

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA**



TESE DE DOUTORADO

**TEORIA DA RESPOSTA AO ITEM:
APLICAÇÃO NA AVALIAÇÃO DA
INTENSIDADE DE SINTOMAS DEPRESSIVOS**

Stela Maris de Jezus Castro

Orientador: Prof. Dr. João Riboldi

Porto Alegre, 28 de novembro de 2008

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA



TESE DE DOUTORADO

**TEORIA DA RESPOSTA AO ITEM:
APLICAÇÃO NA AVALIAÇÃO DA
INTENSIDADE DE SINTOMAS DEPRESSIVOS**

Stela Maris de Jezus Castro

Orientador: Prof.Dr. João Riboldi

A apresentação desta tese é exigência do Programa de Pós-graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul, para obtenção do título de Doutor.

Porto Alegre, Brasil.
2008

C355t Castro, Stela Maris de Jezus

Teoria da resposta ao item : aplicação na avaliação da intensidade de sintomas depressivos / Stela Maris de Jezus Castro ; orient. João Riboldi. – 2008. 153 f.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul. Faculdade de Medicina. Programa de Pós-Graduação em Epidemiologia. Porto Alegre, BR-RS, 2008.

1. Depressão 2. Epidemiologia 3. Estudos de validação 4. Escalas de graduação psiquiátrica 5. Modelos estatísticos 6. Interpretação estatística de dados I. Riboldi, João II. Título.

NLM: WM 171

Catálogo Biblioteca FAMED/HCPA

BANCA EXAMINADORA

Prof. Dr. Dalton Francisco de Andrade

Universidade Federal de Santa Catarina – UFSC

Profa. Dra. Sídia Maria Callegari Jacques

Universidade Federal do Rio Grande do Sul – UFRGS

Profa. Dra. Suzi Alves Camey

Universidade federal do Rio Grande do Sul - UFRGS

AGRADECIMENTOS

Em uma caminhada de muitos anos, particularmente nos últimos quatro, vamos encontrando inúmeras pessoas que contribuem para nosso trabalho e nossa vida. Chegada a hora de escrever os agradecimentos, pode acontecer de alguém, não menos importante, deixar de ser mencionado neste texto, porém, desde já, desculpo-me pelo esquecimento.

Em primeiro lugar, quero expressar minha profunda gratidão ao meu marido Álvaro e ao meu filho Pedro por estarem comigo em todos os momentos da minha vida. Ao Álvaro, em especial, pelo carinho e paciência nos momentos difíceis que já passaram e nos que ainda estão por vir. Devo incluir aqui meu pai Amaury e minha mãe Maria por fazerem parte da minha vida e me apoiarem sempre incondicionalmente.

Agradeço, especialmente também, ao meu orientador e amigo Professor João Riboldi, que me acompanha e orienta desde a graduação, sempre incentivando e acreditando na minha capacidade. Espero que possamos continuar trabalhando juntos ainda por muito tempo.

Quero agradecer imensamente a minha amiga Suzi por ter tão prontamente disponibilizado o banco de dados utilizado nesta tese, sem o qual ela não teria sido realizada.

Não posso deixar de agradecer também à amiga e colega Vanessa, por sua valiosa contribuição ao segundo artigo, à Mariana Cúri, por disponibilizar um dos programas utilizados no segundo artigo e à Clarissa, por sua participação imprescindível no primeiro artigo.

Agradeço também ao Programa de Pós-Graduação em Epidemiologia, na pessoa do Professor Bruce Duncan, que disponibilizou os recursos necessários à compra do software PARSCALE utilizado na obtenção dos resultados do primeiro artigo desta tese.

Finalmente, agradeço aos amigos e colegas nas aulas do curso de Doutorado e aos meus colegas do Departamento de Estatística, que de alguma forma me apoiaram nesta etapa da minha vida.

SUMÁRIO

| | |
|--|-----|
| Abreviaturas e Siglas | 6 |
| Lista de Quadros | 7 |
| Lista de Tabelas | 7 |
| Lista de Figuras | 8 |
| 1. APRESENTAÇÃO | 11 |
| 2. RESUMO E ABSTRACT | 12 |
| 3. INTRODUÇÃO | 14 |
| 4. REVISÃO DE LITERATURA | 16 |
| 4.1 Depressão | 16 |
| 4.2 Inventário de Depressão Beck | 20 |
| 4.3 Teoria da Resposta ao Item | 24 |
| 4.4 Funcionamento Diferencial do Item | 38 |
| 5. OBJETIVOS | 42 |
| 6. REFERÊNCIAS BIBLIOGRÁFICAS | 43 |
| 7. ARTIGOS | 53 |
| 7.1 Artigo 1 – Teoria da Resposta ao Item: aplicação na avaliação da intensidade de sintomas depressivos | 54 |
| 7.2 Artigo 2 – Modelo da Teoria da Resposta ao Item para itens viesados: Funcionamento Diferencial do Item no Inventário de Depressão Beck | 80 |
| 8. CONCLUSÕES E CONSIDERAÇÕES FINAIS | 103 |
| 9. ANEXOS | 107 |
| ANEXO A – Projeto de Pesquisa | 108 |
| ANEXO B – Inventário de Depressão Beck (BDI) | 125 |
| ANEXO C – Curvas de categoria de resposta e Curvas de informação do item para cada item do BDI, e Curva de Informação do Teste – artigo 1 | 129 |
| ANEXO D – Gráficos resultantes do procedimento de Análise Paralela utilizado nos dois artigos | 152 |

ABREVIATURAS E SIGLAS

| | |
|-------|--|
| BAI | Inventário de Ansiedade Beck |
| BDI | Inventário de Depressão Beck |
| BHS | Escala de Desesperança Beck |
| BSI | Escala de Ideação Suicida Beck |
| CES-D | <i>Centre for Epidemiological Studies – Depression Scale</i> |
| CID | Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde |
| DALY | <i>Disability-adjusted life years</i> |
| DCPP | <i>Disease Control Priorities Project</i> |
| DIF | Funcionamento Diferencial do Item (<i>Diferencia Item Functioning</i>) |
| GBD | <i>Global Burden of Disease</i> |
| G-PCM | Modelo de Crédito Parcial Generalizado (<i>Generalized Partial Credit Model</i>) |
| GRM | Modelo de Resposta Gradual (<i>Graded-Response Model</i>) |
| HADS | <i>Hospital Anxiety Depression Scale</i> |
| IRTLR | <i>Item Response Theory Log-Likelihood Ratio</i> |
| IRC | Curvas de Resposta (<i>Item Response Curves</i>) |
| LIDO | <i>Longitudinal Investigation of Depression Outcomes</i> |
| M-GRM | Modelo de Resposta Gradual Modificado (<i>Modified Graded Response Model</i>) |
| MMSE | <i>Mini Mental State Examination</i> |
| MPQ | <i>Multidimensional Personality Questionnaire Stress Reaction Scale</i> |
| PCM | Modelo de Crédito Parcial (<i>Partial Credit Model</i>) |
| PHQ | <i>Patient Health Questionnaire</i> |
| RSM | Modelo de Escala Gradual (<i>Rating Scale Model</i>) |
| TCC | Teoria Clássica de Teste |
| TRI | Teoria da Resposta ao Item |
| WHO | Organização Mundial da Saúde |
| WHS | <i>World Health Survey</i> |
| 3MS | <i>Modified Mini Mental State</i> |

LISTA DE QUADROS

I - Revisão de literatura

| | |
|--|----|
| QUADRO 1 – Exemplos de estudos onde o BDI foi utilizados para medir a intensidade de sintomas depressivos | 22 |
| QUADRO 2 – Exemplos de estudos onde os modelos de TRI foram utilizados na avaliação da intensidade de sintomas depressivos | 26 |
| QUADRO 3 – Simbologia e interpretações dos parâmetros dos modelos TRI dicotômicos na área educacional e no contexto do BDI | 28 |
| QUADRO 4 – Distribuição dos modelos TRI de acordo com a natureza do item e o número de parâmetros | 37 |

II – Artigo 1

| | |
|--|----|
| QUADRO 1 – Quadro comparativo do respectivo escore TRI para indivíduos com escore total 10, com diferentes padrões de resposta | 79 |
|--|----|

LISTA DE TABELAS

I – Artigo 1

| | |
|---|----|
| TABELA 1 – Descrição da amostra de acordo com grupo de origem | 73 |
| TABELA 2 – Estimativa dos parâmetros do modelo de Resposta Gradual | 74 |
| TABELA 3 – Descrição dos indivíduos com estimativa da intensidade de sintomas depressivos acima do percentil 95 ($\hat{\theta} = 1,60$) | 75 |

II – Artigo 2

| | |
|--|-----|
| TABELA 1 – Descrição da amostra de acordo com grupo de origem | 98 |
| TABELA 2 – Estimativa dos parâmetros do modelo logístico de 2 parâmetros | 99 |
| TABELA 3 – Estimativa dos parâmetros do modelo para itens constrangedores .. | 100 |
| TABELA 4 – Descrição dos indivíduos com nível alto de sintomas depressivos estimado acima do percentil 95 ($\hat{\theta} = 1,60$ para o modelo 1 e $\hat{\theta} = 1,62$ para o modelo 2) | 101 |

LISTA DE FIGURAS

I – Revisão de literatura

| | |
|--|----|
| FIGURA 1 – Curva Característica do Item BDI01, relativo ao sintoma depressivo tristeza | 31 |
| FIGURA 2 – Curvas de categoria de resposta do item BDI01 relativo ao sintoma depressivo tristeza | 35 |

II – Artigo 1

| | |
|---|----|
| FIGURA 1 – Representação gráfica do item 1 (tristeza) do BDI segundo o modelo de resposta gradual | 76 |
| FIGURA 2 – Curvas de categoria de resposta para os itens 2 e 6 do BDI | 77 |
| FIGURA 3 – Curvas de informação do item para os itens 1, 2, 3, 4, 7, 13, 15 e 11 do BDI | 78 |

III – Artigo 2

| | |
|---|-----|
| FIGURA 1 – Curva característica do item (CCI) para o sintoma fadigabilidade (item 17) segundo o modelo logístico de 2 parâmetros (1) e o modelo para itens constrangedores (2), este para os sexos feminino e masculino | 102 |
|---|-----|

IV – ANEXO C

| | |
|--|-----|
| FIGURA 1 – Curvas de categoria de resposta e informação do item para o item 1 (TRISTEZA) | 130 |
| FIGURA 2 – Curvas de categoria de resposta e informação do item para o item 2 (PESSIMISMO) | 131 |
| FIGURA 3 – Curvas de categoria de resposta e informação do item para o item 3 (SENTIMENTO DE FRACASSO) | 132 |
| FIGURA 4 – Curvas de categoria de resposta e informação do item para o item 4 (INSATISFAÇÕES) | 133 |

| | |
|--|-----|
| FIGURA 5 – Curvas de categoria de resposta e informação do item para o item 5 (CULPA) | 134 |
| FIGURA 6 – Curvas de categoria de resposta e informação do item para o item 6 (PUNIÇÃO) | 135 |
| FIGURA 7 – Curvas de categoria de resposta e informação do item para o item 7 (AUTO-AVERSÃO) | 136 |
| FIGURA 8 – Curvas de categoria de resposta e informação do item para o item 8 (AUTO-ACUSAÇÕES) | 137 |
| FIGURA 9 – Curvas de categoria de resposta e informação do item para o item 9 (IDÉIAS SUICIDAS) | 138 |
| FIGURA 10 – Curvas de categoria de resposta e informação do item para o item 10 (CHORO) | 139 |
| FIGURA 11 – Curvas de categoria de resposta e informação do item para o item 11 (IRRITABILIDADE) | 140 |
| FIGURA 12 – Curvas de categoria de resposta e informação do item para o item 12 (RETRAIMENTO SOCIAL) | 141 |
| FIGURA 13 – Curvas de categoria de resposta e informação do item para o item 13 (INDECISÃO) | 142 |
| FIGURA 14 – Curvas de categoria de resposta e informação do item para o item 14 (MUDANÇA NA AUTO-IMAGEM) | 143 |
| FIGURA 15 – Curvas de categoria de resposta e informação do item para o item 15 (DIFICULDADE DE TRABALHAR) | 144 |
| FIGURA 16 – Curvas de categoria de resposta e informação do item para o item 16 (INSÔNIA) | 145 |
| FIGURA 17 – Curvas de categoria de resposta e informação do item para o item 17 (FATIGABILIDADE) | 146 |
| FIGURA 18 – Curvas de categoria de resposta e informação do item para o item 18 (PERDA DE APETITE) | 147 |
| FIGURA 19 – Curvas de categoria de resposta e informação do item para o item 19 (PERDA DE PESO) | 148 |
| FIGURA 20 – Curvas de categoria de resposta e informação do item para o item 20 (PREOCUPAÇÕES SOMÁTICAS) | 149 |

| | |
|---|-----|
| FIGURA 21 – Curvas de categoria de resposta e informação do item para o item 21 (PERDA DA LIBIDO) | 150 |
| FIGURA 22 – Curva de informação do teste para o BDI | 151 |

V – ANEXO D

| | |
|---|-----|
| FIGURA 1 – Resultado da Análise Paralela para a amostra total, com 4025 observações e 21 itens | 153 |
| FIGURA 2 – Resultado da Análise Paralela para a amostra total dicotomizada, com 4025 observações e 21 itens | 153 |

1. APRESENTAÇÃO

Este trabalho consiste na tese de doutorado intitulada “TEORIA DA RESPOSTA AO ITEM: APLICAÇÃO NA AVALIAÇÃO DE INTENSIDADE DE SINTOMAS DEPRESSIVOS”, apresentada ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul. O trabalho é apresentado em três partes, na ordem que segue:

1. Resumo
2. Introdução, Revisão de Literatura e Objetivos
3. Artigos:
 - Teoria da Resposta ao Item: aplicação na avaliação da intensidade de sintomas depressivos
 - Modelo da Teoria da Resposta ao Item para itens viesados: Funcionamento Diferencial do Item no Inventário de Depressão Beck
4. Conclusões e Considerações Finais

Documentos de apoio estão apresentados nos anexos: Projeto de Pesquisa (Anexo A), Inventário de Depressão Beck (Anexo B), Curvas de categoria de resposta e curvas de informação do item para cada item do BDI, e curva de informação do teste (Anexo C) e Gráficos resultantes do procedimento de Análise Paralela utilizado nos dois artigos (Anexo D).

2. RESUMO E ABSTRACT

RESUMO

CONTEXTO: A depressão é uma doença com alta prevalência no mundo todo e se manifesta através de diversos sintomas observáveis, os chamados sintomas depressivos. Determinar a intensidade dos sintomas depressivos pode ser importante para verificar o estágio da depressão e avaliar seu desfecho, e quanto mais acurada e rápida for esta medida mais benefícios podem ser alcançados. A intensidade dos sintomas depressivos é um traço latente que pode ser medido através de instrumentos compostos por itens representativos destes sintomas observáveis, como o Inventário de Depressão Beck (BDI). É importante que a metodologia para analisar instrumentos do tipo do BDI considere que nem todos os sintomas depressivos têm a mesma importância em relação ao traço latente que pretendem medir. A Teoria da Resposta ao Item (TRI) compreende um grupo de modelos lineares generalizados e procedimentos estatísticos associados, que descrevem a associação entre o nível de um indivíduo sobre o traço latente e a probabilidade de uma resposta a um item. Estes modelos têm como uma de suas características especiais que os níveis estimados do traço latente sendo medido incorporam as diferenças em discriminação e gravidade de cada item constante no instrumento de medida, isto é, os itens entram com diferentes pesos na estimativa do traço latente dos indivíduos avaliados. **OBJETIVOS:** Este trabalho tem por objetivo mostrar a potencialidade dos modelos da TRI e o total aproveitamento das informações quando do uso destes modelos na análise de dados oriundos do BDI para a medida de intensidade de sintomas depressivos. **MÉTODO:** Os dados 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 (PUCRS) e publicado em 2001; os modelos TRI utilizados na análise destes dados foram o modelo de Resposta Gradual de Samejima (1969) e o modelo para Itens Constrangedores de Cúri (2006). **RESULTADOS:** Os sintomas depressivos que melhor discriminam a população quanto ao nível de intensidade de sintomas depressivos são sentimento de fracasso, insatisfações, tristeza, auto-aversão, indecisão, dificuldade de trabalhar e pessimismo; e os que menos discriminam são perda de peso, irritabilidade e auto-acusações. Os sintomas mais graves são perda de peso, retraimento social, idéias suicidas, sentimento de fracasso apenas para as mulheres e perda da libido apenas para os homens (estes dois últimos são itens com funcionamento diferencial). **CONCLUSÕES:** Este estudo mostrou os inúmeros ganhos advindos da utilização de modelos TRI na avaliação da intensidade de sintomas depressivos, pois sua utilização aproveita totalmente a informação, considerando o perfil de cada indivíduo que responde ao instrumento, contribuindo na identificação daqueles que apresentam potencial depressivo.

Descritores: Depressão, Intensidade de Sintomas Depressivos, Inventário de Depressão Beck, Teoria da Resposta ao Item, Modelo TRI para Itens Constrangedores.

ABSTRACT

CONTEXT: Depression is a disease with high prevalence worldwide and manifests itself through various symptoms observed, so-called depressive symptoms. To determine the intensity of depressive symptoms may be important to determine the stage of depression and evaluate its outcome, and the more rapid and accurate is this more benefits can be achieved. The intensity of depressive symptoms is a latent trait that can be measured by instruments consisting of items representative of observable symptoms, as the Beck Depression Inventory (BDI). It is important that the methodology for analyzing instruments of the type of BDI considers that not all depressive symptoms have the same importance in relation to the latent trait they wish to measure. The Item Response Theory (IRT) comprises a group of generalized linear models and statistical procedures involved, which describe the association between the level of an individual on the latent trait and the likelihood of a response to an item. These models have as one of its special characteristics that the estimated levels of latent trait being measured incorporate the differences in discrimination and severity of each item contained in the measuring instrument, that is, those items come with different weights in the estimation of latent trait of individuals evaluated. **OBJECTIVES:** This paper aims to show the capability of the models of the IRR and total utilization of information when using these models to analyze data from the BDI to measure the intensity of depressive symptoms. **METHOD:** The data come from a cross-sectional study conducted for the adaptation, standardization and validation of Beck scales for the portuguese, in a study conducted by Dr. Alcides Jurema Cunha (PUCRS) and published in 2001; the TRI models used in the analysis of these data was the Graded-Response model of Samejima (1969) and the model IRT for embarrassing items of Cúri (2006). **RESULTS:** The depressive symptoms that best depict the population about the level of intensity of depressive symptoms are feeling of failure, dissatisfaction, sadness, self-hatred, indecision, difficulty of work and pessimism; and those who are less discriminating are weight loss, irritability and self-accusations. The symptoms are more severe weight loss, social withdrawal, suicidal thoughts, feelings of failure only for women and loss of libido only for men (the latter two items are working with differential functioning). **CONCLUSIONS:** This study showed the many gains resulting from use of IRT models in the assessment of the intensity of depressive symptoms, because their use completely takes the information, considering the profile of each person who responds to the instrument, helping to identify those which have the potential depression.

Keywords: Depression, Intensity of Depressive Symptoms, Beck Depression Inventory, Item Response Theory, Model IRT for embarrassing items.

3. INTRODUÇÃO

Depressão é uma doença do organismo como um todo, a qual compromete o funcionamento físico, o humor e, conseqüentemente, o pensamento. Ela altera a maneira como o indivíduo vê o mundo e sente a realidade, entende as coisas, manifesta emoções, sente a disposição e o prazer com a vida. Logo a depressão é uma doença afetiva ou do humor e é considerada como um transtorno psiquiátrico que exige tratamento.

A depressão é uma doença com alta prevalência e se manifesta através de diversos sintomas nomeados de sintomas depressivos. Os sintomas depressivos são muito variados, indo desde as sensações de tristeza, passando pelos pensamentos negativos até as alterações da sensação corporal como dores e enjôos. Determinar a intensidade dos sintomas depressivos pode ser importante para verificar o estágio da doença e avaliar o desfecho, e quanto mais acurada e rápida for esta medida mais benefícios podem ser alcançados.

A intensidade dos sintomas depressivos é uma variável que não é diretamente mensurável, logo pode ser considerada como um traço latente. O traço latente não pode ser observado, mas pode ser inferido baseado em observações do comportamento da pessoa, que neste caso são observações sobre a presença e intensidade dos diversos sintomas depressivos, as quais são obtidas através de instrumentos compostos de itens que representam os sintomas depressivos. O Inventário de Depressão Beck (BDI), revisado por Beck e Steer em 1993, é uma das medidas de auto resposta mais amplamente utilizadas para avaliar a intensidade de sintomas depressivos em pacientes diagnosticados como tendo depressão e para detectar depressão em populações não clínicas.

A análise de dados provenientes de instrumentos de medida, tal como o BDI, bem como procedimentos de avaliação de características psicométricas para o desenvolvimento destes instrumentos, tem sido realizada tradicionalmente através da chamada Teoria Clássica de Teste (TCT), cujo nome tem origem na área educacional. No entanto, nas últimas três décadas, o uso da Teoria da Resposta ao Item (TRI) para medida auto-referida tem florescido. TRI é freqüentemente também chamada de “teoria psicométrica moderna” para distingui-la da teoria clássica de teste (TCT). TRI compreende um grupo de modelos lineares generalizados e procedimentos estatísticos associados, que conectam as respostas aos itens de um instrumento ao traço latente subjacente ao mesmo e às características destes itens, descrevendo, assim, a associação

entre o nível deste indivíduo em relação ao traço latente e a probabilidade de uma resposta a um item. Estes modelos têm como uma de suas características especiais o fato de que os níveis estimados do traço latente medido incorporam as diferenças em discriminação e gravidade de cada item constante no instrumento de medida, isto é, os itens entram com diferentes pesos na estimativa do traço latente dos indivíduos avaliados.

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 mesmo, como também traz informações sobre os instrumentos de medida como um todo e, principalmente, item a item. O uso dos modelos TRI na área psiquiátrica, onde inúmeros desfechos se caracterizam como traços latentes e dependem de instrumentos de medida compostos de itens para serem medidos, têm aumentado cada vez mais, porém, no Brasil, em especial no Rio Grande do Sul, a sua utilização é relativamente restrita.

Acredita-se que a utilização desta metodologia, com características altamente vantajosas, em um desfecho tão prevalente quanto Depressão é importante e relevante, e é baseada nesta convicção que esta tese foi concebida.

4. REVISÃO DE LITERATURA

4.1 DEPRESSÃO

A depressão é essencialmente um transtorno psiquiátrico de humor episódico e recorrente. Cada episódio depressivo dura, em média, 20 semanas, seguido de um período normal entre os mesmos, sendo que em torno de 12% dos pacientes não apresentam remissão de sintomas. A taxa de recorrência para aqueles que se recuperam do primeiro episódio depressivo fica em torno de 80%, em uma média de quatro anos (Fleck et al., 2003).

A depressão, de um modo geral, pode ser classificada em dois tipos: depressão maior (ou unipolar) e transtorno bipolar. A idéia de depressão maior é primariamente médica, isto é, envolve um modo particular de se olhar o distúrbio psicológico. Este centraliza sobre a noção de uma síndrome que é distinta de outras síndromes psiquiátricas. Algumas destas podem ser relativamente fáceis de serem distinguidas, por exemplo, esquizofrenia paranóide, enquanto outras são reconhecidas por serem relacionadas. As desordens que mais se assemelham com depressão maior são outras desordens afetivas, ou seja, condições que são caracterizadas por distúrbio de humor. Elas cobrem um número de desordens de ansiedade, outras condições depressivas e desordem de humor bipolar (Bebbington, 2004).

Transtorno bipolar é identificado pela presença de duas espécies de episódio nos quais o humor associado é depressivo ou predominantemente alegre. Ele se distingue da desordem unipolar de diversas maneiras (tais como herança, curso e desfecho), e a distinção é bastante útil. Entretanto, episódios depressivos no transtorno bipolar não podem ser distinguidos sintomaticamente daqueles da depressão maior. A metade dos casos de transtorno bipolar começa com um episódio depressivo fazendo com que os pacientes sejam classificados como tendo depressão maior. Somente em 5% dos casos, estes pacientes são reclassificados como sofrendo de transtorno bipolar (Ramana and Bebbington, 1995).

Entre as causas estudadas da depressão estão fatores genéticos (Sullivan et al., 2000) e desequilíbrio bioquímico (Marot, 2004). Os eventos desencadeadores são muito estudados e de fato encontra-se relação entre certos acontecimentos estressantes na vida das pessoas e o início de um episódio depressivo (Marot, 2004).

Os sintomas da depressão são muito variados, indo desde as sensações de tristeza, passando pelos pensamentos negativos até as alterações da sensação corporal como dores e enjôos. Porém, segundo a CID-10, a fim de se fazer o diagnóstico é necessário um grupo de sintomas fundamentais tais como perda de energia ou interesse, humor deprimido ou fadigabilidade, e um grupo de sintomas acessórios que são: dificuldade de atenção e concentração, auto-estima e auto-confiança reduzidas, idéias de culpa e inutilidade, visões desoladas e pessimistas do futuro, idéias ou atos autolesivos ou suicídio, alterações do sono e diminuição do apetite (Fleck et al., 2003).

A depressão está associada com prejuízo cognitivo. Perda de memória ocasional parece ser o principal aspecto do funcionamento cognitivo que é vulnerável aos efeitos negativos da depressão (Airaksinen et al., 2004). Também, ela prejudica a habilidade de execução das funções normais, levando à diminuição da capacidade em 50% ou mais dos pacientes (Kessler et al., 2003; Ebmeier et al., 2006). Diminuição de capacidade é provavelmente uma consequência natural da depressão (Ormel et al., 2004).

Um aspecto crucial da epidemiologia da depressão é o aumento da mortalidade associado com esta condição. Uma metanálise de 25 estudos com 1,3 a 16 anos de seguimento de mais de 100.000 pessoas registrou um risco relativo geral de morte em sujeitos depressivos de 1,81 [IC95% (1,58;2,07)] comparados com pessoas que não tinham depressão (Cuijpers and Smit, 2002). Este aumento do risco de morte é, em parte, devido ao alto risco de suicídio (Kessler et al., 2003).

Desordens depressivas são as que mais ocorrem entre as desordens mentais e de comportamento. Estudos epidemiológicos mostram que 15% a 33% dos pacientes internados sofrem de desordens depressivas, comparados com 2% a 4% da população geral, indicando que a depressão é um dos mais frequentes fenômenos psicopatológicos encontrados no ambiente hospitalar (Fliege et al., 2005). Outros estudos apresentam a prevalência anual da depressão, na população geral, variando entre 3% e 11% (Fleck et al., 2003).

A depressão é duas vezes mais comum em mulheres do que em homens e pode começar em qualquer idade (a idade média de início é 25 anos). O estudo de *National Comorbidity Survey Replication* (NCS –R) realizado entre 2001 e 2002 nos Estados Unidos com pessoas maiores de 18 anos estimou a prevalência anual da depressão em 6,6% [IC95% (5,9;7,3)] e em 16,2% [IC95% (15,1;17,3)] para a vida inteira (Kessler et al., 2003).

Entre os adolescentes também foi encontrado um aumento na taxa de depressão na maioria das coortes de nascimento recentes (Patel et al., 2007). Contudo, uma boa parte da evidência que dá suporte a esta conclusão está baseada em dados recordatórios, por exemplo, um aumento na proporção de adultos em coortes recentes que teve seu primeiro episódio depressivo aos 18 anos (Kessler et al., 2003).

Existem alguns estudos sobre a prevalência de depressão na população brasileira. Almeida-Filho *et al.* (1997), reportam as prevalências encontradas para estados depressivos no Estudo Multicêntrico de Morbidez Psiquiátrica (*Multicentric Study of Psychiatric Morbidity*) conduzido, em 1991, nas áreas metropolitanas de Brasília, São Paulo e Porto Alegre, com uma amostra de 6476 adultos, com idade mínima de quatorze anos. As prevalências para a vida inteira, ajustadas para idade, são 2,8, 1,9 e 10,2 para Brasília, São Paulo e Porto Alegre, respectivamente. Já as prevalências para a vida inteira, estratificadas por sexo, são 1,9 para homens e 3,8 para mulheres em Brasília, 3,8 para mulheres em São Paulo (não tendo um resultado para homens), e 5,9 para homens e 14,5 para mulheres em Porto Alegre.

O projeto LIDO (*Longitudinal Investigation of Depression Outcomes*) (Fleck et al., 2002), desenvolvido para avaliar a relação entre sintomas depressivos, qualidade de vida e utilização de recursos em pacientes que procuram serviços de cuidados primários de saúde em seis centros de diferentes países, incluindo o Brasil (Porto Alegre), detectou que os pacientes com mais sintomas depressivos tinham maior comprometimento do funcionamento físico e psicológico e avaliavam sua qualidade de vida como pior. Também, os indivíduos com maior intensidade de sintomas depressivos avaliaram sua saúde como pior e estavam menos satisfeitos com ela do que aqueles com menor intensidade de sintomas depressivos. Outra informação relevante deste estudo, foi a observação de que os pacientes brasileiros com mais sintomas depressivos consultam mais os serviços de saúde, permanecem mais tempo internados e faltam mais dias ao trabalho do que os menos deprimidos. Além disso, o estudo mostrou que as mulheres têm sintomatologia depressiva mais intensa do que os homens que procuram serviço de cuidados primários em saúde (Fleck et al., 2002).

O *World Health Survey* (WHS) conduzido no Brasil, em 2003, como uma parte integrante de um projeto da Organização Mundial da Saúde (WHO) trouxe dados relevantes sobre o estado de saúde da população brasileira. Este levantamento populacional com representação nacional cobriu vários aspectos da saúde da população através de entrevistas pessoais. Entre os vários módulos pesquisados, um foi

especificamente dirigido às doenças crônicas, entre elas, a depressão. O trabalho de Theme-Filha et al. (2005) apresenta os resultados deste levantamento onde foram entrevistados 5000 indivíduos com 18 anos ou mais. Deste estudo resultou que a depressão foi a condição crônica com prevalência geral mais alta (19,2%), e também em cada faixa de idade afetando 14,5% de indivíduos entre 18 e 29 anos, 21,8% entre 30 e 49 anos, e 23,3% para indivíduos com 50 anos de idade ou mais. Outro resultado relevante deste foi a prevalência da depressão mais alta entre as mulheres (25,1% para mulheres versus 12,3% para homens) (Theme-Filha et al., 2005).

Um estudo de grande importância sobre a população brasileira foi conduzido por Gorenstein e colaboradores (2005), para detectar a prevalência de sintomas depressivos e sua expressão em uma amostra não-clínica de estudantes adolescentes. A amostra foi composta por 1555 adolescentes brasileiros (796 meninas e 759 meninos) com idades variando entre 13 e 17 anos, aos quais foi aplicado o Inventário de Depressão Beck (BDI), sendo a prevalência estimada de depressão igual a 7,6%.

A alta prevalência da depressão, tanto no Brasil como no mundo como um todo, acarreta prejuízos econômicos para os mesmos. Estes prejuízos podem ser quantificados através medidas como DALY (*Disability-adjusted life years*), que quantifica a carga global de doenças considerando os anos de vida vividos sem plena saúde e anos de vida perdidos devido à morte prematura.

O estudo chamado *Global Burden of Disease* (GBD) quantificou os efeitos da saúde de mais de 100 doenças para o mundo como um todo e, também, estratificado para oito regiões, utilizando como medida o DALY. O estudo de Lopez et al. (2006), apresenta os resultados do GBD para o ano de 2001. Entre as doenças não fatais, a depressão unipolar (ou depressão maior) foi a causa líder de anos vividos com incapacidade para homens e mulheres. Entretanto, a carga da doença foi 50% maior para as mulheres (Lopez et al., 2006).

Considerando uma faixa de idade mais restrita, Patel e colaboradores (Patel et al., 2007), identificaram em um estudo em Victoria, Austrália, que desordens mentais na população jovem entre 15 e 24 anos de idade contribuem entre 60% a 70% do DALY total.

Dentro desta linha, o DCPP (*Disease Control Priorities Project*), uma realização conjunta do *Fogarty International Center* do *US National Institutes of Health* (NIH), WHO e *The World Bank* (e com substancial financiamento da Fundação Bill&Melinda Gates), foi lançado em 2001 para identificar mudanças de políticas e estratégias de

intervenção para problemas de saúde de países em desenvolvimento. Em sua segunda edição (DCP2) (Laximminaryan et al., 2006), foi focalizada a avaliação do custo-efetividade de estratégias (ou intervenções) que visam ao melhoramento da saúde para as condições responsáveis pelas maiores cargas de doença. De acordo com o DCP2, o tratamento para a depressão (como também para ansiedade) é mais custo-efetivo do que o tratamento para outras desordens mentais mais severas, isto é, as intervenções são mais baratas e a redução de incapacidade é maior.

4.2 INVENTÁRIO DE DEPRESSÃO BECK

Determinar a intensidade de sintomas depressivos pode ser importante para verificar o estágio da depressão e avaliar seu desfecho. Uma medida acurada, rápida e custo-efetiva de sintomas depressivos serve tanto para propósitos clínicos ou de pesquisa. Clínicos podem aferir o benefício do tratamento e fazer ajustes adequados no plano do mesmo. Pesquisas, por outro lado, podem ser feitas com menor custo se tais medidas estão disponíveis. Ainda, a medida de intensidade de sintomas depressivos é importante para a implementação apropriada de diretrizes de tratamentos recomendados para depressão unipolar e outras desordens de humor.

A intensidade de sintomas depressivos é uma variável que não é diretamente mensurável, logo pode ser considerada como um construto latente. Este é aquele que não pode ser observado, mas pode ser inferido baseado em observações do comportamento da pessoa. Por exemplo, dor não pode ser observada diretamente, mas pode ser inferida através da observação de comportamentos (caretas, precauções, respostas a itens sobre dor). Em avaliações educacionais e psicológicas, uma medida é definida como uma “amostra de comportamento”. A amostra de comportamento observado são as escolhas de categorias de resposta feitas por pessoas respondendo a itens (Cook et al., 2005).

Questionários são amplamente usados para medir a intensidade de sintomas depressivos, sendo que grande parte destes usam escalas multi-item. Estes questionários são compostos de itens que contemplam os mais diversos sintomas depressivos observáveis. Dentre os mais conhecidos e frequentemente utilizados estão o Inventário de Depressão Beck (*Beck Depression Inventory – BDI*), a Escala de Depressão do Centro para Estudos Epidemiológicos – (*Centre for Epidemiological Studies – Depression Scale – CES-D*), o Questionário de Saúde do Paciente (*Patient Health Questionnaire PHQ*), e a Escala Hospitalar de Ansiedade e Depressão (*Hospital Anxiety*

Depression Scale – HADS), a Escala de Auto-Avaliação de Depressão de Zung (*Zung Self-Rating Depression Scale*) (Fliege et al., 2005).

O Inventário de Depressão Beck, revisado por Beck e Steer (1993), é uma das medidas de auto resposta mais amplamente utilizadas para avaliar a intensidade de sintomas depressivos em pacientes diagnosticados como tendo depressão e para detectar depressão em populações não clínicas (Kim et al., 2002), como pode ser visto no Quadro 1.

O BDI foi originalmente criado por Beck, Ward, Mendensol, Mock e Erbaugh (Beck et al., 1961). Ele faz parte das chamadas Escalas Beck. Estas são quatro medidas escalares - o Inventário de Depressão (BDI), o Inventário de Ansiedade (BAI), a Escala de Desesperança (BHS) e a Escala de Ideação Suicida (BSI) - que foram todas desenvolvidas por A. T. Beck e seus colegas no *Center for Cognitive Therapy* da Universidade de Pennsylvania, na Philadelphia, Estados Unidos. O manual do BDI no idioma original é de Beck e Steer (Beck and Steer, 1993) e a versão autorizada em português foi publicada em 2001 (Cunha, 2001). De acordo com os autores, o BDI é indicado para sujeitos de 17 a 80 anos (Beck and Steer, 1993). É uma escala de auto-relato, de 21 itens, cada um com quatro alternativas, subentendendo graus crescentes de gravidade da depressão (Williams et al., 1972), com escores variando de 0 a 3. O escore total permite a classificação de níveis de intensidade de sintomas depressivos. Ele é obtido somando os escores da cada item, correspondentes às alternativas assinaladas pelos sujeitos nos 21 itens. Cada item apresenta quatro alternativas, que podem ter escore 0, 1, 2 ou 3. Para a versão em português os pontos de corte, para classificar os respondentes quanto à intensidade dos sintomas depressivos, são: mínima (0–11), leve (12–19), moderada (20–35) e grave (36–63).

Os itens do BDI foram selecionados com base em observações e relatos de sintomas e atitudes mais frequentes em pacientes psiquiátricos, com transtornos depressivos (Cunha, 2001). Eles se referem a: (1) Tristeza; (2) Pessimismo; (3) Sentimento de fracasso; (4) Insatisfação; (5) Culpa; (6) Punição; (7) Auto-aversão; (8) Auto-acusações; (9) Idéias suicidas; (10) Choro; (11) Irritabilidade; (12) Retraimento social; (13) Indecisão; (14) Mudança na auto-imagem; (15) Dificuldade de trabalhar; (16) Insônia; (17) Fatigabilidade; (18) Perda de apetite; (19) Perda de peso; (20) Preocupações somáticas; (21) Perda da libido.

