

Escala Baptista de Depressão (Versão Adulto) – EBADEP-A: evidências de validade de construto e de critério¹

Makilim Nunes Baptista² – Universidade São Francisco, Itatiba, Brasil
Juliana Oliveira Gomes – Universidade São Francisco, Itatiba, Brasil

Resumo

O presente estudo teve como objetivo analisar as qualidades psicométricas da Escala Baptista de Depressão (Versão Adulto) – EBADEP-A, com base na Teoria da Resposta ao Item – TRI e na Teoria Clássica dos Testes – TCT. Inicialmente, sobre os parâmetros de ajuste ao modelo, tanto no que concerne aos itens quanto às pessoas, os valores foram considerados de boa adequação, com percentual de desajuste considerado baixo. Em relação à precisão, tanto o alfa de Cronbach quanto o índice gerado pelo modelo de Rasch foram considerados excelentes. O estudo do funcionamento diferencial apresentou 17 itens com viés de resposta, sendo 11 favorecendo o grupo feminino e 6 o masculino. Quanto às análises com base na TCT, foi realizada ANOVA para análise dos grupos critério, sendo que a EBADEP-A foi capaz de discriminar os grupos de não-depressivos, universitários, psiquiátricos e depressivos. Estes resultados foram considerados como evidências de validade de construto e critério, respectivamente, complementando as diversas evidências já encontradas para a escala.

Palavras-chave: Funcionamento diferencial do item, Teoria da resposta ao item, Teoria clássica dos testes, EBADEP-A, Validade.

Escala Baptista de Depressão (Versão Adulto) – EBADEP-A: construct and criterion validity evidences

Abstract

This study aimed to analyze the psychometric qualities of the Escala Baptista de Depressão (Versão Adulto) – EBADEP-A based on Item Response Theory – IRT and Classical Test Theory – TCT. Initially, adjustment parameters on the model, both regarding to items and to people, were considered good fit values, with a low percentage of mismatches. In relation to the reliability, both Cronbach's alpha and the Rasch index were considered great. The study of differential functioning presented 17 items with response bias, 11 favoring the female group and 6 the male one. As for the analysis based on TCT, it was performed an ANOVA to analyze the criterion groups, and the EBADEP-A was able to discriminate the non-depressed, college students, psychiatric and depressive groups. These results were considered as evidence of construct validity and criterion, respectively, complementing the variety of evidence ever found for the scale.

Keywords: Differential item functioning, Item response theory, Classical test theory, EBADEP-A, Validity.

A depressão pode ser entendida como sintoma, característico de distúrbios emocionais, como por exemplo, o transtorno de estresse pós-traumático ou a demência; pode ser classificada como síndrome, quando inclui alterações de humor, apatia, falta de capacidade de sentir prazer, etc. e ainda pode ser considerada como uma doença, recebendo diferentes classificações, como o transtorno depressivo maior, melancolia, distímia, etc. (Del Porto, 2002). Para os manuais diagnósticos internacionais da Associação Americana de Psiquiatria (APA, 2002), o DSM-IV-TR e da Organização Mundial de Saúde (OMS, 1993), o CID.10, a depressão é considerada como um transtorno de humor e transtorno afetivo, respectivamente.

Embora seja característica da depressão a manifestação por episódios recorrentes, os quais somente devem ser classificados como tal quando o

rebaixamento de humor e outros sintomas perdurarem por no mínimo duas semanas, em torno de 20% dos casos são registrados como crônicos, sem remissão. Trata-se de um transtorno que pode acometer a pessoas de diferentes classes econômicas e idades, ainda que o número de depressivos seja relativamente maior em jovens (OMS, 2001).

Os critérios para diagnóstico da doença são baseados nos manuais diagnósticos internacionais. Embora eles tragam uma classificação diferente para o transtorno, ambos estabelecem critérios baseados na quantidade e qualidade dos sintomas (APA, 2002; OMS, 1993). Ambos os manuais orientam para a identificação obrigatória de humor deprimido e/ou a incapacidade de experimentar prazer em atividades antes consideradas prazerosas, de forma constante, por no mínimo duas semanas, além de vários sintomas adicionais, os quais incluem alterações significativas no sono, apetite, agitação ou retardo psicomotor, fadiga, perda de energia, sentimento de inutilidade, culpa, dificuldade de concentração, pensamentos de morte, ideação suicida ou até mesmo tentativa de suicídio (APA, 2002; OMS, 1993).

Vários são os fatores que podem estar relacionados com a depressão, que podem funcionar tanto como fatores protetivos quanto de risco. Dentre

¹ Apoio: CAPES, CNPq.

²Endereço para correspondência:

Diretoria Acadêmica de Pós-Graduação
Rua Alexandre Rodrigues Barbosa, 45 – Centro – 13251-900 –
Itatiba-SP – Brasil
E-mail: makilim.baptista@saofrancisco.edu.br

estes, podem ser citados o histórico familiar da doença, conviver em família disfuncional, vivenciar grande número de eventos estressores, ter problemas graves de saúde, baixa autoeficácia, locus de controle, dentre outros (Baptista, 2004). Outra característica marcante, que também se caracteriza como um fator de risco, envolve o número maior de casos registrados no sexo feminino, sendo estimada prevalência em 3,2% em mulheres, enquanto o valor para os homens é de 1,9% (Justo & Calil, 2006; OMS, 2001).

A tendência acentuada do sexo feminino tem sido alvo de estudos no sentido de tentar explicar tal fenômeno. Pesquisas como as do *National Institute of Mental Health* têm demonstrado que parte da explicação pode estar respaldada pelo tipo de resposta, tanto somática quanto de humor, à doença, além de oscilações hormonais, naturais das mulheres (Kessler, 2003; NIMH, 2000). Nesses casos, estudam-se não apenas a menarca, tensão pré-menstrual, depressão pós-parto e menopausa, mas também fatores genéticos, suporte social, instabilidades hormonais resultantes da vivência de eventos estressantes e fatores culturais próprios de cada gênero (Baptista, Baptista & Oliveira, 2004; Kessler, 2003; Piccinelli & Wilkinson, 2000).

