

Teoria da resposta ao item aplicada ao Inventário de Depressão Beck

Item response theory applied to the Beck Depression Inventory

Resumo

O Inventário de Depressão Beck (BDI), uma escala que mede o traço latente de intensidade de sintomas depressivos, pode ser avaliado através da Teoria da Resposta ao Item (TRI). Este estudo utilizou o modelo TRI de Resposta Gradual na avaliação da intensidade de sintomas depressivos de 4.025 indivíduos que responderam ao BDI, de modo a explorar eficientemente a informação disponível nos diferentes aspectos possibilitados pelo uso desta metodologia. O ajuste foi efetuado no software PARSCALE. Foram identificados 13 itens do BDI nos quais pelo menos uma categoria de resposta não tinha chance maior que as demais de ser escolhida, de modo que estes itens tiveram de ser recategorizados. Os itens com maior capacidade de discriminação são relativos à tristeza, pessimismo, sentimento de fracasso, insatisfação, auto-aversão, indecisão e dificuldade para trabalhar. Os itens mais graves são aqueles relacionados com perda de peso, retraimento social e idéias suicidas. O grupo dos 202 indivíduos com as maiores intensidades de sintomas depressivos foi composto por 74% de mulheres, e praticamente 84% possuíam diagnóstico de algum transtorno psiquiátrico. Os resultados evidenciam alguns dos inúmeros ganhos advindos da utilização da TRI na análise de traços latentes.

Palavras-chave: Teoria da Resposta ao Item. Traço latente. Intensidade de sintomas depressivos. Inventário de Depressão Beck.

Stela Maris de Jezus Castro

Clarissa Trentini

João Riboldi

Universidade Federal do Rio Grande do Sul

O estudo foi aprovado pelo Comitê de Ética em Pesquisa da UFRGS na reunião nº 37, ata nº 117, de 30 de outubro de 2008.

Correspondência: Stela Maris de Jezus Castro. Rua João Mendes Ouriques, 650, Ipanema, Porto Alegre, RS - CEP: 91760-450. E-mail: stela.castro@ufrgs.br.

Abstract

The Beck Depression Inventory (BDI), a scale that measures the latent trait intensity of depression symptoms, can be assessed by the Item Response Theory (IRT). This study used the Graded-Response model (GRM) to assess the intensity of depressive symptoms in 4,025 individuals who responded to the BDI, in order to efficiently use the information available on different aspects enabled by the use of this methodology. The fit of this model was done in PARSCALE software. We identified 13 items of the BDI in which at least one response category was not more likely than others to be chosen, so that these items had to be categorized again. The items with greater power of discrimination were sadness, pessimism, feeling of failure, dissatisfaction, self-hatred, indecision, and difficulty of work. The most serious items were weight loss, suicidal ideas, and social withdrawal. The group of 202 individuals with the highest levels of depressive symptoms was comprised by 74% of women and almost 84% had a diagnosis of a psychiatric disorder. The results show gains resulting from use of IRT in the analysis of latent traits.

Keywords: Item Response Theory. Latent trait. Intensity of depressive symptoms. Beck Depression Inventory.

Introdução

Variáveis latentes, referidas por traços latentes, são entidades não observáveis e devem ser inferidas a partir da observação de variáveis secundárias que tenham relação com ela. Para isso são em geral utilizados instrumentos de medidas (escalas) compostos por um conjunto de itens, cujas respostas são categorias (ordenadas ou não) que servem para estimar as variáveis secundárias, podendo, assim, estimar os traços latentes dos indivíduos.

O Inventário de Depressão Beck (BDI)¹ é um exemplo deste tipo de instrumento de medida, pois é composto por um grupo de itens que pretende medir o traço latente intensidade de sintomas depressivos, o qual é de extrema relevância para se verificar o estágio da depressão e avaliar o seu desfecho. A depressão é um construto psicológico, biológico e social de alta prevalência. Estudos epidemiológicos apresentam a prevalência anual da depressão na população geral, variando entre 3% e 11%^{2,3}, e a prevalência para a vida inteira em 16,2%³. No Brasil, os estudos apontam a prevalência da depressão para a vida inteira entre 2,8% e 19,2%^{4,5}.

Até recentemente, a modelagem estatística mais utilizada na estimação do traço latente era a Teoria Clássica de Teste (TCT)⁶, que utiliza o escore total como estimativa para o traço latente do indivíduo. Uma revisão desta metodologia pode ser encontrada no trabalho de DeVellis⁷. Apesar de a TCT ter tido grande importância e ter sido extremamente útil, diversos autores têm citado várias limitações do método⁸, as quais estão solucionadas na teoria de medida chamada Teoria da Resposta ao Item⁹.

A Teoria da Resposta ao Item (TRI), já bastante utilizada na área psiquiátrica¹⁰⁻¹², compreende um grupo de modelos lineares generalizados e procedimentos estatísticos associados que descrevem a relação entre as respostas aos itens (amostra de comportamento) e um traço latente. O alvo de um modelo TRI é o encontro de um indivíduo com um item. O padrão de respostas do

indivíduo a um particular grupo de itens fornece a base para a estimativa do traço latente. Nos modelos da TRI, os parâmetros dos itens e os níveis do traço latente do indivíduo são independentes; eles são expressos no nível da resposta observada do item; a contribuição de cada item para a escala pode ser determinada através da informação TRI; estão disponíveis métodos poderosos para detectar funcionamento diferencial de item (DIF) ou viés do item entre populações ou subgrupos; e os escores dos indivíduos podem ser equacionados (comparados) mesmo quando eles respondem a itens diferentes¹³.

A TRI é de extrema relevância na análise de traços latentes, pois propicia um maior aproveitamento da informação, trabalhando não somente com a classificação dos indivíduos quanto ao traço latente, como também trazendo informações sobre os instrumentos de medida como um todo e, principalmente, item a item. Sendo assim, a TRI é um procedimento mais sofisticado que aproveita melhor a informação disponível nos itens, podendo melhorar sensivelmente a medida do traço latente, pois este é calculado considerando diferentes pesos para os itens de acordo com sua importância em relação ao traço que está sendo medido.