Quadro 1: Exemplos de estudos onde o BDI foi utilizado para medir a intensidade dos sintomas depressivos.

| REFERÊNCIA | OBJETIVO | LOCAL |
|--|--|---|
| (Knekt et al., 2008) | Comparar a efetividade de uma terapia de longo prazo e duas terapias de curto prazo no tratamento de distúrbios de humor e ansiedade. | Helsinki Finlândia |
| (Hellerstein et al., 2008) | Investigar se Aripiprazole pode ser um tratamento adjunto eficaz em pacientes não internados com depressão maior que têm sido imunes a tratamento com antidepressivos SSRI ou SNRI. | New York USA |
| (Georgiades et al., 2007) | Investigar se mudanças nos sintomas depressivos estariam associadas com mudanças no controle glicêmico em um período de 12 meses em pacientes com diabetes Tipo I e Tipo II. | Carolina do Norte USA |
| (Zimmermann et al., 2006) | Avaliar o impacto da depressão sobre mortalidade e qualidade de vida em pacientes com doença renal. | Rio Grande do Sul Brasil |
| (Yakar et al., 2007) | Avaliação da depressão e sua influência em pacientes asmáticos. | Osmaniye Turquia |
| (Schrag et al., 2007) | Revisão Sistemática a fim de explorar o uso de escalas de depressão em pacientes com Doença de Parkinson. | Londres Reino Unido |
| (Sinikallio et al., 2007) | Investigar o efeito da depressão em um desfecho rápido após estenose espinhal lombar (LSS). | Siilinjärvi Finlândia |
| (Harrington and Greene-Harrington, 2007) | Examinar depressão perinatal no norte e nordeste de Omaha, USA. | Omaha USA |
| (Moreira et al., 2007) | Investigar qual indicador antropométrico de sobrepeso (índice de massa corporal, circunferência da cintura ou razão cintura quadril) melhor se correlaciona com a presença de distúrbios de humor e intensidade de sintomas depressivos em mulheres obesas. | Rio de Janeiro Brasil |
| (Pirente et al., 2007) | Melhorar a Qualidade de Vida relacionada à Saúde (HRQOL) relacionada à depressão, ansiedade, dor funcionamento físico e aspectos sociais para sobreviventes de trauma severo por iniciar cedo a terapia cognitiva comportamental aplicada em cirurgia preventiva. | Cologne Alemanha |
| (Demir et al., 2007) | Determinar a prevalência de depressão, e avaliar a associação com fatores sócio-demográficos e aspectos do ambiente de trabalho, entre médicos residentes em um hospital de Istambul. | Istambul Turquia |
| (Hagen, 2007) | Revisão das origens, propósito e formato do BDI, discussão de alguns aspectos do poder e limitações do BDI, e conclusão com algumas implicações para enfermagem de saúde mental. | Alberta Canadá |
| (Schmidt et al., 2006) | Determinar a reprodutibilidade e validade de uma versão brasileira do Inventário de Deficiência Física de Zumbido no Ouvido (<i>Tinnitus Handicap Inventory – THI</i>), um questionário auto-aplicável que avalia o impacto do zumbido no ouvido na qualidade de vida de pacientes. A validade foi avaliada através de médias de BDI, calculando coeficiente de correlação de Pearson. | Porto Alegre Rio Grande do Sul Brasil |
| (Fleck et al., 2006) | Descrição do desenvolvimento e validação da versão para o português do módulo WHOQOL – OLD (<i>World Health Organization Quality of Life for Older Persons</i>). Além do módulo WHOQOL – OLD, os instrumentos WHOQOL – BREF, BDI e BHS foram aplicados. | Porto Alegre Rio Grande do Sul Brasil |
| (Jelinski et al., 2007) | Avaliar a relação entre características demográficas selecionadas e características clínicas em pacientes com dor de cabeça e depressão. | Calgary Canadá |

| REFERÊNCIA | OBJETIVO | LOCAL |
|--------------------------------|--|--------------------------------|
| (da Silva et al., 2007) | Registro da adaptação cultural e propriedades psicométricas do QOLIE – 31 (<i>Quality of Life in Epilepsy – 31 Inventory</i>) para a versão para a língua portuguesa e para a cultura brasileira. Além do QOLIE – 31 foram utilizados o NHP (<i>Nottingham Health Profile</i>), BDI e AEP (<i>Adverse Events Profile</i>). | São Paulo Brasil |
| (Brands et al., 2007) | Comparação do perfil cognitivo e bem estar psicológico entre pacientes com Diabetes Mellitus 2 e um grupo controle. | Utrecht Holanda |
| (Lee et al., 2007) | Determinar os efeitos de câncer de mama recorrente sobre a HRQOL. Foram utilizados os instrumentos EORTC (<i>European Organization for Research and Treatment of Cancer</i>), QLC-C30 e QLQ-BR23, MQOL (<i>McGill Quality of Life Questionnaire</i>), BDI, e BFI (<i>Brief Fatigue Inventory</i>). | Korea |
| (Miller et al., 2007) | Determinar se a alta prevalência de sintomas depressivos vista em doentes de Parkinson é também encontrada em outras desordens de humor. | Gainesville, Florida USA |
| (Kessels et al., 2007) | Examinar o efeito de sintomas depressivos auto-registrados na função de memória neurológica em uma amostra de pacientes de ambulatório não-psiquiátricos e identificar quais testes de memória podem ser mais suscetíveis para declínio relacionado à depressão. | Nijmegen Holanda |
| (Lesperance et al., 2007) | Documentar a eficácia em curto prazo de citalopram e psicoterapia interpessoal na redução de sintomas depressivos em pacientes com doença coronária arterial e depressão maior. Foram utilizados o HAM-D (<i>24-item Hamilton Depression Rating Scale</i>) e o BDI-II. | Montreal Canadá |
| (Andalib et al., 2006) | Estabelecer uma nova relação entre atividade imunológica e parâmetros psicológicos (stress/depressão) em mulheres sofrendo de aborto espontâneo recorrente. | Isfahan Irã |
| (Revah-Levy et al., 2007) | Examinar as propriedades psicométricas do ADRS (<i>Adolescent Depression Rating Scale</i>), uma nova medida que foi planejada especificamente para avaliar depressão em adolescentes. | Argenteuil França |
| (Laget et al., 2006) | Comparação de características de personalidade de sujeitos com desordens de dependência que já tenham tentado o suicídio | Lausanne Suíça |
| (Lamberton and Oei, 2007) | Testar a hipótese de especificidade de conteúdo cognitivo (CCSH) para avaliar se ansiedade e depressão podem ser diferenciadas na base de transtorno cognitivo. | Brisbane Austrália |
| (Caulin-Glaser et al., 2007) | Avaliar a independência dos efeitos de depressão e sexo sobre reabilitação cardíaca completa. | Columbus USA |
| (Frasure-Smith et al., 2007) | Investigar o impacto da depressão e marcadores inflamatórios, avaliados dois meses após síndrome coronária aguda, sobre eventos cardíacos adversos maiores num período de 2 anos. | Montreal Canadá |
| (Rodriguez-Gomez et al., 2006) | Registro das propriedades psicométricas do BDI-II aplicado a idosos de Porto Rico. | Porto Rico |
| (Uguz et al., 2007) | Investigar o impacto da depressão maior na qualidade de vida de pacientes com doença de Behcet. | Konya Turquia |
| (Thombs et al., 2007) | Investigar a prevalência e correlação clínica de sintomas de depressão entre pacientes de reconstrução por queimadura. | Baltimore USA |

| REFERÊNCIA | OBJETIVO | LOCAL |
|-------------------------------|---|---------------------|
| (Mogge et al., 2006) | Avaliação da validade do ADI (<i>Assessment os depression Inventory</i>) usando pacientes de laboratório. O ADI Depression Scale foi comparado com outras medidas usadas para avaliar depressão: BDI-II, <i>Zung Self-rating Depression Scale</i> , e <i>Personality Assessment Inventory</i> . | West Virgina USA |
| (Faisal-Cury and Rossi, 2007) | Estimar a prevalência e fatores de risco para ansiedade antenatal e depressão antenatal. | São Paulo Brasil |

4.3 TEORIA DA RESPOSTA AO ITEM

Nas últimas três décadas, o uso da Teoria da Resposta ao Item (TRI) para medida auto-referida tem florescido. A TRI é freqüentemente também chamada de “teoria psicométrica moderna” para distingui-la da teoria clássica de teste (TCT) (Cook et al., 2005). Pesquisadores da área de saúde têm usado a TRI para avaliar as características de medidas existentes (Kirisci et al., 1996), desenvolver novas medidas (Veloze e Peterson, 2001; Cook et al., 2003) e comparar vários grupos de itens a uma métrica matemática comum (McHorney e Cohen, 2000). Os diferentes modelos de TRI propiciam métodos para que algumas das tradicionais limitações das escalas em saúde mental sejam reparadas, especialmente a impossibilidade de comparação de escores inter-escalas, a obtenção de escores intervalares e o desenvolvimento de versões abreviadas de instrumentos originais (mantendo uma performance psicométrica desejada) (Embretson e Reise, 2000).

Duas linhas de desenvolvimento separadas em TRI são as bases de aplicações correntes. Nos Estados Unidos, o começo de TRI é freqüentemente traçado por F. Lord e M. Novick (1968) no seu livro *Statistical Theories of Mental Test Scores*, o qual inclui quatro capítulos sobre TRI, escritos por A. Birnbaum. Em anos subseqüentes, diversos pesquisadores, dentre eles F. Samejima, R. Bock, D. Thissen e E. Muraki desenvolveram modelos e métodos de estimação para os parâmetros dos mesmos, incluindo programas computacionais (BILOG, TESTFACT, MULTILOG e PARSCALE) (Embretson and Reise, 2000). Na Europa, este início é devido a George Rasch (Rasch, 1960). Rasch, um matemático dinamarquês, desenvolveu uma família de modelos de TRI nos quais os parâmetros de indivíduos e de itens estavam totalmente separados. Os métodos de estimação para estes parâmetros foram desenvolvidos por Andersen (1972).

A Teoria da Resposta ao Item (TRI) compreende um grupo de modelos lineares generalizados e procedimentos estatísticos associados que conectam as respostas aos itens, de um determinado instrumento de medida ao qual um indivíduo se submete, a um traço latente subjacente ao mesmo. Estes modelos são equações matemáticas que descrevem a associação entre a probabilidade de uma resposta a um item em particular e o nível de um respondente quanto a um traço latente e as características deste item, usando uma função monotônica não linear. A correspondência entre as respostas previstas a um item e o traço latente é conhecida como curva característica do item.

A maioria das aplicações de TRI assume unidimensionalidade e todos os modelos assumem independência local. Unidimensionalidade significa que somente um traço latente está sendo medido, ou ao menos um é preponderante, pelos itens em um instrumento (ou escala), e independência local significa que os itens são não correlacionados uns com os outros condicionados ao nível do traço latente. As suposições de unidimensionalidade e de independência local estão relacionadas de modo que se uma está satisfeita a outra também está (Hays et al., 2000). Existe evidência de que a suposição de unidimensionalidade pode ser relaxada, bastando que a mesma seja suficiente (Lai et al., 2006; Chan et al., 2004; Kim et al., 2002; McHorney and Cohen, 2000), 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% - McHorney and Cohen, 2000) para que os modelos TRI possam ser utilizados. Um dos procedimentos que podem verificar a unidimensionalidade é a Análise Paralela, que 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. As amostras simuladas são não correlacionadas e são geradas através do método de Monte Carlo.

Modelos TRI têm potenciais vantagens em relação aos métodos da TCT na avaliação de respostas de saúde auto-referidas. Modelos de TRI fornecem estimativas para os itens e para o traço latente que não variam com as características da população com respeito ao traço subjacente (isto é, medidas invariantes podem ser obtidas, independente do instrumento utilizado e dos indivíduos avaliados (Tejadas and Rojas, 2005)), estimativas dos erros padrões condicionais ao nível do traço, e estimativas do traço conectadas ao conteúdo do item.

Teoria da Resposta ao Item (TRI) oferece algumas das melhores alternativas para projetar e otimizar escalas, testes e pesquisas, e realizar análise de itens. Métodos de TRI têm vantagens distintas sobre métodos clássicos, algumas delas são: (a) parâmetros do item e níveis do traço latente do sujeito são independentes, (b) o modelo é expresso ao nível da resposta observada do item e não ao nível do escore de teste observado, (c) a contribuição de cada item para a escala pode ser determinada através da informação TRI, (d) métodos poderosos para detectar funcionamento diferencial de item (DIF) ou viés do item entre populações ou subgrupos estão disponíveis, (e) formas curtas, paralelas e adaptadas são facilmente criadas, e (f) escores de sujeitos podem ser equacionados (ou comparados) mesmo se respondentes respondem questões diferentes (Uttaro and Lehman, 1999). Nos modelos da TRI a precisão da medida é estimada especificamente (medida de erro padrão) para cada nível do traço latente (Tejadas and Rojas, 2005).

Além disso, sendo os modelos TRI focados na medida conjunta de sujeito e itens, significa que eles são colocados na mesma medida continuum com a mesma métrica, ao contrário do TCT que somente coloca pessoas dentro do continuum e supõe que todos os itens contribuem igualmente para a medida do traço latente.

Apesar disso tudo, ainda encontram-se muitos estudos no qual a TCT é utilizada (Richter et al., 1998; Sprohge et al., 2002; Barrick et al., 2002; Streiner, 2003). Já, no Quadro 2, encontram-se alguns trabalhos onde os modelos de TRI foram utilizados para a avaliação de sintomas depressivos utilizando os mais variados instrumentos de medida.

Quadro 2: Exemplos de estudos onde os modelos de TRI foram utilizados na avaliação da intensidade de sintomas depressivos.

| REFERÊNCIA | OBJETIVO | LOCAL |
|----------------------|---|---|
| (Chachamovich, 2008) | Explorar a aplicação do Modelo De Rasch no desenvolvimento de instrumentos de qualidade de vida e atitudes frente ao envelhecimento e utilizar este modelo na validação cultural de um instrumento de depressão geriátrica, assim como testar a adequação de uma escala de resposta Likert de acordo com o nível educacional. | Porto Alegre Brasil |
| (Nuevo et al., 2008) | Analisar a equivalência dos itens e invariância de medida do BDI entre cinco amostras de cinco países europeus. | Europa (Espanha, Reino Unido, Irlanda, Noruega, Finlândia). |
| (Uher et al., 2007) | Examinar e integrar as propriedades psicométricas de três escalas de classificação da intensidade da depressão – <i>17-item Hamilton Depression Rating Scale</i> (HAMD-17), <i>Montgomery-Asberg Depression Rating Scale</i> (MADRS) e <i>Beck Depression Inventory</i> (BDI) - utilizando TRI. | Europa (Bélgica, Croácia, Dinamarca, Alemanha, Itália, Polónia, Eslovênia, Reino Unido). |

| REFERÊNCIA | OBJETIVO | LOCAL |
|--------------------------|--|--------------------|
| (Edelen and Reeve, 2007) | Fornecer uma visão geral da metodologia TRI a fim de revisar algumas das questões críticas associadas com aplicações da mesma, e demonstrar as características básicas do método através de um exemplo. | Providence USA |
| (Carmody et al., 2006) | Comparação das propriedades psicométricas das escalas de depressão HRSD(17) e MADRS usando a teoria clássica (TCT) e TRI. | Dallas USA |
| (Bernstein et al., 2007) | Comparação dos instrumentos QIDS-SR(16) e QIDS-c(16) de sintomas depressivos, em uma amostra de pessoas resistentes a tratamento que inclui minorias étnicas/raciais, utilizando TCT e TRI. | Texas USA |
| (Sharp et al., 2006) | Primeiro trabalho de análise combinada entre TRI e CDFA (análise fatorial de dados categóricos) de uma medida clínica (<i>Short Mood and Feelings Questionnaire – SMFQ</i>) em uma amostra de crianças de 7 a 11 anos de idade. | Texas USA |
| (Pickard et al., 2006) | Ilustrar o uso de TRI na comparação de diferentes subgrupos de pacientes utilizando o CES-D (<i>Epidemiologic Studies-Depression Scale</i>) e informar clínicos e pesquisadores da saúde sobre o funcionamento desta escala e dos sintomas depressivos em pacientes com derrame. | Chicago USA |
| (Stansbury et al., 2006) | Comparação de CTT e TRI na análise da escala para depressão CES-D. | Baltimore USA |
| (Choi et al., 2006) | Usando o modelo de Rasch, este estudo examinou a equivalência de item de duas medidas psiquiátricas, o <i>Harvard Trauma Questionnaire</i> (HTQ), que mede experiência traumática, e o <i>Hopkins Symptom Checklist</i> (HSCL), que avalia sintomas depressivos, entre mães vietnamitas e cambojanas americanas. | Chicago USA |
| (Fliege et al., 2005) | Desenvolver uma aplicação de CAT (Computer Adaptive Tests) baseado em TRI para medir sintomas depressivos. | Berlin Alemanha |

Considerando apenas o uso do BDI para a avaliação da intensidade de sintomas depressivos, o trabalho de Gibbons et al., publicado em 1985, foi o primeiro desta área a trabalhar com TRI, já bastante utilizada na educação.

Entretanto, nem todos os modelos TRI são ótimos para quantificar as mesmas variáveis. Pelo contrário, eles devem ser selecionados de acordo com a natureza da variável que precisa ser quantificada e no modo no qual os testes coletam os dados (ex: formato dos itens) (Tejadas and Rojas, 2005).

Os modelos TRI podem ser divididos em modelos cumulativos e modelos de desdobramento. Detalhes sobre modelos de desdobramento podem ser encontrados no trabalho de Silvana Bortolotti e Dalton Andrade, publicado em 2007.

Os modelos TRI cumulativos podem ser classificados, de um modo geral, de acordo com o tipo de resposta ao item (dicotômica ou politômica), com o número de grupos (ou populações) envolvidos, e de acordo com sua dimensionalidade. Em modelos unidimensionais um único traço latente é considerado suficiente para caracterizar diferenças pessoais ou um traço latente é considerado preponderante.

Modelos de TRI unidimensionais descrevem a relação entre respostas observadas ao item e um traço latente (isto é, construto), usualmente simbolizado por θ , o qual forma a base destas respostas. Eles são apropriados para dados nos quais um único fator comum está sendo avaliado pelos itens. Entretanto, modelos unidimensionais não são adequados para dados nos quais dois ou mais traços latentes têm diferente impacto sobre os itens e pessoas diferem sistematicamente nas estratégias, estruturas de conhecimento ou interpretações que elas aplicam aos itens. Nestes casos, um modelo TRI multidimensional é mais apropriado (Embretson and Reise, 2000).

Entre os modelos unidimensionais encontram-se os modelos logísticos para dados dicotômicos (apresenta o sintoma depressivo ou não apresenta o sintoma depressivo, sucesso ou fracasso). Cabe ressaltar que os modelos TRI dicotômicos utilizam uma simbologia, cuja interpretação vem da área da educação. No contexto do BDI esta simbologia passa a ser referida de outra forma, de acordo com o Quadro 3.

Quadro 3: Simbologia e interpretações dos parâmetros dos modelo TRI dicotômicos na área educacional e no contexto do BDI.

| Símbolo | Parâmetro | Interpretação na educação (Andrade et al., 2000) | Interpretação no contexto do BDI (Cúri, 2006) |
|------------|--|--|--|
| X_{ij} | Resposta do indivíduo j para o item i , podendo assumir valores 0 ou 1 | Quando o indivíduo j acerta a resposta do item i , $X_{ij} = 1$, quando erra $X_{ij} = 0$. | Quando o indivíduo j apresenta o sintoma depressivo descrito pelo item i , $X_{ij} = 1$, quando não apresenta $X_{ij} = 0$. |
| θ_j | Parâmetro do indivíduo. | Nível de habilidade (traço latente) do indivíduo j . | Intensidade de sintomas depressivos do indivíduo j , |
| b_i | Parâmetro de posição do item i . | Dificuldade do item i . | Gravidade do sintoma depressivo descrito pelo item i . |
| a_i | Parâmetro de inclinação do item i . | Discriminação do item i . | Discriminação do item i . |
| c_i | Parâmetro de item | Probabilidade de indivíduos com baixa habilidade responderem corretamente o item i . | Probabilidade de indivíduos apresentarem o sintoma depressivo descrito pelo item i quando têm baixa intensidade de sintomas depressivos. |

A definição destes modelos está baseada em Andrade et. al., 2000 e em Embretson & Reise, 2000.

- Modelo Logístico de 3 Parâmetros.

O modelo de 3 parâmetros prediz a probabilidade de um indivíduo j , com intensidade de sintomas depressivos θ_j , apresentar o sintoma depressivo descrito pelo item i , (isto é, $P(X_{ij} = 1 / \theta_j)$), como segue:

$$P(X_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (1)$$

com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, n$,

onde X_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j

apresenta o sintoma descrito pelo item i , ou 0 caso contrário;

θ_j representa a intensidade de sintomas depressivos do j -ésimo indivíduo, isto é, nível do traço latente do j -ésimo indivíduo;

b_i parâmetro de posição do item i , estimado na mesma escala de θ , representando a gravidade do sintoma depressivo medido pelo item i ;

a_i parâmetro de discriminação (ou inclinação) do item i ;

c_i parâmetro do item que representa a probabilidade de indivíduos apresentarem o sintoma depressivo descrito pelo item i mesmo quando têm baixa intensidade de sintomas depressivos, isto é, baixo nível do traço latente;

D é um fator de escala, constante e igual a 1. Utiliza-se 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

Os resultados do ajuste do modelo TRI para cada item podem ser melhor visualizados através das chamadas Curvas Características do Item. A forma de uma Curva Característica do Item descreve o quanto mudanças no nível do traço latente se relacionam com mudanças na probabilidade de uma resposta específica, ou seja, o quanto mudanças na intensidade de sintomas depressivos se relacionam com a probabilidade de se apresentar o sintoma depressivo descrito por um determinado item.

A escala da intensidade de sintomas depressivos (nomeada *ability* na Figura 1) é arbitrária onde as relações de ordem entre seus pontos é que são importantes. O

parâmetro b é medido na mesma unidade da intensidade de sintomas depressivos e o parâmetro c não depende da escala, pois trata-se de uma probabilidade, assumindo sempre valores entre 0 e 1. O parâmetro b representa a intensidade de sintomas depressivos (nível do traço latente) necessária para uma probabilidade de se apresentar o sintoma depressivo, descrito pelo item, igual a $(1+c)/2$. Já o parâmetro c representa a probabilidade de um indivíduo com baixa intensidade de sintomas depressivos apresentar o sintoma depressivo descrito por um determinado item.

O parâmetro a indica o poder de discriminação do item quanto à intensidade de sintomas depressivos. Baixos valores de a indicam que o sintoma depressivo descrito por um determinado item tem pouco poder de discriminação, isto é, indivíduos com intensidade de sintomas depressivos diferentes têm aproximadamente a mesma probabilidade de apresentarem o sintoma descrito por este item. Já valores muito altos do parâmetro a discriminam os indivíduos basicamente no grupo dos que têm intensidade de sintomas depressivos abaixo do valor do parâmetro b e no grupo dos que possuem intensidade de sintomas depressivos acima do valor do parâmetro b (nestes casos, a curva característica do item é bastante íngreme).

A Figura 1 mostra a Curva Característica do item 1 do BDI, relativo a apresentar o sintoma depressivo tristeza. Ela indica que: a probabilidade dos indivíduos apresentarem o sintoma tristeza quando têm baixa intensidade de sintomas depressivos é igual a 0,02; os indivíduos com intensidade de sintomas depressivos 0,187 têm uma probabilidade igual a $0,51 \left(\frac{1+0,020}{2} \right)$ de apresentarem o sintoma tristeza.

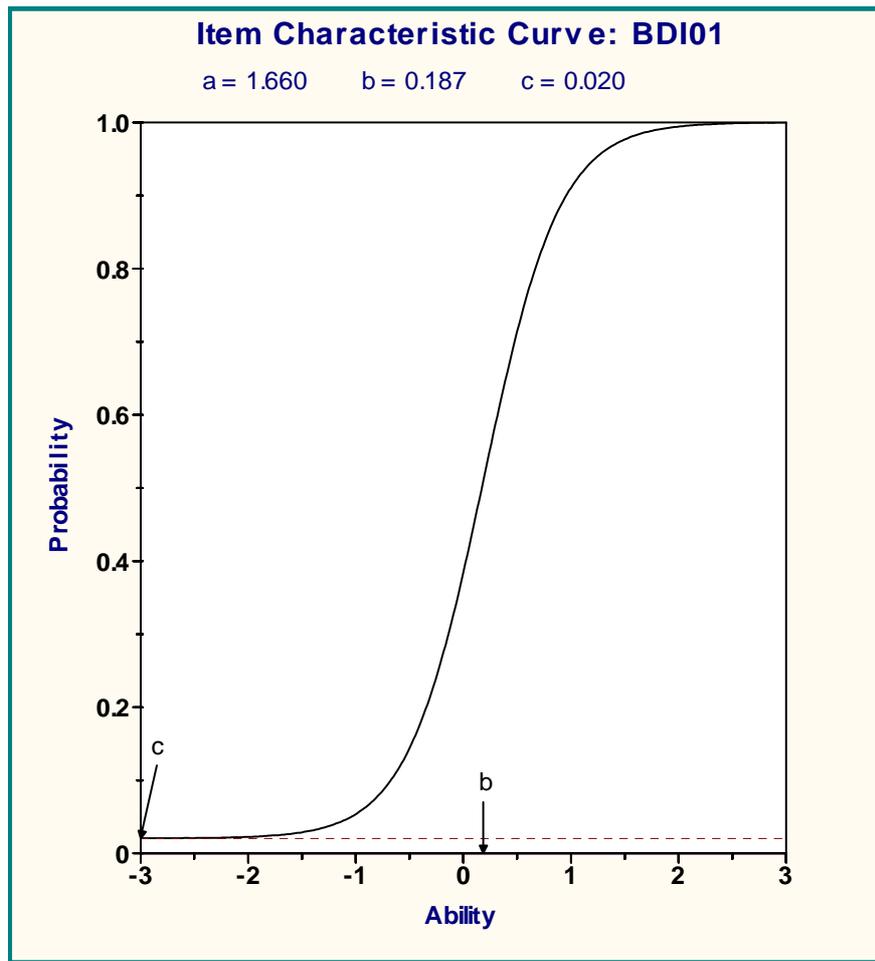


Figura 1: Curva Característica do Item BDI01, relativo ao sintoma depressivo tristeza.

As curvas características do item também são representações gráficas dos modelos TRI que serão definidos a seguir.

- Modelo Logístico de 2 Parâmetros.

O modelo de 2 parâmetros exclui o parâmetro c_i , mantendo os parâmetros de gravidade e discriminação do item. Neste caso,

$$P(X_{ij} = 1 | \theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (2)$$

com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, n$.

- Modelo Logístico de 1 Parâmetro ou Modelo de Rasch (Rasch, 1960).

O modelo de Rasch considera apenas o parâmetro de gravidade do item e é dado por:

$$P(X_{ij} = 1 | \theta_j) = \frac{1}{1 + e^{-D(\theta_j - b_i)}}, \quad (3)$$

com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, n$.

Estes modelos TRI que representam a relação entre um traço latente e uma resposta de item dicotômica são de certa forma limitados, porque muitos dos instrumentos utilizados por psiquiatras não são compostos por itens com respostas do tipo presente *versus* ausente ou verdadeiro *versus* falso. Grande parte dos instrumentos, em especial na área da psiquiatria, inclui itens com múltiplas categorias ordenadas. Os pesquisadores empregam estes formatos por uma variedade de razões, mas principalmente porque eles são mais informativos do que itens com resposta dicotômica.

A natureza da variável intensidade de sintomas depressivos requer o uso de escalas de classificação com preferência a qualquer outro método. Estas escalas apresentam uma série de categorias de resposta graduadas e as categorias representam quantidades crescentes ou decrescentes da variável sendo medida no item. Indivíduos com suas respostas graduem sua concordância ou discordância com a declaração expressada no item. A pontuação dos itens é feita através de números consecutivos inteiros dados às consecutivas categorias. Por exemplo, o item 1 do BDI, relativo à tristeza apresenta quatro afirmações cujos escores atribuídos vão de 0 a 3:

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

Para estes tipos de dados modelos TRI politômicos são necessários para representar a relação não-linear entre o nível de traço e a probabilidade de responder em uma determinada categoria. Entre os modelos politômicos mais utilizados estão o Modelo de Resposta Gradual proposto por Samejima (Samejima, 1969), o Modelo de escala Gradual de Andrich (Andrich, 1978a; Andrich, 1978b), o Modelo de Crédito Parcial de Masters (Masters, 1982) e o Modelo de Crédito Parcial Generalizado (Muraki, 1993).

- Modelo de Resposta Gradual (*Graded-Response Model* - GRM) (Samejima, 1969; 1996)

O modelo de Resposta Gradual, proposto por Samejima em 1969, é apropriado quando as respostas do item podem ser caracterizadas como respostas categóricas ordenadas, como é o caso das escalas de classificação de Likert. O GRM é uma generalização do modelo logístico de 2 parâmetros e é considerado como um modelo TRI “indireto” porque o cálculo da probabilidade condicional de um indivíduo responder uma determinada categoria requer um procedimento de dois passos. Para ajustar o GRM para uma medida para a intensidade de sintomas depressivos, por exemplo, 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 $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, $k = 0, 1, 2, \dots, m_i$ 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, que 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,

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}} \quad (4)$$

As curvas geradas pela expressão (4) 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 e as características do item, isto é, é possível identificar qual a categoria de resposta mais provável de ser escolhida para cada nível de intensidade de sintomas depressivos.

A Figura 2 mostra a curva de categoria de resposta do item 1 do BDI, o qual se refere ao sintoma depressivo tristeza, cujas estimativas dos parâmetros são: $\hat{a}_1 = 1,478$, $\hat{b}_{1,1} = 0,153$, $\hat{b}_{1,2} = 1,280$ e $\hat{b}_{1,3} = 1,897$. Pode-se observar nesta figura que indivíduos com intensidade de sintomas depressivos até 0,153 têm a maior probabilidade de responder a categoria 1 (representada pela curva de cor preta), cuja afirmativa é 'Não me sinto triste'. Indivíduos com intensidade de sintomas depressivos entre 0,153 e 1,280 têm maior probabilidade de responder a categoria 2 (curva de cor azul), cuja afirmativa é 'Eu me sinto triste'. Já os indivíduos com intensidade de sintomas depressivos entre 1,280 e 1,897 mais provavelmente vão escolher a categoria 3 (curva de cor rosa) que afirma 'Estou sempre triste e não consigo sair disto'. Por último, a maior probabilidade de responder a categoria 4 (curva de cor verde) é dos indivíduos com intensidade de sintomas depressivos a partir de 1,897, que afirma 'Estou tão triste ou infeliz que não consigo sair disto'.

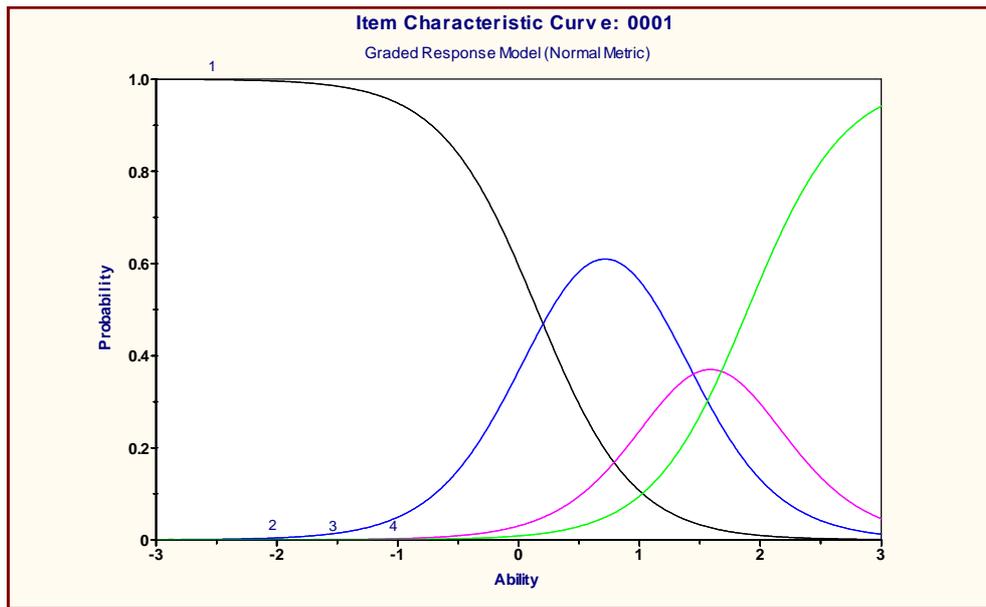


Figura 2: Curva de Categoria de Resposta do Item BDI01, relativo ao sintoma depressivo tristeza.

É importante salientar que aqui, diferente do que ocorre nos modelos dicotômicos, o parâmetro de inclinação (a_i) não pode ser interpretado diretamente como um índice de discriminação do item. Para avaliar diretamente a quantidade de discriminação que os itens fornecem faz-se necessário calcular curvas de informação do item (Embretson and Reise, 2000).

- Modelo de Escala Gradual (*Rating Scale Model – RSM*) (Andrich, 1978)

O modelo de Escala Gradual introduzido por Andrich em 1978 é um caso particular do modelo de Resposta Gradual de Samejima, no qual se faz a suposição que os escores das categorias de resposta são igualmente espaçados. Ele é dado por:

$$P_{i,k}(\theta) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i + d_k)}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_i + d_{k+1})}} \quad (5)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, $k = 0, 1, 2, \dots, m$, onde:

b_i é o parâmetro de posição do item i , ou seja, parâmetro de gravidade do sintoma depressivo descrito pelo item i ;

d_k é um parâmetro de categoria.

Como $P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j) \geq 0$, então $d_k - d_{k+1} \geq 0$, logo $d_1 \geq d_2 \geq \dots \geq d_m$.

A relação entre o modelo de resposta gradual e o modelo de escala gradual é que $b_{i,k} = b_i - d_k$, ou seja, o parâmetro de $b_{i,k}$ do primeiro é decomposto em um parâmetro de gravidade do sintoma descrito pelo item i e em um parâmetro categoria d_k no último. É importante salientar que os parâmetros de categoria d_k não dependem do item, ou seja, são comuns a todos os itens. Se os itens dentro de uma escala têm diferentes formatos, o modelo de escala gradual não é uma escolha apropriada.

- Modelo de Crédito Parcial (*Partial Credit Model* – PCM) (Masters, 1982)

O modelo de crédito parcial de Masters (1982) é também um modelo para análise de categorias de respostas ordenadas. Ele difere, no entanto, do gradual por pertencer à família dos modelos de Rasch, sendo, de fato, uma extensão do modelo de Rasch para itens dicotômicos. Deste modo, se pressupõe que todos os itens têm o mesmo poder de discriminação.

Supondo que o item i tem $(m_i + 1)$ categorias de resposta ordenadas ($k = 0, 1, \dots, m_i$), o modelo de crédito parcial é dado por:

$$P_{i,k}(\theta) = \frac{\exp\left[\sum_{u=0}^k (\theta_j - b_{i,u})\right]}{\sum_{u=0}^{m_i} \exp\left[\sum_{v=0}^u (\theta_j - b_{i,v})\right]} \quad (6)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, $k = 0, 1, 2, \dots, m_i$ e $b_{i,0} \equiv 0$, onde:

$P_{i,k}(\theta_j)$ é a probabilidade de um indivíduo com nível do traço latente θ_j escolher a categoria de resposta k , dentre as $(m_i + 1)$ categorias do item i ;

$b_{i,k}$ é o parâmetro do item que regula a probabilidade de escolher a categoria de resposta k em vez da categoria adjacente $(k - 1)$ no item i . Cada parâmetro $b_{i,k}$ corresponde ao valor do traço latente no qual o indivíduo tem a mesma probabilidade de escolher as categorias k e $(k - 1)$, isto é, onde $P_{i,k}(\theta_j) = P_{i,k-1}(\theta_j)$. Pode ser interpretado como um parâmetro de interseção entre as categorias de resposta do item i .

- Modelo de Crédito Parcial Generalizado (Generalized *Partial Credit Model* – G-PCM) (Muraki, 1992; 1993)

Muraki (1992) desenvolveu uma generalização do PCM que permite que os itens dentro de uma escala tenham diferentes parâmetros de inclinação. A expressão para as curvas de categoria de resposta para o G-PCM é:

$$P_{i,k}(\theta) = \frac{\exp\left[\sum_{u=0}^k Da_i(\theta_j - b_{i,u})\right]}{\sum_{u=0}^{m_i} \exp\left[\sum_{v=0}^u Da_i(\theta_j - b_{i,v})\right]} \quad (7)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, $k = 0, 1, 2, \dots, m_i$.