Tendo em vista essa peculiar característica do transtorno, fazem-se necessárias pesquisas que contemplem a investigação sobre a diferença entre gêneros e no estudo da depressão, principalmente quando o foco é a construção e validação de instrumentos de medida psicológicos e a possibilidade de haver vieses de resposta nos mesmos. De fato, são vários os estudos envolvendo a construção e/ou validação de instrumentos para a identificação da depressão, como por exemplo, o Inventário de Depressão de Beck (Cunha, 2001; Goreinstein, Pang, Argimoni & Werlang, 2011; Gorestein & Andrade, 1998), a Escala de Depressão Geriátrica (Almeida & Almeida, 1999), o Inventário de Depressão Infantil (Gouveia, Barbosa, Almeida & Gaião, 1995), a Escala de Rastreamento Populacional para a Depressão, CES-D (Batistoni, Neri & Cupertino, 2007; Silveira & Jorge, 1998) e as escalas de depressão Montgomery & Asberg e de Hamilton (Moreno & Moreno, 1998), dentre outros. Entretanto, são poucos os que se prestaram a estudar sobre os possíveis vieses de resposta entre homens ou mulheres nesses instrumentos, bem como analisar psicometricamente o instrumento por intermédio da TRI.

Preenchendo tal lacuna, a Escala Baptista de Depressão (Versão Adulto) – EBADEP-A foi construída no Brasil e tem o objetivo de avaliar a intensidade de depressão em adolescentes e adultos (Baptista, 2011). A escala tem sido estudada em diferentes contextos tanto no sentido de busca por

evidências validade pela Teoria Clássica dos Testes – TCT (Baptista, Souza & Alves, 2008; Baptista, Carneiro & Sisto, 2010; Gomes & Baptista, 2010) quanto pela Teoria da Resposta ao Item – TRI (Baptista, Gomes, Carneiro & Borine, 2011).

A TRI consiste em um conjunto de modelos, os quais buscam representar os parâmetros de um teste, focando nos itens, ao invés do teste como um todo, como acontece na TCT (Pasquali & Primi, 2003; Rueda, 2007). Dentro dos modelos estatísticos propostos, podem ser considerados três parâmetros: a dificuldade, discriminação e o acerto casual dos itens. No presente trabalho, focaliza-se o modelo de um parâmetro (dificuldade) de Georg Rasch, no qual se considera que a possibilidade de acerto de uma pessoa nas questões de um teste depende tanto do nível de dificuldade do item apresentado pelo instrumento quanto do nível de habilidade de quem o responde (Embretson & Reise, 2000; Pasquali & Primi, 2003; Rueda, 2007).

Embora a maioria dos testes venha sendo desenvolvida com base na TCT, acredita-se que instrumentos diagnósticos, como o caso da EBADEP-A, podem ser beneficiados com a utilização de modelos estatísticos mais modernos, como a TRI. Forkmann e colaboradores (2009) afirmam que a utilização de TRI em instrumentos de avaliação de saúde mental tem aumentado nos últimos anos e, por exemplo, desenvolveram, com base no Modelo de Rasch, um instrumento denominado *Aachen Depression Item Bank* (ADIB). Participaram 161 pacientes em tratamento para a depressão e 333 não-pacientes, com idade média de 43,6 anos (DP=14), sendo 46,2% mulheres. Os itens mostraram variação na confiabilidade, com indicadores entre 0,48 e 0,82 e a escala apresentou adequadas sensibilidade e especificidade (0,81 e 0,77).

No Brasil, foi realizado estudo com base no modelo politômico de Resposta Gradual para o Inventário de Depressão de Beck, a partir de um banco de dados composto por 4.025 sujeitos, divididos em três grupos, quais sejam, pacientes psiquiátricos (n=1.138), pacientes de clínica-médica (n=490) e amostra não-clínica (n=2.397). Do total de participantes, 45,4 % eram homens e 54,6% mulheres. Embora a suposição de unidimensionalidade tenha se mantido, pelo ajuste do modelo, em 13 dos 21 itens, pelo menos um dos itens não apresentou probabilidade maior que outros de ser respondido, quaisquer que fossem os níveis de depressão, o que foi relacionado a possíveis problemas no dimensionamento da escala. Os itens referentes à tristeza, pessimismo, pensamentos de fracasso, insatisfação, autoaversão, indecisão e dificuldade para trabalhar foram os que mais

forneceram informações à medida de intensidade dos sintomas, e o item de irritabilidade foi o que menos contribuiu para o traço latente. Por fim, os autores concluíram que o BDI possui melhor funcionamento para a intensidade equivalente aos escores entre 22 e 40 pontos (Castro, Trentini & Roboldi, 2010).

Como dito anteriormente, modelo de Rasch permite o estudo do viés de resposta a um item, ou seja, qual a equivalência dos itens quando aplicados em dois grupos distintos, como, por exemplo, grupos de mulheres e homens (Rueda, 2007; Sisto, 2006) e com base nessa premissa é possível verificar se os itens do teste apresentam um Funcionamento Diferencial (DIF). Dessa forma, os objetivos deste estudo se concentram em estudar as qualidades psicométricas da EBADEP-A com base tanto na TRI quanto na TCT. Especificamente, pretende-se verificar a estrutura interna, distribuição da escala e categorias de resposta, a distribuição dos itens e a confiabilidade, com base no modelo de um parâmetro de Rasch, com o objetivo de buscar evidências de validade de construto, além de verificar a distribuição dos grupos critério, gerando outra evidência de validade.

Método

Participantes

A presente pesquisa foi realizada com base em um banco de dados derivado de diferentes aplicações do instrumento em instituições de ensino superior e de saúde nos estados de São Paulo (83,4%) e Minas Gerais (16,6%). O banco foi composto por 1.467 pessoas, com idades variando entre 17 e 81 anos ($M=26,44$; $DP=9,55$), sendo 382 homens (26%) e 1.085 mulheres (74%). Do total de participantes, 1.307 (89,1%) eram universitários, 40 (2,7%) depressivos, 40 (2,7%) não-depressivos e 80 (5,5%) pacientes psiquiátricos, os quais foram diagnosticados por médico psiquiatra, não necessariamente apresentando a depressão como doença principal.