Os modelos da TRI podem ser divididos em modelos cumulativos e modelos de desdobramento. Os modelos da TRI cumulativos podem ser classificados de acordo com sua dimensionalidade (unidimensionais ou multidimensionais). Os modelos da TRI unidimensionais descrevem a relação entre as respostas observadas ao item e um traço latente, usualmente simbolizado por θ , que forma a base destas respostas. Eles são apropriados para dados nos quais um único fator comum está sendo avaliado pelos itens. Entre os modelos unidimensionais encontram-se os modelos para dados dicotômicos (que apresentam o sintoma depressivo ou não apresentam o sintoma depressivo, o sucesso ou o fracasso) e os modelos para dados politômicos (itens com mais de duas categorias de resposta,

conforme os itens do BDI). Estes modelos também são diferenciados de acordo com seu número de parâmetros. Eles podem ter 1, 2 ou 3 parâmetros, sendo: um parâmetro de dificuldade do item (gravidade do sintoma depressivo, no caso do BDI), um parâmetro de discriminação do item, e um parâmetro que representa a probabilidade de os indivíduos apresentarem o sintoma depressivo descrito pelo item, mesmo quando têm baixo nível do traço latente.

Todos os itens do BDI têm quatro categorias de resposta e, por isso, um modelo unidimensional para resposta politômica se mostra mais adequado. Além disso, dificilmente os 21 itens do BDI discriminam a população quanto ao traço latente com a mesma intensidade; portanto, um modelo que permita que cada item tenha um valor diferente para a discriminação, isto é, contenha o parâmetro de discriminação do item, parece se adequar melhor. Diante disso, o presente estudo tem por objetivo caracterizar o modelo da TRI de Resposta Gradual de Samejima (1969)¹⁴, aplicando-o na avaliação da intensidade de sintomas depressivos de indivíduos que responderam ao BDI, de modo a explorar a informação disponível nos diferentes aspectos possibilitados pelo uso desta metodologia.

Método

Fonte de dados

Os indivíduos são provenientes de um estudo transversal conduzido para realizar a adaptação, normatização e validação para o português das Escalas Beck, em um estudo conduzido pela Dra. Jurema Alcides Cunha e publicado em 2001¹⁵. Os 4.025 sujeitos do estudo estão distribuídos em três grandes grupos: grupo 1 - pacientes psiquiátricos (n = 1.138); grupo 2 - pacientes de clínica médica (n = 490); grupo 3 - amostra não-clínica, constituída por grupos da população geral (n = 2.397).

Todos os participantes responderam ao BDI, composto por uma escala de auto-relato de 21 itens, cada um com quatro afir-

mativas, subentendendo graus crescentes de intensidade de sintomas depressivos¹⁵, e escores variando de 0 a 3. Os itens do BDI foram selecionados com base em observações e relatos de sintomas e atitudes mais frequentes em pacientes psiquiátricos, com transtornos depressivos¹⁵: Tristeza; Pessimismo; Sentimento de fracasso; Insatisfação; Culpa; Punição; Auto-aversão; Auto-acusações; Idéias suicidas; Choro; Irritabilidade; Retraimento social; Indecisão; Mudança na auto-imagem; Dificuldade para trabalhar; Insônia; Fatigabilidade; Perda de apetite; Perda de peso; Preocupações somáticas; Perda da libido. Neste estudo¹⁵ foi utilizado o escore total como estimativa da intensidade de sintomas depressivos, o qual varia entre 0 e 63. Como resultado, foram definidos os seguintes intervalos para classificar os indivíduos quanto à intensidade de sintomas depressivos: mínima (0 – 11), leve (12 – 19), moderada (20 – 35) e grave (36 – 63). Como os escores da TRI (estimativas de intensidade de sintomas depressivos, segundo o modelo de resposta gradual) têm média zero e desvio padrão um (isto é, estão na escala (0,1)), pode-se fazer uma transformação linear propiciando uma mudança de escala, de modo a se poder relacionar o escore TRI com o escore total, para os quais já existe uma classificação dos indivíduos quanto à intensidade dos sintomas depressivos.

Teoria da Resposta ao Item^{16,17}

Um modelo TRI politômico adequado a dados oriundos do BDI é o Modelo de Resposta Gradual (*Graded-Response Model* - GRM), proposto por Samejima em 1969¹⁴, pois este pressupõe que as categorias de resposta de um item podem ser ordenadas entre si. Para ajustar o GRM, os itens do instrumento não precisam ter o mesmo número de categorias de resposta.

Considerando que as categorias de resposta de um item estejam arranjadas com seus escores em ordem crescente, e denotados por $k = 0, 1, 2, \dots, m_i$; onde $(m_i + 1)$ é o número de categorias do i -ésimo item,

a probabilidade de um indivíduo j escolher uma determinada categoria ou outra de ordem mais alta do item i pode ser dada por:

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}},$$

com $i=1, 2, \dots, I$, $j=1, 2, \dots, n$, $k=1, 2, \dots, m_i$ e $b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i}$, onde:

- θ_j representa a intensidade de sintomas depressivos (traço latente) do j -ésimo indivíduo;
- a_i é o parâmetro de inclinação comum a todas as categorias do item i ;
- $b_{i,k}$ é o parâmetro de posição (um ponto do *continuum* traço latente) da k -ésima categoria do item i , isto é, cada $b_{i,k}$ é o ponto de interseção entre as categorias de resposta ordenadas, e representa o nível de intensidade de sintomas depressivos necessário para a escolha da categoria de resposta k , ou acima de k , com probabilidade igual a 0,50;
- D é um fator de escala, que pode ser igual a 1 ou 1,7, este último quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

A probabilidade de um indivíduo j responder a categoria k no item i é dada pela diferença:

$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j)$$

Por definição, $P_{i,0}^+(\theta_j) = 1$ e $P_{i,m_i+1}^+(\theta_j) = 0$. Logo,

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}} \quad (1)$$

As curvas geradas pela expressão (1) são chamadas Curvas de Categoria de Resposta. São estas curvas que mostram a relação entre as probabilidades de respostas nas categorias de cada item e o nível do traço latente, isto é, é possível identificar qual a categoria de resposta com maior probabilidade de ser escolhida para cada nível do traço latente.

O ajuste do modelo GRM foi realizado utilizando o software PARSCALE¹⁸, versão

4.1. Neste, a estimativa dos parâmetros do modelo GRM é feita de modo que os $b_{i,k}$ entre categorias são particionados em dois termos: um parâmetro de posição (b_i) para cada item, e um grupo de parâmetros $c_{i,k}$ para cada item, ou seja, $b_{i,k} = b_i - c_{i,k}$. No caso do traço latente Intensidade de Sintomas Depressivos, os parâmetros de posição (b_i) podem ser interpretados como uma medida de gravidade do sintoma avaliado por um determinado item¹⁹ e os parâmetros $c_{i,k}$ representam as distâncias entre os pontos de interseção entre as curvas de categorias de resposta para cada item.