Os parâmetros de intersecção das categorias $(b_{i,k})$ neste modelo são interpretados do mesmo modo que no PCM, isto é, como o ponto de intersecção de curvas de categoria de resposta adjacentes. Eles são os pontos sobre a escala do traço latente onde uma categoria de resposta torna-se relativamente mais provável do que a resposta precedente, dado que o indivíduo tenha completado os passos prévios. Entretanto, os parâmetros de inclinação (a_i) não são interpretados da mesma forma que no caso dos modelos TRI dicotômicos. A razão disso é que em modelos politômicos a discriminação do item depende de uma combinação de parâmetros de inclinação com a distribuição dos parâmetros de intersecção das categorias.

Os modelos descritos no presente trabalho, que podem conduzir a uma adequada escolha de um modelo de TRI, encontram-se no Quadro 4.

Quadro 4: Distribuição dos modelos TRI de acordo com a natureza do item e o número de parâmetros.

| Número de parâmetros | Natureza do item | |
|--------------------------------------|---|--|
| | Respostas Dicotômicas (sim/não; certo/errado) | Respostas Politômicas (nunca/às vezes/sempre) |
| Dificuldade (Modelos Rasch) | Modelo Logístico de 1 Parâmetro (Rasch) | Modelo de Escala Gradual de Andrich. Modelo de Crédito Parcial. |
| Dificuldade e Discriminação | Modelo Logístico de 2 Parâmetros | Modelo de Resposta Gradual (Samejima). Modelo de Crédito Parcial Generalizado (Muraki). |
| Dificuldade, Discriminação e “Chute” | Modelo Logístico de 3 Parâmetros | |

4.4 FUNCIONAMENTO DIFERENCIAL DO ITEM

Estabelecer equivalência de medidas entre grupos que diferem em características como educação, gênero e raça, por exemplo, é importante em avaliação de saúde. Os grupos podem ser comparados em termos de suas medidas dos construtos de interesse, tais como intensidade de sintomas depressivos, funcionamento físico ou satisfação com cuidado (Teresi and Fleishman, 2007).

Antes de comparar grupos de respondentes (em relação ao gênero, por exemplo) em termos do construto sendo medido, deve-se estar confiante de que os itens que compreendem a medida operam equivalentemente entre os diferentes grupos de respondentes (Teresi and Fleishman, 2007). Em outras palavras, existe a possibilidade que itens, em especial em medidas psicológicas e/ou psiquiátricas, funcionem diferentemente ou de forma viesada para diferentes grupos de respondentes (Embretson and Reise, 2000).

Na literatura sobre TRI, o termo viés do item tem sido essencialmente substituído pela expressão funcionamento diferencial do item (*differential item functioning* – DIF). DIF ocorre quando um item do instrumento não tem a mesma relação com o traço latente entre dois ou mais grupos de respondentes. Mais concretamente, DIF ocorre quando um item apresenta uma Curva de Resposta (*Item Response Curves* – IRC) diferente para cada grupo ou, equivalentemente, qualquer parâmetro do item difere entre os grupos. Se um item está livre deste viés, as respostas para aquele item serão relacionadas somente com o nível do traço latente que o item está tentando medir. Se o item apresenta viés, as respostas ao mesmo serão relacionadas também a algum outro fator, bem como ao nível do traço latente.

Duas formas distintas de DIF são reconhecidas. Elas são chamadas de DIF uniforme (ou unidirecional) e DIF não uniforme. DIF uniforme é aquele que se aplica quando diferenças entre grupos em respostas ao item são encontradas em todos os níveis do traço latente, enquanto que DIF não uniforme é aquele que se aplica quando diferenças entre grupos em respostas ao item dependem dos níveis de traço latente, ou seja, é encontrada uma interação entre o nível do traço latente, designação de grupo e respostas ao item. DIF uniforme e não uniforme são diretamente análogos aos conceitos de confundimento e modificação de efeito, respectivamente, em pesquisa epidemiológica (Crane et al., 2004).

A presença de um grande número de itens com DIF é uma severa ameaça à validade do construto de medidas e a conclusões baseadas nos escores derivados destes

itens com e sem DIF. Diversos estudos examinaram a presença de potencial viés em medidas cognitivas considerando diferentes variáveis demográficas. Estas variáveis demográficas abrangem educação (Jagger et al., 1992; Schmand et al., 1995; Kraemer et al., 1998; Black et al., 1999; Jones and Gallo, 2001), classe social (Jagger et al., 1992), tipo de vizinhança (Espino et al., 2001), etnia (Blesa et al., 2001) e idade (Kraemer et al., 1998).

Vários estudos têm sido publicados com diferentes metodologias para detecção de itens com DIF (Teresi.J.A., 2001; Birgenheimer et al., 2005; Stark et al., 2006; Crane et al., 2006a; Teresi.J.A. et al., 2007; Crane et al., 2007c), enquanto outros tantos aplicam estas metodologias nos muitos instrumentos de medida em que estão trabalhando (Smith and Reise, 1998; Edelen.M.O. et al., 2006; Jones, 2006; Crane et al., 2006b; Bernstein et al., 2007; Reichenheim et al., 2007; Jane et al., 2007; Crane et al., 2007a; Crane et al., 2007b). Considerando o BDI, foram publicados estudos onde foi avaliada a presença de itens com DIF em relação a dois diferentes grupos de idade (Kim et al., 2002), e em relação aos dois grupos de gênero (Hammen and Padesky, 1977; Schaeffer, 1988; Santor et al., 1994; Salokangas et al., 2002; Wenzel et al., 2005).

A avaliação de DIF procede controlando para uma estimativa do traço latente e então examinando se os respondentes em grupos diferentes têm uma distribuição similar das respostas ao item em questão. Primeiro faz-se uma estimativa provisória do construto latente, nomeada de “variável condicionante”. No caso mais simples, isto poderia ser o escore total para o grupo de itens do instrumento de medida do construto. Respondentes em diferentes grupos (por exemplo, homens e mulheres) com o mesmo escore total (ou amplitude do escore total) são comparados para ver se eles diferem em suas respostas para cada item. Se uma diferença de grupo em respostas ao item aparece após o condicionamento a uma estimativa do construto latente, o item é considerado como manifestando DIF (Teresi and Fleishman, 2007).

Uma distinção entre os métodos de detecção de DIF é a natureza da variável condicionante. Alguns métodos usam uma variável condicionante que é baseada em respostas observadas aos itens, tais como amplitude do escore total ou um escore bruto total (ou ponderado). Outros procedimentos assumem a existência de uma variável latente; modelos estatísticos, estimados usando máxima verossimilhança marginal ou outros procedimentos, avaliam o DIF controlando para esta variável.

É importante perceber que, independente do método, benefícios surgem da “purificação” da variável condicionante. Considerando o uso do escore total para um

grupo de itens como uma medida condicionante, se alguns destes itens têm DIF, então o escore total pode ser uma estimativa pobre do construto latente; em tais casos pode resultar uma detecção de DIF pouco acurada. Isto implica que análises DIF são feitas de uma maneira iterativa ou por estágios. Itens que são identificados como tendo DIF são removidos e um novo escore total é calculado para os itens restantes. Este escore é “purificado”, pois não sofre influência de um item com DIF. Este procedimento é repetido até que apenas itens sem DIF compoñham o escore total “purificado”. O novo escore total é usado como uma variável condicionante para detectar DIF nos itens restantes. Eventualmente, para alguns métodos, um grupo de itens “âncora” que não tem DIF significativa é identificado e estes (junto com o item estudado) formam uma medida purificada que é usada como uma estimativa da variável condicionante em estágios subsequentes das análises DIF. Todos os métodos que utilizam itens âncora usam itens livres de DIF para ancorar as relações entre os itens na comparação de grupos, de modo que outros itens possam ser comparados durante detecção de DIF, entretanto as estimativas finais do nível do construto latente podem ser baseadas na inclusão de todos os itens.

Os métodos de detecção de DIF podem ser agrupados em paramétricos ou não-paramétricos. Métodos paramétricos assumem um modelo estatístico (tal como um modelo de regressão) e examinam DIF em termos de estimativas de parâmetros específicos para aquele modelo (por exemplo, coeficientes de regressão). Já métodos não-paramétricos não estimam parâmetros de um modelo estatístico.

Diversos métodos paramétricos, por exemplo, regressão logística, e alguns baseados em TRI, comparam modelos hierárquicos. Um primeiro modelo (completo) contém termos adicionais que são testados para DIF. Um segundo modelo (reduzido) exclui estes termos. Quando os modelos são estimados usando máxima verossimilhança, primeiro se obtém um valor de verossimilhança para cada modelo. A diferença no logaritmo da verossimilhança associado com os modelos reduzido e completo é examinada e forma a base para um teste Qui-Quadrado onde a hipótese nula é que os termos extras no modelo completo não são significativos, isto é, se faz o teste da razão da verossimilhança. Se o modelo ajustado é melhor com os termos extras (modelo completo), o teste da razão da verossimilhança será significativo, indicando a presença de DIF.

Um procedimento paramétrico, que segue esta metodologia para detecção de itens com DIF, é apresentado por Crane e colaboradores (Crane et al., 2006a). Neste

procedimento o modelo de regressão logística ordinal é utilizado empregando estimativas de nível de traço latente através de um modelo de TRI. São gerados três modelos de regressão logística ordinal hierárquicos para cada item e os testes dos modelos começam com o exame da significância estatística do termo de interação entre o nível do traço latente estimado e o indicador de grupo, consistente com DIF não uniforme. Então se volta a atenção para o coeficiente do termo do nível do traço latente em modelos com e sem o termo de grupo. Se incluindo o termo de grupo tem-se um efeito marcante sobre aquele coeficiente, declara-se que ele apresenta DIF uniforme.

Em um contexto diferenciado, um item pode apresentar DIF (viés) para um determinado grupo em função do tipo de pergunta ou afirmação que apresenta. Muitos instrumentos de medida, em especial na área da psiquiatria, apresentam itens que podem funcionar de maneira diferente para diferentes grupos (principalmente no caso de variáveis como gênero). Um estudo de 1988 (Schaeffer, 1988) mostra a diferença entre as distribuições das respostas de homens e mulheres a um item relativo ao choro, tornando claro o funcionamento diferencial deste item quando considerando como grupo de comparação o gênero.

A presença de um grande número de itens com DIF em instrumentos tais como o BDI é uma severa ameaça à validade da medida da intensidade de sintomas depressivos e a conclusões baseadas nos escores derivados dos itens com e sem DIF. Uma possível solução para este problema poderia ser a eliminação dos mesmos do instrumento de medida. No entanto, isto poderia comprometer a medida do traço latente, pois estes itens provavelmente contêm informação relevante à medida do mesmo, pois o BDI foi construído procurando abranger todos os sintomas depressivos observáveis (Cunha, 2001). O uso de um modelo que permita a manutenção de todos os itens no instrumento e, ao mesmo tempo, contemple as diferenças entre os grupos trabalhados vem a ser uma ótima alternativa para a análise de dados oriundos do BDI.

O 'Modelo TRI para itens constrangedores', proposto por Cúri (Cúri, 2006), insere-se dentro desta perspectiva, pois preserva estas características. Este modelo possibilita diferenciar a gravidade da presença do sintoma entre indivíduos de diferentes grupos para itens nos quais existe a presença de DIF quando comparando os mesmos. O modelo TRI para itens constrangedores procura levar em consideração possíveis diferenças na probabilidade de que os indivíduos dos dois grupos expressem a presença do sintoma (item com funcionamento diferencial).

5. OBJETIVOS

Objetivo geral

Este trabalho tem por objetivo evidenciar a potencialidade dos modelos da Teoria da Resposta ao Item caracterizado pelo aproveitamento global das informações quando do uso destes modelos na análise de dados oriundos do Inventário de Depressão Beck para a avaliação da intensidade de sintomas depressivos.

Objetivos específicos

- ARTIGO 1 – apresentar o modelo TRI de Resposta Gradual de Samejima (1969), 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.
- ARTIGO 2 – identificar itens do BDI que apresentem DIF para gênero, isto é, que apresentem viés comparando-se homens e mulheres, através da análise diferencial do item, ajustar o modelo para itens constrangedores, para a amostra de indivíduos considerada, e comparar estes resultados com os do ajuste do modelo logístico de dois parâmetros, que não está habilitado a trabalhar com itens viesados.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Airaksinen, E., M. Larsson, I. Lundberg, and Y. Forsell, 2004, Cognitive functions in depressive disorders: evidence from a population-based study.: *Psychological Medicine*, v. 34, p. 83-91.

Almeida-Filho, N., J. J. Mari, E. Coutinho, J. F. Franca, J. Fernandes, S. B. Andreoli, and E. D. Busnello, 1997, Brazilian multicentric study of psychiatric morbidity. Methodological features and prevalence estimates.: *The British Journal of Psychiatry*, v. 171, no. 12, p. 524-529.

Andalib, A., A. Rezaie, F. Oreizy, K. Shafiei, and S. Baluchi, 2006, A study on stress, depression and NK cytotoxic potential in women with recurrent spontaneous abortion: *Iran J Allergy Asthma Immunol.*, v. 5, no. 1, p. 9-16.

Andersen, E. B., 1972, The numerical solution of a set of conditional estimation equations: *Journal of the Royal Statistical Society, Series B*, v. 34, p. 42-54.

Andrade,DF, H R Tavares, R C Valle. Teoria da Resposta ao Item: conceitos e aplicações. ABE - Associação Brasileira de Estatística. 14º SINAPE - Caxambu. 1-154. 2000.

Ref Type: Serial (Book,Monograph)

Andrich, D., 1978a, A rating formulation for ordered response categories.: *Psychometrika*, v. 43, p. 561-573.

Andrich, D., 1978b, Application of a psychometric model to ordered categories which are scored with successive integers: *Applied Psychological Measurement*, v. 2, p. 581-594.

Barrick, C. B., D. E. Taylor, and E. I. Correa, 2002, Psychometric assessment and clinical application of the Correa-Barrick Depression Rating Scale: *Clin.Nurs.Res.*, v. 11, no. 4, p. 363-381.

Bebbington, P., 2004, The classification and Epidemiology of Unipolar Depression, in M Power ed., *Mood Disorders: Chichester, John Wiley & Sons Ltd.*, p. 3-28.

Beck,AT, R A Steer. Beck Depression Inventory. Manual. San Antonio. TX: Psychological Corporation. 1993.

Beck, A. T., C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh, 1961, An inventory for measuring depression: *Arch.Gen.Psychiatry*, v. 4, p. 561-571.

Bernstein, I. H., A. J. Rush, T. J. Carmody, A. Woo, and M. H. Trivedi, 2007, Clinical vs. self-report versions of the quick inventory of depressive symptomatology in a public sector sample: *J Psychiatr.Res.*, v. 41, no. 3-4, p. 239-246.

Birgenheimer, J. B., S. W. Raudenbush, T. Leventhal, and J. Brooks-Gunn, 2005, Measurement Equivalence and Differential Item Functioning in Family Psychology.: *Journal of Family Psychology*, v. 19, no. 3, p. 441-455.

Black, S. A., D. V. Espino, R. Mahurin, M. J. Lichtenstein, H. P. Hazuda, D. Fabrizio, L. A. Ray, and K. S. Markides, 1999, The influence of noncognitive factors on the Mini-Mental State Examination in older Mexican-Americans: findings from the Hispanic EPESE. Established Population for the Epidemiologic Study of the Elderly: *J Clin.Epidemiol.*, v. 52, no. 11, p. 1095-1102.

Blesa, R., M. Pujol, M. Aguilar, P. Santacruz, I. Bertran-Serra, G. Hernandez, J. M. Sol, and J. Pena-Casanova, 2001, Clinical validity of the 'mini-mental state' for Spanish speaking communities: *Neuropsychologia*, v. 39, no. 11, p. 1150-1157.

Bortolotti, S. L. V. & D. F. Andrade, 2007, Aplicação de um modelo de desdobramento graduado generalizado – GGUM da Teoria da Resposta ao Item: Estudos em Avaliação Educacional, v. 18, no. 37, p. 157-187.

Brands, A. M., B. E. Van den, S. M. Manschot, G. J. Biessels, L. J. Kappelle, E. H. de Haan, and R. P. Kessels, 2007, A detailed profile of cognitive dysfunction and its relation to psychological distress in patients with type 2 diabetes mellitus: *J Int.Neuropsychol.Soc.*, v. 13, no. 2, p. 288-297.

Carmody, T. J., A. J. Rush, I. Bernstein, D. Warden, S. Brannan, D. Burnham, A. Woo, and M. H. Trivedi, 2006, The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures: *Eur Neuropsychopharmacol.*, v. 16, no. 8, p. 601-611.

Caulin-Glaser, T., P. K. Maciejewski, R. Snow, M. LaLonde, and C. Mazure, 2007, Depressive symptoms and sex affect completion rates and clinical outcomes in cardiac rehabilitation: *Prev.Cardiol.*, v. 10, no. 1, p. 15-21.

Chan, K. S., M. Orlando, B. Ghosh-Dastidar, N. Duan, C. D. Sherbourne, 2004, The interview mode effect on the Center for Epidemiological Studies Depression (CES-D) scale: an item response theory analysis: *Med. Care*, v. 42, p. 281-289.

Chachamovich, E., 2008, Teoria de Resposta ao Item: Aplicação do modelo Rasch em desenvolvimento e validação de instrumentos em saúde mental, Universidade Federal do Rio Grande do Sul, 327 p. [tese de doutorado]

Choi, Y., A. Mericle, and T. W. Harachi, 2006, Using Rasch analysis to test the cross-cultural item equivalence of the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist across Vietnamese and Cambodian immigrant mothers: *J Appl.Meas*, v. 7, no. 1, p. 16-38.

Cook, K. F., K. J. O'Malley, and T. S. Roddey, 2005, Dynamic Assessment of Health Outcomes: Time to Let the CAT Out of the Bag?: *Health Services Research*, v. 40, no. 5 Part II, p. 1694-1711.

Cook, K. F., T. S. Roddey, G. M. Gartsman, and S. L. Olson, 2003, Development and psychometric evaluation of the Flexilevel Scale of Shoulder Function: *Med.Care*, v. 41, no. 7, p. 823-835.

Crane, P. K., K. E. Cook, K. Johnson, R. Deyo, and D. Amtmann, 2007a, Differential item functioning impact in a modified version of the Roland-Morris Disability Questionnaire.: *Qual.Life Res*, v. on-line first.

Crane, P. K., L. E. Gibbons, L. Jolley, and B. G. van, 2006a, Differential item functioning analysis with ordinal logistic regression techniques. DIFdetect and difwithpar: *Med.Care*, v. 44, no. 11 Suppl 3, p. S115-S123.

Crane, P. K., L. E. Gibbons, K. Narasimhalu, J. S. Lai, and D. Cella, 2007b, Rapid detection of differential item functioning in assessments of health-related quality of life: The Functional Assessment of Cancer Therapy: *Qual.Life Res*, v. 16, no. 1, p. 101-114.

Crane, P. K., L. E. Gibbons, K. Ocepek-Welikson, and et al., 2007c, A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression.: *Qual.Life Res*, v. on-line first.

Crane, P. K., D. L. Hart, L. E. Gibbons, and K. F. Cook, 2006b, A 37-item shoulder functional status item pool had negligible differential item functioning: *J Clin.Epidemiol.*, v. 59, no. 5, p. 478-484.

Crane, P. K., B. G. van, and E. B. Larson, 2004, Test bias in a cognitive test: differential item functioning in the CASI: *Stat.Med.*, v. 23, no. 2, p. 241-256.

Cuijpers, P., and F. Smit, 2002, Excess mortality in depression: a meta-analysis of community studies: *J Affect.Disord.*, v. 72, no. 3, p. 227-236.

Cunha, J. A., 2001, Manual da versão em português das ESCALAS BECK. Casa do psicólogo: São Paulo.

Cúri, M., 2006, Análise de questionários com itens constrangedores, Universidade de São Paulo, 100 p. [tese de doutorado]

da Silva, T. I. et al., 2007, Validity and reliability of the Portuguese version of the quality of life in epilepsy inventory (QOLIE-31) for Brazil: *Epilepsy Behav.*, v. 10, no. 2, p. 234-241.

Demir, F., P. Ay, M. Erbas, M. Ozdil, and E. Yasar, 2007, [The prevalence of depression and its associated factors among resident doctors working in a training hospital in Istanbul]: *Turk.Psikiyatri.Derg.*, v. 18, no. 1, p. 31-37.

DeVellis, R. F., 2006, Classical test theory: *Med.Care*, v. 44, no. 11 Suppl 3, p. S50-S59.

Dodd, B. G., 1990, The effect of item selection procedure and stepsize on computerized adaptive attitude measurement using the rating scale model.: *Applied Psychological Measurement*, v. 14, p. 355-366.

Ebmeier, K. P., C. Donaghey, and J. D. Steele, 2006, Recent developments and current controversies in depression: *Lancet*, v. 367, no. 9505, p. 153-167.

Edelen, M. O., and B. B. Reeve, 2007, Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement: *Qual.Life Res.*

Edelen.M.O., D. Thissen, Teresi.J.A., M. Kleinman, and K. Ocepek-Welikson, 2006, Identification of Differential Item Functioning Using Item Response Theory and the Likelihood-Based Model Comparison Approach. Application to the Mini-Mental State Examination.: *Medical Care*, v. 44, no. Supplement, p. S134-S142.

Embretson, S. E., and S. P. Reise, 2000, *Item Response Theory for Psychologists*: New Jersey, Lawrence Erlbaum Associates, Inc..

Espino, D. V., M. J. Lichtenstein, R. F. Palmer, and H. P. Hazuda, 2001, Ethnic differences in mini-mental state examination (MMSE) scores: where you live makes a difference: *J Am.Geriatr.Soc.*, v. 49, no. 5, p. 538-548.

Faisal-Cury, A., and M. P. Rossi, 2007, Prevalence of anxiety and depression during pregnancy in a private setting sample: *Arch.Womens Ment.Health*, v. 10, no. 1, p. 25-32.

Fleck, M. P., E. Chachamovich, and C. Trentini, 2006, Development and validation of the Portuguese version of the WHOQOL-OLD module: *Rev.Saude Publica*, v. 40, no. 5, p. 785-791.

Fleck, M. P. A., B. Lafer, E. B. Sougey, J. A. Porto, M. A. Brasil, and M. F. Juruena, 2003, Diretrizes da Associação Médica Brasileira para o tratamento da depressão (versão integral): *Revista Brasileira de Psiquiatria*, v. 25, no. 2, p. 114-122.

Fleck, M. P. A., A. F. B. S. Lima, S. Louzada, G. Schestasky, A. Henriques, V. R. Borges, S. Camey, and LIDO, 2002, Associação entre sintomas depressivos e funcionamento social em cuidados primários à saúde.: *Revista de Saúde Pública*, v. 36, no. 4, p. 431-438.

Fliege, H., J. Becker, O. B. Walter, J. B. Bjorner, B. F. Klapp, and M. Rose, 2005, Development of a computer-adaptive test for depression (D-CAT): *Qual.Life Res*, v. 14, no. 10, p. 2277-2291.

Frasure-Smith, N., F. Lesperance, M. R. Irwin, C. Sauve, J. Lesperance, and P. Theroux, 2007, Depression, C-reactive Protein and Two-year Major Adverse Cardiac Events in Men after Acute Coronary Syndromes: *Biol.Psychiatry*.

Georgiades, A., N. Zucker, K. E. Friedman, C. J. Mosunic, K. Applegate, J. D. Lane, M. N. Feinglos, and R. S. Surwit, 2007, Changes in depressive symptoms and glycemic control in diabetes mellitus: *Psychosom.Med.*, v. 69, no. 3, p. 235-241.

Gibbons, R. D., D. C. Clark, C. S. VonAmmon, and J. M. Davis, 1985, Application of modern psychometric theory in psychiatric research: *J Psychiatr.Res.*, v. 19, no. 1, p. 43-55.

Gorenstein, C., L. Andrade, E. Zanolo, and R. Artes, 2005, Expression of depressive symptoms in a nonclinical Brazilian adolescent sample: *Can.J Psychiatry*, v. 50, no. 3, p. 129-136.

Hagen, B., 2007, Measuring melancholy: a critique of the Beck Depression Inventory and its use in mental health nursing: *Int.J Ment.Health Nurs.*, v. 16, no. 2, p. 108-115.

Hammen, C. L., C. A. Padesky, 1977, Sex differences in the expression of depressive responses on the Beck Depression Inventory: *J. Abnorm. Psychol.*, v. 86, p. 609-614.

Harrington, A. R., and C. C. Greene-Harrington, 2007, Healthy Start screens for depression among urban pregnant, postpartum and interconceptional women: *J Natl.Med.Assoc.*, v. 99, no. 3, p. 226-231.

Hays, R. D., L. S. Morales, and S. P. Reise, 2000, Item response theory and health outcomes measurement in the 21st century: *Med.Care*, v. 38, no. 9 Suppl, p. II28-II42.

Hellerstein, D. J., S. Batchelder, S. Hyler, B. Arnaout, V. Corpuz, L. Coram, and G. Weiss, 2008, Aripiprazole as an adjunctive treatment for refractory unipolar depression: *Prog.Neuropsychopharmacol.Biol.Psychiatry*, v. 32, no. 3, p. 744-750.

Jagger, C., M. Clarke, J. Anderson, and T. Battcock, 1992, Misclassification of dementia by the mini-mental state examination--are education and social class the only factors?: *Age Ageing*, v. 21, no. 6, p. 404-411.

Jane, J. S., T. F. Oltmanns, S. C. South, and E. Turkheimer, 2007, Gender bias in diagnostic criteria for personality disorders: an item response theory analysis: *J Abnorm.Psychol*, v. 116, no. 1, p. 166-175.

Jelinski, S. E., J. E. Magnusson, and W. J. Becker, 2007, Factors associated with depression in patients referred to headache specialists: *Neurology*, v. 68, no. 7, p. 489-495.

Jones, R. N., 2006, Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination. Detecting differential item functioning using MIMIC modeling: *Med.Care*, v. 44, no. 11 Suppl 3, p. S124-S133.

Jones, R. N., and J. J. Gallo, 2001, Education bias in the mini-mental state examination: *Int.Psychogeriatr.*, v. 13, no. 3, p. 299-310.

Kessels, R. P., C. Ruis, and L. J. Kappelle, 2007, The impact of self-reported depressive symptoms on memory function in neurological outpatients: *Clin.Neurol.Neurosurg.*, v. 109, no. 4, p. 323-326.

Kessler, R. C., P. Berglund, O. Demler, R. Jin, D. Koretz, K. R. Merikangas, A. J. Rush, E. E. Walters, and P. S. Wang, 2003, The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R): *JAMA*, v. 289, no. 23, p. 3095-3105.

Kim, Y., P. A. Pilkonis, E. Frank, M. E. Thase, and C. F. Reynolds, 2002, Differential Functioning of the Beck Depression Inventory in Late-Life Patients: Use of Item Response Theory.: *Psychology and Aging*, v. 17, no. 3, p. 379-391.

Kirisci, L., H. B. Moss, and R. E. Tarter, 1996, Psychometric evaluation of the Situational Confidence Questionnaire in adolescents: fitting a graded item response model: *Addict.Behav.*, v. 21, no. 3, p. 303-317.

Knekt, P., O. Lindfors, T. Harkanen, M. Valikoski, E. Virtala, M. A. Laaksonen, M. Marttunen, M. Kaipainen, and C. Renlund, 2008, Randomized trial on the effectiveness of long-and short-term psychodynamic psychotherapy and solution-focused therapy on psychiatric symptoms during a 3-year follow-up: *Psychol Med.*, v. 38, no. 5, p. 689-703.

Kraemer, H. C., D. J. Moritz, and J. Yesavage, 1998, Adjusting Mini-Mental State Examination scores for age and educational level to screen for dementia: correcting bias or reducing validity?: *Int.Psychogeriatr.*, v. 10, no. 1, p. 43-51.

Laget, J., B. Plancherel, P. Stephan, M. Bolognini, M. Corcos, P. Jeammet, and O. Halfon, 2006, Personality and repeated suicide attempts in dependent adolescents and young adults: *Crisis*, v. 27, no. 4, p. 164-171.

Lai, J. S., P.K. Crane, D. Cella. 2006, Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue: *Qual. Life Res.*, v. 15, p. 1179-1190.

Lamberton, A., and T. P. Oei, 2007, A test of the cognitive content specificity hypothesis in depression and anxiety: *J Behav.Ther.Exp.Psychiatry*.

Laximminaryan, R., A.J.Mills, J.G. Breman, et al., 2006, Advancement of global health: key messages from the Disease Control Priorities Project.: *The Lancet*, v. 367, p. 1193-1208.

Lee, M. K., B. H. Son, S. Y. Hwang, W. Han, J. H. Yang, S. Lee, and Y. H. Yun, 2007, Factors affecting health-related quality of life in women with recurrent breast cancer in Korea: *Qual.Life Res*, v. 16, no. 4, p. 559-569.

Lesperance, F. et al., 2007, Effects of citalopram and interpersonal psychotherapy on depression in patients with coronary artery disease: the Canadian Cardiac Randomized Evaluation of Antidepressant and Psychotherapy Efficacy (CREATE) trial: *JAMA*, v. 297, no. 4, p. 367-379.

Lopez, A. D., C. D. Mathers, M. Ezzati, D. T. Jamison, and C. J. L. Murray, 2006, Global and regional burden of disease and risk factors, 2001: systematic analysis of population data.: *The Lancet*, v. 367, p. 1747-1757.

Lord, F. N., and M. R. Novick, 1968, Lord, F. N., and M. R. Novick Statistical theories of mental test scores: Reading, MA.

Marot, R., 08/10/2004, <http://www.psicosite.com.br/tra/hum/depressao.htm>

Masters, G. N., 1982, A Rasch model for partial credit scoring: *Psychometrika*, v. 47, p. 149-174.

McHorney, C. A., and A. S. Cohen, 2000, Equating health status measures with item response theory: illustrations with functional status items: *Med.Care*, v. 38, no. 9 Suppl, p. II43-II59.

Miller, K. M., M. S. Okun, H. F. Fernandez, C. E. Jacobson, R. L. Rodriguez, and D. Bowers, 2007, Depression symptoms in movement disorders: Comparing Parkinson's disease, dystonia, and essential tremor: *Mov Disord.*, v. 22, no. 5, p. 666-672.

Mogge, N. L., J. S. Steinberg, W. Fremouw, and J. Messer, 2006, The Assessment of Depression Inventory (ADI): an appraisal of validity in an outpatient sample: *Depress.Anxiety*.

Moreira, R. O., K. F. Marca, J. C. Appolinario, and W. F. Coutinho, 2007, Increased waist circumference is associated with an increased prevalence of mood disorders and depressive symptoms in obese women: *Eat.Weight.Disord.*, v. 12, no. 1, p. 35-40.

Muraki, E., 1990, Fitting a polytomous item response model to Likert-type data.: *Applied Psychological Measurement*, v. 14, p. 159-176.

Muraki, E., 1992, A generalized partial credit model: Application of an EM algorithm: *Applied Psychological Measurement*, v. 16, p. 159-176.

Muraki, E., 1993, Information functions of the generalized partial credit model.: *Applied Psychological Measurement*, v. 17, p. 351-363.

Nuevo, R., G. Dunn, C. Dowrick, J. L. Vazquez-Barquero, P. Casey, O. S. Dalgard, V. Lehtinen, and J. L. yuso-Mateos, 2008, Cross-cultural equivalence of the Beck Depression Inventory: A five-country analysis from the ODIN study: *J Affect.Disord.*

Ormel, J., A. J. Oldehinkel, W. A. Nolen, and W. Vollebergh, 2004, Psychosocial disability before, during, and after a major depressive episode: a 3-wave population-based study of state, scar, and trait effects: *Arch.Gen.Psychiatry*, v. 61, no. 4, p. 387-392.

Patel, V., A. J. Flisher, S. Hetrick, and P. McGorry, 2007, Mental health of young people: a global public-health challenge.: *The Lancet*, v. 369, p. 1302-1313.

Pickard, A. S., M. R. Dalal, and D. M. Bushnell, 2006, A comparison of depressive symptoms in stroke and primary care: applying Rasch models to evaluate the center for epidemiologic studies-depression scale: *Value.Health*, v. 9, no. 1, p. 59-64.

Pirente, N., C. Blum, S. Wortberg, S. Bostanci, E. Berger, R. Lefering, B. Bouillon, K. E. Rehm, and E. A. Neugebauer, 2007, Quality of life after multiple trauma: the effect of early onset psychotherapy on quality of life in trauma patients: *Langenbecks Arch.Surg.*

Ramana, R., and P. Bebbington, 1995, Social influences on bipolar affective disorders: *Soc.Psychiatry Psychiatr.Epidemiol.*, v. 30, no. 4, p. 152-160.

Rasch, G., 1960. Probabilistic models for some intelligence and attainment tests: Chicago.

Reichenheim, M. E., R. Klein, and C. L. Moraes, 2007, Assessing the physical violence component of the Revised Conflict Tactics Scales when used in heterosexual couples: An item response theory analysis: *Cad.Saude Publica*, v. 23, no. 1, p. 53-62.

Revah-Levy, A., B. Birmaher, I. Gasquet, and B. Falissard, 2007, The Adolescent Depression Rating Scale (ADRS): a validation study: *BMC.Psychiatry*, v. 7, p. 2.

Richter, P., J. Werner, A. Heerlein, A. Kraus, and H. Sauer, 1998, On the validity of the Beck Depression Inventory. A review: *Psychopathology*, v. 31, no. 3, p. 160-168.

Rodriguez-Gomez, J. R., M. G. Davila-Martinez, and L. C. Collazo-Rodriguez, 2006, Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) with Puerto Rican elderly.: *Puerto Rico Health Sciences Journal*, v. 25, no. 2, p. 137-132.

Salokangas R. K., K. Vaahtera, S. Pacriev, B. Sohlman, V. Lehtinen, 2002, Gender differences in depressive symptoms. An artifact caused by measurement instruments? *J. Affect. Disord.*, v. 68, p. 215-220.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph* 17. 1969.

Ref Type: Journal (Full)

Samejima, F.,1996, The graded response model. In W.J. van der Linden & Hambleton, R.K. (Eds.), *Handbook of modern item response theory*: New York, Springer.

Santor, D., J. Ramsay, D. Zuroff, 1994, Nonparametric Item Analysis of the Beck Depression Inventory: Evaluating Gender Item Bias and Response Option Weights: *Psychological Assessment*, v.6, p. 255-270.

Schaeffer, N. C., 1988, An application of item response theory to the measurement of depression.: *Sociological Methodology.*, v. 18, p. 271-307.

Schmand, B., J. Lindeboom, C. Hooijer, and C. Jonker, 1995, Relation between education and dementia: the role of test bias revisited: *J Neurol.Neurosurg.Psychiatry*, v. 59, no. 2, p. 170-174.

Schmidt, L. P., V. N. Teixeira, C. Dall'Igna, D. Dallagnol, and M. M. Smith, 2006, Brazilian Portuguese Language version of the "Tinnitus Handicap Inventory": validity and reproducibility: *Rev.Bras.Otorrinolaringol.(Engl.Ed)*, v. 72, no. 6, p. 808-810.

Schrag, A. et al., 2007, Depression rating scales in Parkinson's disease: Critique and recommendations: *Mov Disord*.

Sharp, C., I. M. Goodyer, and T. J. Croudace, 2006, The Short Mood and Feelings Questionnaire (SMFQ): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children: *J Abnorm.Child Psychol*, v. 34, no. 3, p. 379-391.

Sinikallio, S., T. Aalto, O. Airaksinen, A. Herno, H. Kroger, S. Savolainen, V. Turunen, and H. Viinamaki, 2007, Depression is associated with poorer outcome of lumbar spinal stenosis surgery: *Eur Spine J*.

Smith, L. L., and S. P. Reise, 1998, Gender differences on negative affectivity: an IRT study of differential item functioning on the Multidimensional Personality Questionnaire Stress Reaction Scale
1: *J Pers.Soc.Psychol*, v. 75, no. 5, p. 1350-1362.

Sprohge, E., L. Handler, D. D. Plant, and D. Wicker, 2002, A Rorschach study of oral dependence in alcoholics and depressives: *J Pers.Assess.*, v. 79, no. 1, p. 142-160.

Stansbury, J. P., L. D. Ried, and C. A. Velozo, 2006, Unidimensionality and bandwidth in the Center for Epidemiologic Studies Depression (CES-D) Scale: *J Pers.Assess.*, v. 86, no. 1, p. 10-22.

Stark, S., O. S. Chernyshenko, F. Drasgow, and B. A. Williams, 2006, Examining assumptions about item responding in personality assessment: should ideal point methods be considered for scale development and scoring?: *J Appl. Psychol*, v. 91, no. 1, p. 25-39.

Streiner, D. L., 2003, Being inconsistent about consistency: when coefficient alpha does and doesn't matter: *J. Pers. Assess.*, v. 80, no. 3, p. 217-222.

Sullivan, P. F., M. C. Neale, and K. S. Kendler, 2000, Genetic epidemiology of major depression: review and meta-analysis: *Am. J. Psychiatry*, v. 157, no. 10, p. 1552-1562.

Tejadas, A. J. R., and O. M. L. Rojas, 2005, Application of an IRT Model for measuring Health Related Quality of Life.: *Social Indicators Research*, v. 74, p. 369-394.