Instrumentos

Gosto muito da minha vida	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Não dou mais valor à minha vida
Tenho me sentido muito bem	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Tenho estado mais angustiado
O futuro será melhor	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Não acredito que as coisas melhorem

Escala Baptista de Depressão (Versão Adulto) – EBADEP-A (Baptista, 2011).

Trata-se de um instrumento construído no Brasil, de rastreio de sintomatologia depressiva, direcionado tanto a amostras psiquiátricas quanto a não-psiquiátricas. Sua construção teve como base indicadores sintomáticos de depressão derivados de teorias sobre a depressão, como a Cognitiva (Beck & cols., 1997) e a Comportamental (Ferster, Culbertson & Boren, 1977), além dos manuais internacionais de diagnóstico da Associação Americana de Psiquiatria, DSM-IV-TR e da Organização Mundial de Saúde, CID-10 (APA, 2002; OMS, 1993). A escala é constituída por 90 frases, as quais são apresentadas em pares, formando 45 itens. Cada item apresenta um indicador de sintomatologia com uma frase de cunho positivo e outra de cunho negativo. Ela é estruturada em formato *Likert* de quatro pontos, variando de zero a três, com pontuação mínima de zero e máxima de 135 pontos. Para sua interpretação, considera-se quanto menor a pontuação, menor sintomatologia em depressão.

A Escala possui estudos de qualidades psicométricas desde sua primeira versão, com 75 itens, com diferentes tipos de amostra. Os estudos psicométricos finais foram realizados com base em um banco de dados composto por 1.676 participantes, sendo 78,5% universitários, 2,4% pacientes depressivos, com diagnóstico confirmado pela *Structured Clinical Interview Diagnostic* (SCID), 2,4% não-depressivos (também confirmados pela SCID), 6% hospitalizados, 6% acompanhantes dos participantes hospitalizados e 4,8% pacientes psiquiátricos. Foram realizadas pesquisas com base na Teoria Clássica dos Testes e também na Teoria da Resposta ao Item, a partir das quais o instrumento recebeu evidências de validade de conteúdo, construto e critério, além de ter sido verificada sua fidedignidade e análises de sensibilidade e especificidade (Baptista, 2011). Como exemplos dos itens têm-se:

Resultados

Para que os objetivos propostos para a seguinte pesquisa pudessem ser cumpridos, dois tipos diferentes de análises foram realizadas. Inicialmente, utilizou-se o programa estatístico Winsteps para as análises baseadas no modelo de Rasch, tanto as de caráter descritivo quanto as de estrutura interna da escala, e a análise do funcionamento diferencial dos itens. Para responder ao segundo objetivo, relativo às diferenças de média em relação aos grupos característicos que compuseram o banco de dados, recorreu-se à análise de variância (ANOVA), com o Teste Post-Hoc de Tukey, baseada na Teoria Clássica dos Testes.

Primeiramente, o instrumento foi analisado de modo a verificar seu ajuste ao modelo Rasch, sendo verificados os parâmetros de ajuste dos itens e pessoas

(Tabela 1). Pode-se dizer que os itens foram respondidos dentro de um padrão esperado, uma vez que a média do *Infit* foi próxima a 1,00, valor considerado como referência (Linacre, 2002). Percebe-se que os valores variaram entre 0,62 a 1,74, observando-se que de fato somente um item se apresentou com *Infit* no valor acima de 1,5, já que os valores considerados como de bom ajuste variam entre 0,70 e 1,30.

Da mesma forma, a média dos *outfits* foi de 1,06 (DP=0,33), valor próximo ao padrão esperado, que é 1,0, apontando boa adequação ao modelo Rasch (Linacre, 2002). A variação dos *outfits* foi de 0,54 a 2,23, apontando que alguns dos itens não se enquadraram dentro do intervalo esperado, porém, apenas quatro itens se apresentaram com *outfit* acima de 1,5, especificamente três itens entre 1,5 e 2,0 (4,4%) e apenas um item com valor acima de 2,0 (2,2%).

Tabela 1. Parâmetros de ajuste de itens e pessoas

Parâmetros	Itens			Pessoas		
	<i>Infit</i>	<i>Outfit</i>	Erro	<i>Infit</i>	<i>Outfit</i>	Erro
Média	1,01	1,06	0,03	1,05	1,06	0,19
DP	0,24	0,33	0,00	0,40	0,51	0,07
Máximo	1,74	2,23	0,04	2,86	8,02	0,99
Mínimo	0,62	0,54	0,03	0,21	0,22	0,15
Entre 1,3 e 1,5	5 (11,1%)	6 (13,3%)		135 (9,2%)	134 (9,13%)	
Entre 1,5 e 2,0	1 (2,2%)	3 (6,7%)		145 (9,89%)	134 (9,13%)	
Maior que 2,0	0 (0,0%)	1 (2,2%)		45 (3,1%)	61 (4,15%)	

No que tange às pessoas, observou-se que a maioria apresentou um padrão de resposta esperado, já que os valores das médias foram próximos a 1,0 tanto para os *infits* quanto os *outfits*. De forma mais específica, percebeu-se que poucas foram as pessoas com valores de *infit* e *outfit* maiores que 2,0. Tais resultados levam a crer que os percentuais de desajuste podem ser considerados baixos.

Ainda em relação aos itens e às pessoas, foram também verificadas a distribuição e o agrupamento dos mesmos, conforme visualizado na Tabela 2. A letra M, situada na régua que divide o mapa entre itens e pessoas, refere-se à posição média de um para outro. Percebe-se que embora a média dos itens tenha sido maior que a média das pessoas, o que geralmente indica tendência da amostra a responder com maior frequência as opções de resposta cujos sentidos indicam pouca sintomatologia depressiva, não foram

observados itens não-adequados, pois não há nenhum item com distância discrepante em relação aos outros.