A escala das estimativas de intensidade de sintomas depressivos (nomeadas aqui de escores da TRI) é arbitrária, onde o que importa são as relações de ordem existentes entre seus pontos e não necessariamente sua magnitude. Assim, pode assumir qualquer valor real entre $-\infty$ e $+\infty$, havendo necessidade de se definir uma origem e uma unidade de medida para a mesma. Neste estudo, foi definido que os escores TRI tivessem média zero e desvio padrão um (escala (0,1)). O parâmetro de gravidade do sintoma (b_i) é medido na mesma unidade dos escores TRI, sendo assim comparável a eles.

Na Figura 1a está a representação gráfica deste modelo para o item 1 (tristeza) do BDI, com 4 categorias de resposta medindo a intensidade de sintomas depressivos com as seguintes estimativas dos parâmetros: $\hat{a}_1 = 1,478$, $\hat{b}_{1,1} = 0,153$, $\hat{b}_{1,2} = 1,280$ e $\hat{b}_{1,3} = 1,897$, onde as categorias de resposta são:

- 1 Não me sinto triste.
- 2 Eu me sinto triste.
- 3 Estou sempre triste e não consigo sair disto.
- 4 Estou tão triste ou infeliz que não consigo suportar.

Pode-se observar nesta figura que os indivíduos com intensidade de sintomas depressivos até 0,153 têm a maior probabilidade de responder à categoria 1 (curva 1), indivíduos com intensidade de sintomas depressivos entre 0,153 e 1,280 têm maior probabilidade de responder à categoria 2

(curva 2) e os indivíduos com intensidade de sintomas depressivos entre 1,280 e 1,897 são mais prováveis de escolherem a categoria 3 (curva 3). Por último, a maior probabilidade de responder a categoria 4 (curva 4) é dos indivíduos com intensidade de sintomas depressivos a partir de 1,897.

A partir dos modelos da TRI também são produzidas as Curvas de Informação do Item, que são bastante utilizadas em conjunto com as Curvas de Categoria de Resposta, e a Curva de Informação do Teste. As Curvas de Informação do Item (construídas a partir das funções de informações do item) permitem analisar quanto um item contém de informação para a medida do traço latente; isto é, aqui elas indicam a quantidade de informação psicométrica com que um determinado sintoma depressivo contribui para a medida de intensidade de sintomas depressivos e, mais ainda, em que intervalo desta medida este sintoma é mais informativo. É com esta característica que se pode avaliar quais sintomas depressivos discriminam melhor a população quanto à intensidade de sintomas depressivos¹⁷. Nos modelos da TRI politômicos, a quantidade de informação que um item fornece depende tanto da magnitude do parâmetro de inclinação (a_i) como da distribuição dos pontos de interseção entre as categorias de resposta $b_{i,k}$ ao longo do *continuum* traço latente. Por exemplo, a curva de informação do item 1 do BDI (Figura 1b) mostra que a região de maior concentração da informação está entre, aproximadamente, 0,7 e 1,7 na escala de intensidade de sintomas depressivos. Isto significa que o sintoma depressivo tristeza discrimina melhor os indivíduos que têm sua intensidade de sintomas depressivos neste intervalo.

A Curva de Informação do Teste é uma representação gráfica da função de informação do teste. Esta é uma função aditiva do grupo de itens que compõe o teste (por exemplo, o BDI), de modo que resume a contribuição de cada item deste para a informação total. A quantidade total de informação fornecida por um grupo de itens para cada nível do traço latente está

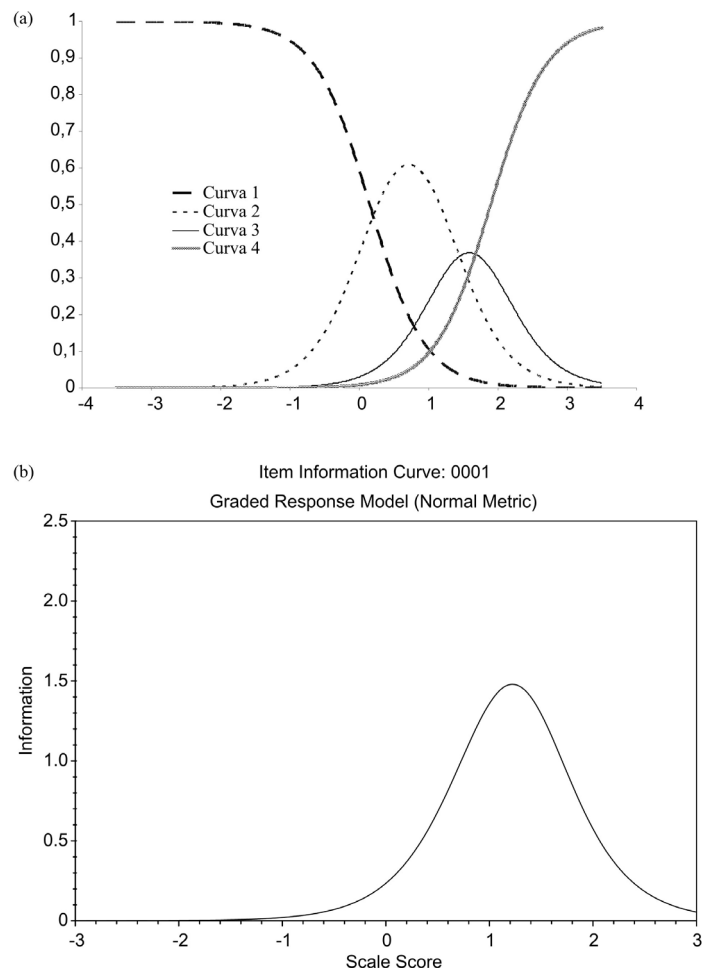


Figura 1 - Representação gráfica do item 1 (tristeza) do BDI segundo o modelo de resposta gradual: (a) Curva de Categoria de Resposta; (b) Curva de Informação do item.

Figure 1 - Graphic representation of item 1 (sadness) of the BDI on the graded response model: (a) Category of Response Curve; (b) Item Information Curve.

inversamente relacionada ao erro padrão associado com a estimativa do mesmo. A função de informação do teste fornece uma alternativa viável para os conceitos de confiabilidade e erro padrão da Teoria Clássica do Teste. Através da Curva de Informação do Teste pode-se verificar para qual intervalo do traço latente o teste funciona melhor.