Teresi, J. A., and J. A. Fleishman, 2007, Differential item functioning and health assessment: *Qual.Life Res*.

Teresi.J.A., 2001, Statistical methods for examination of differential item functioning (DIF) with applications to cross-cultural measurement of functional, physical and mental health.: *Journal of Mental Health and Aging*, v. 7, p. 31-40.

Teresi.J.A., K. Ocepek-Welikson, M. Kleinman, and et al., 2007, Evaluating measurement equivalence using the item response theory log-likelihood ratio (IRTLR) method to assess differential item functioning (DIF): applications (with illustrations) to

measure of physical functioning ability and general distress.: *Qual.Life Res*, v. on-line first.

Theme-Filha, M. M., C. L. Szwarcwald, and P. R. Souza-Junior, 2005, Socio-demographic characteristics, treatment coverage, and self-rated health of individuals who reported six chronic diseases in Brazil, 2003: *Cad.Saude Publica*, v. 21 Suppl, p. 43-53.

Thombs, B. D., J. M. Haines, M. G. Bresnick, G. Magyar-Russell, J. A. Fauerbach, and R. J. Spence, 2007, Depression in burn reconstruction patients: symptom prevalence and association with body image dissatisfaction and physical function: *Gen.Hosp.Psychiatry*, v. 29, no. 1, p. 14-20.

Uguz, F., R. Dursun, N. Kaya, and A. S. Cilli, 2007, Quality of life in patients with Behcet's disease: the impact of major depression: *Gen.Hosp.Psychiatry*, v. 29, no. 1, p. 21-24.

Uher, R. et al., 2007, Measuring depression: comparison and integration of three scales in the GENDEP study: *Psychol Med.*, p. 1-12.

Uttaro, T., and A. Lehman, 1999, Graded response modeling of the Quality of Life Interview: *Evaluation and Program Planning*, v. 22, p. 41-52.

Velozo, C. A., and E. W. Peterson, 2001, Developing meaningful Fear of Falling Measures for community dwelling elderly: *Am.J Phys.Med.Rehabil.*, v. 80, no. 9, p. 662-673.

Wenzel, A., R.A. Steer, A. T. Beck, 2005, Are there any gender differences in frequency of self-reported somatic symptoms of depression? *J. Affect. Disord.*, v. 89, p. 177-181.

Williams, J. G., D. H. Barlow, and W. S. Agras, 1972, Behavioral measurement of severe depression: *Arch.Gen.Psychiatry*, v. 27, no. 3, p. 330-333.

Yakar, T., A. Baran, S. Gungor, B. Altinsoy, M. Yalcinsoy, G. Can, and E. Akkaya, 2007, [The factors affecting Beck depression scale in asthmatic patients.]: *Tuberk.Toraks.*, v. 55, no. 1, p. 11-17.

Zimmermann, P. R., S. A. Camey, and J. J. Mari, 2006, A cohort study to assess the impact of depression on patients with kidney disease: *Int.J Psychiatry Med.*, v. 36, no. 4, p. 457-468.

7. ARTIGOS

- 7.1 Artigo 1: Teoria da Resposta ao Item: aplicação na avaliação da intensidade de sintomas depressivos 54**
- 7.2 Artigo 2: Modelo da Teoria da Resposta ao Item para itens viesados: Funcionamento Diferencial do Item no Inventário de Depressão Beck 80**

7.1 Artigo 1

**TEORIA DA RESPOSTA AO ITEM:
APLICAÇÃO NA AVALIAÇÃO DA
INTENSIDADE DE SINTOMAS DEPRESSIVOS**

Stela Maris de Jezus Castro

Clarissa Trentini

João Riboldi

RESUMO

A intensidade de sintomas depressivos é um traço latente, que pode ser medido através do Inventário de Depressão Beck (BDI), e que pode ser avaliada 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 4025 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 com problema de entendimento da escala, de modo que estes itens tiveram que ser recategorizados. Os itens com maior capacidade de discriminação são relativos à tristeza, pessimismo, sentimento de fracasso, insatisfação, auto-aversão, indecisão e dificuldade de trabalhar. Os itens mais graves são os relacionados com perda de peso, retraimento social e idéias suicidas. O grupo dos 202 indivíduos com as maiores intensidades de sintomas depressivos é composto por 74% de mulheres e praticamente 84% possuem 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.

ABSTRACT

The intensity of depressive symptoms is considered a latent trait that can be measured by Beck Depression Inventory (BDI), and that can be assessed through Item Response Theory (IRT). This study used the Graded-Response model (GRM) in assessing the intensity of depressive symptoms in 4025 individuals who responded to BDI, in order to efficiently exploit the available information on different aspects enabled by the use of this methodology. The fit of this model was made in the software PARSCALE. We identified 13 items of BDI with problem of understanding of the scale, so that these items had to be categorized again. The items with greater power of discrimination are 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 is composed of 74% of women and almost 84% have a diagnosis of some psychiatric disorder. The results show gains resulting from use of IRT in the analysis of latent traits.

Keyword: Item Response Theory, Latent Trait, Intensity of Depressive Symptoms, Beck Depression Inventory.

INTRODUÇÃO

Construtos psicológicos são usualmente definidos como variáveis latentes que são a base do comportamento. Estas variáveis, 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, em geral, são 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 de indivíduos.

Um construto psicológico de alta prevalência é a depressão maior (transtorno psiquiátrico de humor). Estudos epidemiológicos apresentam a prevalência anual da depressão maior, na população geral, variando entre 3% e 11%^{1,2} e a prevalência para a vida inteira em 16,2%². No Brasil, os estudos apontam a prevalência da depressão para a vida inteira entre 2,8% e 19,2%^{3,4}. Entre as doenças não fatais, a depressão maior é a causa líder de anos vividos com incapacidade⁵ e, em 2020, só será superada por doenças cardíacas⁶. Nos países em desenvolvimento, depressão e doenças crônicas, tais como doenças cardíacas e câncer, estão se tornando rapidamente as causas líderes não só de incapacidade como também de morte⁷.

Um traço latente que é de extrema relevância para se verificar o estágio da depressão e avaliar seu desfecho é a medida da intensidade de sintomas depressivos, a qual pode ser estimada a partir de instrumentos compostos por itens que representam os mais variados sintomas depressivos observáveis, como o Inventário de Depressão Beck⁹ (BDI).

Até recentemente, a modelagem estatística mais utilizada na estimação do traço latente foi a Teoria Clássica de Teste (TCT)¹⁰, que utiliza o escore total como estimativa para o traço latente do indivíduo. Uma revisão desta metodologia pode ser encontrada no trabalho de DeVellis¹¹, publicado em 2006. Apesar da 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 de Teoria da Resposta ao Item¹².

A Teoria da Resposta ao Item (TRI), já bastante utilizada na área psiquiátrica^{13,14,15}, 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 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; formas curtas, paralelas e adaptadas são facilmente criadas, e escores de indivíduos podem ser equacionados (comparados) mesmo quando eles respondem itens diferentes¹⁶.

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, 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 sendo medido. Diante disso, o presente estudo tem por objetivo caracterizar o modelo TRI de Resposta Gradual de Samejima (1969)¹⁷, 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.

MÉTODOS

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 2001¹⁸. Os 4025 sujeitos do estudo estão distribuídos em três grandes grupos:

Grupo 1: Pacientes psiquiátricos ($n = 1138$) - pacientes com transtornos depressivos, transtornos de ansiedade, transtorno de pânico sem ou com agorafobia, transtorno obsessivo-compulsivo, fobia social, fobia simples, agorafobia, transtorno de dependência do álcool, dependência de outras substâncias ou transtornos psiquiátricos não especificados.

Grupo 2: Pacientes de clínica médica ($n = 490$) - pacientes com cardiopatia, com dor crônica, portadores de HIV assintomáticos e sintomáticos, com disfunção erétil, com obesidade, com diabetes mellitus tipo II, ou pacientes de unidades de cuidados primários em saúde.

Grupo 3: Amostra não-clínica, constituída por grupos da população geral ($n = 2397$) - composta por universitários, reclusos de penitenciária de segurança máxima, bombeiros, funcionários de hospital geral, adolescentes, idosos.

Todos os participantes responderam ao BDI, composto por uma escala de auto-relato, de 21 itens, cada um com quatro afirmativas, subentendendo graus crescentes de intensidade de sintomas depressivos¹⁸, com escores variando de 0 a 3 (Anexo B). Os itens do BDI foram selecionados com base em observações e relatos de sintomas e atitudes mais freqüentes em pacientes psiquiátricos, com transtornos depressivos¹⁸: Tristeza; Pessimismo; Sentimento de fracasso; Insatisfação; Culpa; Punição; Auto-aversão; Auto-acusações; Idéias suicidas; Choro; Irritabilidade; Retraimento social; Indecisão; Mudança na auto-imagem; Dificuldade de trabalhar; Insônia; Fatigabilidade; Perda de apetite; Perda de peso; Preocupações somáticas; Perda da libido. Neste estudo¹⁸ 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 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 de sintomas depressivos.

Teoria da Resposta ao Item^{19,20}: Um modelo TRI politômico adequado a dados oriundos do BDI é o Modelo de Resposta Gradual (*Graded-Response Model - GRM*) proposto por Samejima em 1969¹⁷, 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 $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, $k = 0, 1, 2, \dots, m_i$ 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 PARSCALE²¹, 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_{ik} para cada item, ou seja, $b_{i,k} = b_i - c_{ik}$. 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 item²² e os parâmetros c_{ik} representam as distâncias entre 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 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 que se defina 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:

- 4 Não me sinto triste.
- 5 Eu me sinto triste.
- 6 Estou sempre triste e não consigo sair disto.
- 7 Estou tão triste ou infeliz que não consigo suportar.

Pode-se observar nesta figura que indivíduos com intensidade de sintomas depressivos até 0,153 têm a maior probabilidade de responder a categoria 1 (representada pela curva de cor preta), indivíduos com intensidade de sintomas depressivos entre 0,153 e 1,280 têm maior probabilidade de responder a categoria 2 (curva de cor azul) e os indivíduos com intensidade de sintomas depressivos entre 1,280 e 1,897 são mais prováveis de escolherem a categoria 3 (curva de cor rosa). Por último, a maior probabilidade de responder a categoria 4 (curva de cor verde) é dos indivíduos com intensidade de sintomas depressivos a partir de 1,897.

A partir dos modelos 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 pode-se avaliar quais sintomas depressivos discriminam melhor a população quanto à intensidade de sintomas depressivos²⁰. Nos modelos 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* do 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 é 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 de 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 é não relacionada à resposta do item anterior) e unidimensionalidade (todos os itens do instrumento medem 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^{19,23}. Existe evidência de que a suposição de unidimensionalidade pode ser relaxada, bastando

que a mesma seja suficiente²⁴⁻²⁷, 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%²⁴) para que os modelos TRI possam ser utilizados. Em vista disso, foi avaliada a unidimensionalidade do BDI através do procedimento conhecido como Análise Paralela^{23,28-31}, viabilizada através do uso de uma macro³² 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, 4025 observações e 21 variáveis). As amostras simuladas são não correlacionadas e são geradas através do método de Monte Carlo (foram geradas 5000 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 totalidade dos indivíduos se declarou como 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 freqüentes 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^{22,33}, pois foi encontrado um fator preponderante com 38,7% de explicação da variação total (Figura 1 – Anexo D).

O ajuste do modelo de Resposta Gradual mostrou, a partir das curvas produzidas pela expressão (1), que treze dos vinte e um itens do BDI apresentam problemas de entendimento da escala, pois 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 nas Figuras 2a e 2c para os itens relativos ao pessimismo (item 2) e à punição (item 6), respectivamente. 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 (Figuras 2b e 2d). Considerando o ajuste 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 (Figura 3), que são fortemente influenciadas pelo parâmetro de inclinação (a_i - Tabela 2), observa-se que, considerando como ponto de corte o número um^{19,22} 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 a sua curva de informação do item revela que estes itens são pobres em relação à capacidade de discriminação (Anexo C). O item relativo à irritabilidade (item 11) é o que menos contribui com informação à medida de intensidade de sintomas depressivos (item 11 – Figura 3h). 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 (item 3 - Figura 3c) tem maior poder de discriminação da população quanto à intensidade de sintomas depressivos quando o escore TRI está no intervalo de $[0,7;2]$, o que equivale aproximadamente ao intervalo $[21;35]$ na escala do escore total. O item relativo a tristeza (item 1 – Figura 6a), discrimina melhor a população quando o escore TRI vai de 0,7 a 1,7, o equivalente a ir de 21 a 32 no escore total. As curvas de informação do item, assim como as curvas de categoria de resposta, para os 21 itens do BDI estão apresentadas no anexo C. Para trazer os escores 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 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 TRI. A Tabela 1 traça um perfil deste grupo: praticamente 84% destes indivíduos estão no grupo psiquiátrico, têm em média 39 anos de idade, 74% são mulheres, aproximadamente metade estudou menos de cinco anos, e quase 40% são casados.

A Curva de Informação do Teste (Figura 22 – Anexo C) 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 depressão maior¹⁸. A avaliação destes itens, quanto a sua quantidade de informação psicométrica e seu grau de gravidade do traço latente sendo medido, é um ganho importante dos modelos 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 entram com o mesmo peso^{8,16}.

Outro ganho substancial dos modelos 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, isto é, 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^{19,20} 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 moderada¹⁸. Apenas o item relativo à auto-aversão (item 7 – Figura 6e) se mantém discriminando bem também na região de intensidade de sintomas depressivos grave, 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 Uher¹⁴ e colaboradores, publicado em 2007, 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 e insatisfação (Tabela 2), sendo estes itens também classificados neste grupo no trabalho de Uher¹⁴ e colaboradores. Indecisão e dificuldade de trabalhar ficam, agora, no grupo com moderada discriminação junto a todos os outros, com exceção de irritabilidade que é o único item classificado como tendo baixa discriminação (Tabela 2).

O valor encontrado para o percentil 95 do escore TRI, equivalente a 31 no escore total, também se enquadra na classificação grave para a intensidade de sintomas depressivos. Neste grupo, onde todos os 202 indivíduos são classificados como tendo intensidade de sintomas depressivos grave (seus escores 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 homens³⁴.

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 inclinação e posição), quando aplicados ao mesmo instrumento. No estudo de Cúri²², onde foi ajustado o modelo logístico de 3 parâmetros¹⁹ a 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, o qual é um dos mais graves e também um dos que menos discrimina (Tabela 2). Apesar de 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 razões: pode 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 artrite³⁵⁻⁴²), ou pode ser uma conseqüência física de outros sintomas, isto é, antes o indivíduo sente, por exemplo, tristeza, desânimo, entre outros, em seguida deixa de se cuidar, inclusive de sua alimentação, para conseqüentemente, diminuir de peso.

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 que os níveis estimados para a intensidade de sintomas depressivos (score TRI) são calculados considerando as diferentes contribuições de cada item (sintoma depressivo) para o traço latente medido, ao contrário da TCT, onde os respondentes com o mesmo score 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 TRI diferenciados, este último calculado considerando os diferentes perfis de resposta a itens que têm pesos diferentes no cálculo (exceção para os 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 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.

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;25:114-122.
- (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;289:3095-3105.
- (3) Almeida-Filho N, MariJ.J., Coutinho E, et al. Brazilian multicentric study of psychiatric morbidity. Methodological features and prevalence estimates. *The British Journal of Psychiatric.* 1997;171:524-529.
- (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. *Caderno de 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;367:1747-1757.
- (6) Innovative care for chronic conditions: building blocks for action: global report. World Health Organization. 2002. Disponível em: www.who.int/ . (Acessado em 30 de junho de 2008)
- (7) Health and ageing: a discussion paper. World Health Organization. 2001. Disponível em: www.who.int/ . (Acessado em 30 de junho de 2008)
- (8) Tejadas AJR, Rojas OML. Application on an IRT Model for measuring Health Related Quality of Life. *Social Indicators Research.* 2005;74:369-394.
- (9) Beck, AT & Steer, RA. *Beck Depression Inventory. Manual.* San Antonio, TX: Psychological Corporation: 1993.
- (10) Lord, FM. & Novick MR. *Statistical theories of mental test scores.* Reading, MA: 1968.
- (11) DeVellis RF. Classical test theory. *Med Care.* 2006; 44: S50-S59.
- (12) Lord, FM. *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Erlbaum, 1980.

- (13) Chachamovich, E. *Teoria de Resposta ao Item: Aplicação do modelo Rasch em desenvolvimento e validação de instrumentos em saúde mental*. [Universidade Federal do Rio Grande do Sul. 327 p. 2008].
- (14) 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.
- (15) Nuevo R, Dunn G, Dowrick C, Vazquez-Barquero JL, Casey P, Dalgard O S, et al. Cross-cultural equivalence of the Beck Depression Inventory: A five-country analysis from the ODIN study. *J Affect. Disord*. 2008.
- (16) Uttaro T, Lehman A. Graded response modeling of the Quality of Life Interview. *Evaluation and Program Planning*. 1999; 22: 41-52.
- (17) Samejima, F. *Estimation of latent ability using a response pattern of graded scores*. Psychometrika Monograph 17. 1969.
- (18) Cunha JA. *Manual da versão em português das ESCALAS BECK*. São Paulo: 2001.
- (19) Andrade, D. F., Tavares, H. R., and Valle, R. C. Teoria da Resposta ao Item: conceitos e aplicações. 28-7-2000. Caxambu, ABE - Associação Brasileira de Estatística. 14º SINAPE.
- (20) Embretson SE, Reise SP. *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc.;2000.
- (21) *PARSCALE*. [Chicago: Scientific Software International, Inc.: 2003].
- (22) Cúri M. *Análise de questionários com itens constrangedores*. [São Paulo: Universidade de São Paulo. 2006].
- (23) Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000;38:II28-II42.
- (24) McHorney CA, Cohen AS. Equating health status measures with item response theory: illustrations with functional status items. *Med Care*. 2000;38:II43-II59.
- (25) 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;42:281-289.
- (26) 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;17:379-391.

- (27) 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.
- (28) Glorfeld LW. An improvement on Horn's Parallel Analysis Methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement.* 1995;55:377-393.
- (29) 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.
- (30) 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.
- (31) Franklin SB, Gibson DJ, Robertson PA, Pohlmann JT, Fralish JS. Parallel Analysis: a method for determining significant principal components. *Journal of Vegetation Science.* 1995;6:99-106.
- (32) Determining the Dimensionality of Data: A SAS Macro for Parallel Analysis. [Portland.: SUGI 28, Paper 90-28; 2007].
- (33) 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.
- (34) 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.: 2000:288-324.
- (35) 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;49:1834-1840.
- (36) 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.
- (37) 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.

- (38) Golden SH, Lee HB, Schreiner PJ et al. Depression and type 2 diabetes mellitus: the multiethnic study of atherosclerosis. *Psychosom Med.* 2007;69:529-536.
- (39) Kamphuis MH, Geerlings MI, Tijhuis MA et al. Physical inactivity, depression, and risk of cardiovascular mortality. *Med Sci Sports Exerc.* 2007;39:1693-1699.
- (40) Knol MJ, Heerdink ER, Egberts AC et al. Depressive symptoms in subjects with diagnosed and undiagnosed type 2 diabetes. *Psychosom Med.* 2007;69:300-305.
- (41) 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;31:105-107.
- (42) 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;370:851-858.

TABELAS:

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

| Características sócio-demográficas** | 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

** O n° de indivíduos varia de acordo com as características sócio-demográficas devido a ocorrência de valores faltantes.

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

| 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;

a_i parâmetro de inclinação comum a todas as categorias do item i ;

b_i parâmetro de gravidade do sintoma avaliado pelo item i ;

$b_{i,1}$ ponto de interseção entre as categorias de resposta 1 e 2;

$b_{i,2}$ ponto de interseção entre as categorias de resposta 2 e 3;

$b_{i,3}$ ponto de interseção entre as categorias de resposta 3 e 4.

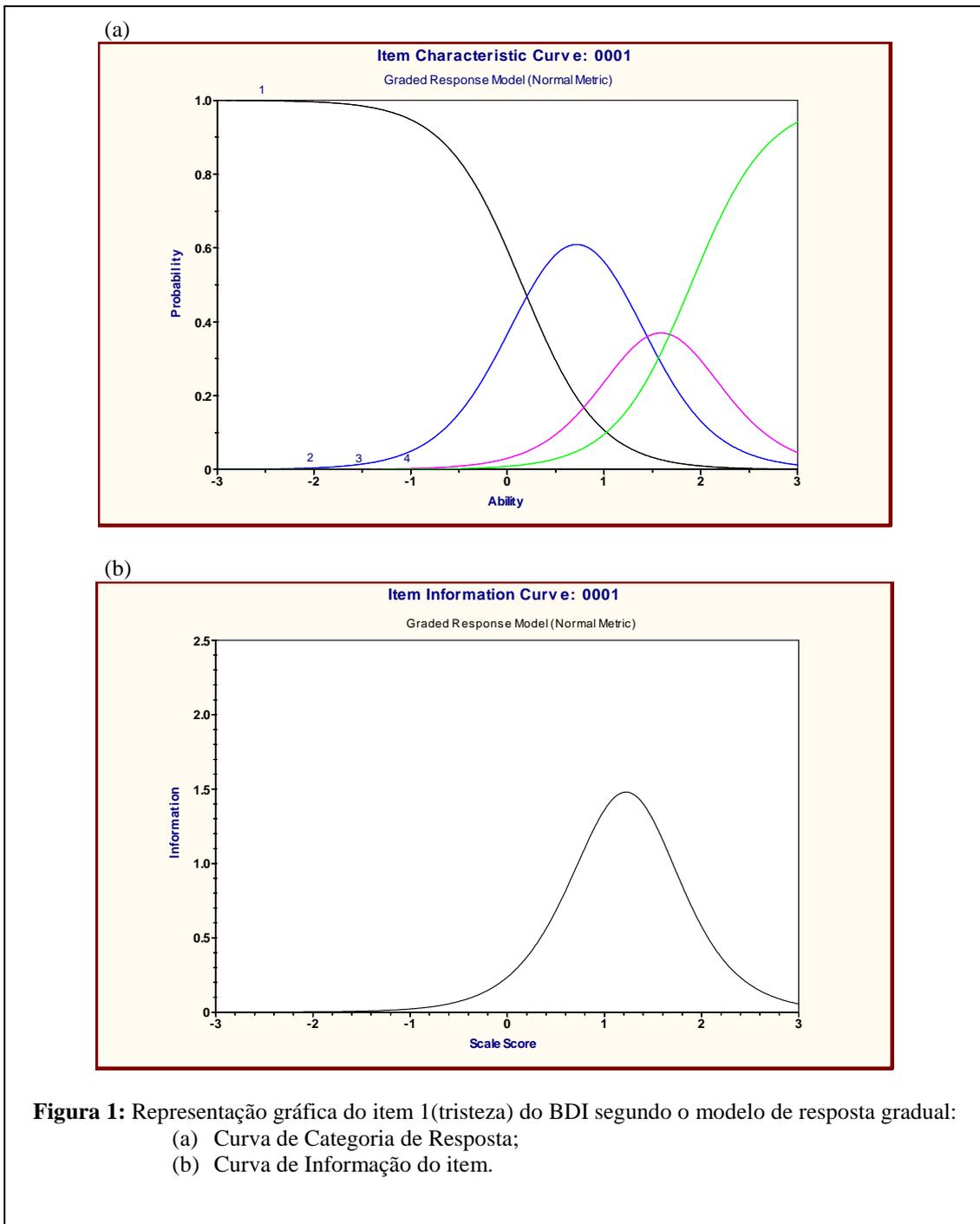
Tabela 3: Descrição dos indivíduos com estimativa de intensidade de sintomas depressivos acima do percentil 95 ($\hat{\theta} = 1,6^{}$).**

| 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

** Na escala do escore total este percentil 95 equivale a 31.

FIGURAS:



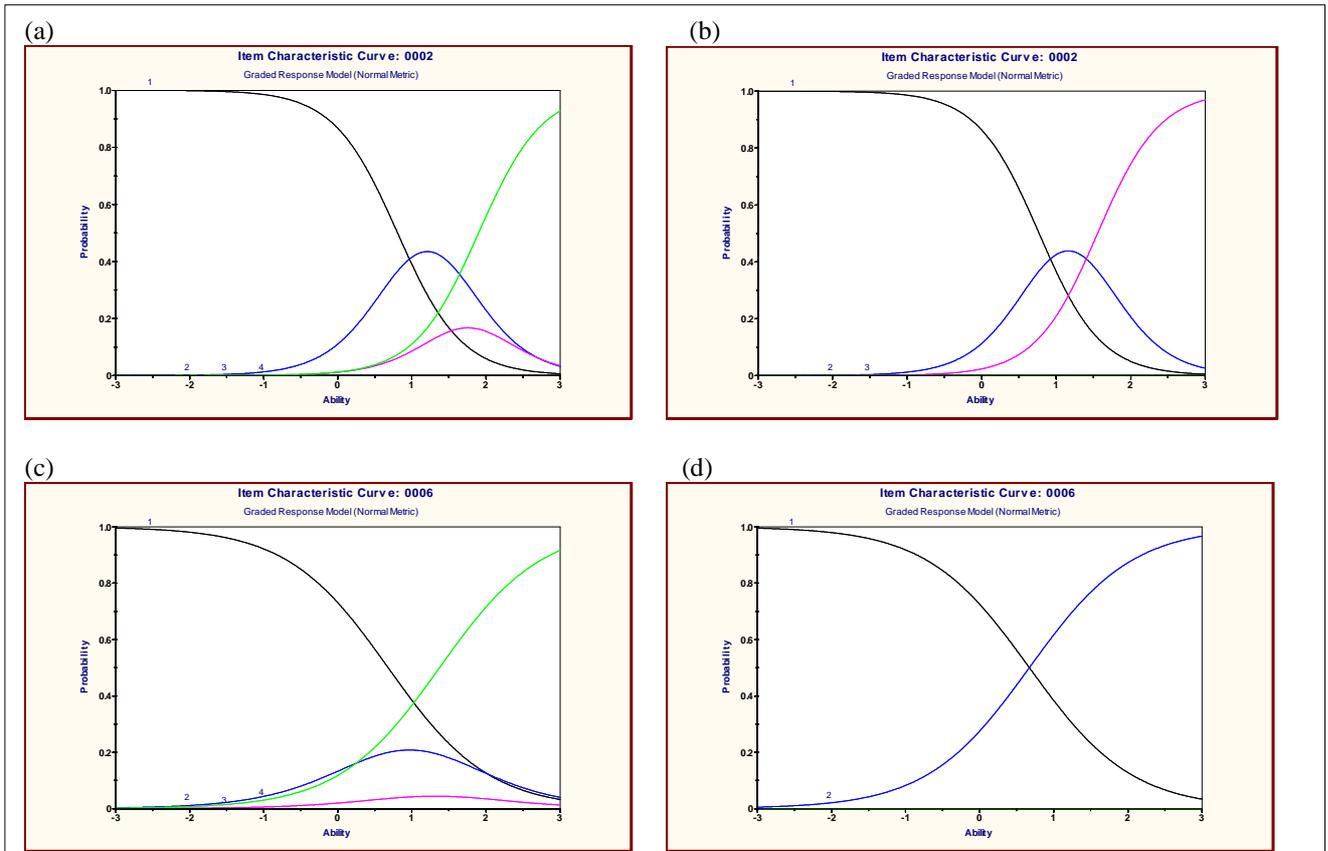


Figura 2: Curvas de categoria de resposta para os itens 2 e 6 do BDI, cujos escores significam:

(a)

- 1 (preto) Não estou especialmente desanimado quanto ao futuro.
- 2 (azul) Eu me sinto desanimado quanto ao futuro.
- 3 (rosa) Acho que nada tenho a esperar.
- 4 (verde) Acho o futuro sem esperança e tenho a impressão de que as coisas não podem melhorar.

(b)

- 1 (preto) Não estou especialmente desanimado quanto ao futuro.
- 2 (azul) Eu me sinto desanimado quanto ao futuro.
- 3 (rosa) {
 - Acho que nada tenho a esperar.
 - Acho o futuro sem esperança e tenho a impressão de que as coisas não podem melhorar.

(c)

- 1 (preto) Não acho que esteja sendo punido.
- 2 (azul) Acho que posso ser punido.
- 3 (rosa) Creio que vou ser punido.
- 4 (verde) Acho que estou sendo punido.

(d)

- 1 (preto) Não acho que esteja sendo punido.
- 2 (azul) {
 - Acho que posso ser punido.
 - Creio que vou ser punido.
 - Acho que estou sendo punido.

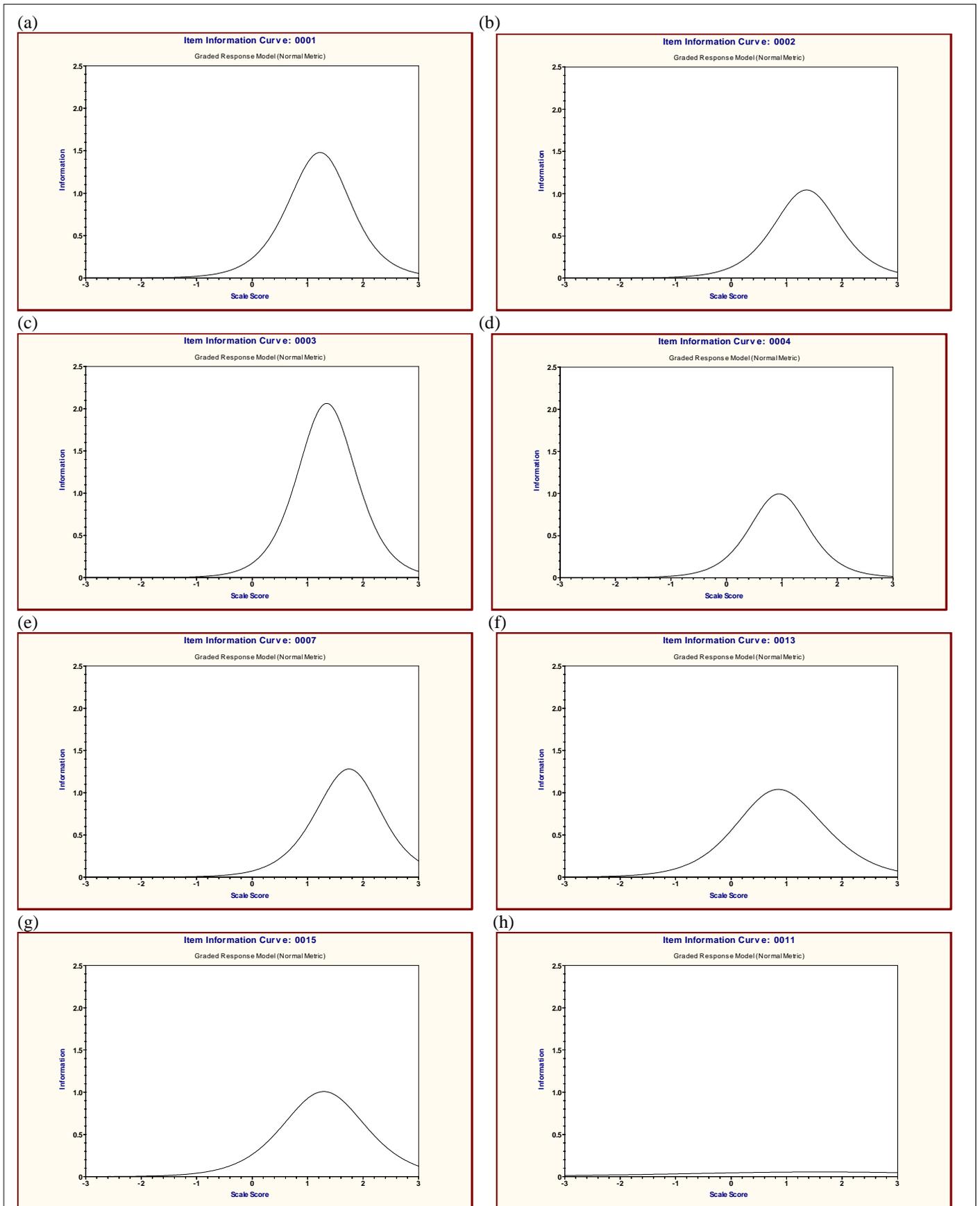


Figura 3: Curvas de Informação do Item para os itens 1 - tristeza (a), 2 - pessimismo (b), 3 - sentimento de fracasso (c), 4 - insatisfação (d), 7 - auto-aversão (e), 13 - indecisão (f), 15 - dificuldade de trabalhar (g) e 11 - irritabilidade (h) do BDI.

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

| Itens | Padrão de resposta dos indivíduos n° | | | | | | | | |
|-----------------------------|--------------------------------------|-------------|--------------|--------------|-------------|--------------|--------------|--------------|--------------|
| | 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.

** Erro padrão da estimativa da intensidade de sintomas depressivos.

*** Valores equivalentes dos escores TRI na escala do escore total.

7.2 Artigo 2

**MODELO DA TEORIA DA RESPOSTA AO ITEM
PARA ITENS VIESADOS:
FUNCIONAMENTO DIFERENCIAL DO ITEM NO
INVENTÁRIO DE DEPRESSÃO BECK**

Stela Maris de Jezus Castro

Mariana Cúri

Vanessa Leotti

João Riboldi

RESUMO

Existem diversos estudos demonstrando a presença de Funcionamento Diferencial do Item (DIF) em alguns itens do Inventário de Depressão Beck (BDI), quando comparando homens e mulheres. O modelo TRI para Itens Constrangedores (modelo 2) comporta a presença de itens com DIF. Este estudo analisou a intensidade de sintomas depressivos através deste modelo. Para tanto, foi ajustado este modelo a dados do BDI e comparou-se com o ajuste do modelo logístico de 2 parâmetros (modelo 1). Os itens com DIF foram: tristeza, sentimento de fracasso, insatisfações, culpa, punição, choro, fatigabilidade e perda da libido. Os resultados do ajuste dos dois modelos são similares quanto à discriminação, gravidade (exceção para os itens com DIF) e no cálculo de escores para os indivíduos. Apesar disso, o modelo 2 ainda é vantajoso, pois mostra as diferenças em gravidade do sintoma depressivo para os grupos avaliados trazendo, desta forma, mais informação ao pesquisador sobre a população estudada. Este modelo, que tem um alcance mais amplo em termos de população alvo, pode ser uma ótima alternativa na identificação de indivíduos com potencial depressivo.

Palavras-chave: Teoria da Resposta ao Item, Funcionamento Diferencial do Item, Inventário de Depressão Beck, Modelo TRI para Itens Constrangedores.

ABSTRACT

There are several studies showing the presence of Differential Item Functioning (DIF) in some of the Beck Depression Inventory (BDI), when comparing men and women. The model IRT for embarrassing items (model 2) involves the presence of items with DIF. This study examined the intensity of depressive symptoms by using this model. It was adjusted to data from the BDI and compared with the adjustment of the two-parameter logistic model (model 1). Items with DIF were: sadness, feeling of failure, dissatisfaction, guilty, punishment, crying, fatigability and loss of libido. The results of the adjustment of the two models are similar in discrimination, gravity (except for items with DIF), and in the calculation of scores for individuals. Nevertheless, the model 2 is still beneficial because it shows the differences in gravity of depressive symptoms for groups evaluated, thus providing more information to the researcher on the study population. This model which has a broader scope in terms of population, may be a good alternative to the identification of individuals with potential depression.

Keyword: Item Response Theory, Differential Item Functioning, Beck Depression Inventory, Model IRT for embarrassing items.

INTRODUÇÃO

Na tentativa de se medir um traço latente, faz-se necessária a utilização de um instrumento composto por itens que presumivelmente o refletem. O traço latente não é observado diretamente, mas sim indiretamente através das respostas ao grupo de itens do instrumento utilizado.

Estabelecer equivalência de medidas entre grupos que diferem em características como educação, gênero e raça, por exemplo, é importante em avaliação de saúde mental. Os grupos podem ser comparados em termos de suas medidas dos traços de interesse, tais como intensidade de sintomas depressivos, funcionamento físico ou satisfação com cuidado, por exemplo¹. Sendo assim, antes de comparar grupos de respondentes (em relação à idade ou gênero, por exemplo) em termos do traço sendo medido, deve-se estar confiante de que os itens que compreendem a medida operam equivalentemente entre os diferentes grupos¹. Em outras palavras, existe a possibilidade de que itens, em especial em medidas psicológicas e/ou psiquiátricas, funcionem diferentemente ou de forma viesada para diferentes grupos de respondentes², se um item tem uma função de resposta diferente para dois grupos, este item é dito viesado³.

Estudos epidemiológicos psiquiátricos, realizados em diversos países desenvolvidos utilizando amostras aleatórias de comunidades, têm consistentemente mostrado uma maior prevalência de depressão maior (transtorno psiquiátrico de humor) em mulheres quando comparadas com homens⁴. Esta diferença de gênero em depressão está cotada como um dos mais robustos achados em pesquisa epidemiológica⁵ e as possíveis explicações para esta diferença de gênero são muitas. Grande parte delas pode pertencer a um destes dois grupos: aquele que enfatiza os diferentes papéis sociais de homens e mulheres, ou aquele que foca sobre diferenças biológicas subjacentes⁴.