Nota-se que mesmo com a variação da escala aproximadamente entre -5 e 2, a maior parte dos itens se concentrou entre -1 e 1. De fato, os itens mais difíceis foram os de número 44 (Hipocondria), 29 (Ideação Suicida), 30 (Autoestima rebaixada e Autocrítica exacerbada) e 34 (Autoestima rebaixada e Humor deprimido), enquanto os mais fáceis foram o 37 (Irritabilidade), 1 (Humor deprimido e Choro) e 38 (Fadiga/perda de energia). Neste caso, por não se tratar de uma escala de habilidade, em que não há itens certos ou errados, fáceis ou difíceis, interpreta-se dificuldade do item como a frequência de resposta, ou seja, os itens denominados como difíceis são aqueles que apresentaram baixa frequência de respostas para as opções mais próximas da frase de cunho negativa, enquanto as mais fáceis tiveram esta opção amplamente assinalada.

Tabela 2. Mapa de itens e pessoas (N=1.467)

Pessoas - Mapa - Itens	
	<mais> <raro>
2	. +
	.
	.
	.
	.
1	.# T+T It44
	.# It29 It30 It34
	## It10 It33
	.#### S It28
	.### It13 It20 It21 It32
	.#### S It12 It35
	.##### It05 It08 It09 It11 It19
0	.##### +M It03 It18 It31 It41 It42
	.##### It04 It06 It17 It22 It26 It40 It43
	.##### It07 It15 It16 It24
	.##### It23 It25 It39 It45
	##### M S It14 It27 It36
	.##### It02
	.##### It37
-1	.##### +T It01 It38
	.###
	.#####
	.#### S
	.###
	##
	.##
-2	### +
	. T
	.
	.#
	.
	.
-3	. +
	.
	.
	.
-4	. +
	.
	.
-5	. +
	<menos> <frequente>

Cada “#” equivale a 11 pessoas. Cada “.” equivale de 1 a 10 pessoas

Após análise das características gerais do instrumento, procederam-se às análises referentes à estrutura interna e precisão da EBADEP-A. A Tabela 3 apresenta o somatório das pontuações para cada categoria de resposta do instrumento, com suas respectivas porcentagens, médias observadas, *infits*, *outfits* e limiares de resposta. Verificou-se que, à medida

que as opções de resposta da escala aumentaram de valor, ou seja, de zero a um, dois e três, observou-se aumento nas médias observadas dos níveis de teta, ou seja, dos níveis de sintomatologia associadas à probabilidade da amostra em assinalar cada uma das opções de resposta.

Tabela 3. Médias observadas em cada categoria de resposta, *infits*, *outfits* e limiares

Categoria de resposta	Pontuação	%	Média observada	Infit MNSQ	Outfit MNSQ	Limiares
0	28.971	44	-1,16	1,00	1,06	(-1,71)
1	14.475	22	-0,50	0,89	0,77	-0,47
2	11.966	18	-0,05	0,91	0,89	0,45
3	10.595	16	0,34	1,11	1,34	(1,74)

Em relação aos limiares, percebe-se leve equivalência na distância entre eles, uma vez que os valores variam entre -1,71 e 1,74, havendo valores negativos em uma das categorias intermediárias. Entretanto, como essas distâncias não são discrepantes não demonstram acentuados problemas de distribuição da escala.

A distribuição dos itens do instrumento podem ser observada de forma gráfica na Figura 1.

Neste gráfico é possível observar, nas curvas características de cada categoria de resposta, como os limiares ou intersecções (*Thresholds*) se apresentaram e as áreas por eles formadas. Percebeu-se que, embora não tenha sido formada uma delimitação perfeita entre todos os limiares, há uma boa representação das categorias extremas, ou seja, das opções de resposta mais próximas às frases que compõem os itens da escala.

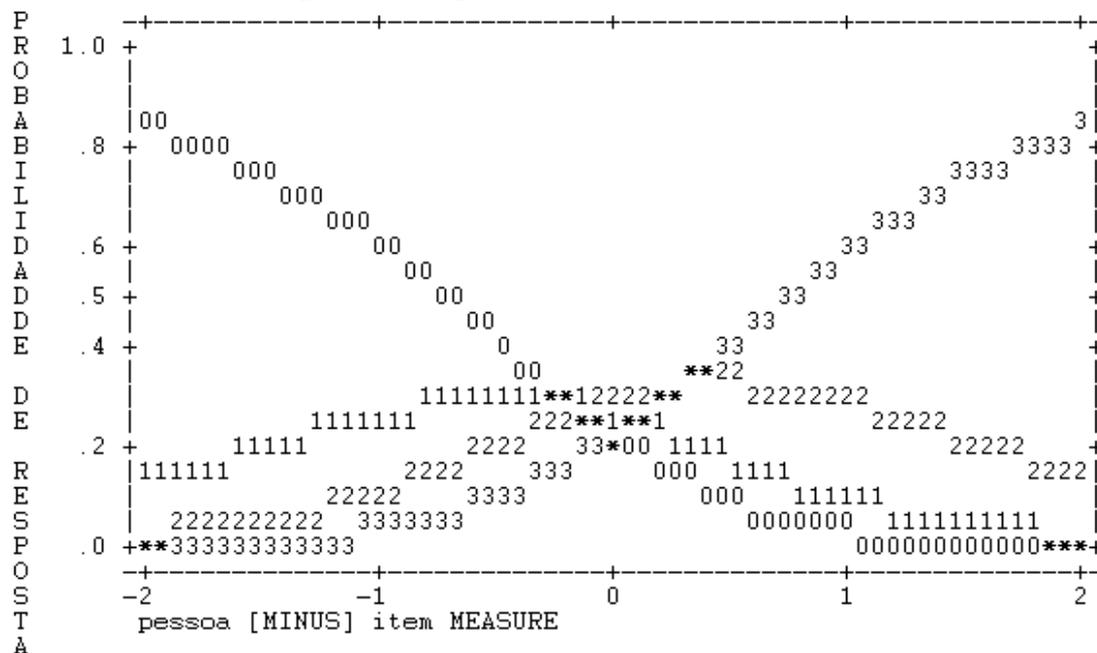


Figura 1. Categorias de resposta da EBADEP-A

Quando ao estudo da precisão da escala, ressalta-se que no modelo de Rasch não se considera um único índice de confiabilidade, mas sim as médias das precisões locais, calculadas com base na variância e nos erros padrão de medida. Assim, são averiguadas as

variações ao longo das pontuações da escala. Como apresentado na Tabela 4, observaram-se os valores de precisão de acordo com o modelo (*Model RMSE*), isto é, sem as respostas dos sujeitos que compuseram a

amostra, e o real (*Real RMSE*), considerando possíveis resíduos.