Suposições do modelo

A adequabilidade do uso do modelo GRM requer que duas suposições sejam satisfeitas: independência local (significa que, controlando para os níveis do traço latente, a resposta para qualquer item não está relacionada à resposta do item anterior)

e unidimensionalidade (todos os itens do instrumento apresentam o mesmo traço latente). Estas duas suposições estão relacionadas, implicando que uma vez que um instrumento é considerado unidimensional, ele apresenta independência local; isto é, se a suposição de unidimensionalidade está satisfeita, então somente um único traço latente está influenciando as respostas dos itens e a independência local é obtida^{16,20}. Existe evidência de que a suposição de unidimensionalidade pode ser relaxada, bastando que a mesma seja suficiente²¹⁻²⁴; isto é, bastando que um fator seja preponderante (proporção de explicação do primeiro fator em uma análise de componentes principais no mínimo igual a 20%²¹) para que os

modelos TRI possam ser utilizados. Em vista disso, foi avaliada a unidimensionalidade do BDI através do procedimento conhecido como Análise Paralela^{20,25-28}, viabilizada através do uso de uma macro²⁹ do SAS versão 9.1.3 (SAS Institute, Cary, NC, USA). Este procedimento envolve a comparação dos autovalores de uma análise de componentes principais feita com os dados reais com uma estatística resumo dos autovalores de amostras de dados simulados com o mesmo número de observações e variáveis dos dados reais (neste caso, 4.025 observações e 21 variáveis). As amostras simuladas são não correlacionadas e são geradas através do método de Monte Carlo (foram geradas 5.000 amostras e a estatística resumo utilizada foi a mediana).

Resultados

As características demográficas da amostra encontram-se na Tabela 1. Os indivíduos estão divididos quase que igualmente entre homens e mulheres, com pequena vantagem para as últimas. A quase totali-

dade dos indivíduos declarou-se branca e pouco mais da metade é solteira. Quanto à escolaridade, eles se distribuem quase que igualmente nas categorias até ensino médio completo, sendo menos frequentes aqueles com ensino superior completo. A idade média dos indivíduos é de 32 anos (desvio padrão de 15,1 anos), sendo que o grupo da clínica médica é, em média, o mais velho.

O resultado da análise paralela mostrou que a suposição de unidimensionalidade pode ser considerada suficiente^{19,30}, pois foi encontrado um fator preponderante com 38,7% de explicação da variação total.

O ajuste do modelo de Resposta Gradual mostrou, a partir das curvas produzidas pela expressão (1), que, para treze dos vinte e um itens do BDI, pelo menos uma das categorias de resposta não tem probabilidade maior do que as outras de ser respondida para nenhum nível de intensidade de sintomas depressivos, como pode ser visualizado na Figura 2a para o item relativo ao pessimismo (item 2). Uma das possíveis explicações para este fato é que os itens relativos a pessimismo, insatisfações, culpa,

Tabela 1 - Descrição da amostra de acordo com grupo de origem.

Table 1 - Description of the sample according to group of origin.

Características sociodemográficas**	Psiquiátrica (%)	Clínica (%)	Não-clínica (%)	Total
Sexo (n=4025)				
Masculino	43,3	36,7	48,2	45,4
Feminino	56,7	63,3	51,8	54,6
Cor de pele (n=3767)				
Branco	88,6	93,3	92,3	91,4
Não branco	11,4	6,7	7,7	8,6
Escolaridade (n=3816)				
Menos de 5 anos	28,3	31,8	27,5	28,2
Ensino Fundamental Completo	25,6	19,8	28,3	26,5
Ensino Médio Completo	33,9	27,2	25,2	28,0
Ensino Superior Completo	12,2	21,2	19,0	17,3
Estado Civil (n=3898)				
Solteiro	36,7	33,3	77,7	60,3
Casado	42,4	50,1	15,3	27,5
Desquitado, divorciado, viúvo	20,9	16,6	7,0	12,2
Idade (n=4014)				
Média (DP*)	38,4 (12,3)	44,1 (14,4)	26,4 (13,6)	32,0 (15,1)

* Desvio padrão / Standard deviation

** O nº de indivíduos varia de acordo com as características sociodemográficas devido a ocorrência de valores faltantes.
The number of individuals varies according to socio-demographic characteristics due to the occurrence of missing values.

punição, idéias suicidas, choro, irritabilidade, mudança na auto-imagem, insônia, perda de apetite, perda de peso, preocupações somáticas e perda da libido possam estar apresentando problemas de entendimento da escala, Em vista disto, foi realizada uma recategorização para estes itens a partir da junção de categorias adjacentes àquela que apresentou problemas de entendimento,

produzindo assim uma escala contendo itens com diferente número de categorias de resposta. Ajustou-se novamente o modelo de Resposta Gradual, obtendo-se deste modo curvas de categoria de resposta que mostram que todas as categorias de resposta têm chance de serem escolhidas para algum intervalo no *continuum* traço latente (Figura 2b). Considerando o ajuste

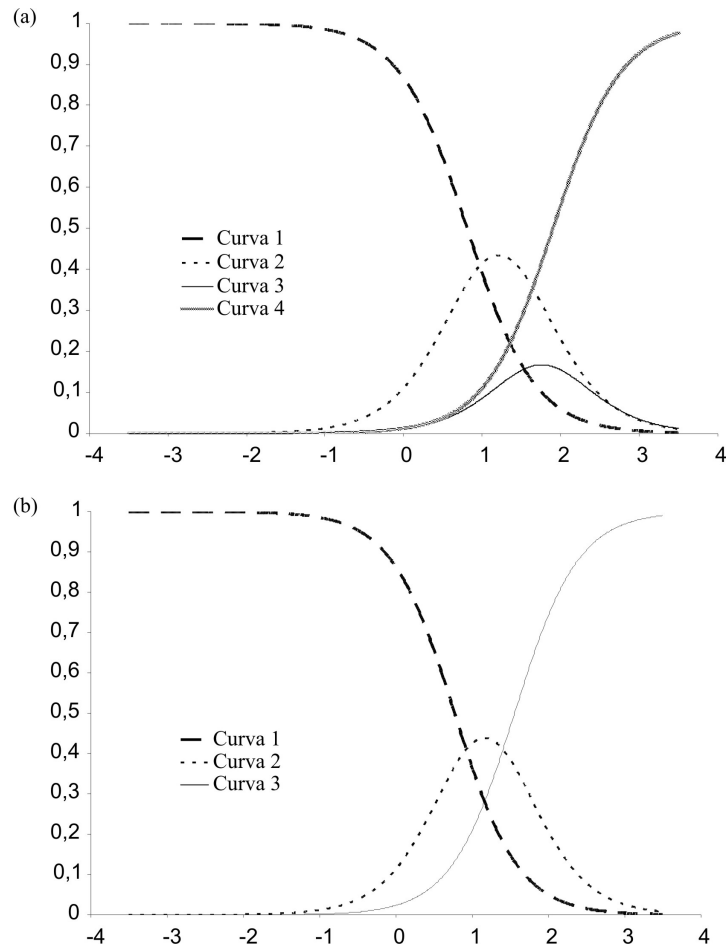


Figura 2 - Curvas de categoria de resposta para o item 2 do BDI, cujos escores significam: (a) Curva 1: Não estou especialmente desanimado quanto ao futuro. Curva 2: Eu me sinto desanimado quanto ao futuro. Curva 3: Acho que nada tenho a esperar. Curva 4: Acho o futuro sem esperança e tenho a impressão de que as coisas não podem melhorar. (b) Curva 1: Não estou especialmente desanimado quanto ao futuro. Curva 2: Eu me sinto desanimado quanto ao futuro. Curva 3: Acho que nada tenho a esperar. Acho o futuro sem esperança e tenho a impressão de que as coisas não podem melhorar.