Na literatura sobre Teoria da Resposta ao Item (TRI), o termo viés do item tem sido essencialmente substituído pela expressão funcionamento diferencial do item (*Differential Item Functioning* – DIF). DIF ocorre quando um item do instrumento não tem a mesma relação com o traço latente em dois ou mais grupos de respondentes. Mais concretamente, DIF ocorre quando um item apresenta uma Curva Característica do Item (CCI) diferente para cada grupo ou, equivalentemente, quando qualquer parâmetro do item difere entre os grupos. Se um item está livre deste viés, as respostas para aquele item serão relacionadas somente com o nível do traço latente que o item está tentando

medir. Se o item apresenta viés, as respostas ao mesmo serão relacionadas a algum outro fator além do nível do traço latente.

Diversos estudos examinaram a presença de viés em medidas cognitivas considerando diferentes variáveis demográficas, tais como educação⁶⁻¹⁰, classe social⁹, tipo de vizinhança¹¹, etnia¹² e idade¹⁰. Estudos mais recentes já utilizam a denominação DIF para este viés¹³⁻²¹. Estes aplicam diferentes metodologias para a detecção de DIF, muitas das quais podem ser verificadas em outras tantas publicações²²⁻²⁷.

Muitos instrumentos de medida, em especial na área da psiquiatria, apresentam itens que podem funcionar de maneira diferente para diferentes grupos. Em 2002, foi publicado um estudo onde foi avaliada a presença de itens com DIF no Inventário de Depressão Beck (BDI) em relação a dois diferentes grupos de idade²⁸ e outros estudos relatam o mesmo em relação ao gênero²⁹⁻³¹. A diferença entre as distribuições das respostas de homens e mulheres se dão em itens referentes a choro, sentimento de fracasso, punição, perda da libido, insatisfação, culpa e fadabilidade.

A presença de um grande número de itens com DIF no BDI é uma severa ameaça à validade da medida da intensidade de sintomas depressivos e a conclusões baseadas nos escores derivados dos itens com e sem DIF, e uma possível solução para este problema poderia ser a eliminação dos mesmos do instrumento de medida. No entanto, isto poderia comprometer a medida do traço latente, pois estes itens provavelmente contêm informação relevante à medida do mesmo, já que o BDI foi construído procurando abranger todos os sintomas depressivos observáveis³². O uso de um modelo que permita a manutenção de todos os itens no instrumento e, ao mesmo tempo, contemple as diferenças entre os grupos trabalhados vem a ser uma ótima alternativa para a análise de dados oriundos do BDI.

O ‘Modelo TRI para itens constrangedores’, proposto por Cúri³³, insere-se dentro desta perspectiva, pois preserva estas características. Sendo assim, este estudo tem por objetivos identificar itens do BDI que apresentem DIF para gênero, isto é, que apresentem viés comparando-se homens e mulheres, através da análise diferencial do item, ajustar o modelo para itens constrangedores para a amostra de sujeitos considerada, e comparar estes resultados com os do ajuste do modelo logístico de dois parâmetros tradicional.

MÉTODO

Amostra: Os dados são provenientes de um estudo transversal conduzido para realizar a adaptação, normatização e validação para o português das Escalas Beck - Inventário de Depressão Beck (BDI), Inventário de Ansiedade Beck (BAI), Escala de Desesperança Beck (BHS) e Escala de Ideação Suicida Beck (BSI) – em um estudo conduzido pela Dra. Jurema Alcides Cunha e publicado em 2001³². Os 4025 sujeitos do estudo estão distribuídos em três grandes grupos:

Grupo 1: Pacientes psiquiátricos ($n = 1138$) - pacientes com transtornos depressivos, transtornos de ansiedade, transtorno de pânico sem ou com agorafobia, transtorno obsessivo-compulsivo, fobia social, fobia simples, agorafobia, transtorno de dependência do álcool, dependência de outras substâncias ou transtornos psiquiátricos não especificados.

Grupo 2: Pacientes de clínica médica ($n = 490$) - pacientes com cardiopatia, com dor crônica, portadores de HIV assintomáticos e sintomáticos, com disfunção erétil, com obesidade, com diabetes mellitus tipo II, ou pacientes de unidades de cuidados primários em saúde.

Grupo 3: Amostra não-clínica, constituída por grupos da população geral ($n = 2397$) - composta por universitários, reclusos de penitenciária de segurança máxima, bombeiros, funcionários de hospital geral, adolescentes, idosos.

Todos os participantes responderam ao BDI (Anexo B), que é composto por uma escala de auto-relato, de 21 itens, cada um com quatro alternativas com escores variando de 0 a 3, subentendendo graus crescentes de intensidade de sintomas depressivos³². Os itens do BDI foram selecionados com base em observações e relatos de sintomas e atitudes mais frequentes em pacientes psiquiátricos, com transtornos depressivos³². São eles: Tristeza; Pessimismo; Sentimento de fracasso; Insatisfação; Culpa; Punição; Auto-aversão; Auto-acusações; Idéias suicidas; Choro; Irritabilidade; Retraimento social; Indecisão; Mudança na auto-imagem; Dificuldade de trabalhar; Insônia; Fatigabilidade; Perda de apetite; Perda de peso; Preocupações somáticas; Perda da libido.

A escala do BDI, originalmente com quatro pontos, para os objetivos do presente trabalho, foi dicotomizada de modo que a resposta assume o valor 1 ($X_{ij} = 1$), quando o indivíduo relata apresentar o sintoma descrito no item i (isto é, escolhe uma

das categorias com escores 1, 2 ou 3 do determinado item) e 0 ($X_{ij} = 0$) caso não apresente aquele sintoma.

Teoria da Resposta ao Item: Foram ajustados dois modelos TRI para variáveis dicotômicas (aqui sendo do tipo ausência ou presença de sintoma depressivo).

a) **Modelo logístico unidimensional de 2 parâmetros**³⁴: este é um modelo TRI para resposta dicotômica, apropriado para medidas nas quais os itens não são igualmente relacionados ao traço latente. O modelo de 2 parâmetros prediz a probabilidade do indivíduo j apresentar o sintoma medido no item i , condicionado a sua intensidade de sintomas depressivos (isto é, $P(X_{ij} = 1 | \theta_j)$), como segue:

$$P(X_{ij} = 1 | \theta_j, b_i, a_i) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (1)$$

onde θ_j é a intensidade de sintomas depressivos (traço latente) do indivíduo j (parâmetro de indivíduo); b_i é o parâmetro de posição do item i e representa a gravidade do sintoma depressivo descrito pelo item i ; a_i é o parâmetro de discriminação (ou inclinação) do item i ; 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).

Os resultados do ajuste do modelo TRI para cada item podem ser melhor visualizados através das chamadas Curvas Características do Item (CCI). Neste estudo, onde os itens são dicotômicos, ela indica a probabilidade de presença do sintoma para cada nível de intensidade de sintomas depressivos considerado.

b) **Modelo TRI para itens constrangedores**: este é um modelo para itens dicotômicos, proposto por Cúri³³, que possibilita diferenciar a gravidade da presença do sintoma depressivo entre indivíduos de diferentes grupos que possam apresentar comportamentos diferenciados frente a um determinado item. A probabilidade do indivíduo j apresentar o sintoma medido no item i e pertencer a um destes diferentes grupos, isto é, $P(X_{ij} = x_{ij}, C_{ij} = c_{ij} | \theta_j, \zeta)$, é:

$$P(X_{ij} = x_{ij}, C_{ij} = c_{ij} | \theta_j, \zeta) = P(X_{ij} = x_{ij} | C_{ij} = c_{ij}, \theta_j, \zeta) \times P(C_{ij} = c_{ij} | \theta_j, \zeta)$$

$$= \left[\left(P_{ij}^* \right)^{x_{ij}} \left(1 - P_{ij}^* \right)^{(1-x_{ij})} \delta_i \right]^{c_{ij}} \times \left[P_{ij}^{x_{ij}} \left(1 - P_{ij} \right)^{(1-x_{ij})} (1 - \delta_i) \right]^{1-c_{ij}} \quad (2)$$

com

$$\zeta = (\zeta_1, \dots, \zeta_I)^t \text{ e } \zeta_i = (a_i, b_{1i}, b_{2i}, \gamma_i, \delta_i)^t \text{ e}$$

$$c_{ij} = \begin{cases} 0 & \text{quando o indivíduo pertence ao grupo com comportamento padrão,} \\ & \text{que neste estudo são as mulheres.} \\ 1 & \text{quando o indivíduo pertence ao grupo com comportamento diferenciado,} \\ & \text{que neste estudo são homens.} \end{cases}$$

$$P_{ij}^* = \frac{\gamma_i}{1 + e^{-a_i(\theta_j - b_{2i})}} = P(X_{ij} = 1 | C_{ij} = 1, \theta_j, a_i, b_{2i}, \gamma_i);$$

$$P_{ij} = \frac{1}{1 + e^{-a_i(\theta_j - b_{1i})}} = P(X_{ij} = 1 | C_{ij} = 0, \theta_j, a_i, b_{1i});$$

onde θ_j é a intensidade de sintomas depressivos (traço latente) do indivíduo j (parâmetro de indivíduo); b_{1i} é o parâmetro de gravidade do item i para indivíduos do grupo considerado com comportamento padrão (mulheres); b_{2i} é o parâmetro de gravidade do item i para indivíduos do grupo considerado com comportamento diferenciado (homens); γ_i é a probabilidade do indivíduo do grupo com comportamento diferenciado afirmar que tem o sintoma depressivo, isto é, probabilidade de presença de sintoma em indivíduos com alta intensidade de sintomas depressivos neste grupo; δ_i é a probabilidade do item i apresentar funcionamento diferencial (DIF) para os grupos considerados.

Este modelo, além do parâmetro de discriminação do item, comum aos demais modelos TRI, estima outros parâmetros que contemplam o funcionamento diferencial daqueles itens com DIF.

Estratégia de análise:

A análise de funcionamento diferencial do item foi realizada através da técnica conhecida como Razão de Verossimilhança em Teoria da Resposta ao Item²⁴ (*Item Response Theory Log-Likelihood Ratio* – IRTLRL, versão 2.0b), utilizando o software IRTLRLDIF desenvolvido por Dave Thissen e disponibilizado em sua *homepage*³⁵. Este procedimento parte da definição de F. Lord sobre DIF (então chamado de viés do item) e usa o teste da razão de verossimilhança como um teste de significância para a hipótese

de nulidade de que os parâmetros de uma função de resposta de um item não diferem entre grupos – um resultado significativo aponta a detecção de DIF. Como, para modelos TRI paramétricos, o grupo de parâmetros do item é isomorfo (tem a mesma forma) com a função de resposta do item, o exame das diferenças entre os parâmetros do item é equivalente ao exame da diferença nas funções de resposta do item. O software IRTLRDIF tem implementado dois dos modelos TRI mais comumente usados: o modelo logístico de 3 parâmetros e o modelo politômico de resposta gradual de Samejima³⁶. Em ambos os casos o modelo logístico de 2 parâmetros (utilizado para a identificação de itens com DIF) é um caso especial e, neste software, ele está implementado como um modelo de resposta gradual com duas categorias de resposta, pois, na perspectiva computacional, é mais simples executar desta forma³⁵. Em função do tamanho da amostra, o nível de significância adotado para a identificação dos itens com DIF foi 0,1%.

O ajuste dos modelos TRI de 2 parâmetros (1) e para itens constrangedores (2) foi realizado através de rotinas elaboradas no WinBUGS, versão 1.4.3 (*WinBUGS 1996-2007: Imperial College and MRC, UK*). As rotinas referentes ao modelo 2, de autoria de Mariana Cúri³³, utilizam os métodos de estimação de parâmetros conhecidos como máxima verossimilhança marginal e bayesiano.

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 totalidade dos indivíduos se declarou como 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.

Os itens que apresentaram DIF foram: tristeza, sentimento de fracasso, insatisfações, culpa, punição, choro, fadigabilidade e perda da libido. Os resultados do ajuste do modelo 1 encontram-se na Tabela 2 e os resultados do ajuste do modelo 2,

considerando os oito itens mencionados apresentando DIF e o grupo masculino como o grupo com comportamento diferenciado perante os mesmos, na Tabela 3.

O modelo 1 apresentou estimativas dos parâmetros de discriminação (Tabela 2), que apontam que praticamente todos os itens discriminam razoavelmente bem ($a_i > 1$ ^{33,34}), à exceção de perda de peso e auto-acusações. Os itens com maior poder de discriminação são relacionados a sintomas de sentimento de fracasso e insatisfação.

A partir das estimativas de gravidade (b_i) dos sintomas depressivos (Tabela 2), observa-se que os sintomas auto-acusações e irritabilidade são os menos graves e os sintomas perda de peso e idéias suicidas os mais graves. É interessante ressaltar que perda de peso é o sintoma depressivo mais grave e ao mesmo tempo o que menos discrimina a população ($\hat{a}_{19} = 0,93$). No entanto, o sintoma idéias suicidas é o segundo mais grave ($\hat{b}_9 = 1,20$) e discrimina bem a população quanto ao nível de intensidade de sintomas depressivos ($\hat{a}_9 = 1,71$).

De acordo com o modelo 2 (Tabela 3), praticamente todos os itens discriminam razoavelmente bem, à exceção de perda de peso e auto-acusações ($\hat{a}_{19} = 0,93$ e $\hat{a}_8 = 0,97$, respectivamente). Os itens que se destacam em discriminação são sentimento de fracasso ($\hat{a}_3 = 6,83$), insatisfação ($\hat{a}_4 = 4,13$) e fadigabilidade ($\hat{a}_{17} = 3,24$).

Quanto à gravidade dos sintomas, os resultados são os mesmos do modelo 1 para todos os itens do BDI que não apresentam DIF. As diferenças ocorrem apenas nos oito itens restantes. Observa-se que a presença de tristeza tem maior probabilidade de ser observada em níveis mais altos de intensidade de sintomas depressivos ($\hat{b}_{1,1} = 0,24$) nas mulheres e mais baixos nos homens ($\hat{b}_{2,1} = -0,50$). Esta direção se mantém para os sintomas sentimento de fracasso, insatisfações, culpa, choro e fadigabilidade. Isto se inverte para os sintomas punição e perda da libido. Por exemplo, a presença de perda da libido tem maior probabilidade de ser observada em níveis mais baixos de intensidade de sintomas depressivos ($\hat{b}_{1,21} = 0,47$) nas mulheres e mais altos nos homens ($\hat{b}_{2,21} = 1,01$). Ainda como um resultado do modelo 2, estima-se que a probabilidade de um homem com alta intensidade de sintomas depressivos expresse a presença dos sintomas relativos à tristeza, sentimento de fracasso, insatisfações, culpa, punição,

choro, fatigabilidade e perda da libido, é maior ou igual a 85% ($\hat{\gamma}_i \geq 0,85$). A diferença entre as CCI's produzidas pelos modelos 1 e 2 (este último diferenciando homens e mulheres) para o item 17, relativo à fatigabilidade, pode ser melhor visualizada na Figura 1.

Os níveis de intensidade de sintomas depressivos estimados sob os modelos TRI estão na mesma escala da gravidade do sintoma estimada para cada item do BDI, logo são comparáveis. Os percentis 95 dos níveis de intensidade de sintomas depressivos são 1,60 e 1,62 para os modelos 1 e 2, respectivamente. Dos 201 indivíduos com intensidade de sintomas depressivos acima do percentil 95 para cada modelo, 194 são classificados de forma igual pelos dois modelos. As características (Tabela 4) deste grupo mostram que quase 80% são oriundos do grupo psiquiátrico, aproximadamente 68% são mulheres, a maioria (mais de 58%) não tem um companheiro(a) e têm, em média, 37 anos de idade.

As estimativas de intensidade de sintomas depressivos obtidas segundo os modelos 1 e 2 apresentam alta associação, com coeficiente de correlação igual a 0,99.

DISCUSSÃO

O modelo TRI de dois parâmetros (modelo 1) foi escolhido como modelo de comparação ao modelo para itens constrangedores (modelo 2) porque ambos incluem parâmetros de discriminação e gravidade do sintoma depressivo. Outros estudos já aplicaram o modelo de 2 parâmetros a dados psiquiátricos: Schaeffer³⁷, em 1988, ajustou este modelo a respostas para 11 sintomas de depressão para os quais foram oferecidas quatro categorias de resposta ('nunca', 'uma vez até agora', 'com relativa frequência' e 'muitas vezes'), e Kessler et al.³⁸ aplicaram-no na construção de duas escalas (uma com 10 itens e outra com seis itens) sobre saúde mental.

Os achados do ajuste do modelo 1, quanto à presença de DIF em oito itens do BDI, mostram que homens e mulheres com o mesmo nível de intensidade de sintomas depressivos responderam de modo diferente aos itens tristeza, sentimento de fracasso, insatisfações, culpa, punição, choro, fatigabilidade e perda da libido. Diversos estudos^{4,29-31,39-44} corroboram estes achados, entretanto, o funcionamento diferencial do item choro em relação ao gênero é o que aparece na maioria deles. Boa parte dos estudos que mostram a diferença de gênero em relação ao choro salienta que mulheres

tendem a chorar mais que homens^{30,31,42}. Isto pode ser mais um reflexo da bem conhecida tendência das mulheres chorarem mais prontamente e mais intensamente que os homens em uma variedade de situações angustiantes do que um indicador de uma diferença de gênero na prevalência da depressão⁴. Esta constatação sugere que o choro em resposta a angústia é, em grande medida, determinado pelo gênero, logo homens e mulheres com o mesmo nível de intensidade de sintomas depressivos provavelmente não responderão ao item choro da mesma forma. No entanto, neste estudo constatou-se uma direção inversa, isto é, constatou probabilidade maior da presença do sintoma choro em níveis de intensidade de sintomas depressivos mais baixos nos homens que nas mulheres. Este resultado pode ser devido ao fato de se ter dicotomizado a escala do BDI, transformando-a em ausência ou presença dos sintomas avaliados pelos 21 itens, o que sabidamente pode gerar perda de informação relevante. Originalmente, a escala do BDI tem quatro pontos, sendo que, em especial no item sobre choro, a categoria de ordem mais alta afirma que o indivíduo perdeu a capacidade de chorar, mesmo que tenha vontade, enquanto as três primeiras categorias determinam um crescente na quantidade de vezes que se costuma chorar. De todos os homens que receberam o valor 1 na escala dicotomizada, mais da metade respondeu a categoria 3, o mesmo ocorrendo quando observa-se apenas os homens que pertencem ao grupo dos 5% maiores valores estimados para nível de intensidade de sintomas depressivos, mostrando que são sérios candidatos a um diagnóstico positivo para depressão. Esta perda da capacidade de chorar masculina também aparece no estudo de Hammen e Padesk²⁹, onde se trabalhou com o BDI na escala original.

Comparando-se os resultados encontrados para os modelos 1 e 2 em relação à discriminação dos sintomas depressivos medidos pelos itens, pode-se perceber que, considerando-se itens com valores de $a_i \geq 1$ ^{33,34} como tendo razoável discriminação, os mesmos 19 itens nos dois modelos estão nesta categoria, com exceção apenas de perda de peso e auto-acusações. No estudo de Cúri³³, onde foi ajustado um modelo logístico de 3 parâmetros a dados do BDI, apenas perda de apetite obteve estimativa para discriminação abaixo deste ponto de corte, porém o sintoma perda de peso está muito próximo desta região. Já entre os sintomas que mais discriminam sentimento de fracasso e insatisfação são os sintomas que aparecem nos modelos 1, 2 e no ajustado por Cúri³³ mostrando que estes são sintomas importantes na discriminação da população quanto à intensidade de sintomas depressivos.

Um resultado que aparece no modelo 2 é a maior gravidade do sintoma perda da libido para os homens que para as mulheres, pois a maior probabilidade da presença do mesmo ocorre em níveis de intensidade de sintomas depressivos mais altos para homens ($\hat{b}_{2,21} = 1,01$) que para as mulheres ($\hat{b}_{1,21} = 0,47$). A importância da perda da libido para os homens aparece em diversos estudos. Em um ensaio clínico randomizado sobre os efeitos sexuais (tais como melhora na perda da libido e na disfunção erétil) da reposição de testosterona em homens com diagnóstico de depressão maior⁴⁵, os autores pretendiam verificar se o tratamento seria eficaz nesta população da mesma forma que o é na população geral. No entanto, a reposição de testosterona não surtiu o efeito já conhecido, indicando que talvez o impedimento tenha sido a condição de depressão da população alvo.

Os grupos formados pelos 5% de indivíduos com maiores estimativas de intensidade de sintomas depressivos (traço latente sendo medido), obtidas a partir dos modelos 1 e 2, evidenciam a superioridade feminina no grupo psiquiátrico, pois mais de 75% destes grupos é formado por mulheres. Estes dados são consistentes com a evidência de que a depressão é de duas a três vezes mais comum em mulheres adolescentes e adultas do que em homens adolescentes e adultos⁴⁶, pois estas mulheres têm os níveis de intensidade de sintomas depressivos mais altos, sendo fortes candidatas a terem um diagnóstico positivo para depressão.

É importante ressaltar que os modelos 1 e 2 rastreiam praticamente os mesmos indivíduos como pertencentes a estes grupos com as maiores estimativas de intensidade de sintomas depressivos. Dos 201, apenas 4 mulheres e 3 homens apresentam classificações discordantes, sendo que todos eles têm nível de intensidade de sintomas depressivos estimada abaixo do percentil 95 para o modelo 1 e acima para o modelo 2. Estas diferenças parecem ocorrer devido ao fato de que os níveis de intensidade de sintomas depressivos estimados para estes indivíduos estão no limite de seu respectivo valor do percentil 95.

Os resultados encontrados entre os modelos 1 e 2 são bastante similares, especialmente no caso das estimativas de intensidade de sintomas depressivos para cada indivíduo, comprovado pela alta correlação entre os escores TRI. Apesar disso, o modelo 2 ainda é vantajoso, pois mostra as diferenças em gravidade do sintoma depressivo para os grupos avaliados trazendo, desta forma, mais informação ao pesquisador sobre a população estudada. Por exemplo, a Figura 1 evidencia a vantagem

do uso do modelo 2 em relação ao modelo 1, pois mostra claramente as diferenças no comportamento de um item com DIF, em relação a sua gravidade, para os dois grupos comparados. A utilização de um modelo de alcance mais amplo em termos de população alvo pode ser uma alternativa bastante proveitosa também na área clínica, onde a existência de modelos validados pode contribuir para a identificação de indivíduos com potencial depressivo.

Uma limitação deste trabalho é a de que ele consiste de uma comparação empírica, fazendo-se necessário um estudo mais amplo com dados simulados.

Ainda, como Curi³³ comenta, faz-se necessária a extensão do modelo 2 a itens com resposta ordinal, pois assim como o BDI, inúmeros instrumentos de medidas psiquiátricas contemplam itens de resposta ordinal, e a sua transformação em itens dicotômicos (do tipo ausência ou presença, por exemplo) não aproveita totalmente a informação disponível podendo produzir resultados com alguma inconsistência.

REFERÊNCIAS

- (1) Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Qual Life Res.* 2007.
- (2) Embretson SE, Reise SP. *Item Response Theory for Psychologists.* New Jersey: Lawrence Erlbaum Associates, Inc.; 2000.
- (3) Lord F. *Applications of item response theory to practical testing problems.* Hillsdale, NJ: 1980.
- (4) Romans SE, Tyas J, Cohen MM, Silverstone T. Gender differences in the symptoms of major depressive disorder. *J Nerv Ment Dis.* 2007; 195: 905-911.
- (5) Kuehner C. Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. *Acta Psychiatr Scand.* 2003; 108: 163-174.
- (6) Jones RN, Gallo JJ. Education bias in the mini-mental state examination. *Int Psychogeriatr.* 2001; 13: 299-310.
- (7) Black SA, Espino DV, Mahurin R et al. The influence of noncognitive factors on the Mini-Mental State Examination in older Mexican-Americans: findings from the Hispanic EPESE. Established Population for the Epidemiologic Study of the Elderly. *J Clin Epidemiol.* 1999; 52: 1095-1102.

- (8) Schmand B, Lindeboom J, Hooijer C, Jonker C. Relation between education and dementia: the role of test bias revisited. *J Neurol Neurosurg Psychiatry*. 1995; 59: 170-174.
- (9) Jagger C, Clarke M, Anderson J, Battcock T. Misclassification of dementia by the mini-mental state examination--are education and social class the only factors? *Age Ageing*. 1992; 21: 404-411.
- (10) Kraemer HC, Moritz DJ, Yesavage J. Adjusting Mini-Mental State Examination scores for age and educational level to screen for dementia: correcting bias or reducing validity? *Int Psychogeriatr*. 1998; 10: 43-51.
- (11) Espino DV, Lichtenstein MJ, Palmer RF, Hazuda HP. Ethnic differences in mini-mental state examination (MMSE) scores: where you live makes a difference. *J Am Geriatr Soc*. 2001; 49: 538-548.
- (12) Blesa R, Pujol M, Aguilar M et al. Clinical validity of the 'mini-mental state' for Spanish speaking communities. *Neuropsychologia*. 2001; 39: 1150-1157.
- (13) Crane PK, Gibbons LE, Narasimhalu K, Lai JS, Cella D. Rapid detection of differential item functioning in assessments of health-related quality of life: The Functional Assessment of Cancer Therapy. *Qual Life Res*. 2007; 16: 101-114.
- (14) Crane PK, Cook KF, Johnson K, Deyo R, Amtmann D. Differential item functioning impact in a modified version of the Roland-Morris Disability Questionnaire. *Qual Life Res*. 2007; on-line first.
- (15) Jane JS, Oltmanns TF, South SC, Turkheimer E. Gender bias in diagnostic criteria for personality disorders: an item response theory analysis. *J Abnorm Psychol*. 2007; 116: 166-175.
- (16) Reichenheim ME, Klein R, Moraes CL. Assessing the physical violence component of the Revised Conflict Tactics Scales when used in heterosexual couples: An item response theory analysis. *Cad Saude Publica*. 2007; 23: 53-62.
- (17) Jones RN. Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination. Detecting differential item functioning using MIMIC modeling. *Medical Care*. 2006; 44: S124-S133.
- (18) Edelen.M.O., Thissen D, Teresi.J.A., Kleinman M, Ocepek-Welikson K. Identification of Differential Item Functioning Using Item Response Theory and the Likelihood-Based Model Comparison Approach. Application to the Mini-Mental State Examination. *Medical Care*. 2006; 44: S134-S142.

(19) Crane PK, Hart DL, Gibbons LE, Cook KF. A 37-item shoulder functional status item pool had negligible differential item functioning. *J Clin Epidemiol.* 2006; 59: 478-484.

(20) Smith LL, Reise SP. Gender differences on negative affectivity: an IRT study of differential item functioning on the Multidimensional Personality Questionnaire Stress Reaction Scale 1. *J Pers Soc Psychol.* 1998; 75: 1350-1362.

(21) 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.

(22) Crane PK, Gibbons LE, Jolley L, van BG. Differential item functioning analysis with ordinal logistic regression techniques. DIFdetect and difwithpar. *Medical Care.* 2006; 44: S115-S123.

(23) Crane PK, Gibbons LE, Ocepek-Welikson K, et al. A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Qual Life Res.* 2007; on-line first.

(24) Teresi.J.A., Ocepek-Welikson K, Kleinman M, et al. Evaluating measurement equivalence using the item response theory log-likelihood ratio (IRTLR) method to assess differential item functioning (DIF): applications (with illustrations) to measure of physical functioning ability and general distress. *Qual Life Res.* 2007; on-line first.

(25) Teresi.J.A. Statistical methods for examination of differential item functioning (DIF) with applications to cross-cultural measurement of functional, physical and mental health. *Journal of Mental Health and Aging.* 2001; 7: 31-40.

(26) Stark S, Chernyshenko OS, Drasgow F, Williams BA. Examining assumptions about item responding in personality assessment: should ideal point methods be considered for scale development and scoring? *J Appl Psychol.* 2006; 91: 25-39.

(27) Birgenheimer JB, Raudenbush SW, Leventhal T, Brooks-Gunn J. Measurement Equivalence and Differential Item Functioning in Family Psychology. *Journal of Family Psychology.* 2005; 19: 441-455.

(28) 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. *Psychology and Aging.* 2002; 17: 379-391.

- (29) Hammen CL, Padesky CA. Sex differences in the expression of depressive responses on the Beck Depression Inventory. *J Abnorm Psychol.* 1977; 86: 609-614.
- (30) Santor D, Ramsay J, Zuroff D. Nonparametric Item Analyses of the Beck Depression Inventory: Evaluating Gender Item Bias and Response Option Weights. *Psychological Assessment.* 1994; 6: 255-270.
- (31) Salokangas RK, Vaahtera K, Pacriev S, Sohlman B, Lehtinen V. Gender differences in depressive symptoms. An artefact caused by measurement instruments? *J Affect Disord.* 2002; 68: 215-220.
- (32) Cunha JA. *Manual da versão em português das ESCALAS BECK.* São Paulo: 2001.
- (33) Cúri M. *Análise de questionários com itens constrangedores.* [São Paulo: Universidade de São Paulo, 2006].
- (34) Andrade, D. F., Tavares, H. R., and Valle, R. C. Teoria da Resposta ao Item: conceitos e aplicações. 28-7-2000. Caxambu, ABE - Associação Brasileira de Estatística. 14º SINAPE.
- (35) Thissen, D. Dave Thissen's Front Page. Disponível em: www.unc.edu/~dthissen/dl.html . (Acessado em 26 de julho de 2008)
- (34) Wenzel A, Steer RA, Beck AT. Are there any gender differences in frequency of self-reported somatic symptoms of depression? *J Affect Disord.* 2005; 89: 177-181.
- (36) Samejima, F. *Estimation of latent ability using a response pattern of graded scores.* Psychometrika Monograph 17. 1969.
- (37) Schaeffer NC. An application of item response theory to the measurement of depression. *Sociological Methodology.* 1988; 18: 271-307.
- (38) Kessler, RC, Andrews, G, Colpe, LJ, Hiripi, E, Mroczek, DK, Normand, SLT, et al. *Psychological Medicine.* 2002; 32: 959-976.
- (39) Stommel M, Given BA, Given CW, Kalaian HA, Schulz R, McCorkle R. Gender bias in the measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D). *Psychiatry Res.* 1993; 49: 239-250.
- (40) Wilhelm K, Parker G, Asghari A. Sex differences in the experience of depressed mood state over fifteen years. *Soc Psychiatry Psychiatr Epidemiol.* 1998; 33: 16-20.

(41) Carter JD, Joyce PR, Mulder RT, Luty SE, McKenzie J. Gender differences in the presentation of depressed outpatients: a comparison of descriptive variables. *J Affect Disord.* 2000; 61: 59-67.

(42) Gelin M, Zumbo B. Differential item functioning results may change depending on how an item is scored: an illustration with the Center for Epidemiologic Studies Depression Scale. *Educational and Psychological Measurement.* 2003; 63: 65-74.

(43) Wenzel A, Steer RA, Beck AT. Are there any gender differences in frequency of self-reported somatic symptoms of depression? *J Affect Disord.* 2005; 89: 177-181.

(44) Angst J, Gamma A, Gastpar M, Lepine JP, Mendlewicz J, Tylee A. Gender differences in depression. Epidemiological findings from the European DEPRES I and II studies. *Eur Arch Psychiatry Clin Neurosci.* 2002; 252: 201-209.

(45) Seidman SN, Roose SP. The sexual effects of testosterone replacement in depressed men: randomized, placebo-controlled clinical trial. *J Sex Marital Ther.* 2006; 32: 267-273.

(46) 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.: 2000:288-324.

TABELAS

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

| Características sócio-demográficas** | 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

** O nº de indivíduos varia de acordo com as características sócio-demográficas devido a ocorrência de valores faltantes.

Tabela 2: Estimativa dos parâmetros do modelo logístico de 2 parâmetros (1).

| ITEM | a_i (EP) | b_i (EP) |
|-----------------------------|-------------|--------------|
| 1 Tristeza | 2,38 (0,09) | 0,16 (0,02) |
| 2 Pessimismo | 2,41 (0,10) | 0,76 (0,03) |
| 3 Sentimento de Fracasso | 2,90 (0,12) | 0,71 (0,03) |
| 4 Insatisfações | 2,79 (0,11) | 0,10 (0,02) |
| 5 Culpa | 1,98 (0,08) | 0,57 (0,03) |
| 6 Punição | 1,48 (0,06) | 0,66 (0,04) |
| 7 Auto-aversão | 2,50 (0,10) | 0,51 (0,02) |
| 8 Auto-acusações | 0,97 (0,05) | -1,19 (0,06) |
| 9 Idéias Suicidas | 1,71 (0,08) | 1,20 (0,04) |
| 10 Choro | 1,69 (0,06) | 0,36 (0,03) |
| 11 Irritabilidade | 1,07 (0,05) | -0,65 (0,04) |
| 12 Retraimento Social | 1,51 (0,06) | 0,71 (0,03) |
| 13 Indecisão | 1,99 (0,08) | 0,13 (0,03) |
| 14 Mudança na auto-imagem | 1,72 (0,07) | 0,52 (0,03) |
| 15 Dificuldade de trabalhar | 1,94 (0,07) | 0,30 (0,03) |
| 16 Insônia | 1,44 (0,06) | 0,06 (0,03) |
| 17 Fatigabilidade | 1,41 (0,06) | -0,25 (0,03) |
| 18 Perda de apetite | 1,22 (0,06) | 0,96 (0,04) |
| 19 Perda de peso | 0,93 (0,06) | 2,00 (0,11) |
| 20 Preocupações Somáticas | 1,19 (0,05) | 0,35 (0,04) |
| 21 Perda da libido | 1,43 (0,06) | 0,74 (0,04) |

a_i - parâmetro de discriminação do item i ;

b_i - parâmetro de gravidade do item i ;

EP – erro-padrão.

Tabela 3: Estimativa dos parâmetros do modelo para itens constrangedores (2).

| ITEM | a_i (EP) | b_{1i} (EP) | b_{2i} (EP) | γ_i (EP) | δ_i (EP) |
|-----------------------------|-------------|---------------|---------------|-----------------|-----------------|
| 1 Tristeza | 2,68 (0,30) | 0,24 (0,10) | -0,50 (0,59) | 0,89 (0,10) | 0,21 (0,16) |
| 2 Pessimismo | 2,40 (0,10) | 0,76 (0,03) | - | - | - |
| 3 Sentimento de Fracasso | 6,83 (1,67) | 1,07 (0,06) | 0,08 (0,06) | 0,97 (0,03) | 0,39 (0,04) |
| 4 Insatisfações | 4,13 (1,15) | 0,34 (0,15) | -0,39 (0,29) | 0,96 (0,04) | 0,33 (0,14) |
| 5 Culpa | 2,32 (0,32) | 0,74 (0,23) | 0,21 (0,46) | 0,91 (0,09) | 0,35 (0,17) |
| 6 Punição | 1,61 (0,17) | 0,64 (0,23) | 0,71 (1,13) | 0,89 (0,11) | 0,23 (0,15) |
| 7 Auto-aversão | 2,47 (0,10) | 0,52 (0,02) | - | - | - |
| 8 Auto-acusações | 0,97 (0,05) | -1,19 (0,06) | - | - | - |
| 9 Idéias Suicidas | 1,69 (0,75) | 1,21 (0,04) | - | - | - |
| 10 Choro | 2,03(0,33) | 0,49 (0,45) | 0,29 (0,54) | 0,90 (0,09) | 0,42 (0,19) |
| 11 Irritabilidade | 1,07 (0,05) | -0,65 (0,04) | - | - | - |
| 12 Retraimento Social | 1,51 (0,06) | 0,72 (0,03) | - | - | - |
| 13 Indecisão | 1,98 (0,08) | 0,13 (0,02) | - | - | - |
| 14 Mudança na auto-imagem | 1,72 (0,07) | 0,52 (0,03) | - | - | - |
| 15 Dificuldade de trabalhar | 1,93 (0,07) | 0,30 (0,03) | - | - | - |
| 16 Insônia | 1,43 (0,06) | 0,07 (0,03) | - | - | - |
| 17 Fatigabilidade | 3,24 (0,64) | 0,82 (0,12) | -0,89 (0,09) | 0,97 (0,02) | 0,63 (0,04) |
| 18 Perda de apetite | 1,21 (0,05) | 0,96 (0,04) | - | - | - |
| 19 Perda de peso | 0,93 (0,06) | 2,01 (0,11) | - | - | - |
| 20 Preocupações Somáticas | 1,19 (0,05) | 0,35 (0,03) | - | - | - |
| 21 Perda da libido | 1,92 (0,40) | 0,47 (0,48) | 1,01 (0,84) | 0,85 (0,11) | 0,37 (0,15) |

EP : erro padrão;

a_i : parâmetro de discriminação do item i ;

b_{1i} : parâmetro de gravidade do item i para indivíduos do grupo feminino;

b_{2i} : parâmetro de gravidade do item i para indivíduos do grupo masculino;

γ_i : probabilidade do indivíduo do grupo masculino dizer que tem o sintoma depressivo, isto é, probabilidade de presença de sintoma em homens com um alto nível de intensidade de sintomas depressivos;

δ_i : probabilidade do item i apresentar funcionamento diferencial (DIF) para os grupos considerados.