Tabela 4. Medidas de confiabilidade para a EBADEP-A

Pessoas (N=1467)						
RMSE Real	0,24	DP	0,81	α		0,92*
RMSE do Modelo	0,22	DP	0,81	α		0,93*
Itens(N=45)						
RMSE Real	0,03	DP	0,5	α		0,99
RMSE do Modelo	0,03	DP	0,5	α		0,99

*Cronbach Alpha (kr-20) person raw score reliability = 0,95

Pode-se dizer, segundo os critérios de Prieto e Muñiz (2000), que a EBADEP-A possui índices excelentes de precisão, tanto em relação ao alfa de Cronbach ($\alpha=0,95$) quanto pela precisão real encontrada de acordo com o modelo de Rasch ($\alpha=0,92$). Ressalta-se que esta variação entre os dois valores se deve ao fato de o modelo Rasch considerar os possíveis resíduos encontrados, o que faz com que os índices sejam relativamente menores, porém mais

confiáveis que o valor clássico obtido pelo alfa de Cronbach.

No que tange ao estudo sobre o funcionamento diferencial dos itens, propriamente dito, os resultados são sumarizados na Tabela 5. Do total de 45 itens, 17 (37,7%) indicaram a presença de DIF, segundo o critério de $t > 2,40$ (Draba, 1977). Tais resultados indicam uma quantidade razoável de itens com funcionamento diferencial em razão do sexo.

Tabela 5. Medidas de DIF por sexo, mudança e valores de t por item

Item	Sexo	DIF acrescido	Sexo	DIF acrescido	Mudança	t
ebadep1	masculino	-0,57	feminino	-1,08	0,51	7,79
ebadep6	masculino	0,36	feminino	-0,33	0,69	9,57
ebadep9	masculino	0,07	feminino	0,25	-0,17	2,48
ebadep11	masculino	-0,06	feminino	0,27	-0,33	4,85
ebadep12	masculino	0,15	feminino	0,39	-0,24	3,41
ebadep16	masculino	-0,36	feminino	-0,17	-0,19	2,89
ebadep17	masculino	-0,22	feminino	-0,03	-0,20	2,96
ebadep20	masculino	0,11	feminino	0,53	-0,42	5,86
ebadep21	masculino	0,26	feminino	0,54	-0,28	3,74
ebadep26	masculino	-0,27	feminino	-0,11	-0,16	2,44
ebadep31	masculino	-0,11	feminino	0,08	-0,19	2,76
ebadep32	masculino	0,32	feminino	0,53	-0,20	2,72
ebadep35	masculino	0,02	feminino	0,30	-0,28	4,03
ebadep36	masculino	-0,31	feminino	-0,61	0,29	4,52
ebadep37	masculino	-0,70	feminino	-0,93	0,22	3,43
ebadep38	masculino	-0,89	feminino	-1,12	0,22	3,36
ebadep42	masculino	0,52	feminino	-0,10	0,62	8,14

Dos 45 itens, 11 apresentaram viés de resposta pelo grupo feminino, a saber, os itens 1 (Humor deprimido e choro), 9 (Sentimento de Incapacidade e Carência/Dependência), 11 (Sentimento de inadequação e Autocrítica exacerbada), 12 (Desesperança e Sentimento de incapacidade), 20 (Sentimento de inutilidade e Autocrítica exacerbada), 21 (Esquiva de situações sociais), 32 (Sentimento de

incapacidade e Queda de produtividade), 35 (Queda de produtividade), 37 (Irritabilidade) e 38 (Fadiga/perda de energia). Isso quer dizer que para esses itens as mulheres apresentam maior probabilidade de escolherem a opção “três”, isto é, a frase imediatamente à direita da régua.

Por sua vez, foram seis os itens nos quais os homens estiveram mais propensos a escolherem a

opção mais negativa do par de frases. São eles, os itens 6 (Choro), 16 (Carência/dependência e Sentimento de incapacidade), 17 (Anedonia), 26 (Humor deprimido), 31 (Alterações de sono) e 42 (Perda de libido).

Ressalta-se neste ponto que as informações até aqui apresentadas, as quais auxiliam no estudo da maneira como a depressão, neste instrumento, é mensurada de acordo com um grupo (favorecendo-o ou não), podem ser também consideradas como uma

evidência de validade, do tipo construto (AERA, APA & NCME, 1999).

Como última análise, foi realizada análise de variância (ANOVA) conforme métodos da Teoria Clássica dos Testes, em relação aos grupos de participantes que compuseram a amostra (Tabela 6). O instrumento discriminou cada um dos grupos: Não-depressivos, Universitários, Psiquiátricos e Depressivos, com médias crescentes, nesta ordem.

Tabela 6. ANOVA com prova *Post hoc* de Tuckey para os grupos diagnósticos da EBADEP-A

Grupos	n	Alfa = 0,05			
		1	2	3	4
Não-Depressivos	40	29,93			
Universitários	1.304		45,73		
Psiquiátricos	75			58,15	
Depressivos	40				92,30
<i>p</i>		1,000	1,000	1,000	1,000

Destaca-se que esta característica discriminativa, na qual os participantes com diagnóstico de depressão apresentaram as maiores médias no instrumento, mesmo estando em um número consideravelmente reduzido de pessoas, caracteriza-se como evidência de validade de critério.