Figure 2 - Category of Response Curve of item 2 of the BDI, mean scores: (a) Curve 1: I am not particularly discouraged about the future. Curve 2: I feel discouraged about the future. Curve 3: I think I have nothing to expect. Curve 4: I think there is no hope in the future and I have the impression that things can not improve. (b) Curve 1: I am not particularly discouraged about the future. Curve 2: I feel discouraged about the future. Curve 3: I think I have nothing to expect OR I think there is no hope for the future and I have the impression that things can not improve.

do modelo de Resposta Gradual para o BDI com itens recategorizados, as estimativas dos parâmetros do modelo estão apresentadas na Tabela 2.

A partir da Curvas de Informação do Item, que são fortemente influenciadas pelo parâmetro de inclinação (a_i - Tabela 2), observa-se que, considerando como ponto de corte o número um^{16,19} nas estimativas deste parâmetro para identificar os itens como tendo boa discriminação ($a_i > 1$), os itens relativos à tristeza, pessimismo, sentimento de fracasso, insatisfação, auto-aversão, indecisão e dificuldade de trabalhar (itens 1, 2, 3, 4, 7, 13 e 15, respectivamente) são aqueles que mais informação fornecem à medida de intensidade de sintomas depressivos, podendo, deste modo, melhor discriminar a população quanto a este traço latente. Cabe ressaltar

que os itens relacionados à culpa (item 5) e idéias suicidas (item 9) também têm suas estimativas do parâmetro de inclinação maiores do que 1 ($\hat{a}_5 = 1,172$ e $\hat{a}_9 = 1,078$ - Tabela 2); no entanto, sua curva de informação do item revela que estes itens são pobres em relação à capacidade de discriminação. O item relativo à irritabilidade (item 11) é o que menos contribui com informação para a medida de intensidade de sintomas depressivos. Traçando uma linha horizontal no ponto igual a 1 da função de informação do item, observa-se que o item relativo a sentimento de fracasso tem maior poder de discriminação da população quanto à intensidade de sintomas depressivos quando o escore da TRI está no intervalo de [0,7;2], o que equivale aproximadamente ao intervalo [21;35] na escala do escore total. O item

Tabela 2 - Estimativa dos parâmetros do modelo de Resposta Gradual.

Table 2 - Estimation of the parameters of the graded response model.

ITEM		a_i (EP)	b_i (EP)	$b_{i,1}$ (EP)	$b_{i,2}$ (EP)	$b_{i,3}$ (EP)
1	Tristeza	1,478 (0,036)	1,110 (0,023)	0,153 (0,028)	1,280 (0,031)	1,897 (0,039)
2	Pessimismo	1,408 (0,041)	1,163 (0,028)	0,770 (0,033)	1,556 (0,038)	-
3	Sentimento de Fracasso	1,684 (0,052)	1,359 (0,025)	0,720 (0,030)	1,339 (0,033)	2,020 (0,041)
4	Insatisfações	1,574 (0,042)	0,667 (0,023)	0,102 (0,028)	1,232 (0,031)	-
5	Culpa	1,172 (0,032)	1,271 (0,028)	0,568 (0,035)	1,974 (0,045)	-
6	Punição	0,850 (0,035)	0,671 (0,036)	0,671 (0,036)	-	-
7	Auto-aversão	1,393 (0,038)	1,596 (0,026)	0,529 (0,032)	1,826 (0,040)	2,433 (0,053)
8	Auto-acusações	0,702 (0,014)	0,965 (0,031)	-1,048 (0,045)	1,150 (0,045)	2,793 (0,065)
9	Idéias Suicidas	1,078 (0,036)	1,726 (0,036)	1,152 (0,044)	2,300 (0,056)	-
10	Choro	0,792 (0,021)	1,067 (0,035)	0,447 (0,044)	1,687 (0,050)	-
11	Irritabilidade	0,326 (0,007)	0,734 (0,060)	-1,011 (0,084)	2,479 (0,092)	-
12	Retraimento Social	0,988 (0,027)	1,767 (0,030)	0,650 (0,038)	1,790 (0,045)	2,861 (0,068)
13	Indecisão	1,185 (0,029)	1,022 (0,025)	0,120 (0,032)	0,814 (0,033)	2,132 (0,046)
14	Mudança na auto-imagem	0,950 (0,025)	1,336 (0,031)	0,549 (0,039)	2,123 (0,051)	-
15	Dificuldade de trabalhar	1,179 (0,027)	1,300 (0,026)	0,287 (0,032)	1,288 (0,036)	2,324 (0,051)
16	Insônia	0,832 (0,021)	0,827 (0,030)	0,065 (0,039)	1,589 (0,045)	-
17	Fatigabilidade	0,955 (0,022)	1,226 (0,028)	-0,254 (0,036)	1,535 (0,041)	2,397 (0,054)
18	Perda de apetite	0,711 (0,032)	0,961 (0,048)	0,961 (0,048)	-	-
19	Perda de peso	0,547 (0,033)	1,995 (0,104)	1,995 (0,104)	-	-
20	Preocupações Somáticas	0,745 (0,020)	1,087 (0,033)	0,318 (0,043)	1,856 (0,051)	-
21	Perda da libido	0,856 (0,024)	1,400 (0,035)	0,733 (0,044)	2,067 (0,044)	-

EP erro padrão da estimativa; (standard error of estimate)

a_i parâmetro de inclinação comum a todas as categorias do item i / parameter of slope common to all categories of item i

b_i parâmetro de gravidade do sintoma avaliado pelo item i / measure of severity of symptoms assessed by item i

$b_{i,1}$ ponto de interseção entre as categorias de resposta 1 e 2 / point of intersection between response categories 1 and 2

$b_{i,2}$ ponto de interseção entre as categorias de resposta 2 e 3 / point of intersection between response categories 2 and 3

$b_{i,3}$ ponto de interseção entre as categorias de resposta 3 e 4 / point of intersection between response categories 3 and 4

relativo à tristeza discrimina melhor a população quando o escore da TRI vai de 0,7 a 1,7, o equivalente a ir de 21 a 32 no escore total. Para trazer os escores da TRI para a escala do escore total, os primeiros foram multiplicados pelo desvio padrão dos escores totais e depois somados com a sua média.

