., Roy, M., & Meadows, E. A. (2018). Why we need a formal systematic approach to validating psychological tests: The case of the Rorschach Comprehensive System. *Journal of Personality Assessment*, 101(4), 374–392. doi:10.1080/00223891.2018.1458315
- Nielsen, M. G., Ørnbøl, E., Vestergaard, M., Bech, P., & Christensen, K. S. (2017). The construct validity of the Major Depression Inventory: A Rasch analysis of a self-rating scale in primary care. *Journal of Psychosomatic Research*, 97, 70–81. doi:10.1016/j.jpsychores.2017.04.001
- Olsen, L. R., Jensen, D. V., Noerholm, V., Martiny, K., & Bech, P. (2003). The internal and external validity of the Major Depression Inventory in measuring severity of depressivestates. *Psychological Medicine*, 33, 351–356. doi:10.1017/S0033291702006724

- Parcias, S., Rosario, B. P. do, Sakae, T., Monte, F., Guimarães, A. C. A., & Xavier, A. J. (2011). Validação da versão em português do Inventário de Depressão Maior. *Jornal Brasileiro de Psiquiatria*, 60(3), 164–170. doi:10.1590/S0047-20852011000300003
- Pettersson, A., Bostrom, K. B., Gustavsson, P., & Ekselius, L. (2015). Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nordic Journal of Psychiatry*, 69(7), 497–508. doi:10.3109/08039488.2015.1008568
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, M. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. doi:10.1037/0021-9010.88.5.879
- Santor, D. A., Gregus, M., & Welch, A. (2006). Eight decades of measurement in depression. *Measurement: Interdisciplinary Research and Perspectives*, 4(3), 135–155. doi:10.1207/s15366359mea0403_1
- World Health Organization (2017). *Depression and other common mental disorders. Global health estimates*. Geneva: WHO. Recuperado de <http://apps.who.int/iris/bits-tream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf>

Nota dos autores

Makilim N. Baptista, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco (USF); **Gabriela da S. Cremascom**, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco (USF); **Felipe Augusto Cunha**, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco (USF); **Samira R. Marcon**, Departamento de Enfermagem, Universidade Federal do Mato Grosso (UFMT); **Hugo G. B. dos Santos**, Unidade de Atenção Psicossocial, Hospital Universitário Júlio Muller.

Correspondências referentes a este artigo devem encaminhas para Felipe Augusto Cunha, Rua José de Paula Andrade, 45, Vila Belém, Itatiba, SP, Brasil. CEP 13256320.
E-mail: flpcunha@terra.com.br