Discussão

Dentro da temática da construção de instrumentos de medida em depressão, podem ser encontrados diversos estudos de validação e de qualidades psicométricas para diferentes testes disponíveis do mercado (Almeida & Almeida, 1999; Batistoni, Neri & Cupertino, 2007; Gorestein & Andrade, 1998; Gouveia, Barbosa, Almeida & Gaião, 1995; Moreno & Moreno, 1998; Silveira & Jorge, 1998). Embora não possa ser considerada uma característica necessariamente negativa, tais pesquisas envolvem primordialmente o uso de métodos baseados na Teoria Clássica dos Testes – TCT, e discute-se a possibilidade de utilização de métodos estatísticos mais modernos, como a Teoria da Resposta ao Item – TRI, de modo a beneficiar o estudo e avaliação psicológica (Forkmann & cols., 2009).

O presente trabalho teve como objetivos realizar um estudo das qualidades psicométricas da Escala Baptista de Depressão (Versão Adulto) – EBADEP-A com base na Teoria da Resposta ao Item – TRI, mais especificamente, a teoria de um parâmetro de Rasch, e da Teoria Clássica dos Testes – TCT. Com base no modelo de Rasch, foram considerados tanto o nível de dificuldade dos itens apresentados pelo teste

quanto o nível de habilidade dos sujeitos a respondê-lo, além do estudo dos vieses de resposta em relação aos grupos de homens e mulheres, por meio do funcionamento diferencial do item (DIF), enquanto, com base na TCT, foi realizada análise dos grupos critério.

Inicialmente, sobre os parâmetros de ajuste dos itens e pessoas ao modelo, foi possível perceber que, no que tange aos itens, a maioria foi respondida dentro de um padrão esperado e adequado, embora tenham sido encontrados poucos itens com valores sem boa adequação. Sobre os sujeitos, a maioria apresentou um padrão de resposta dentro do esperado, e, com poucos valores de *infit* e *oufit* maiores que 2,0, os percentuais de desajuste foram considerados baixos. Tais resultados sugerem que, tanto no que se refere aos itens quanto às pessoas, não foi encontrada alta porcentagem de respostas acima ou abaixo do padrão esperado, podendo a escala ser considerada de bom ajuste (Linacre, 2002; Rueda, 2007; Sisto, 2006).

No que diz respeito à distribuição e agrupamento dos itens e pessoas, foi notado que a amostra tendeu a responder com maior frequência às opções de resposta relacionados à sintomatologia baixa de depressão, uma vez que a média dos itens foi maior que a média das pessoas. No entanto, vale ressaltar que trata-se de uma amostra heterogênea, composta tanto por sujeitos entre os quais o diagnóstico foi comprovado ou rejeitado, com base em entrevista diagnóstica estruturada quanto por pessoas as quais não passaram por avaliação diagnóstica, além de pacientes com outros transtornos psiquiátricos, os quais não necessariamente apresentaram depressão

primária. Neste caso, deve ser priorizada a capacidade discriminativa do teste, o que foi observado, já que não foram encontrados itens não adequados no mapa de itens.

Sobre a análise da estrutura interna do instrumento, a EBADEP-A foi considerada adequada, uma vez que as médias observadas dos níveis de teta aumentavam na medida em que as opções de resposta eram acrescidas, além de ter sido observada equivalência entre os limiares (Thresholds), características consideradas adequadas (Pasquali & Primi, 2003; Rueda, 2007). Observou-se também que, embora a escala não tenha apresentado delimitação perfeita entre os limiares, as categorias extremas obtiveram boa representação gráfica.

Quanto ao estudo da precisão, foi realizada comparação entre o alfa de Cronbach, índice de confiabilidade utilizado na TCT ($\alpha=0,95$), e a precisão encontrada pelo modelo de Rasch ($\alpha=0,92$). Observa-se neste ponto que o valor da precisão encontrada pela TRI é relativamente mais baixo que o índice de confiabilidade da TCT. Isso se deve ao fato de que, dentro do modelo de Rasch, é possível a existência de vários índices de precisões locais ao longo da escala enquanto, pelo modelo clássico, há um índice único. Dessa forma, a precisão, pelo modelo de Rasch é na verdade a média das precisões locais, o que faz com que a taxa de resíduos seja menor e, conseqüentemente, o valor seja menor (Linacre, 2002).

O estudo sobre o funcionamento diferencial dos itens indicou a presença de DIF em 17 itens (37,7%), dos quais 11 apresentaram viés de resposta para o grupo feminino e seis para os homens. Isso quer dizer que nesses itens, ou homens ou mulheres foram mais suscetíveis a assinalar a opção três na régua de círculos que compõem os itens, isto é, a opção mais negativa do par de frases apresentado. No caso das mulheres, citam-se os itens referentes aos descritores de humor deprimido, choro, sentimento de incapacidade, carência/dependência, sentimento de inadequação, autocrítica exacerbada, desesperança, esquiva de situações sociais, queda de produtividade, irritabilidade e fadiga/perda de energia. Para os homens, os vieses de resposta se relacionaram a choro, carência/dependência, sentimento de incapacidade, anedonia (incapacidade de experimentar prazer), humor deprimido, alterações de sono e perda de libido.

Algumas observações podem ser feitas neste momento. Primeiramente, mesmo apresentando vieses de resposta, os indicadores humor deprimido, choro, incapacidade e dependência foram considerados como tendo DIFs para ambos os grupos. No entanto, alguns sintomas foram mais típicos de um dos gêneros, o que parece ocorrer com frequência nesse tipo de análise,

como apontam Kessler (2003) e Piccinelli e Wilkinson (2000), mas os estudos não são conclusivos.

Os resultados da presente pesquisa foram parcialmente corroborados pelos encontrados por Wu (2010), com o BDI, em uma amostra de mais de 3.000 adolescentes, utilizando-se do DIF, em que as mulheres também endossaram mais sintomas, tais como autocrítica e irritabilidade, e os homens incapacidade, anedonia e problemas com o sono, além de outros sintomas diferentes dos encontrados nesta pesquisa. Uebelacker, Strong, Weinstock & Miller (2009), ao realizarem análise de DIF em mais de 13 mil participantes de uma pesquisa epidemiológica sobre consumo de álcool, utilizando os critérios do DSM-IV, encontraram, em um mesmo nível de severidade de depressão, mulheres endossando mais sintomas de apetite, distúrbios de peso e fadiga, sendo somente o último sintoma coerente com os atuais achados. Já, Forkmann e cols. (2010), ao utilizarem um instrumento construído com base na TRI, denominado DESC, em uma amostra representativa da Alemanha, não encontraram DIF entre os gêneros.