As estimativas do parâmetro de gravidade do sintoma depressivo (b_i - Tabela 2) apontam que os itens sobre perda de peso, retraimento social e idéias suicidas são os considerados mais graves para a medida de intensidade de sintomas depressivos.

Os níveis de intensidade de sintomas depressivos estimados sob o modelo de Resposta Gradual (escores da TRI) estão na mesma escala da gravidade estimada para os itens do BDI; logo, são comparáveis aos mesmos. O percentil 95 do nível de intensidade de sintomas depressivos é igual a 1,6, o equivalente a 31 no escore total. Existem 202 indivíduos que pertencem ao grupo destes 5% com maiores escores da TRI. A Tabela 3 traça um perfil deste grupo: praticamente 84% destes indivíduos estão no grupo psiquiátrico, têm em média 39 anos de idade, 74% são mulheres, aproximadamente me-

tade estudou menos de cinco anos, e quase 40% são casados.

A Curva de Informação do Teste mostra que o BDI funciona melhor para indivíduos com intensidade de sintomas depressivos variando de 0,8 a 2,4, o equivalente ao intervalo que vai de 22 a 40 na escala do escore total.

Discussão

Os 21 itens do BDI são representativos dos sintomas de maior prevalência na ocorrência de depressão¹⁵. A avaliação destes itens quanto à sua quantidade de informação psicométrica e ao seu grau de gravidade do traço latente que está sendo medido, é um ganho importante dos modelos da TRI em relação à TCT na análise de dados desta natureza, pois possibilita identificar qual sintoma depressivo tem maior ou menor peso na avaliação da intensidade de sintomas depressivos e, mais importante ainda, considera estas características no cálculo deste traço latente para cada respondente do instrumento, diferentemente do cálculo do escore total (proveniente da TCT) onde todos os itens

Tabela 3 - Descrição dos indivíduos com estimativa de intensidade de sintomas depressivos acima do percentil 95 ($\hat{\theta} = 1,6^{**}$).

Table 3 - Description of subjects with estimated intensity of depressive symptoms above the 95th percentile ($\hat{\theta} = 1,6^{**}$).

Características sócio-demográficas**	Psiquiátrica n=169	Clínica n=14	Não-clínica n=19	Total n=202
Sexo (n=202)				
Masculino	34	9	9	25,7%
Feminino	135	5	10	74,3%
Escolaridade (n=196)				
Menos de 5 anos	71	10	8	45,4%
Ensino Fundamental Completo	47	2	4	27,1%
Ensino Médio Completo	34	1	6	20,9%
Ensino Superior Completo	12	0	1	6,6%
Estado Civil (n=201)				
Solteiro	50	8	13	35,4%
Casado	75	4	1	39,8%
Desquitado, divorciado, viúvo	43	2	5	24,8%
Idade (n=201)				
Média (DP*)	40,1 (12,7)	39,8 (18,5)	33,7 (19,7)	39,5 (13,9)

* Desvio Padrão (Standard deviation)

** Na escala do escore total este percentil 95 equivale a 31. (On the total score scale this 95th percentile is equal to 31)

entram com o mesmo peso¹³.

Outro ganho substancial dos modelos da TRI é a geração das curvas de categoria de resposta, as quais permitem que se observe a relação entre o nível de intensidade de sintomas depressivos e a probabilidade de se escolher uma determinada categoria, pois através delas pode-se constatar se algum item está com categorias mal dimensionadas, como foi o caso neste estudo. Este achado sugere que os indivíduos que responderam o BDI não conseguem distinguir algumas afirmações nas categorias de resposta de determinados itens, mostrando a necessidade de se repensar o dimensionamento da escala. Aqui, este problema apareceu em treze itens, entre os quais dois que se referem a sintomas que carregam uma grande quantidade de informação psicométrica sobre a intensidade de sintomas depressivos: pessimismo e insatisfações. Uma solução possível foi a que se fez uso neste estudo, ou seja, juntar categorias de resposta adjacentes àquela com problema, pois é provável que os indivíduos não tenham distinguido os conteúdos das afirmações contidas nestas categorias.

A definição de quais itens carregam maior quantidade de informação psicométrica sobre a intensidade de sintomas depressivos, sendo desta forma aqueles que mais discriminam a população quanto ao traço latente, pode se basear no ponto de corte igual a 1^{16,17} para as estimativas da função de informação do item, pois esta é influenciada pela magnitude das estimativas do parâmetro de inclinação no modelo de Resposta Gradual. Neste contexto, é interessante observar que, dos sete itens que melhor discriminam os indivíduos quanto à intensidade de sintomas depressivos, seis têm sua melhor performance na região deste traço latente classificada como de intensidade de sintomas depressivos moderada¹⁵. Apenas o item relativo à auto-aversão se mantém discriminando bem também na região da intensidade de sintomas depressivos graves, pois seu intervalo no escore TRI

vai de 1,3 a 2,2, o que equivale a 27 a 37 no escore total.

Já no estudo de Uher et al.¹¹ são utilizados outros pontos de corte para o parâmetro de inclinação do item (a_i), este sendo utilizado diretamente como o parâmetro de discriminação do item no modelo de resposta gradual. Estes pontos de corte classificam os itens em três grupos: itens com baixa discriminação ($a_i < 0,65$); itens com discriminação moderada ($0,65 \leq a_i \leq 1,34$), e itens com alta discriminação ($a_i > 1,34$). Considerando este critério, os itens que apresentam alta discriminação são tristeza, pessimismo, sentimento de fracasso, insatisfação e auto-aversão (Tabela 2), sendo que, com exceção do último, estes itens também estão classificados neste grupo no trabalho de Uher et al.¹¹ Indecisão e dificuldade de trabalhar ficam, agora, no grupo com moderada discriminação junto a todos os outros, com exceção de irritabilidade e perda de peso, que são os únicos itens classificados como tendo baixa discriminação por este critério (Tabela 2).

O valor encontrado para o percentil 95 do escore da TRI, equivalente a 31 no escore total, se enquadra na classificação moderada para a intensidade de sintomas depressivos. Neste grupo, onde todos os 202 indivíduos são classificados como tendo intensidade de sintomas depressivos de moderada a grave (seus escores da TRI são no mínimo iguais ao percentil 95), quase 75% são mulheres, o que está de acordo com a evidência de que a prevalência de depressão em mulheres é de duas a três vezes a dos homens³¹.