Tabela 4: Descrição dos indivíduos com nível alto de sintomas depressivos, estimado como um valor acima do percentil 95 ($\hat{\theta} = 1,60$ para o modelo 1 e $\hat{\theta} = 1,62$ para o modelo 2).

| Variável | Psiquiátrica | Clínica | População geral | Total |
|-------------------------------|--------------|-------------|-----------------|--------------|
| Modelo 1 | n=157 | n=18 | n=26 | n=201 |
| Idade (n=200) | | | | |
| média | 39,18 | 34,67 | 29,76 | 37,60 |
| (desvio padrão) | (12,5) | (13,26) | (13,83) | (13,08) |
| mínimo | 15 | 18 | 18 | 15 |
| máximo | 75 | 64 | 67 | 75 |
| Escolaridade (n=194) | | | | |
| Menos de 5 anos | 54 | 9 | 12 | 38,7% |
| Ensino Fundamental Completo | 43 | 3 | 5 | 26,3% |
| Ensino Médio Completo | 43 | 5 | 5 | 27,3% |
| Ensino Superior Completo | 11 | 0 | 4 | 7,7% |
| Estado Civil (n=199) | | | | |
| Solteiro | 45 | 8 | 19 | 36,2% |
| Casado | 74 | 8 | 1 | 41,7% |
| Desquitado, divorciado, viúvo | 36 | 2 | 6 | 22,1% |
| Sexo (201) | | | | |
| Masculino | 36 | 11 | 16 | 31,3% |
| Feminino | 121 | 7 | 10 | 68,7 % |
| Modelo 2 | n=155 | n=20 | n=26 | n=201 |
| Idade (n=201) | | | | |
| média | 39,13 | 35,05 | 28,96 | 37,45 |
| (desvio padrão) | (12,45) | (13,23) | (13,69) | (13,08) |
| mínimo | 15 | 18 | 15 | 15 |
| máximo | 75 | 64 | 67 | 75 |
| Escolaridade (n=193) | | | | |
| Menos de 5 anos | 52 | 10 | 13 | 38,9% |
| Ensino Fundamental Completo | 43 | 3 | 5 | 26,4% |
| Ensino Médio Completo | 43 | 5 | 5 | 27,5% |
| Ensino Superior Completo | 11 | 0 | 3 | 7,3% |
| Estado Civil (n=199) | | | | |
| Solteiro | 46 | 10 | 19 | 37,7% |
| Casado | 72 | 8 | 1 | 40,7% |
| Desquitado, divorciado, viúvo | 35 | 2 | 6 | 21,6% |
| Sexo (201) | | | | |
| Masculino | 36 | 12 | 16 | 31,8% |
| Feminino | 119 | 8 | 10 | 68,2 % |

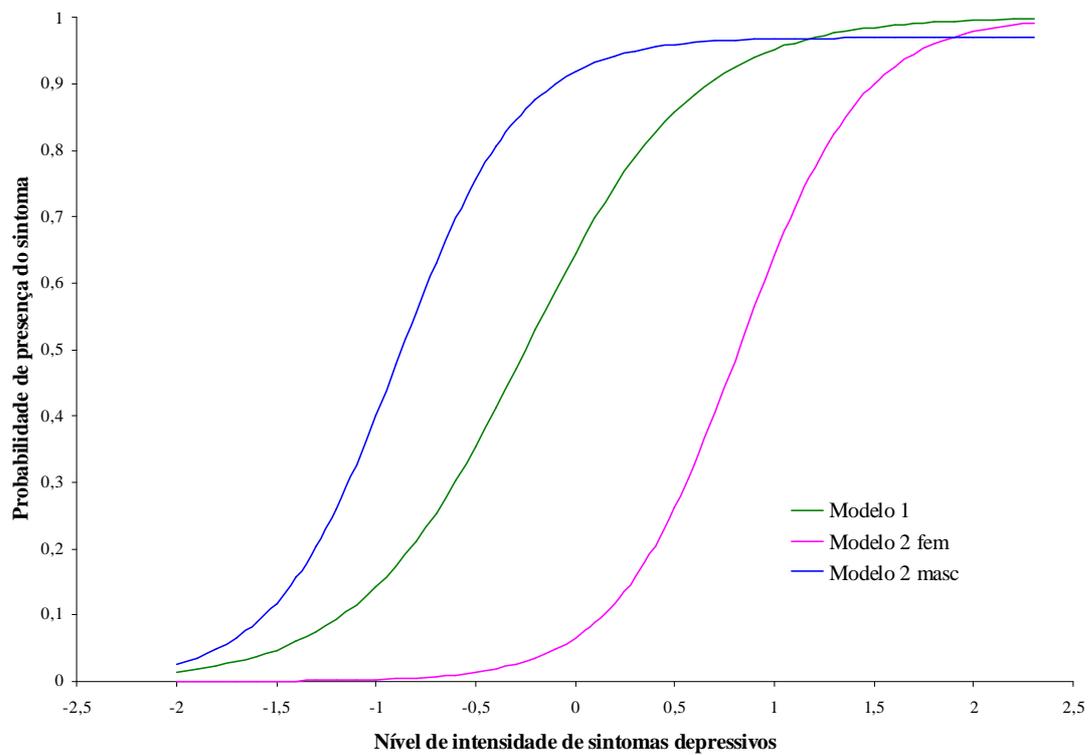
FIGURA

Figura 1: Curva Característica do Item (CCI) para o sintoma fadiga (item 17) segundo o modelo Logístico de 2 parâmetros (1) e o modelo para Itens Constrangedores (2) para os sexos feminino e masculino.

8. CONCLUSÕES E CONSIDERAÇÕES FINAIS

ARTIGO 1:

- O modelo de Resposta Gradual foi a escolha adequada ao tipo de dado oriundo de um instrumento como o BDI, cujos itens medem apenas um traço latente (intensidade de sintomas depressivos) e são de resposta categórica ordinal. O ajuste do modelo, representado pelas Curvas de Categoria de Resposta, mostrou que os itens relacionados aos sintomas depressivos pessimismo, insatisfação, culpa, punição, idéias suicidas, choro, irritabilidade, mudança na auto-imagem, insônia, perda de apetite, perda de peso preocupações somáticas e perda da libido tinham problemas no entendimento da escala, de modo que foram recategorizadas.
- O modelo fornece também dois tipos de informação extremamente relevantes: o quanto cada item discrimina a população quanto à intensidade de sintomas depressivos, mostrando, deste modo, os itens mais importantes na composição do instrumento de medida para atribuir um escore a cada indivíduo que responde ao mesmo; e, o grau de gravidade de cada item (que representa um sintoma depressivo), o qual é calculado na mesma escala do escore atribuído ao indivíduo, permitindo assim que sejam comparáveis. É importante salientar que cada escore individual estimado (nível de intensidade de sintomas depressivos estimado para cada indivíduo) incorpora os diferentes pesos dos itens, em relação a sua discriminação e gravidade, associados ao perfil de respostas que este indivíduo deu ao BDI.
- Os sintomas com maior capacidade de discriminação foram tristeza, pessimismo, sentimento de fracasso, insatisfação, auto-aversão, indecisão e dificuldade de trabalhar. Irritabilidade foi o sintoma que apresentou menor capacidade de discriminação.
- A região do traço latente ‘intensidade de sintomas depressivos’ onde os sete itens que têm maior capacidade de discriminação conseguem discriminar melhor a população fica aproximadamente entre os valores 0,7 e 2, o que corresponde ao intervalo que vai do 21 ao 35 na escala do escore total.
- Os itens mais graves foram os relativos à perda de peso, retraimento social e idéias suicidas.

- Dos 202 indivíduos com os maiores valores para a estimativa de intensidade de sintomas depressivos, praticamente 84% pertencem ao grupo psiquiátrico e 74% são mulheres.

ARTIGO 2:

- Os itens que apresentaram funcionamento diferencial quando comparando-se os dois grupos de gênero foram: tristeza, sentimento de fracasso, insatisfações, culpa, punição, choro, fatigabilidade e perda da libido.
- Considerando os itens com DIF, o ajuste do Modelo TRI para Itens Constrangedores se mostrou perfeitamente adequado, pois ele permite a manutenção destes no instrumento de medida. Este modelo, além do parâmetro de discriminação de cada item similar ao modelo logístico de 2 parâmetros, desdobra o parâmetro de gravidade do item em dois para aqueles itens viesados. Outra característica interessante deste modelo é que ele também estima a probabilidade da presença do sintoma depressivo com funcionamento diferencial nos homens (grupo de gênero considerado com funcionamento diferencial).
- Com exceção dos itens relacionados à perda de peso e auto-acusações, todos os outros têm uma boa capacidade de discriminação da população quanto à intensidade de sintomas depressivos, considerando o Modelo TRI para Itens Constrangedores. Os itens que se destacam são relacionados a sintomas de sentimento de fracasso, insatisfação, fatigabilidade, tristeza, auto-aversão, pessimismo e culpa.
- Quanto à gravidade dos sintomas depressivos, este modelo relaciona como os sintomas mais graves perda de peso, idéias suicidas, sentimento de fracasso (apenas para as mulheres) e perda da libido (somente para os homens).
- A presença de tristeza tem maior probabilidade de ser observada em níveis mais altos de intensidade de sintomas depressivos nas mulheres e mais baixos nos homens. Esta direção se mantém para os sintomas sentimento de fracasso, insatisfações, culpa, choro e fatigabilidade, e se inverte para os sintomas punição e perda da libido.
- Ainda como um resultado do Modelo TRI para Itens Constrangedores, estima-se que a probabilidade de que um homem com alto nível de intensidade de sintomas depressivos expresse que apresenta os sintomas relativos à tristeza, sentimento de fracasso, insatisfações, culpa, punição, choro, fatigabilidade e perda da libido, é pelo menos de 85% .

- Os resultados encontrados no ajuste do modelo logístico de 2 parâmetros a estes dados são relativamente semelhantes em relação ao parâmetro de discriminação e, exatamente iguais com respeito ao parâmetro de gravidade do item, com exceção dos oito itens com funcionamento diferencial.

São inúmeros os ganhos provenientes da utilização da TRI em relação à TCT. Além das vantagens métricas do método, a possibilidade de utilização mais global dos resultados, em especial no que tange aos itens do instrumento, permite que o pesquisador compreenda o comportamento dos mesmos e a sua relação com os seus respondentes. Neste trabalho, constatam-se estes ganhos 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 sintoma depressivo 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 identificação de itens com viés de comportamento (DIF), na possibilidade de se comparar os níveis de intensidade de sintomas depressivos estimados para cada respondente com a gravidade de cada sintoma. Mais importante ainda, é que os níveis estimados para a intensidade de sintomas depressivos são calculados considerando 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.

É importante salientar que ainda se fazem necessários estudos que busquem a obtenção de modelos 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 existência destes pode contribuir muito na identificação de indivíduos com potencial depressivo.

Também são necessários estudos que visem a generalização do Modelo TRI para Itens Constrangedores de modo que possa ser utilizado nos instrumentos em sua escala original, pois a transformação de uma escala que é originalmente categórica ordinal em uma escala dicotômica, do tipo ausência ou presença, pode acarretar perda de informação e conseqüentemente levar a conclusões equivocadas. Outra sugestão que

pode contribuir ainda mais para a eficiência deste modelo, talvez fosse a possibilidade de poder estimar diferentes parâmetros de discriminação para os diferentes grupos sendo comparados, pois, algumas vezes, o viés do item também reside neste parâmetro.

Este trabalho contribuiu para mostrar o potencial que os modelos da Teoria da Resposta ao Item têm quando se está trabalhando com a estimação de traços latentes, em especial, no caso da intensidade de sintomas depressivos, característica muito importante que auxilia os profissionais no fechamento do diagnóstico da Depressão.

9. ANEXOS

| | |
|--|------------|
| ANEXO A – Projeto de Pesquisa | 108 |
| ANEXO B – Inventário de Depressão Beck (BDI) | 125 |
| ANEXO C – Curvas de categoria de resposta, Curvas de informação do item para cada item do BDI e Curva de informação do teste – artigo 1 | 129 |
| ANEXO D – Gráficos resultantes do procedimento de Análise Paralela utilizado nos dois artigos | 152 |

ANEXO A – Projeto de Pesquisa

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA**

PROJETO DE PESQUISA

**TEORIA DA RESPOSTA AO ITEM:
APLICAÇÃO NA AVALIAÇÃO DA SEVERIDADE
DE SINTOMAS DEPRESSIVOS**

Autora: Stela Maris de Jezus Castro

Orientador: Prof. Dr. João Riboldi

1. INTRODUÇÃO

Depressão é uma doença do organismo como um todo, a qual compromete o aspecto físico, o humor e, conseqüentemente, o pensamento. Ela altera a maneira como o indivíduo vê o mundo e sente a realidade, entende as coisas, manifesta emoções, sente a disposição e o prazer com a vida. Logo a depressão é uma doença afetiva ou do humor e é considerada como um transtorno psiquiátrico e que exige tratamento.

A depressão é essencialmente um transtorno episódico e recorrente. Cada episódio depressivo dura, em geral, 20 semanas, seguido de um período normal entre os mesmos, sendo que em torno de 12% dos pacientes não apresentam remissão de sintomas. A taxa de recorrência para aqueles que se recuperam do primeiro episódio depressivo fica em torno de 80%, em uma média de quatro anos¹.

A depressão, de um modo geral, pode ser classificada em dois tipos: depressão unipolar (ou depressão maior) e depressão bipolar. A idéia de depressão unipolar é primariamente médica, isto é, envolve um modo particular de se olhar o distúrbio psicológico. Este centraliza sobre a noção de uma síndrome que é distinta de outras síndromes psiquiátricas. Algumas destas podem ser relativamente fáceis de serem distinguidas, por exemplo, esquizofrenia paranóide, enquanto outras são reconhecidas por serem relacionadas. As desordens que mais se assemelham com depressão unipolar são outras desordens afetivas, ou seja, condições que são caracterizadas por distúrbio de humor. Elas cobrem um número de desordens de ansiedade, outras condições depressivas e desordem de humor bipolar².

Depressão bipolar é identificada pela presença de duas espécies de episódio nos quais o humor associado é depressivo ou predominantemente alegre. Ela se distingue da desordem unipolar de diversas maneiras (tais como herança, curso e desfecho), e a distinção é bastante útil. Entretanto, episódios depressivos na desordem bipolar não podem ser distinguidos sintomaticamente daqueles da depressão unipolar. A metade dos casos de desordem bipolar começa com um episódio depressivo fazendo com que os pacientes sejam classificados como tendo depressão unipolar. Somente em 5% dos casos, estes pacientes são reclassificados como sofrendo de depressão bipolar³.

Entre as causas estudadas da depressão estão fatores genéticos⁴ e desequilíbrio bioquímico⁵. Os eventos desencadeadores são muito estudados e de fato encontra-se relação entre certos acontecimentos estressantes na vida das pessoas e o início de um episódio depressivo⁵.

Os sintomas da depressão são muito variados, indo desde as sensações de tristeza, passando pelos pensamentos negativos até as alterações da sensação corporal como dores e enjôos. Porém, segundo a CID-10, a fim de se fazer o diagnóstico é necessário um grupo de sintomas fundamentais tais como perda de energia ou interesse, humor deprimido ou fadigabilidade, e um grupo de sintomas acessórios que são: dificuldade de atenção e concentração, auto-estima e auto-confiança reduzidas, idéias de culpa e inutilidade, visões desoladas e pessimistas do futuro, idéias ou atos autolesivos ou suicídio, alterações do sono e diminuição do apetite¹.

A depressão está associada com prejuízo cognitivo. Perda de memória ocasional parece ser o principal aspecto do funcionamento cognitivo que é vulnerável aos efeitos negativos da depressão⁶. Também, ela prejudica a habilidade de execução das funções normais, levando à diminuição da capacidade em 50% ou mais dos pacientes^{7,8}. Diminuição de capacidade é provavelmente uma consequência natural da depressão⁹.

Um aspecto crucial da epidemiologia da depressão é o aumento da mortalidade associado com esta condição. Uma metanálise de 25 estudos com 1,3 a 16 anos de seguimento de mais de 100.000 pessoas registrou um risco relativo geral de morte em sujeitos depressivos de 1,81 [IC95% (1,58;2,07)] comparados com pessoas que não tinham depressão¹⁰. Este aumento do risco de morte é, em parte, devido ao alto risco de suicídio⁸.

Desordens depressivas são as que mais ocorrem entre as desordens mentais e de comportamento. Estudos epidemiológicos mostram que 15% a 33% dos pacientes internados sofrem de desordens depressivas, comparados com 2% a 4% da população geral, indicando que a depressão é um dos mais frequentes fenômenos psicopatológicos encontrados no ambiente hospitalar¹¹. Outros estudos apresentam a prevalência anual da depressão, na população geral, variando entre 3% e 11%¹.

Existem alguns estudos sobre a prevalência de depressão na população brasileira. Almeida *et al.*¹², reportam as prevalências encontradas para estados depressivos no Estudo Multicêntrico de Morbidez Psiquiátrica (*Multicentric Study of Psychiatric Morbidity*) conduzido, em 1991, nas áreas metropolitanas de Brasília, São Paulo e Porto Alegre, com uma amostra de 6476 adultos (com no mínimo quatorze anos). As prevalências para a vida inteira, ajustadas para idade, são 2,8, 1,9 e 10,2 para Brasília, São Paulo e Porto Alegre, respectivamente. Já as prevalências para a vida inteira, estratificadas por sexo, são 1,9 para homens e 3,8 para mulheres em Brasília, 3,8

para mulheres em São Paulo, e 5,9 para homens e 14,5 para mulheres em Porto Alegre¹².

A alta prevalência da depressão, tanto no Brasil como no mundo como um todo, acarreta prejuízos econômicos. Estes prejuízos podem ser quantificados através de medidas como DALY (*Disability-adjusted life years*), que quantifica a carga global de doenças considerando os anos de vida vividos sem plena saúde e anos de vida perdidos devido à morte prematura.

O estudo chamado *Global Burden of Disease* (GBD) quantificou os efeitos da saúde de mais de 100 doenças para o mundo como um todo e, também, estratificado para oito regiões, utilizando como medida o DALY. O estudo de Lopez e colaboradores¹³, apresenta os resultados do GBD para o ano de 2001. Entre as doenças não fatais, a depressão unipolar (ou depressão maior) foi a causa líder de anos vividos com incapacidade para homens e mulheres. Entretanto, a carga da doença foi 50% maior para as mulheres¹³.

Considerando uma faixa de idade mais restrita, Patel e colaboradores¹⁴, identificaram em um estudo em Victoria, Austrália, que desordens mentais na população jovem entre 15 e 24 anos de idade contribuem entre 60% a 70% do DALY total.

Determinar a severidade de sintomas depressivos (intensidade da depressão) pode ser importante para verificar o estágio da doença e avaliar o desfecho. A acurada, rápida e custo-efetiva medida de sintomas depressivos serve tanto para propósitos clínicos ou de pesquisa. Clínicos podem aferir o benefício do tratamento e fazer ajustes adequados no plano do mesmo. Pesquisas, por outro lado, podem ser feitas com menor custo se tais medidas estão disponíveis. Ainda, a medida de severidade de sintomas depressivos é importante para a implementação apropriada de diretrizes de tratamentos recomendados para depressão unipolar e outras desordens de humor.

A pesquisa científica na área da psiquiatria visa principalmente medir um construto, isto é, um traço latente não observável, tal como a severidade dos sintomas depressivos. Com este intuito, a psiquiatria se vale do uso de instrumentos de medida compostos por itens que pretendem operacionalizar este construto. Os instrumentos de medida recebem diferentes denominações de acordo com a área de atuação: escala, inventário, questionário ou prova. Basicamente, os instrumentos de medida são compostos por inúmeros itens que podem ser abertos ou fechados, com duas ou mais alternativas de resposta, as quais podem ser nominais ou ordinais. Os itens fechados são

os tipos mais freqüentes nestes instrumentos e, para eles, para cada alternativa de resposta atribui-se um escore de modo que as categorias do item associadas a um maior nível do traço latente devem estar pontuadas no mesmo sentido.

A severidade de sintomas depressivos é uma medida de auto-resposta que não é diretamente mensurável, logo pode ser considerada como um construto (ou traço latente). Questionários são amplamente usados para medir a severidade de sintomas depressivos, sendo que grande parte destes usam escalas multi-item. Estes questionários são compostos de itens que contemplam os mais diversos sintomas depressivos observáveis. Dentre os mais conhecidos e frequentemente utilizados estão o Inventário de Depressão Beck (*Beck Depression Inventory – BDI*), a Escala de Depressão do Centro para Estudos Epidemiológicos – (*Centre for Epidemiological Studies – Depression Scale – CES-D*), o Questionário de Saúde do Paciente (*Patient Health Questionnaire PHQ*), a Escala Hospitalar de Ansiedade e Depressão (*Hospital Anxiety Depression Scale – HADS*), a Escala de Auto-Avaliação de Depressão de Zung (*Zung Self-Rating Depression Scale*)¹¹, entre outros.

Nas últimas três décadas, o uso da Teoria da Resposta ao Item (TRI) para medida auto-referida tem florescido. TRI é frequentemente também chamada de “teoria psicométrica moderna” para distingui-la da teoria clássica de teste (TCC)¹⁵. Pesquisadores da área de saúde têm usado TRI para avaliar as características de medidas existentes¹⁶, desenvolver novas medidas^{17,18} e comparar vários grupos de itens a uma métrica matemática comum¹⁹.

Duas linhas de desenvolvimento separadas em TRI são as bases de aplicações correntes. Nos Estados Unidos, o começo de TRI é freqüentemente traçado por Lord e Novick²⁰ no seu livro *Statistical Theories of Mental Test Scores*, o qual inclui quatro capítulos sobre TRI, escritos por Allan Birnbaum. Em anos subseqüentes, diversos pesquisadores, dentre eles Samejima, Bock, Thissen e Muraki, desenvolveram modelos e métodos de estimação para os parâmetros dos mesmos, incluindo programas computacionais (BILOG, TESTFACT, MULTILOG e PARSCALE)²¹. Na Europa, este início é devido a George Rasch²². Rasch, um matemático dinamarquês, desenvolveu uma família de modelos de TRI nos quais os parâmetros de indivíduos de itens estavam totalmente separados. Os métodos de estimação para estes parâmetros foram desenvolvidos por Andersen²³.

Teoria da Resposta ao Item (TRI) oferece algumas das melhores alternativas para projetar e otimizar escalas, testes e pesquisas, e realizar análise de item. Métodos

de TRI têm vantagens distintas sobre métodos clássicos em que (a) parâmetros do item e níveis do traço latente do sujeito são independentes, (b) o modelo é expresso antes ao nível da resposta observada do item do que ao nível do escore de teste observado, (c) a contribuição de cada item para a escala pode ser determinada através da informação TRI, (d) métodos poderosos para detectar funcionamento diferencial de item (DIF) ou viés do item entre populações ou subgrupos estão disponíveis, (e) formas curtas, paralelas e adaptadas são facilmente criadas, e (f) escores de sujeitos podem ser equacionados mesmo se respondentes respondem questões diferentes²⁴. Mais ainda, não faz sentido referir à confiabilidade do teste onde a TRI é utilizada. Em se tratando de modelos TRI a acurácia da medida é estimada especificamente (medida de erro padrão) para cada nível de habilidade no traço latente²⁵.

Além disso, sendo os modelos TRI focados na medida conjunta de sujeito e itens, significa que eles são colocados na mesma medida continuum com a mesma métrica, ao contrário do TCC que somente coloca pessoas dentro do continuum e supõe que todos os itens contribuem igualmente para a medida do construto.

Em função das inúmeras características dos modelos TRI para traços latentes, pretende-se analisar dados sobre severidade de sintomas depressivos provenientes de sujeitos submetidos ao instrumento BDI, através da utilização desta metodologia.

2. JUSTIFICATIVA

A Teoria da Resposta ao Item é de extrema relevância na análise de construtos latentes, pois propicia um maior aproveitamento da informação, trabalhando não somente com a classificação dos indivíduos quanto ao mesmo, como também traz informações sobre os instrumentos de medida como um todo e, principalmente, item a item. A utilização da Teoria da Resposta ao Item ainda é incipiente no Brasil, nas áreas da psiquiatria e psicologia, muito embora a maioria das publicações a nível internacional, nestas áreas, utilizem a metodologia. No Rio Grande do Sul nenhum estudo na área de psiquiatria usou TRI.

Acredita-se, associando-se as características altamente vantajosas da metodologia e a originalidade de utilização no estado, que o projeto é importante e relevante, justificando seu desenvolvimento.

3. OBJETIVOS

3.1. Objetivo Geral:

Avaliar a severidade de sintomas depressivos aplicando modelos da Teoria da Resposta ao Item em dados, provenientes de sujeitos submetidos ao Inventário de Depressão Beck (BDI), cuja composição é feita por pacientes psiquiátricos, por pacientes de clínica médica e pela população geral (não-clínica).

3.2. Objetivos Específicos dos artigos:

ARTIGO 1 - avaliar a severidade de sintomas depressivos através de um modelo politômico da Teoria da Resposta ao Item quando da aplicação do Inventário de Depressão Beck e identificar itens do mesmo que apresentem viés de resposta quando comparando homens e mulheres, através de uma análise de funcionamento diferencial, fazendo um paralelo, quando possível, com os resultados da análise clássica tradicional.

ARTIGO 2 - Identificar itens que apresentem viés de resposta (DIF) de acordo com o sexo do respondente no Inventário de Depressão Beck através de uma análise de funcionamento diferencial, e ajustar o modelo da Teoria da Resposta ao Item para Itens com DIF.

4. MATERIAL E MÉTODOS

4.1. MATERIAL

Os dados são provenientes de um estudo transversal e referem-se a uma amostra composta por 4026 sujeitos distribuídos em três grandes grupos:

1. Pacientes psiquiátricos ($n = 1138$) - pacientes com transtornos depressivos, transtornos de ansiedade, transtorno de pânico sem ou com agorafobia, transtorno obsessivo-compulsivo, fobia social, fobia simples, agorafobia, transtorno de dependência do álcool, dependência de outras substâncias ou transtornos psiquiátricos não especificados.

2. Pacientes de clínica médica ($n = 491$) - pacientes com cardiopatia, com dor crônica, portadores de HIV assintomáticos e sintomáticos, com disfunção erétil, com obesidade, com diabetes mellitus tipo II, ou pacientes de unidades de cuidados primários em saúde.
3. Amostra não-clínica, constituída por grupos da população geral ($n = 2397$) - composta por universitários, reclusos de penitenciária de segurança máxima, bombeiros, funcionários de hospital geral, adolescentes, idosos.

Esta amostra foi constituída a fim de se realizar a validação para o português das Escalas Beck - Inventário de Depressão Beck (BDI), Inventário de Ansiedade Beck (BAI), Escala de desesperança Beck (BHS) e Escala de ideação Suicida Beck (BSI) – em um estudo conduzido pela psiquiatra Jurema Alcides Cunha e publicado em 2001²⁶.

Os sujeitos da amostra foram submetidos ao BDI (em anexo) cujos dados serão utilizados neste trabalho. O BDI é uma escala de auto-relato, de 21 itens, cada um com quatro alternativas, subentendendo graus crescentes de severidade de sintomas depressivos²⁶, com 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 depressivos²⁶. Eles se referem a: (1) Tristeza; (2) Pessimismo; (3) Sentimento de fracasso; (4) Insatisfação; (5) Culpa; (6) Punição; (7) Auto-aversão; (8) Auto-acusações; (9) Idéias suicidas; (10) Choro; (11) Irritabilidade; (12) Retraimento social; (13) Indecisão; (14) Mudança na auto-imagem; (15) Dificuldade de trabalhar; (16) Insônia; (17) Fatigabilidade; (18) Perda de apetite; (19) Perda de peso; (20) Preocupações somáticas; (21) Perda da libido.

4.2. ANÁLISE ESTATÍSTICA

Na análise dos dados utilizar-se-á a metodologia da Teoria da Resposta ao Item (TRI). A TRI propõe modelos para os traços latentes, os quais devem ser inferidos através da observação de variáveis secundárias que tenham relação com os mesmos. Os modelos TRI sugerem formas de representar a relação entre a probabilidade de um sujeito responder uma determinada categoria (no caso de itens fechados) e seu traço latente²⁷.

Os modelos TRI podem ser classificados, de um modo geral, de acordo com o tipo de resposta ao item (dicotômica ou politômica), com o número de grupos (ou populações) envolvidos, e de acordo com sua dimensionalidade. Em modelos unidimensionais um único traço latente é considerado suficiente para caracterizar diferenças pessoais. Eles são apropriados para dados nos quais um único fator comum forma a base de resposta do item. Entretanto, modelos unidimensionais não são adequados para dados nos quais dois ou mais traços latentes têm diferente impacto sobre os itens e pessoas diferem sistematicamente nas estratégias, estruturas de conhecimento ou interpretações que elas aplicam aos itens. Nestes casos, um modelo TRI multidimensional é mais apropriado²¹.

A severidade dos sintomas depressivos, no presente caso, foi medida através do BDI, logo modelos TRI unidimensionais de resposta politômica são os mais apropriados. Neste trabalho, poderão ser utilizados os seguintes modelos:

- Modelo de Resposta Gradual (*Graded-Response Model - GRM*)

O modelo de Resposta Gradual²¹, proposto por Samejima em 1969, é apropriado quando as respostas do item podem ser caracterizadas como respostas categóricas ordenadas, como é o caso das escalas de classificação de Likert. O GRM é uma generalização do modelo logístico de 2 parâmetros e é considerado como um modelo TRI “indireto” porque o cálculo da probabilidade condicional de um indivíduo responder uma determinada categoria requer um procedimento de dois passos. Para ajustar o GRM para uma medida (por exemplo, intensidade da depressão) os itens do instrumento não precisam ter o mesmo número de categorias de resposta.

No GRM, cada item (i) é descrito por um parâmetro de inclinação do item (α_i) e $j = 1, 2, \dots, m_i$ parâmetros *threshold* (β_{ij}) entre as categorias de resposta (isto é, parâmetros que determinam o ponto de intersecção entre as categorias ordenadas). Sendo assim, $m_i + 1 = K_i$ é o número de categorias de resposta do item i . Um dos objetivos do GRM é determinar a locação destes *thresholds* no continuum do traço latente.

O primeiro passo, para estimar as probabilidades de resposta no GRM envolve o cálculo de m_i curvas para cada item através da seguinte expressão:

$$P_{ix}^*(\theta) = \frac{\exp[\alpha_i(\theta - \beta_{ij})]}{1 + \exp[\alpha_i(\theta - \beta_{ij})]} \quad (1)$$

onde $x = j = 1, 2, \dots, m_i$. Cada curva representa a probabilidade de que uma resposta de um item (x) esteja em uma determinada, ou acima, categoria *threshold* ($j = 1, 2, \dots, m_i$) condicional ao nível do traço latente (θ). As curvas $P_{ix}^*(\theta)$ são chamadas de “curvas características de operação”. No GRM, uma curva característica de operação deve ser estimada para cada *threshold* entre categorias e, conseqüentemente, para um item de resposta gradual com quatro categorias, são estimados três parâmetros β_{ij} e um parâmetro de inclinação comum (α_i). Os parâmetros β_{ij} têm uma interpretação e representam o nível de traço latente necessário para responder, com probabilidade de 0,50, acima do *threshold* j . Essencialmente, o que ocorre no GRM é que o item é tratado como uma série de $m_i = K - 1$ itens dicotômicos (ou seja, 0 vs. 1, 2, 3; 0,1 vs. 2, 3; 0,1, 2 vs. 3) e modelos de 2 parâmetros são estimados para cada um com a restrição que as inclinações das curvas características de operação são as mesmas para um mesmo item politômico. A probabilidade real de responder a categoria x do item i condicional ao nível de traço latente θ é obtida por diferença:

$$P_{ix}(\theta) = P_{ix}^*(\theta) - P_{i(x+1)}^*(\theta) \quad (2)$$

Por definição, $P_{i(x=0)}^*(\theta) = 1$ e $P_{i(x=m+1)}^*(\theta) = 0$.

As curvas produzidas através da expressão (2) são chamadas de curvas de categoria de resposta.

- Modelo de Crédito Parcial Generalizado (*Generalized Partial Credit Model – G-PCM*)

O modelo de Crédito Parcial Generalizado²¹, desenvolvido por Muraki em 1992, é uma generalização do modelo de Crédito Parcial de Masters, que é apropriado para analisar respostas de escalas de atitude ou personalidade onde sujeitos marcam sua opinião, ou respondem afirmações em uma escala de múltiplos pontos. A generalização se deve ao fato de que o G-PCM permite que os itens dentro de uma escala possam diferir em relação ao parâmetro de inclinação.

Diferente do modelo GRM, o modelo de crédito parcial generalizado é considerado “direto”. Isto significa que a probabilidade de um indivíduo responder uma certa categoria será escrita diretamente como uma exponencial dividida pela soma de exponenciais. Assumindo que o item i é escoreado por $x = 0, 1, \dots, m_i$ para um item com $K_i = m_i + 1$ categorias de resposta. Para $x = j$ as curvas de resposta de categoria para o G-PCM podem ser escritas como:

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^x \alpha_i(\theta - \delta_{ij})\right]}{\sum_{x=0}^{m_i} \exp\left[\sum_{j=0}^x \alpha_i(\theta - \delta_{ij})\right]} \quad (3)$$

onde $\sum_{j=0}^0 \alpha_i(\theta - \delta_{ij}) \equiv 0$.

O termo δ_{ij} ($j = 1, 2, \dots, m_i$) é o parâmetro de item que regula a probabilidade de se escolher a categoria j em vez da categoria adjacente ($j-1$) no item i . Cada parâmetro δ_{ij} corresponde ao nível de traço latente em que o indivíduo tem a mesma probabilidade de responder à categoria j e à categoria ($j-1$), isto é, onde $P_{i,j}(\theta) = P_{i,j-1}(\theta)$. Já os parâmetros de inclinação (α_i) indicam o grau em que as respostas categóricas variam entre itens a medida que o nível do traço latente (θ) muda. Se os parâmetros de inclinação forem menores que 1, as curvas de categoria de resposta serão relativamente horizontais, enquanto que se eles forem maiores que 1 as curvas tenderão a ser relativamente mais pontudas, quando comparadas às curvas produzidas pelo modelo de créditos parciais.

O ajuste dos modelos TRI politômicos e a conseqüente análise dos dados deverá ser efetivada no software PARSCALE, versão 4.1.

O ajuste do modelo TRI para itens com DIF deverá ser feito através de rotinas elaboradas no WINBUGS²⁸ – as quais utilizam os métodos de estimação de parâmetros conhecidos como máxima verossimilhança marginal e bayesiano.

5. ASPECTOS ÉTICOS

Dados autorizados para uso através de carta de autorização fornecida pelas proprietárias dos mesmos.

6. CRONOGRAMA

| 2008 | | | | | | |
|-------------------------|-----|-----|-----|-----|-----|-----|
| | JAN | FEV | MAR | ABR | MAI | JUN |
| Apresentação do projeto | X | | | | | |
| Revisão de Literatura | X | X | | | | |
| Análise e interpretação | | | X | X | | |
| Elaboração artigo 1 | | | | | X | X |
| | JUL | AGO | SET | OUT | NOV | DEZ |
| Elaboração do artigo2 | X | X | | | | |
| Encaminhamento ao PPG | | | X | | | |
| Apresentação preliminar | | | | X | | |
| Defesa da tese | | | | | X | |

7. LOCAL DE ORIGEM

Esta pesquisa está vinculada ao Programa de Pós-Graduação em Epidemiologia (Curso de Doutorado) da Faculdade de Medicina da Universidade Federal do Rio Grande do Sul.

8. LOCAL DE REALIZAÇÃO

Este projeto será desenvolvido no Programa de Pós-Graduação em Epidemiologia da Faculdade de Medicina da Universidade Federal do Rio Grande do Sul.

9. FINANCIAMENTO

Este projeto não dispõe de recursos financeiros para sua execução. Todas as despesas para a execução do mesmo ficarão sob responsabilidade dos pesquisadores envolvidos, com exceção da compra do software PARSCALE, adquirido com verba do Programa de Pós-Graduação em Epidemiologia da Faculdade de Medicina da Universidade Federal do Rio Grande do Sul.

REFERÊNCIAS

- (1) Fleck MPA, Lafer B, Sougey EB, Porto JA, Brasil MA, Juruena MF. Diretrizes da Associação Médica Brasileira para o tratamento da depressão (versão integral). *Revista Brasileira de Psiquiatria*. 2003;25:114-122.
- (2) Bebbington P. The classification and Epidemiology of Unipolar Depression. In: Power M, ed. *Mood Disorders*. 1 ed. Chichester: John Wiley & Sons Ltd.; 2004:3-28.
- (3) Ramana R, Bebbington P. Social influences on bipolar affective disorders. *Soc Psychiatry Psychiatr Epidemiol*. 1995;30:152-160.
- (4) Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000;157:1552-1562.
- (5) Marot, R. <http://www.psicosite.com.br/tra/hum/depressao.htm> . 8-10-0004.
Ref Type: Electronic Citation
- (6) Airaksinen E, Larsson M, Lundberg I, Forsell Y. Cognitive functions in depressive disorders: evidence from a population-based study. *Psychological Medicine*. 2004;34:83-91.
- (7) 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;289:3095-3105.
- (8) Ebmeier KP, Donaghey C, Steele JD. Recent developments and current controversies in depression. *Lancet*. 2006;367:153-167.
- (9) Ormel J, Oldehinkel AJ, Nolen WA, Vollebergh W. Psychosocial disability before, during, and after a major depressive episode: a 3-wave population-based study of state, scar, and trait effects. *Arch Gen Psychiatry*. 2004;61:387-392.
- (10) Cuijpers P, Smit F. Excess mortality in depression: a meta-analysis of community studies. *J Affect Disord*. 2002;72:227-236.
- (11) Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res*. 2005;14:2277-2291.