Os resultados encontrados por meio do estudo da TRI em relação à forma como a depressão é mensurada pela EBADEP-A, favorecendo ou não um grupo ou outro, podem ser considerados como uma evidência de validade do tipo construto (AERA, APA & NCME, 1999). Em relação aos grupos critério, analisados por meio da TCT, foi verificada a característica discriminativa da EBADEP-A por meio da análise de variância (ANOVA). O instrumento foi capaz de discriminar cada um dos grupos, ou seja, Não-depressivos, Universitários, Psiquiátricos e Depressivos, com médias crescentes nesta ordem, de forma a oferecer ao teste evidência de validade do tipo critério.

No presente estudo, foram cumpridos os objetivos, sendo apresentadas qualidades psicométricas para a EBADEP-A com base na Teoria da Resposta ao Item e da Teoria Clássica dos Testes, e espera-se ter contribuído neste quesito. Entretanto, também possui limitações, que devem ser consideradas. Primeiramente, a amostra foi composta por 74% de mulheres, o que pode ter influenciado em alguns dos resultados, e também houve número reduzido de pessoas com diagnóstico confirmado de depressão, embora a prova *Post hoc* de Tukey tenha tido força suficiente para discriminar tal grupo.

A identificação de itens com funcionamento diferencial é importante na construção de um instrumento na medida em que apresenta mais detalhadamente as diferenças entre grupos distintos em relação ao construto avaliado, e conseqüentemente subsidia as interpretações encontradas (Sisto, 2006).

Por essa razão, sugere-se a realização de outros estudos do gênero, com outras amostras, para fins comparativos.

Por último, iniciativas de construção de instrumentos brasileiros para avaliações psicopatológicas são de suma importância para o cenário da Avaliação Psicológica nacional, já que proporcionam o aumento de opções de uso para clínicos e pesquisadores. Como apontam Calil e Pires (1998), os diferentes instrumentos de depressão podem estar avaliando, de forma muito particular, o fenômeno em questão. Complementando essa afirmação, Smarr (2003) atenta para o fato de que os diferentes instrumentos podem variar enormemente no conteúdo avaliado, já que podem se basear em definições diferentes do construto ou terem maior ou menor ênfase em algumas dimensões da depressão (ex: sintomas cognitivos, afetivos, somáticos); número de itens (ou descritores); formato de resposta; método e tempo de administração e propriedades psicométricas.

Referências

- Almeida, O. P. & Almeida, S. A. (1999). Confiabilidade da versão brasileira da Escala de Depressão em Geriatria (GDS) versão reduzida. *Arquivos de Neuropsiquiatria*, 57(2B), 421-426.
- American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME) (1999). *Psychological testing and assessment. Standards for education and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychiatric Association (APA). (2002). *DSM-IV-TR – Manual diagnóstico e estatístico de transtornos mentais* (4ª ed). Porto Alegre: Artmed.
- Baptista, M. N. (2004). Depressão: histórico, epidemiologia, fatores de risco e modelos integrativos. Em: M. N. Baptista. *Suicídio e depressão: atualizações* (pp. 35-49). Rio de Janeiro: Guanabara Koogan.
- Baptista, M. N. (2011). *Manual técnico da Escala Baptista de Depressão em Adultos (EBADEP-A)*. Relatório Técnico. Programa de Pós Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba, SP
- Baptista, M. N., Baptista, A. S. D. & Oliveira, M. G. (2004). Depressão e gênero: por que as mulheres se deprimem mais que os homens? Em M. N. Baptista (Org.). *Suicídio e depressão: atualizações*. Rio de Janeiro: Guanabara Koogan.
- Baptista, M. N., Carneiro, A. M. & Sisto, F. F. Estudo Psicométrico de Escalas de Depressão (EDEP e BDI) e o Inventário de Percepção de Suporte Familiar (ISPF). *Revista Psicologia em Pesquisa*, 4(2), 8-22.
- Baptista, M. N., Souza, M. S. & Alves, G. A. S. (2008). Evidências de validade entre a Escala de Depressão (EDEP), o BDI e o Inventário de Percepção de Suporte Familiar (IPSF). *Psico-USF*, 13(2), 211-220.
- Baptista, M. N., Gomes, J. O., Carneiro, A. M. & Borine, B. (2011). Estrutura interna da Escala de Depressão com base na Teoria da Resposta ao Item. *Anais do V Congresso Brasileiro de Avaliação Psicológica*, Bento Gonçalves. Obtido na world wide web: http://www.ibapnet.org.br/congresso2011/trabalhos/Resumo_estruturainternaedep.pdf.
- Batistoní, S. A., Neri, A. L. & Cupertino, A. P. F. B. (2007). Validade da Escala de Depressão do Center for Epidemiological Studies entre idosos brasileiros. *Revista de Saúde Pública*, 41(4), 598-605.
- Beck, A. T., Rush, A. J., Shaw, B. F. & Emery, G. (1997). *Terapia da depressão*. Rio de Janeiro: Zahar.
- Calil, H. M. & Pires, M. L. N. (1998). Aspectos gerais das escalas de avaliação de depressão. *Revista de Psiquiatria Clínica*, 25(5), 240-244. Edição Especial.
- Castro, S. M. J., Trentini, C. & Riboldi, J. (2010). Teoria da Resposta ao Item aplicada ao Inventário de Depressão de Beck. *Revista Brasileira de Epidemiologia*, 13(3), 487-501.
- Cunha, J. (2001). *Manual em português das escalas Beck*. São Paulo: Casa do Psicólogo.
- Del Porto (2002). Depressão: conceito e diagnóstico. Em J. J. Mari e cols. (2002). *Atualização em Psiquiatria 1* (pp. 21-29). São Paulo: Casa do Psicólogo.
- Draba, R. E. (1977). *The identification and interpretation of Item Bias*. Rasch measurement transactions, MESA Memorandum n. 25. Obtido da world wide web: <http://www.rasch.org/rmt/rmt122m.htm>.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for psychologists*. Nova Jersey: Lawrence Erlbaum.
- Ferster, C. B., Culbertson, S. & Boren, C. P. (1977). *Princípios do comportamento*. (Trad: Maria Ignez Rocha e Silva, Maria Alice de Campos Rodrigues e Maria Benedita Lima Pardo). São Paulo: Hucitec.
- Forkman, T., Boecker, M., Wirtz, M., Eberle, N., Westhofen, M., Schauerterm P., Mischke, K., Kircher, T., Gauggel, S. & Norra, C. (2009). Development and validation of the Rasch-based depression screening (DESC) using Rasch analysis and structural equation modeling. *Journal of Behavior Therapy and Experimental Psychiatry*, 40, 468-478.