As estimativas dos pontos de interseção ($b_{i,k}$) são particionadas em dois termos, sendo um deles (b_i) um parâmetro de posição que pode ser interpretado como uma medida da gravidade do sintoma depressivo representado por um determinado item do BDI. Isto possibilita que se façam comparações de resultados com modelos TRI para respostas dicotômicas, tais como os modelos de dois ou três parâmetros (ambos incluem parâmetros de in-

clinação e posição), quando aplicados ao mesmo instrumento. No estudo de Cúri¹⁹, onde foi ajustado o modelo logístico de 3 parâmetros¹⁶ para dados provenientes do BDI, encontraram-se resultados que mostram razoável concordância com aqueles obtidos no presente trabalho quanto à gravidade e discriminação dos itens. Em especial no item sobre perda de peso, que é um dos mais graves e também um dos que menos discrimina (Tabela 2). Apesar de a perda de peso ser um sintoma que não tenha condições de discriminar a população quanto à intensidade de sintomas depressivos, talvez ele seja grave, entre outras, por uma das duas seguintes razões: a possibilidade de ser decorrente de alguma condição clínica que pode levar à depressão (pois existe evidência de que a depressão está fortemente associada a diversas condições médicas crônicas, tais como hipertensão, doença da artéria coronária, diabetes, derrame cerebral, doença renal terminal, doença pulmonar obstrutiva crônica, falha cardíaca congênita, angina, asma e artrite³²⁻³⁹); ou de ser uma consequência física de outros sintomas, isto é, antes o indivíduo sente, por exemplo, tristeza, desânimo, entre outros, e em seguida deixa de se cuidar, inclusive de sua alimentação, e conseqüentemente, seu peso diminui.

São inúmeros os ganhos provenientes da utilização da TRI. Neste estudo, estes ganhos aparecem na possibilidade de classificação e comparação dos sintomas depressivos quanto ao seu poder de discriminação e sua gravidade, na possibilidade de se verificar a relação do nível de intensidade de sintomas depressivos com a probabilidade de resposta a uma categoria de cada sintoma, podendo aí ser verificado o entendimento da escala entre os respondentes, na possibilidade de se comparar os níveis de intensidade de sintomas depressivos estimados para cada respondente com a gravidade de cada sintoma, entre outros. Mais importante ainda é o fato de os níveis estimados para a intensidade de sintomas depressivos (score

TRI) serem calculados considerando-se as diferentes contribuições de cada item (sintoma depressivo) para o traço latente medido, ao contrário da TCT, onde os respondentes com o mesmo escore total são considerados como iguais, mesmo tendo um diferenciado perfil de respostas ao instrumento. Isto pode ser visualizado melhor no Quadro 1, onde nove indivíduos apresentam escore total igual a 10 (no entanto, têm perfis de resposta diferentes aos itens do BDI) e escores da TRI diferenciados, estes últimos calculados considerando-se os diferentes perfis de resposta a itens que têm pesos diferentes no cálculo (com exceção dos modelos de Rasch onde todos os itens entram com o mesmo peso no cálculo do escore).

É importante salientar que ainda se fazem necessários estudos para a obtenção de modelos da TRI validados para populações específicas visando a sua utilização na clínica médica, para os quais possam ser determinados pontos de corte para a classificação dos respondentes de instrumentos, tais como o BDI, quanto ao seu grau de intensidade de sintomas depressivos, pois a relevância da utilização destes modelos na área de pesquisa médica já vem sendo demonstrada. Também seria necessário desenvolver e implementar computacionalmente modelos da TRI (em especial para itens com resposta politômica) que lidem com a presença de itens com DIF (funcionamento diferencial do item), pois muitas vezes diferentes grupos de sujeitos reagem de modo diferente ao responderem um determinado item.

No BDI alguns itens apresentam funcionamento diferencial para a variável sexo, ou seja, os homens reagem a estes itens de forma diferente das mulheres. O fato de o modelo da TRI utilizado não contemplar a presença de itens que apresentam DIF constitui uma limitação do presente estudo. Outra potencial limitação é que a idade média da amostra é mais baixa do que a média de idade da população que pretende representar (em particular na amostra não-clínica).

Quadro 1 - Quadro comparativo do respectivo escore TRI* para indivíduos com escore total 10, com diferentes padrões de resposta.

Chart 1 - Comparative table of IRT scores* for individuals with total score 10 and different patterns of response.

Itens	Padrão de resposta dos indivíduos nº								
	0005	0007	0028	0047	0073	0111	0150	0162	0163
1	1	0	0	0	0	0	1	0	0
2	0	0	0	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0
4	1	1	0	0	1	0	1	1	0
5	1	0	0	0	1	3	0	0	3
6	0	0	0	3	0	0	1	0	0
7	0	0	0	0	1	0	0	0	1
8	1	1	0	0	1	2	2	1	1
9	0	0	0	0	0	0	0	0	0
10	0	3	0	3	0	2	0	0	1
11	0	0	3	0	3	0	0	3	2
12	1	0	3	0	0	0	0	1	0
13	0	2	0	1	2	0	0	1	0
14	0	0	0	2	0	0	0	0	0
15	1	0	1	0	1	0	0	0	0
16	1	0	1	0	0	3	3	1	0
17	1	0	1	0	0	0	1	1	1
18	0	0	0	0	0	0	0	0	0
19	0	0	1	0	0	0	0	0	0
20	1	2	0	0	0	0	1	1	1
21	1	0	0	0	0	0	0	0	0
Escore TRI	0,27	0,01	-0,28	-0,21	0,07	-0,41	-0,09	-0,03	-0,14
EP**	0,23	0,27	0,30	0,29	0,26	0,33	0,26	0,26	0,28
Mudança de escala***	16	13	10	10	14	8	12	13	11
Escore Total	10	10	10	10	10	10	10	10	10

* Intensidade de sintomas depressivos estimado mediante o ajuste do modelo de Resposta Gradual / Intensity of depressive symptoms estimated by the fit of the Graded Response model

** Erro padrão da estimativa da intensidade de sintomas depressivos / Standard error of the estimate of the intensity of depressive symptoms

*** Valores equivalentes dos escores TRI na escala do escore total / Equivalent values of the IRT scores on the total score scale

Referências

1. Fleck MP, Lafer B, Sougey EB, Del Porto JA, Brasil MA, Juruena MF. [Guidelines of the Brazilian Medical Association for the treatment of depression (complete version)]. *Rev Bras Psiquiatr* 2003; 25: 114-22.
2. Kessler RC, Berglund P, Demler O et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003; 289: 3095-105.
3. Almeida-Filho N, Mari J.J., Coutinho E, et al. Brazilian multicentric study of psychiatric morbidity. Methodological features and prevalence estimates. *Br J Psychiatry* 1997; 171: 524-9.
4. Theme-Filha MM, Szwarcwald CL, Souza-Junior PR. Socio-demographic characteristics, treatment coverage, and self-rated health of individuals who reported six chronic diseases in Brazil, 2003. *Cad Saúde Pública* 2005; 21: 43-53.
5. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* 2006; 367: 1747-57.
6. Beck, AT, Steer, RA. *Beck Depression Inventory. Manual*. San Antonio, TX: Psychological Corporation; 1993.
7. Lord, FM, Novick MR. *Statistical theories of mental test scores*. Reading, MA; 1968.