- (12) Almeida-Filho N, Mari JJ, Coutinho E et al. Brazilian multicentric study of psychiatric morbidity. Methodological features and prevalence estimates. *The British Journal of Psychiatry*. 1997;171:524-529.
- (13) Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL. Global and regional burden of disease and risk factors, 2001: systematic analysis of population data. *The Lancet*. 2006;367:1747-1757.
- (14) Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. *The Lancet*. 2007;369:1302-1313.
- (15) Cook KF, O'Malley KJ, Roddey TS. Dynamic Assessment of Health Outcomes: Time to Let the CAT Out of the Bag? *Health Services Research*. 2005;40:1694-1711.
- (16) 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.
- (17) Velozo CA, Peterson EW. Developing meaningful Fear of Falling Measures for community dwelling elderly. *Am J Phys Med Rehabil*. 2001;80:662-673.
- (18) Cook KF, Roddey TS, Gartsman GM, Olson SL. Development and psychometric evaluation of the Flexilevel Scale of Shoulder Function. *Med Care*. 2003;41:823-835.
- (19) McHorney CA, Cohen AS. Equating health status measures with item response theory: illustrations with functional status items. *Med Care*. 2000;38:II43-II59.
- (20) Lord FN, Novick MR. *Statistical theories of mental test scores*. Reading, MA: 1968.
- (21) Embretson SE, Reise SP. *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc.; 2000.
- (22) Rasch G. *Probabilistic models for some intelligence and attainment tests*. Chicago: 1960.
- (23) Andersen EB. The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society, Series B*. 1972;34:42-54.

- (24) Uttaro T, Lehman A. Graded response modeling of the Quality of Life Interview. *Evaluation and Program Planning*. 1999;22:41-52.
- (25) Tejedas AJR, Rojas OML. Application of an IRT Model for measuring Health Related Quality of Life. *Social Indicators Research*. 2005;74:369-394.
- (26) Cunha, J. A. Manual da versão em português das ESCALAS BECK. São Paulo. Casa do Psicólogo. 2001.
Ref Type: Report
- (27) Andrade, D. F., Tavares, H. R., and Valle, R. C. Teoria da Resposta ao Item: conceitos e aplicações. ABE - Associação Brasileira de Estatística. 14° SINAPE - Caxambu. 1-154. 28-7-2000.
Ref Type: Serial (Book, Monograph)
- (28) Cúri M. *Análise de questionários com itens constrangedores*. [Universidade de São Paulo, 2006.

ANEXO B – Inventário de Depressão Beck (BDI)

Inventário de Depressão Beck (BDI)

Este questionário consiste em 21 grupos de afirmações. Depois de ler cuidadosamente cada grupo, faça um círculo em torno do número (0,1,2 ou 3) próximo à afirmação, em cada grupo, que descreve **melhor** a maneira que você tem se sentido na **última semana, incluindo hoje**. Se várias afirmações num grupo parecerem se aplicar igualmente bem, faça um círculo em cada uma. **Tome o cuidado de ler todas as afirmações, em cada grupo, antes de fazer a sua escolha.**

Item 1:

- 8 Não me sinto triste.
- 9 Eu me sinto triste.
- 10 Estou sempre triste e não consigo sair disto.
- 11 Estou tão triste ou infeliz que não consigo suportar.

Item 2:

- 0 Não estou especialmente desanimado quanto ao futuro.
- 1 Eu me sinto desanimado quanto ao futuro.
- 2 Acho que nada tenho a esperar.
- 3 Acho o futuro sem esperança e tenho a impressão de que as coisas não podem melhorar.

Item 3:

- 0 Não me sinto um fracasso.
- 1 Acho que fracassei mais do que uma pessoa comum.
- 2 Quando olho para trás, na minha vida, tudo o que posso ver é um monte de fracassos.
- 3 Acho que, como pessoa, sou um completo fracasso.

Item 4:

- 1 Tenho tanto prazer em tudo como antes.
- 2 Não sinto mais prazer nas coisas como antes.
- 3 Não encontro um prazer real em mais nada.
- 4 Estou insatisfeito ou aborrecido com tudo.

Item 5:

- 0 Não me sinto especialmente culpado.
- 1 Eu me sinto culpado grande parte do tempo.
- 2 Eu me sinto culpado na maior parte do tempo.
- 3 Eu me sinto sempre culpado.

Item 6:

- 1 Não acho que esteja sendo punido.
- 2 Acho que posso ser punido.
- 3 Creio que vou ser punido.
- 4 Acho que estou sendo punido.

Item 7:

- 0 Não me sinto decepcionado comigo mesmo.
- 1 Estou decepcionado comigo mesmo.
- 2 Estou enojado de mim.
- 3 Eu me odeio.

Item 8:

- 0 Não me sinto de qualquer modo pior que os outros.
- 1 Sou crítico em relação a mim por minhas fraquezas ou erros.
- 2 Eu me culpo sempre por minhas falhas.
- 3 Eu me culpo por tudo de mal que acontece.

Item 9:

- 0 Não tenho quaisquer idéias de me matar.
- 1 Tenho idéias de me matar, mas não as executaria.
- 2 Gostaria de me matar.
- 3 Eu me mataria se tivesse oportunidade.

Item 10:

- 0 Não choro mais do que o habitual.
- 1 Choro mais agora do que costumava.
- 2 Agora, choro o tempo todo.
- 3 Costumava ser capaz de chorar, mas agora não consigo, mesmo que o queira.

Item 11:

- 0 Não sou mais irritado agora do que já fui.
- 1 Fico aborrecido ou irritado mais facilmente do que costumava.
- 2 Agora, eu me sinto irritado o tempo todo.
- 3 Não me irrita mais com coisas que costumavam me irritar.

Item 12:

- 0 Não perdi o interesse pelas outras pessoas.
- 1 Estou menos interessado pelas outras pessoas do que costumava estar.
- 2 Perdi a maior parte do meu interesse pelas outras pessoas.
- 3 Perdi todo o interesse pelas outras pessoas.

Item 13:

- 0 Tomo decisões tão bem quanto antes.
- 1 Adio as tomadas de decisões mais do que costumava.
- 2 Tenho mais dificuldades de tomar decisões do que antes.
- 3 Absolutamente não consigo mais tomar decisões.

Item 14:

- 1 Não acho que de qualquer modo pareço pior do que antes.
- 2 Estou preocupado em estar parecendo velho ou sem atrativo.
- 3 Acho que há mudanças permanentes na minha aparência, que me fazem parecer sem atrativo.
- 4 Acredito que pareço feio.

Item 15:

- 0 Posso trabalhar tão bem quanto antes.
- 1 É preciso algum esforço extra para fazer alguma coisa.
- 2 Tenho que me esforçar muito para fazer alguma coisa.
- 3 Não consigo mais fazer qualquer trabalho.

Item 16:

- 0 Consigo dormir tão bem como o habitual.
- 1 Não durmo tão bem como costumava.
- 2 Acordo 1 a 2 horas mais cedo do que habitualmente e acho difícil voltar a dormir.
- 3 Acordo várias horas mais cedo do que costumava e não consigo voltar a dormir.

Item 17:

- 0 Não fico mais cansado do que o habitual.
- 1 Fico cansado mais facilmente do que costumava.
- 2 Fico cansado em fazer qualquer coisa.
- 3 Estou cansado demais para fazer qualquer coisa.

Item 18:

- 0 O meu apetite não está pior do que o habitual.
- 1 Meu apetite não é tão bom como costumava ser.
- 2 Meu apetite é muito pior agora.
- 3 Absolutamente não tenho mais apetite.

Item 19:

- 0 Não tenho perdido muito peso se é que perdi algum recentemente.
- 1 Perdi mais do que 2 quilos e meio.
- 2 Perdi mais do que 5 quilos.
- 3 Perdi mais do que 7 quilos.

Estou tentando perder peso de propósito, comendo menos: Sim_____ Não_____

Item 20:

- 0 Não estou mais preocupado com a minha saúde do que o habitual.
- 1 Estou preocupado com problemas físicos, tais como dores, indisposição do estômago ou constipação.
- 2 Estou muito preocupado com problemas físicos e é difícil pensar em outra coisa.
- 3 Estou tão preocupado com meus problemas físicos que não consigo pensar em qualquer outra coisa.

Item 21:

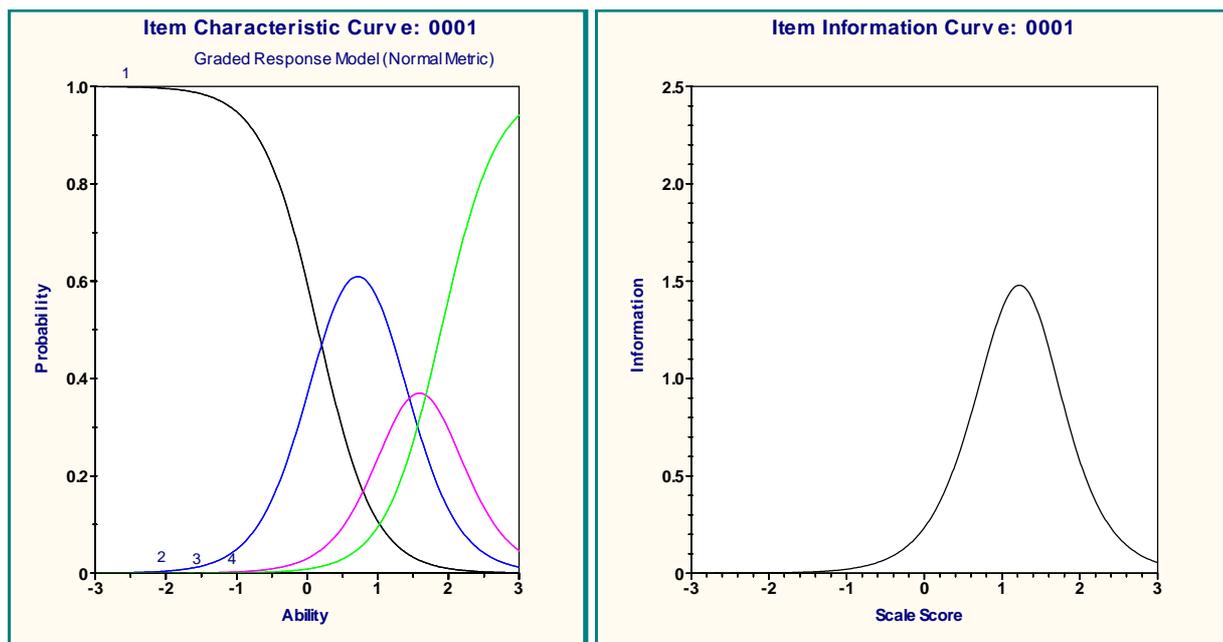
- 0 Não notei qualquer mudança recente no meu interesse por sexo.
- 1 Estou menos interessado por sexo agora.
- 2 Estou muito menos interessado por sexo agora.
- 3 Perdi completamente o interesse por sexo.

ANEXO C

CURVAS DE CATEGORIA DE RESPOSTA, CURVAS DE INFORMAÇÃO DO ITEM PARA CADA ITEM DO BDI E CURVA DE INFORMAÇÃO DO TESTE - ARTIGO 1 -

Este anexo apresenta as curvas de categoria de resposta para cada um dos 21 itens do BDI produzidas no software PARSCALE. Por imposição deste, a numeração das categorias de resposta para cada item tiveram que ser modificadas de modo que foi somado 1 a cada número, pois o PARSCALE não aceita o número zero como rótulo de categoria. Assim, as curvas identificadas com o número 1 nos gráficos referem-se à categoria zero de cada item, as identificadas com o número 2 referem-se à categoria 1, as identificadas pelo número 3 referem-se à categoria 2 e as identificadas pelo número 4 referem-se à categoria 3.

**Figura 1 – Curvas de categoria de resposta e informação do item para o item 1
- TRISTEZA -**

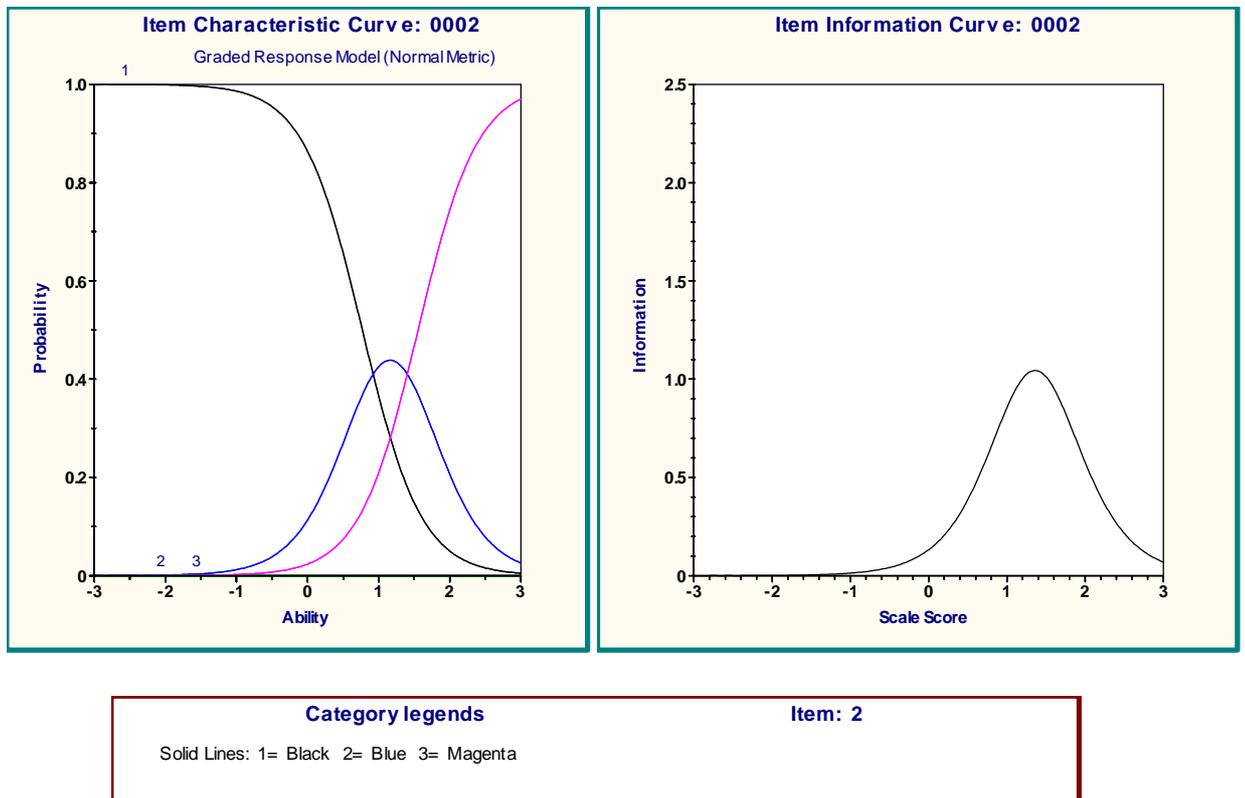


| Category legends | Item: 1 |
|---|---------|
| Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green | |

Categorias de resposta:

- 1 (curva preta) Não me sinto triste.
- 2 (curva azul) Eu me sinto triste.
- 3 (curva rosa) Estou sempre triste e não consigo sair disto.
- 4 (curva verde) Estou tão triste ou infeliz que não consigo suportar.

**Figura 2 – Curvas de categoria de resposta e informação do item para o item 2
- PESSIMISMO -**



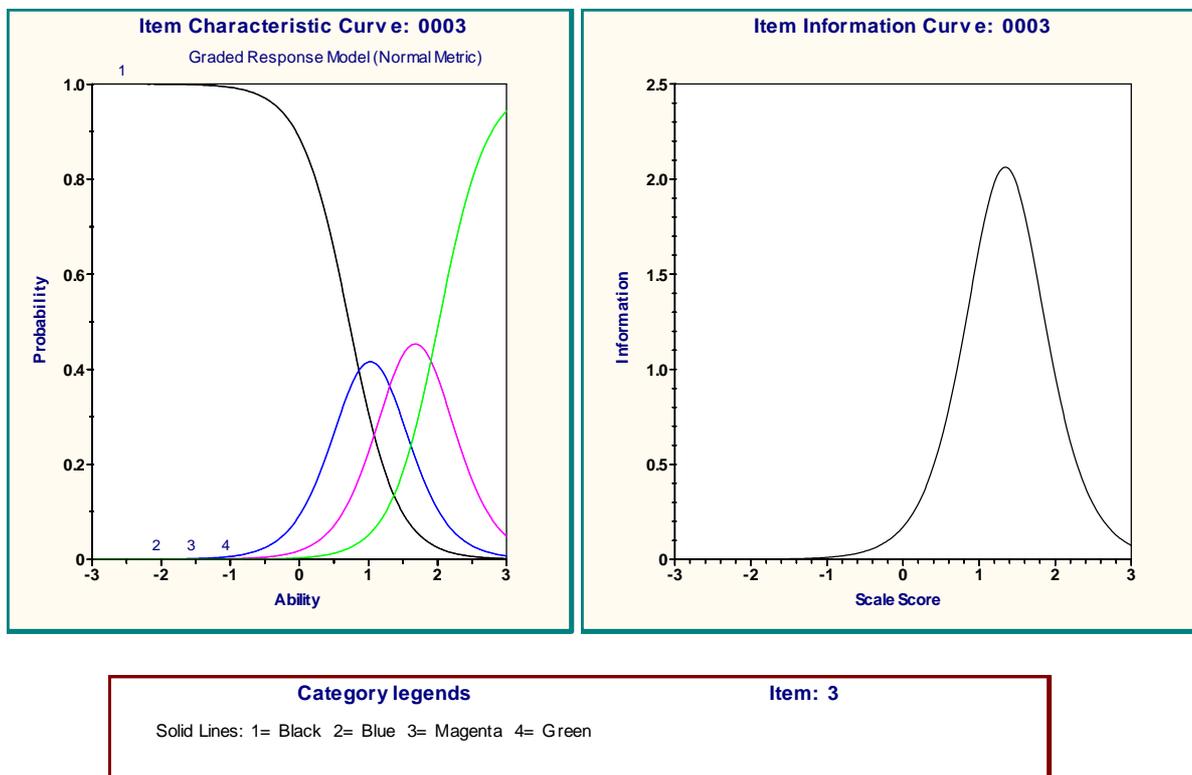
Categorias de resposta:

1 (curva preta) Não estou especialmente desanimado quanto ao futuro.

2 (curva azul) Eu me sinto desanimado quanto ao futuro.

3 (curva rosa) { Acho que nada tenho a esperar.
Acho o futuro sem esperança e tenho a impressão de que as
coisas não podem melhorar.

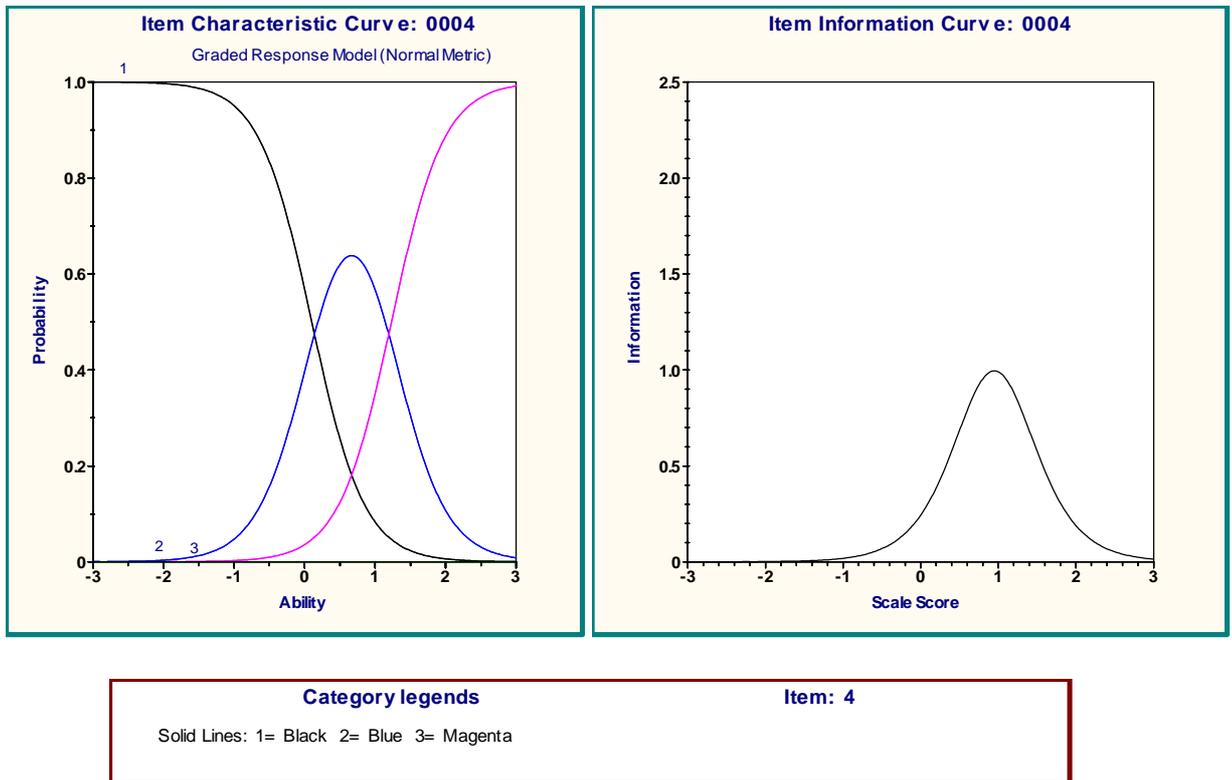
**Figura 3 – Curvas de categoria de resposta e informação do item para o item 3
- SENTIMENTO DE FRACASSO -**



Categorias de resposta:

- 1 (curva preta) Não me sinto um fracasso.
- 2 (curva azul) Acho que fracassei mais do que uma pessoa comum.
- 3 (curva rosa) Quando olho para trás, na minha vida, tudo o que posso ver é um monte de fracasso.
- 4 (curva verde) Acho que, como pessoa, sou um completo fracasso.

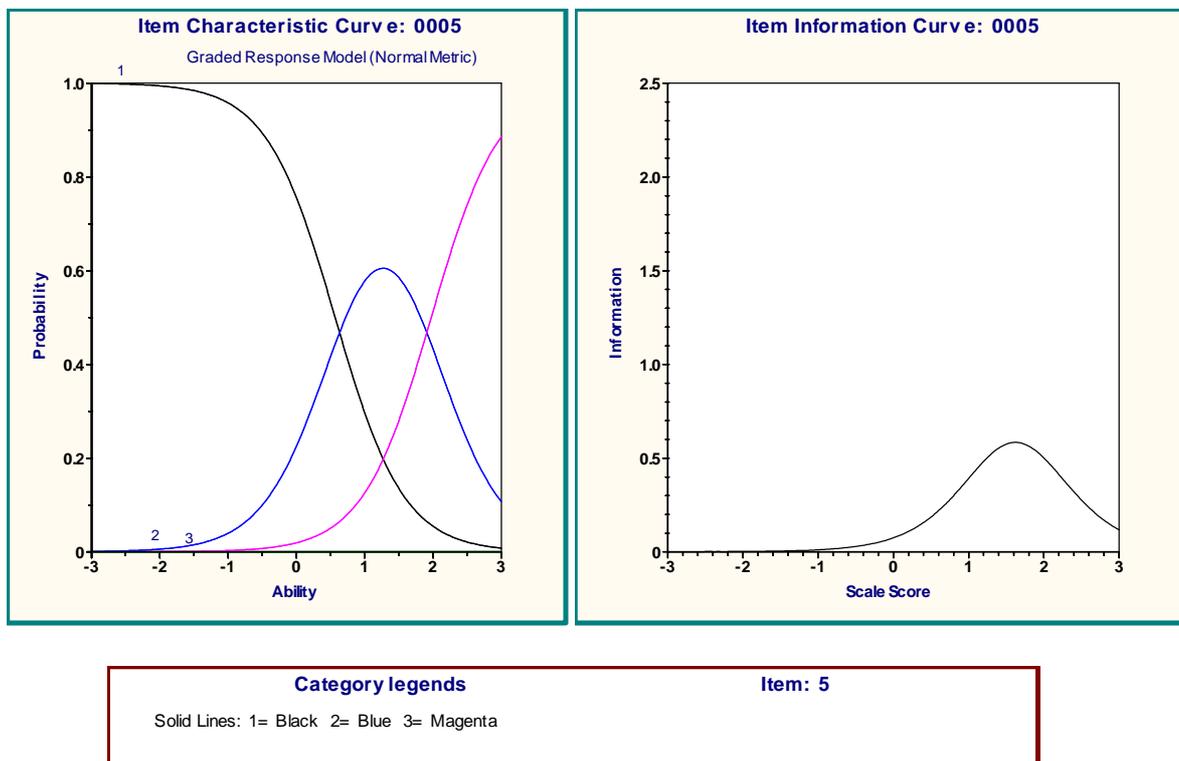
**Figura 4 – Curvas de categoria de resposta e informação do item para o item 4
- INSATISFAÇÕES -**



Categorias de resposta:

- 1 (curva preta) Tenho tanto prazer em tudo como antes.
- 2 (curva azul) Não sinto mais prazer nas coisas como antes.
- 3 (curva rosa) { Não encontro um prazer real em mais nada.
Estou insatisfeito ou aborrecido com tudo.

Figura 5 – Curvas de categoria de resposta e informação do item para o item 5
- CULPA –



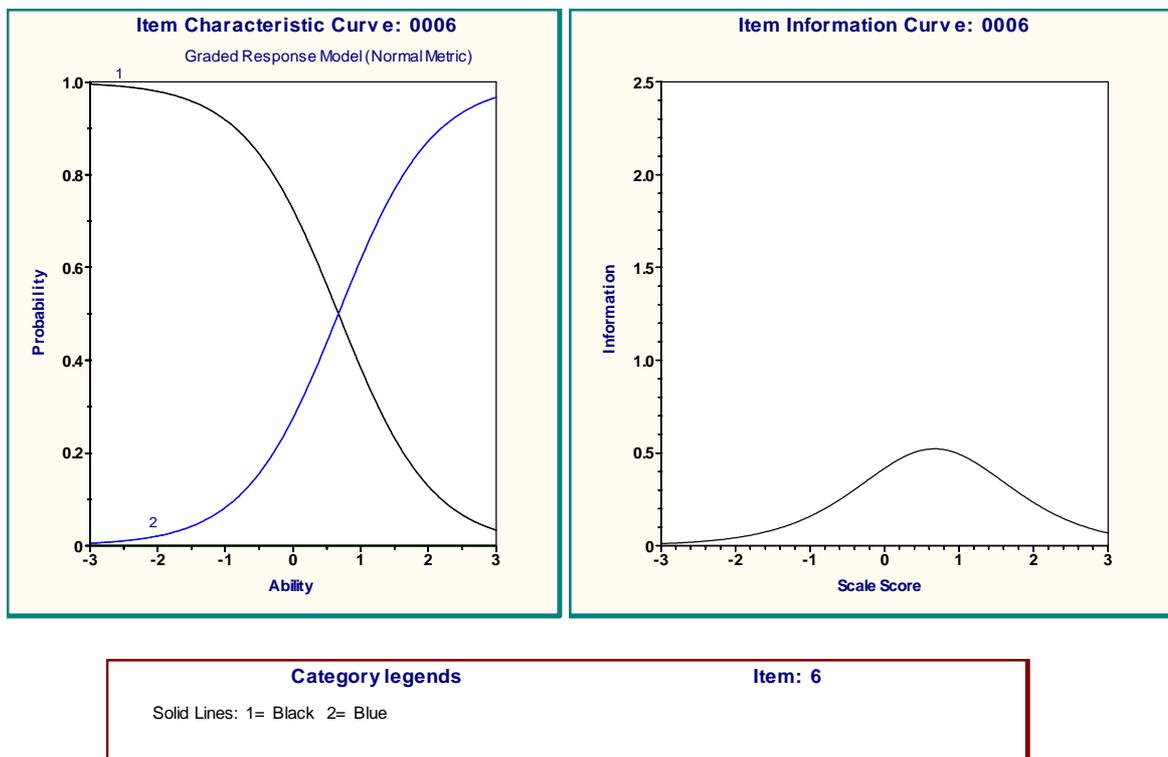
Categorias de resposta:

1 (curva preta) Não me sinto especialmente culpado.

2 (curva azul) { Eu me sinto culpado grande parte do tempo.
 { Eu me sinto culpado na maior parte do tempo.

3 (curva rosa) Eu me sinto sempre culpado.

**Figura 6 – Curvas de categoria de resposta e informação do item para o item 6
- PUNIÇÃO –**

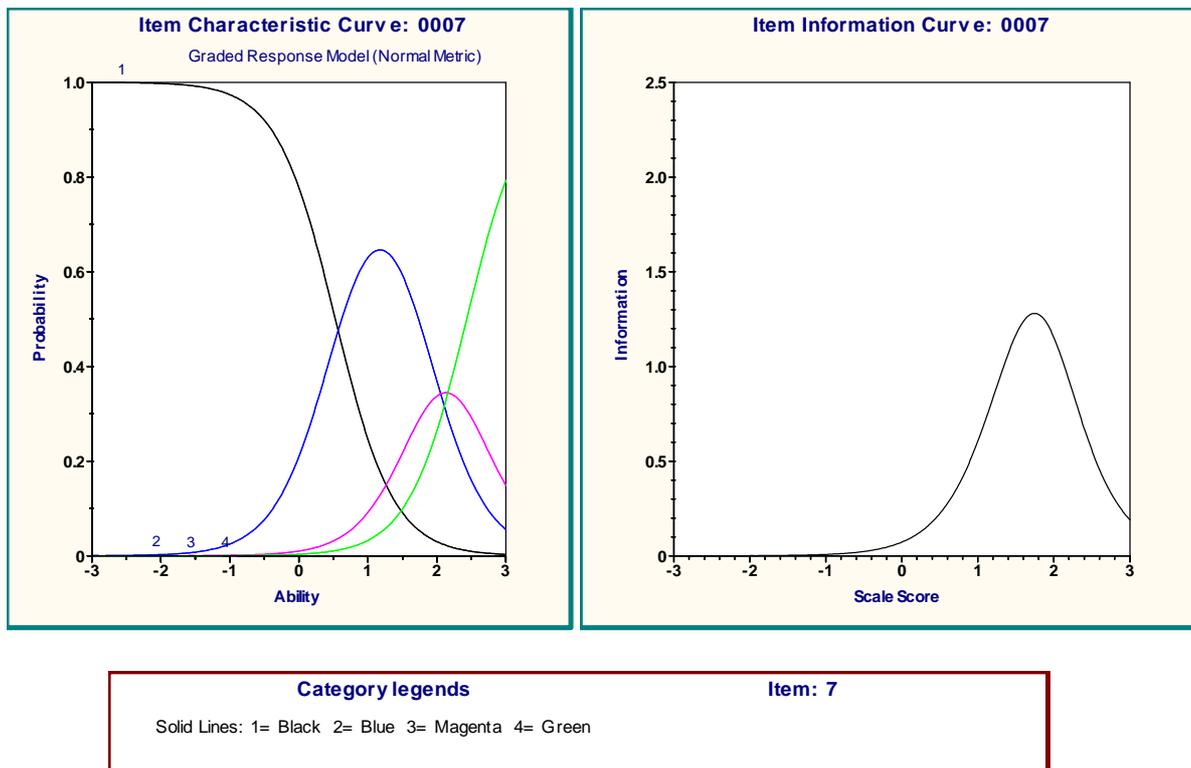


Categorias de resposta:

1 (curva preta) Não acho que esteja sendo punido.

2 (curva azul) {
Acho que posso ser punido.
Creio que vou ser punido.
Acho que estou sendo punido.

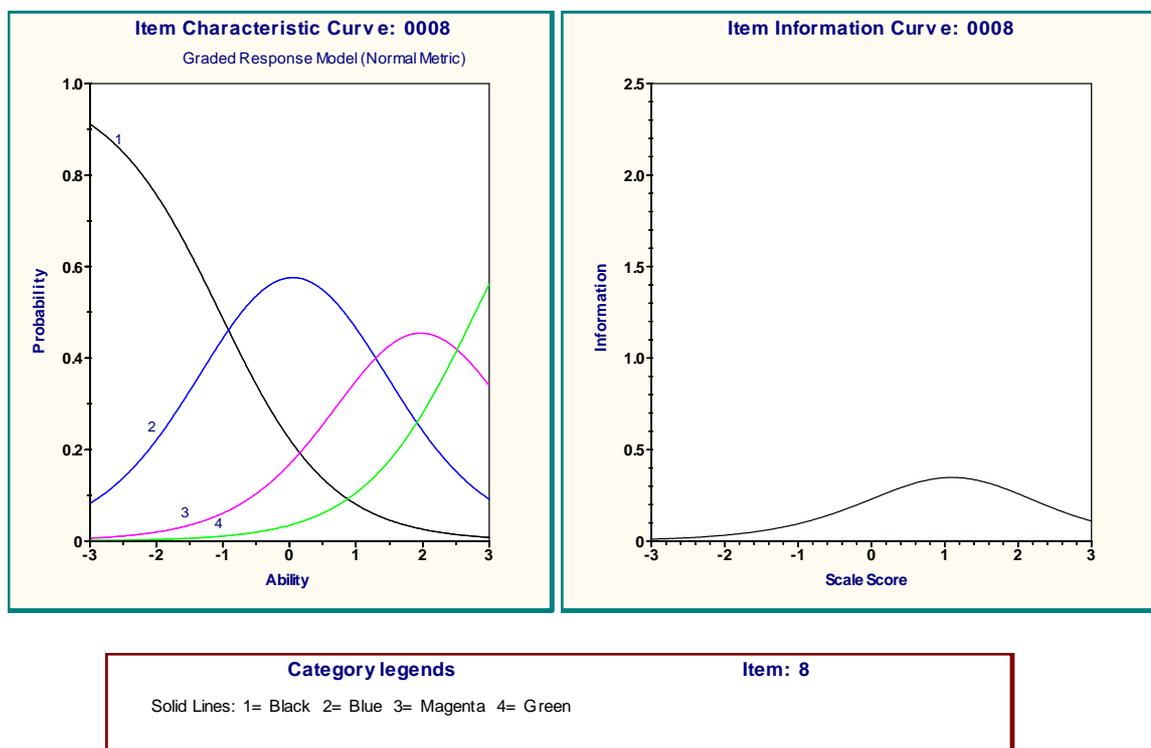
**Figura 7 – Curvas de categoria de resposta e informação do item para o item 7
- AUTO-AVERSÃO -**



Categorias de resposta:

- 1 (curva preta) Não me sinto decepcionado comigo mesmo.
- 2 (curva azul) Estou decepcionado comigo mesmo.
- 3 (curva rosa) Estou enojado de mim.
- 4 (curva verde) Eu me odeio.

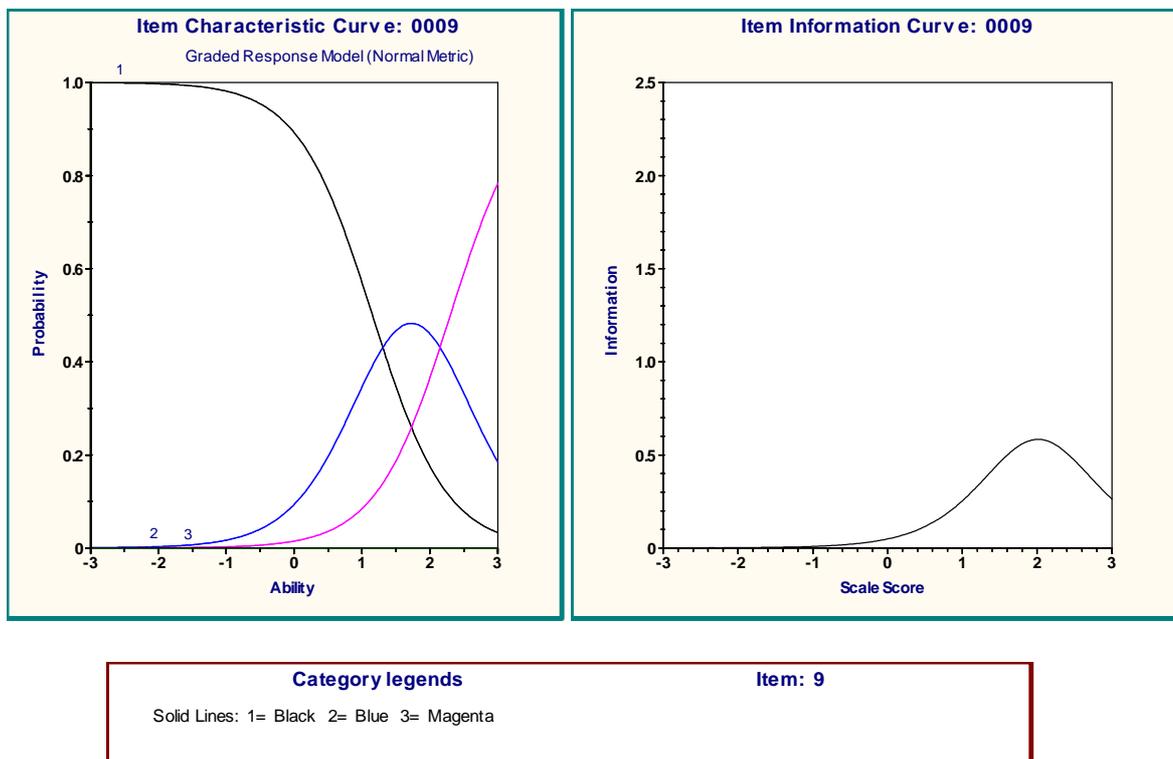
**Figura 8 – Curvas de categoria de resposta e informação do item para o item 8
- AUTO-ACUSAÇÕES -**



Categorias de resposta:

- 1 (curva preta) Não me sinto de qualquer modo pior que os outros.
- 2 (curva azul) Sou crítico em relação a mim por minhas fraquezas ou erros.
- 3 (curva rosa) Eu me culpo sempre por minhas falhas.
- 4 (curva verde) Eu me culpo por tudo de mal que acontece.