- Forkmann, T., Boecker, M., Wirtz, M., Glaesmer, H., Brähler, E., Norra, C. & Gauggel, S. (2010). Validation of the Rasch-based Depression Screening in a large scale German general population sample. *Health and Quality of Life Outcomes*, 8, 105.
- Gomes, J. O. & Baptista, M. N. (2010). Escala de Depressão (EDEP) e medidas de atenção dividida e sustentada em universitários. *Boletim de Psicologia*, 60(133), 191-204.
- Gorestein, C. & Andrade, L. (1998). Inventário de Depressão de Beck: propriedades psicométricas da versão em português. *Revista de Psiquiatria Clínica*, 25(5), 245-250.
- Goreinstein, C., Pang, W. Y., Argimoni, L. & Werlang, B. S. G. (2011). Manual do Inventário de Depressão de Beck – BDI-II. São Paulo: Casa do Psicólogo.
- Gouveia, V. V., Barbosa, G. A., Almeida, H. J. F. & Gaião, A. A. (1995). Inventário de Depressão Infantil – CDI: estudo de adaptação com escolares de João Pessoa. *Jornal Brasileiro de Psiquiatria*, 44(7), 345-349.
- Justo, L. P. & Calil, H. M. (2006). Depressão: o mesmo acometimento para homens e mulheres? *Revista de Psiquiatria Clínica*, 33(2), 74-79.
- Kessler, R. C. (2003). Epidemiology of women and depression. *Journal of Affective Disorders*, 74, 5-13.
- Linacre J. M. (2002). What do infit and outfit, mean-squared and standardized mean? *Rasch Measurement Transactions*, 16(2), 878.
- Moreno, R. A. & Moreno, D. H. (1998). Escalas de Depressão de Montgomery & Asberg (MADRS) e de Hamilton (HAM-D). *Revista de Psiquiatria Clínica*, 25(5), 262-272.
- National Institute of Mental Health (NIMH). (2000). *Depression research*. Office of communications and Public Liaison National Institute of Mental Health.
- Organização Mundial da Saúde (OMS). (1993). *Classificação dos transtornos mentais e do comportamento – CID-10: descrições e diretrizes diagnósticas*. Trad. Dorgival Caetano. (3º Vol., 10ª ed.). Porto Alegre: Artes Médicas.
- Organização Mundial de Saúde (OMS). (2001). *Relatório mundial da saúde: saúde mental: nova concepção, nova esperança*. Obtido em 20 de nov. 2005 da world wide web: http://www.who.int/whr/2001/en/whr01_po.pdf.
- Pasquali, L. & Primi, R. (2003). Fundamentos da Teoria da Resposta ao Item – TRI. *Avaliação Psicológica*, 2(2), 99-110.
- Piccinelli, M. & Wilkinson, G. (2000). Gender differences in depression. *British Journal of Psychiatry*, 177, 486-492.
- Prieto, G. & Muniz, J. (2000). Un modelo para evaluar la calidad de tests usados en España. *Papeles Del Psicólogo. Publicación del Consejo General de Colegios Oficiales de Psicólogos*, 77, 65-75.
- Rueda, F. J. M. (2007). O funcionamento diferencial do item no Teste Pictórico de Memória. *Revista Avaliação Psicológica*, 6(2), 229-237.
- Silveira, D. X., & Jorge, M. R. (1998). Propriedades psicométricas da Escala de Rastreamento Populacional para Depressão CES-D em populações clínica e não-clínica de adolescentes e adultos jovens. *Revista de Psiquiatria Clínica*, 25(5), 251-261.
- Sisto, F. F. (2006). O funcionamento diferencial dos itens. *Psico-USF*, 11(1), 35-43.
- Smar, K. L. (2003). Measures of depression and depressive symptoms. *Arthritis & Rheumatism*, 49(5S): 134-146.
- Uebelacker, L. A., Strong, D., Weinstock, L. M. & Miller, I. W. (2009). Use of Item Response Theory to understand differential functioning of DSM-IV major depression symptoms by race, ethnicity and gender. *Psychological Medicine*, 39(4), 591-601.
- Wu, P. (2010). Differential functioning of the chinese version of Beck Depression Inventory-II in adolescent gender groups: use of a multiple-group mean and covariance structure model. *Social Indicators Research*, 96(3), 535-550.

Recebido em 06/09/2010
Reformulado em 10/02/2011
Aprovado em 13/04/2011

Sobre os autores:

Makilim Nunes Baptista é doutor pelo Depto. de Psiquiatria e Psicologia Médica da Escola Paulista de Medicina, docente do Programa de Pós-Graduação *Stricto Sensu* em Psicologia da Universidade São Francisco, Itatiba, São Paulo e bolsista de Produtividade em Pesquisa do CNPq – Nível 1D.

Juliana Oliveira Gomes é psicóloga, mestre em Psicologia pela Universidade São Francisco, discente do curso de Pós-Graduação *Stricto Sensu* – Doutorado em Psicologia da Universidade São Francisco, Itatiba, SP e bolsista CAPES.