8. DeVellis RF. Classical test theory. *Med Care* 2006; 44: S50-9.
9. Lord, FM. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum; 1980.
10. Chachamovich, E. *Teoria de Resposta ao Item: Aplicação do modelo Rasch em desenvolvimento e validação de instrumentos em saúde mental* [tese de doutorado]. Rio Grande do Sul: Faculdade de Medicina da UFRGS; 2008.
11. Uher R, Farmer A, Maier W et al. Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol Med* 2008; 38: 289-300.
12. Nuevo R, Dunn G, Dowrick C, Vazquez-Barquero JL, Casey P, Dalgard OS et al. Cross-cultural equivalence of the Beck Depression Inventory: A five-country analysis from the ODIN study. *J Affect Disord* 2008.
13. Uttaro T, Lehman A. Graded response modeling of the Quality of Life Interview. *Eval Program Plann* 1999; 22: 41-52.
14. Samejima, F. *Estimation of latent ability using a response pattern of graded scores*. Psychometrika Monograph 17; 1969.
15. Cunha JA. *Manual da versão em português das ESCALAS BECK*. São Paulo; 2001.
16. Andrade, D. F., Tavares, H. R., & Valle, R. C. *Teoria da Resposta ao Item: conceitos e aplicações*. IN: 14º Simpósio Nacional de Probabilidade e Estatística; 2000 jul 28; Caxambu (BR). ABE - Associação Brasileira de Estatística.
17. Embretson SE, Reise SP. *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc.; 2000.
18. PARSCALE. [computer program]. Versão 4.1. Chicago (Illinois): Scientific Software International, Inc.; 2003.
19. Cúri M. *Análise de questionários com itens constrangedores* [tese de doutorado]. São Paulo: Instituto de Matemática e Estatística da USP; 2006.
20. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000; 38: II28-42.
21. McHorney CA, Cohen AS. Equating health status measures with item response theory: illustrations with functional status items. *Med Care* 2000; 38: II43-59.
22. Chan KS, Orlando M, Ghosh-Dastidar B, Duan N, Sherbourne CD. The interview mode effect on the Center for Epidemiological Studies Depression (CES-D) scale: an item response theory analysis. *Med Care* 2004; 42: 281-9.
23. Kim Y, Pilkonis PA, Frank E, Thase ME, Reynolds CF. Differential functioning of the Beck depression inventory in late-life patients: use of item response theory. *Psychol Aging* 2002; 17: 379-91.
24. Bernstein IH, Rush AJ, Carmody TJ, Woo A, Trivedi MH. Clinical vs. self-report versions of the quick inventory of depressive symptomatology in a public sector sample. *J Psychiatr Res* 2007; 41: 239-246.
25. Glorfeld LW. An improvement on Horn's Parallel Analysis Methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement* 1995; 55: 377-93.
26. Hayton JC, Allen DG, Scarpello V. Factor Retention Decisions in Exploratory Factor Analysis: a Tutorial on Parallel Analysis. *Organizational Research Methods* 2004; 7: 191-205.
27. Ledesma RD, Valero-Mora P. Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment, Research & Evaluation* 2007; 12.
28. Franklin SB, Gibson DJ, Robertson PA, Pohlmann JT, Fralish JS. Parallel Analysis: a method for determining significant principal components. *J Vegetation Sci* 1995; 6: 99-106.
29. Determining the Dimensionality of Data: A SAS Macro for Parallel Analysis. [Portland.: SUGI 28, Paper 90-28; 2007].
30. Kirisci L, Moss HB, Tarter RE. Psychometric evaluation of the Situational Confidence Questionnaire in adolescents: fitting a graded item response model. *Addict Behav* 1996; 21: 303-317.
31. Beyer JL, Nash J, Shelton R, Loosen PT. Transtorno Depressivo maior. In: Artmed Editora, ed. *Manual Diagnóstico e Estatístico de Transtornos Mentais*. 4 ed. Porto Alegre; 2000. p. 288-324.
32. Dickens C, McGowan L, Percival C et al. Depression is a risk factor for mortality after myocardial infarction: fact or artifact? *J Am Coll Cardiol* 2007; 49: 1834-40.
33. Bogner HR, Morales KH, Post EP, Bruce ML. Diabetes, depression, and death: a randomized controlled trial of a depression treatment program for older adults based in primary care (PROSPECT). *Diabetes Care* 2007; 30: 3005-3010.
34. Collins-McNeil J, Holston EC, Edwards CL, Carbage-Martin J, Benbow DL, Dixon TD. Depressive symptoms, cardiovascular risk, and diabetes self-care strategies in African American women with type 2 diabetes. *Arch Psychiatr Nurs* 2007; 21: 201-209.
35. Golden SH, Lee HB, Schreiner PJ et al. Depression and type 2 diabetes mellitus: the multiethnic study of atherosclerosis. *Psychosom Med* 2007; 69: 529-36.
36. Kamphuis MH, Geerlings MI, Tjhuis MA et al. Physical inactivity, depression, and risk of cardiovascular mortality. *Med Sci Sports Exerc* 2007; 39: 1693-9.
37. Knol MJ, Heerdink ER, Egberts AC et al. Depressive symptoms in subjects with diagnosed and undiagnosed type 2 diabetes. *Psychosom Med* 2007; 69: 300-5.

38. Li C, Ford ES, Strine TW, Mokdad AH. Prevalence of depression among U.S. adults with diabetes: findings from the 2006 behavioral risk factor surveillance system. *Diabetes Care* 2008; 31: 105-7.

39. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* 2007; 370: 851-8.

Recebido em: 20/07/09

Versão final reapresentada em: 26/06/10

Aprovado em: 12/07/10