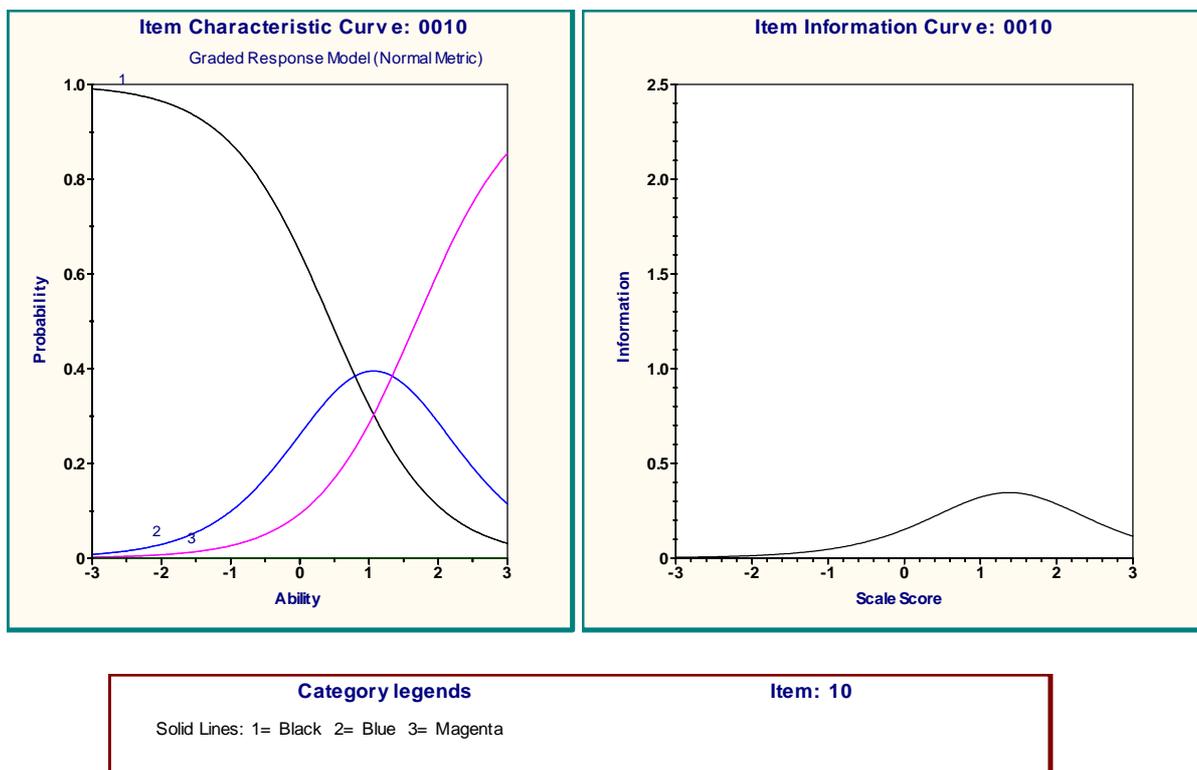
**Figura 9 – Curvas de categoria de resposta e informação do item para o item 9
- IDÉIAS SUICIDAS -**



Categorias de resposta:

- 1 (curva preta) Não tenho quaisquer idéias de me matar.
- 2 (curva azul) Tenho idéias de me matar, mas não as executaria.
- 3 (curva rosa) { Gostaria de me matar.
Eu me mataria se tivesse oportunidade.

**Figura 10 – Curvas de categoria de resposta e informação do item para o item 10
- CHORO -**



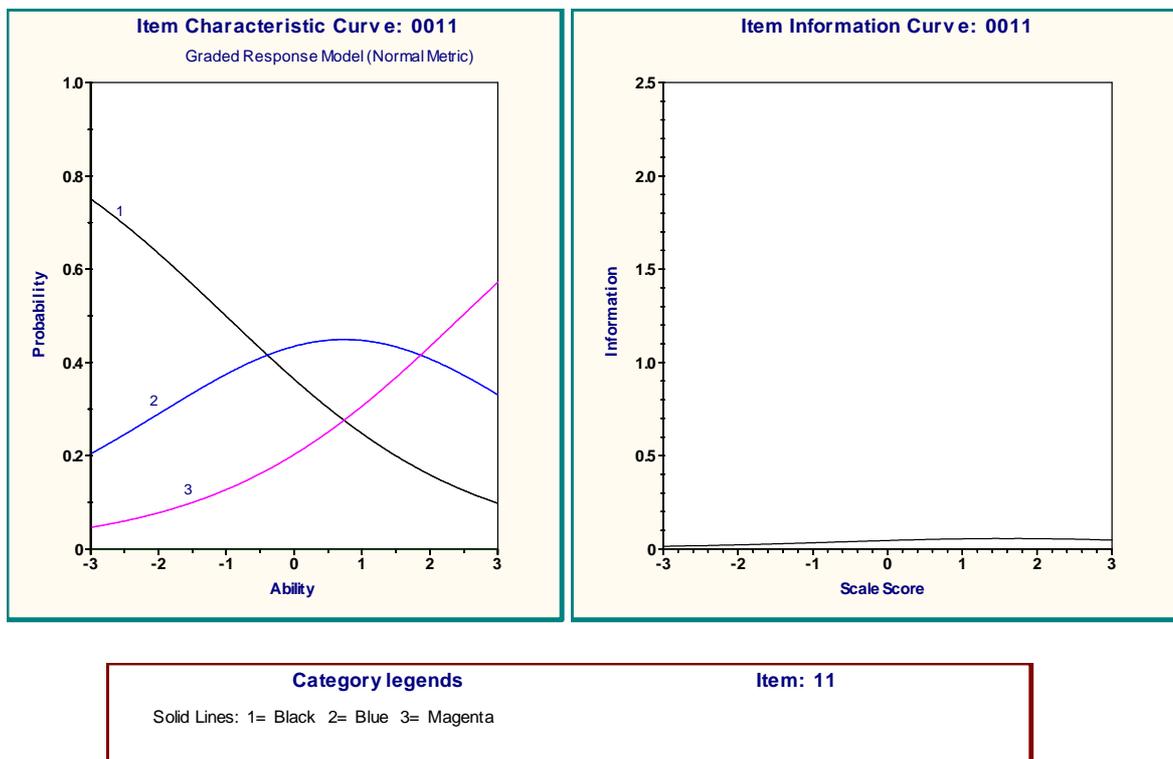
Categorias de resposta:

1 (curva preta) Não choro mais do que costumava.

2 (curva azul) { Choro mais agora do que costumava.
Agora, choro o tempo todo.

3 (curva rosa) Costumava ser capaz de chorar, mas agora não consigo, mesmo que o queira.

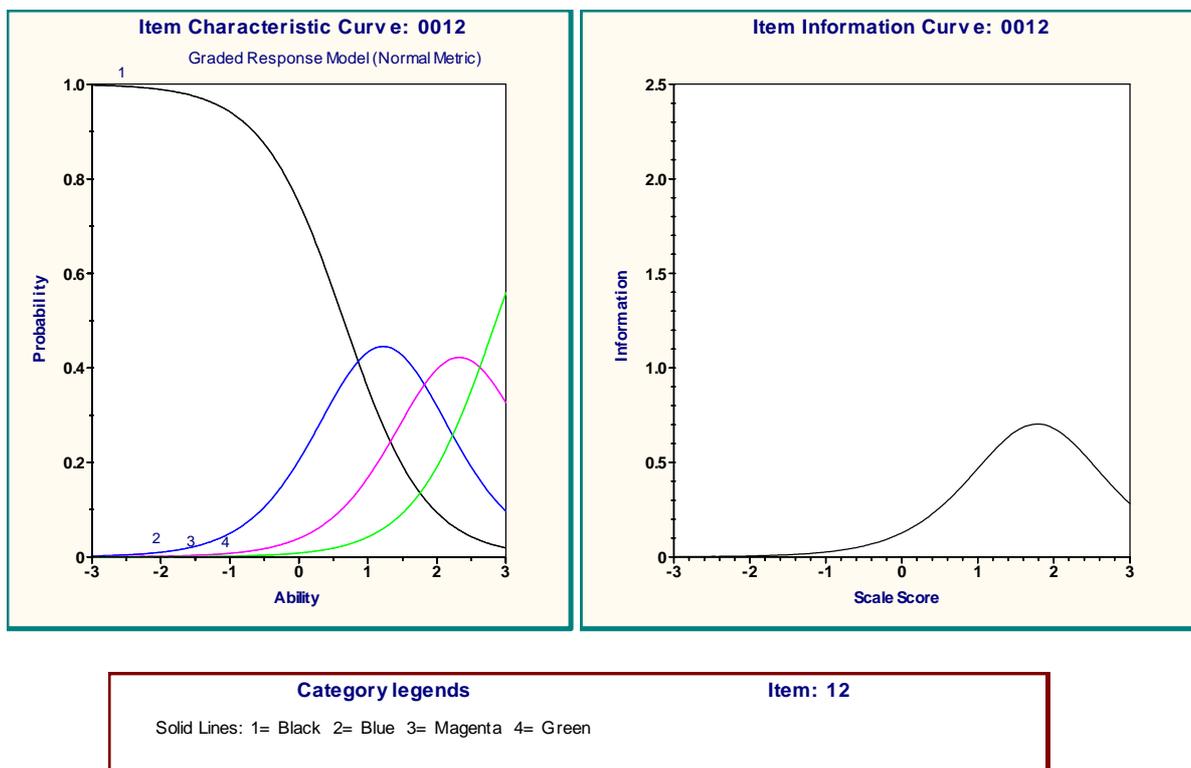
**Figura 11 – Curvas de categoria de resposta e informação do item para o item 1
- IRRITABILIDADE -**



Categorias de resposta:

- 1 (curva preta) Não sou mais irritado agora do que já fui.
- 2 (curva azul) { Fico aborrecido ou irritado mais facilmente do que costumava.
Agora, eu me sinto irritado o tempo todo.
- 3 (curva rosa) Não me irrito mais com coisas que costumavam me irritar.

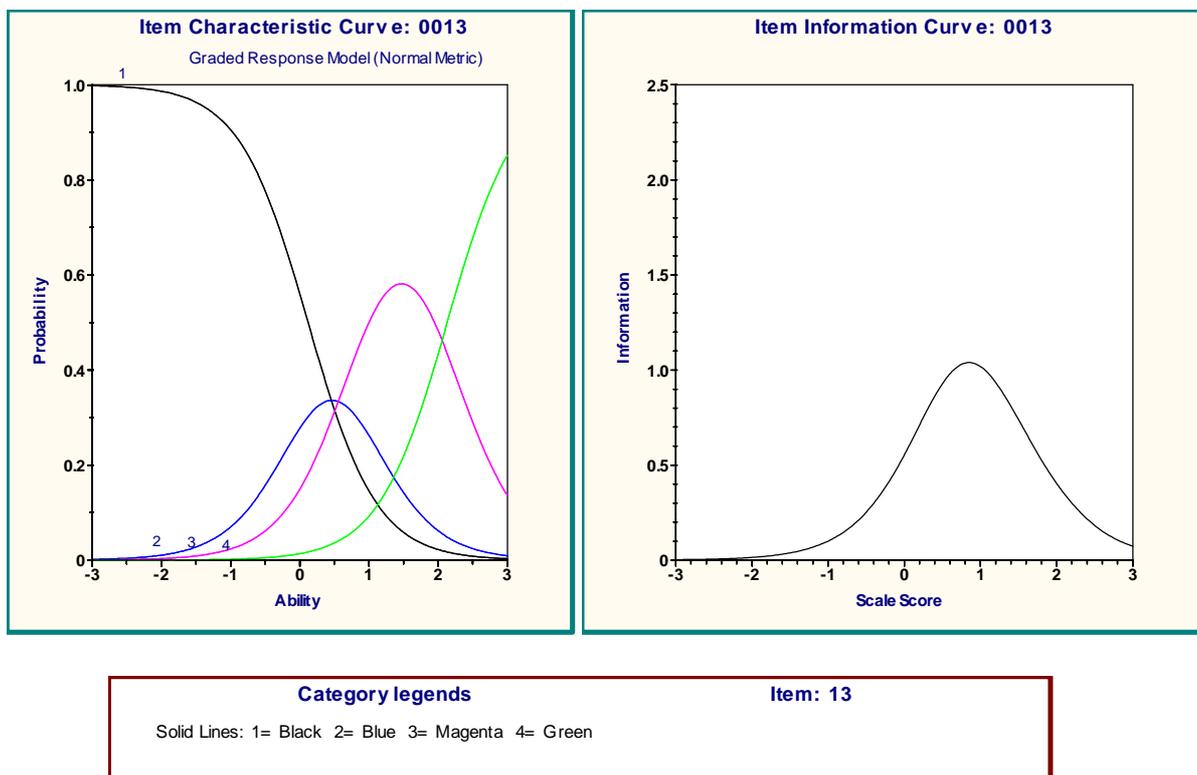
**Figura 12 – Curvas de categoria de resposta e informação do item para o item 12
- RETRAIMENTO SOCIAL -**



Categorias de resposta:

- 1 (curva preta) Não perdi o interesse pelas outras pessoas.
- 2 (curva azul) Estou menos interessado pelas outras pessoas do que costumava estar.
- 3 (curva rosa) Perdi a maior parte do meu interesse pelas outras pessoas.
- 4 (curva verde) Perdi todo o interesse pelas outras pessoas.

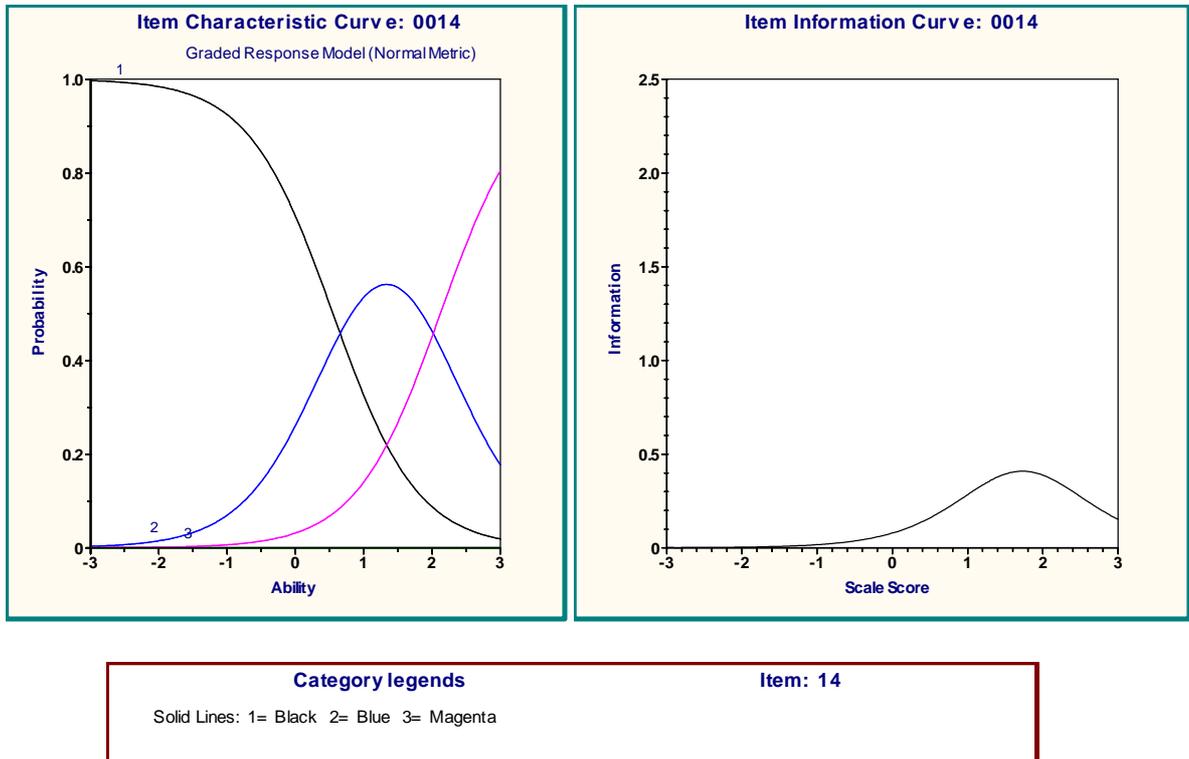
Figura 13 – Curvas de categoria de resposta e informação do item para o item 13
- INDECISÃO -



Categorias de resposta:

- 1 (curva preta) Tomo decisões tão bem quanto antes.
- 2 (curva azul) Adio as tomadas de decisões mais do que costumava.
- 3 (curva rosa) Tenho mais dificuldades de tomar decisões do que antes.
- 4 (curva verde) Absolutamente não consigo mais tomar decisões.

**Figura 14 – Curvas de categoria de resposta e informação do item para o item 14
- MUDANÇA NA AUTO-IMAGEM –**



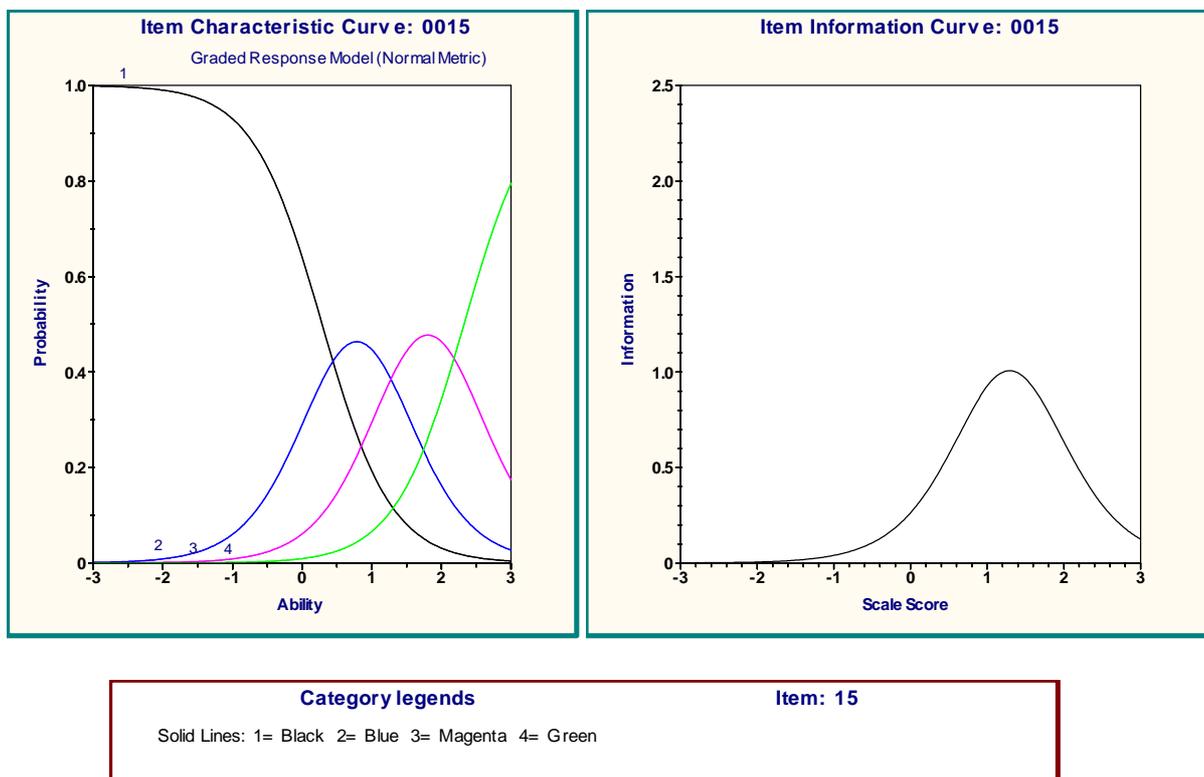
Categorias de resposta:

1 (curva preta) Não acho que de qualquer modo pareço pior do que antes.

2 (curva azul) { Estou preocupado em estar parecendo velho ou sem atrativo.
Acho que há mudanças permanentes na minha aparência, que me fazem parecer sem atrativo.

3 (curva rosa) Acredito que pareço feio.

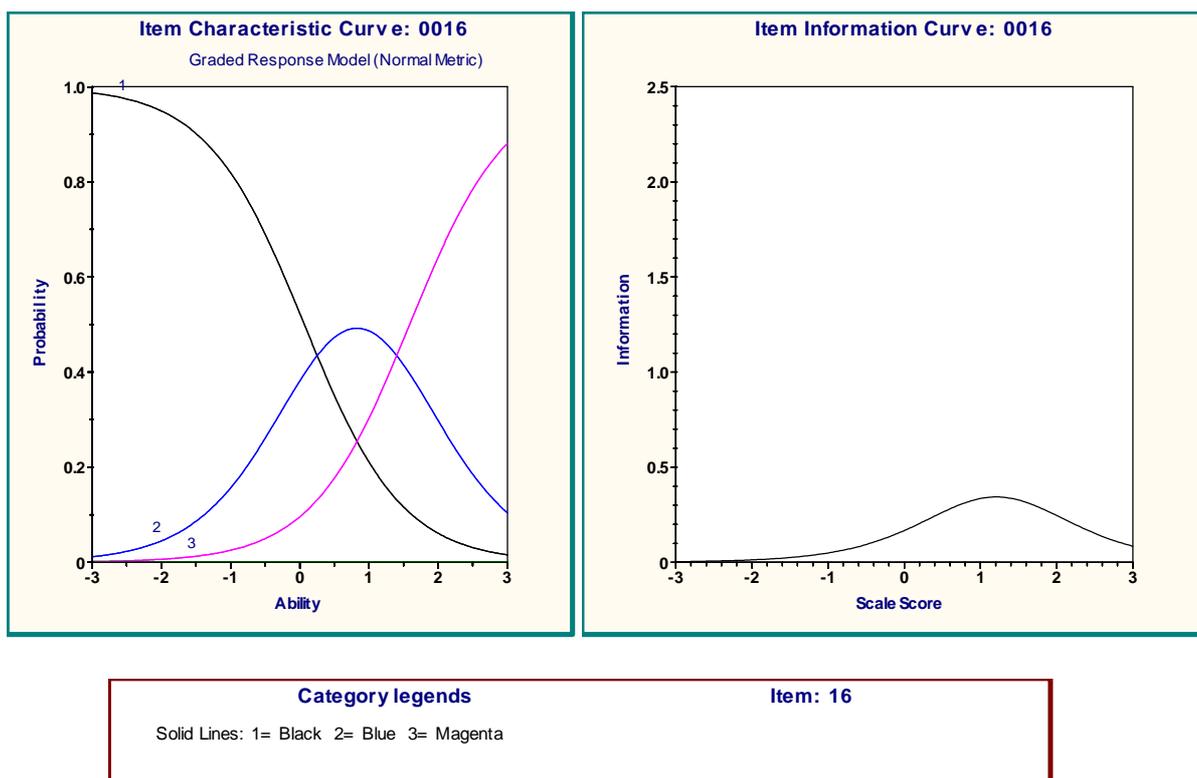
**Figura 15 – Curvas de categoria de resposta e informação do item para o item 15
- DIFICULDADE DE TRABALHAR –**



Categorias de resposta:

- 1 (curva preta) Posso trabalhar tão bem quanto antes.
- 2 (curva azul) É preciso algum esforço extra para fazer alguma coisa.
- 3 (curva rosa) Tenho que me esforçar muito para fazer alguma coisa.
- 4 (curva verde) Não consigo mais fazer qualquer trabalho.

Figura 16 – Curvas de categoria de resposta e informação do item para o item 16
- INSÔNIA –



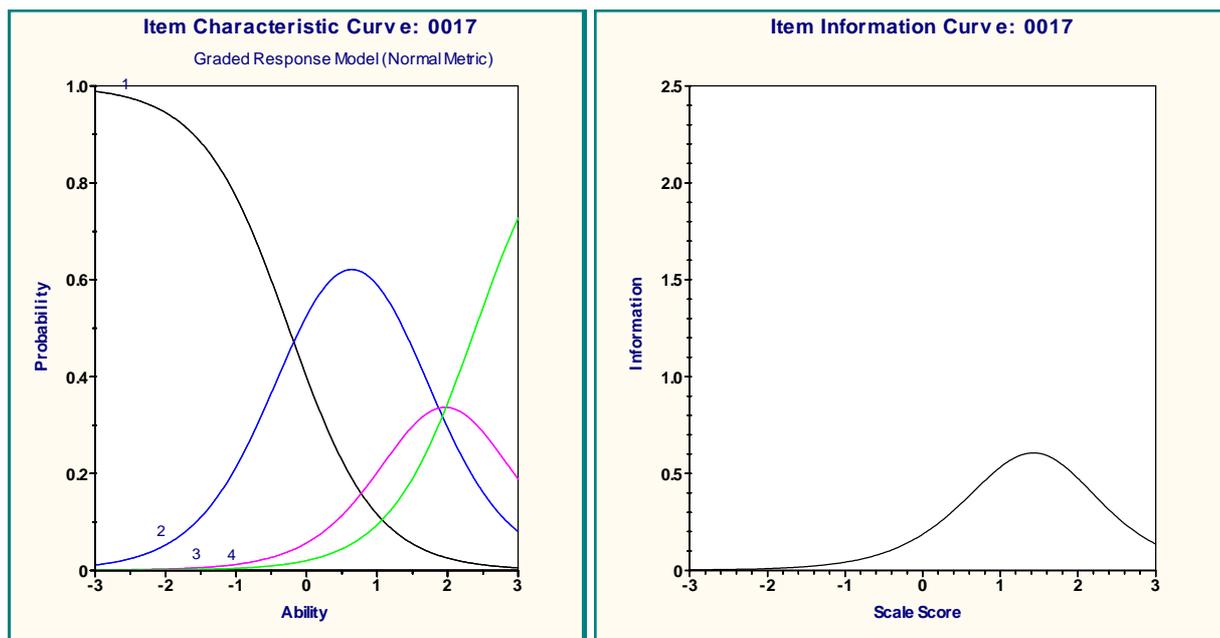
Categorias de resposta:

1 (curva preta) Consigo dormir tão bem como o habitual.

2 (curva azul) Não durmo tão bem como costumava.

3 (curva rosa) { Acordo 1 a 2 horas mais cedo do que costumava e não consigo voltar a dormir.
 Acordo várias horas mais cedo do que costumava e não consigo voltar a dormir.

**Figura 17 – Curvas de categoria de resposta e informação do item para o item 17
- FATIGABILIDADE –**

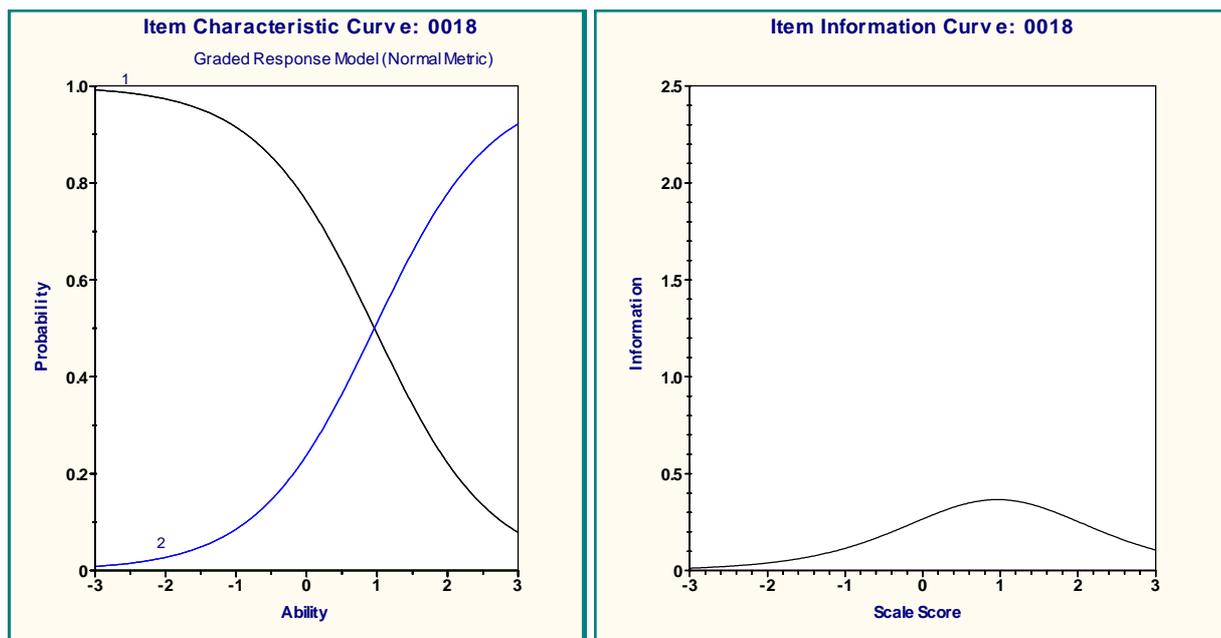


| Category legends | Item: 17 |
|---|----------|
| Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green | |

Categorias de resposta:

- 1 (curva preta) Não fico mais cansado do que o habitual.
- 2 (curva azul) Fico cansado mais facilmente do que costumava.
- 3 (curva rosa) Fico cansado em fazer qualquer coisa.
- 4 (curva verde) Estou cansado demais para fazer qualquer coisa.

**Figura 18 – Curvas de categoria de resposta e informação do item para o item 18
- PERDA DE APETITE –**

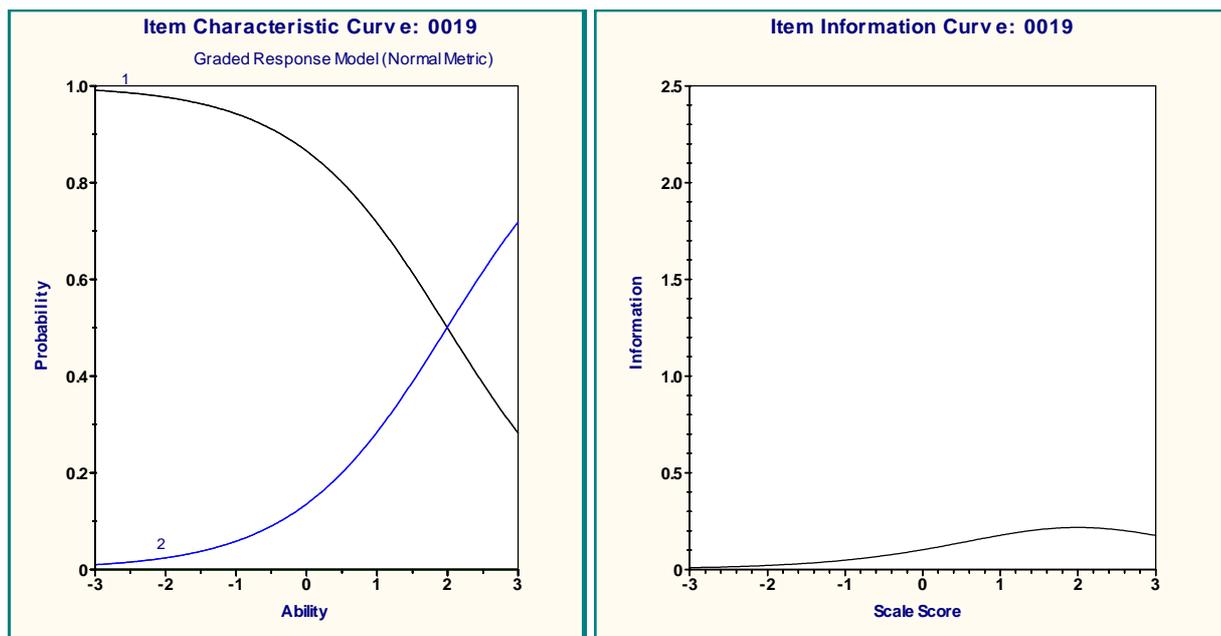


| Category legends | Item: 18 |
|-------------------------------|----------|
| Solid Lines: 1= Black 2= Blue | |

Categorias de resposta:

- 1 (curva preta) O meu apetite não está pior do que o habitual.
- 2 (curva azul) {
 Meu apetite não é tão bom como costumava ser.
 Meu apetite é muito pior agora.
 Absolutamente não tenho mais apetite.

Figura 19 – Curvas de categoria de resposta e informação do item para o item 18
- PERDA DE PESO –



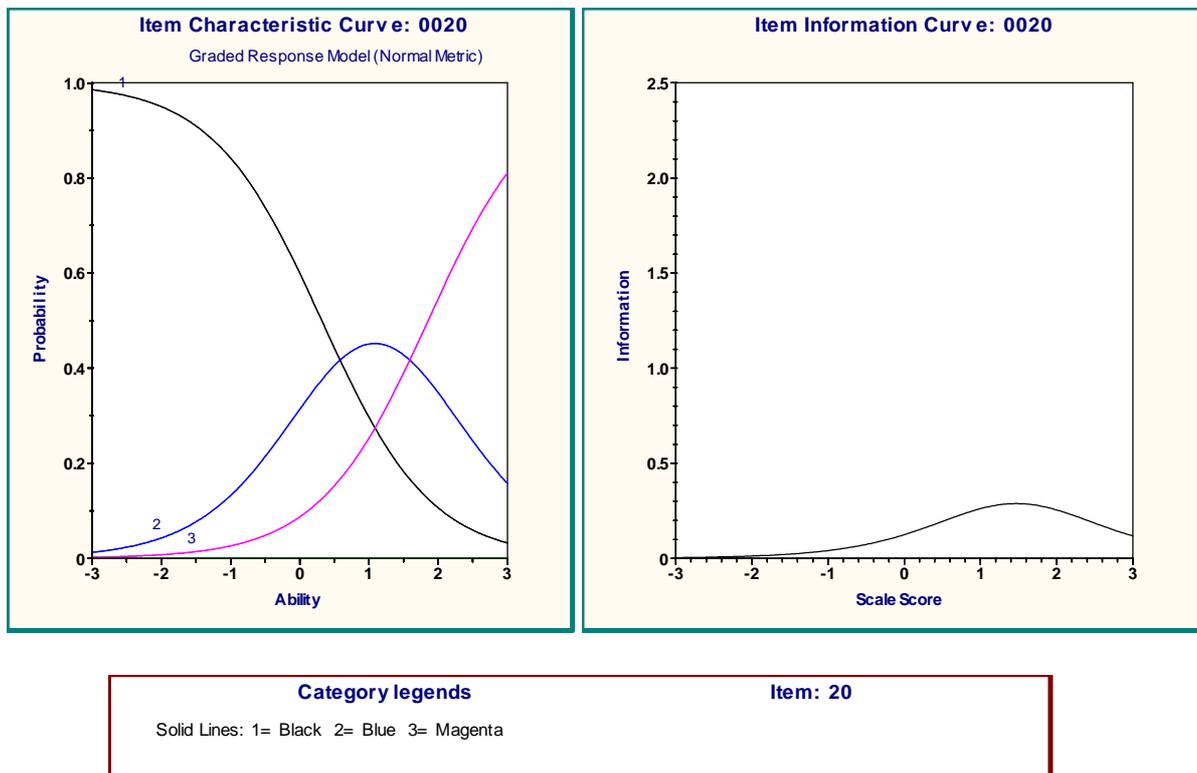
| Category legends | Item: 19 |
|-------------------------------|----------|
| Solid Lines: 1= Black 2= Blue | |

Categorias de resposta:

1 (curva preta) Não tenho perdido muito peso se é que perdi algum recentemente.

2 (curva azul) { Perdi mais do que 2 quilos e meio.
 Perdi mais do que 5 quilos.
 Perdi mais do que 7 quilos.

**Figura 20 – Curvas de categoria de resposta e informação do item para o item 20
- PREOCUPAÇÕES SOMÁTICAS -**



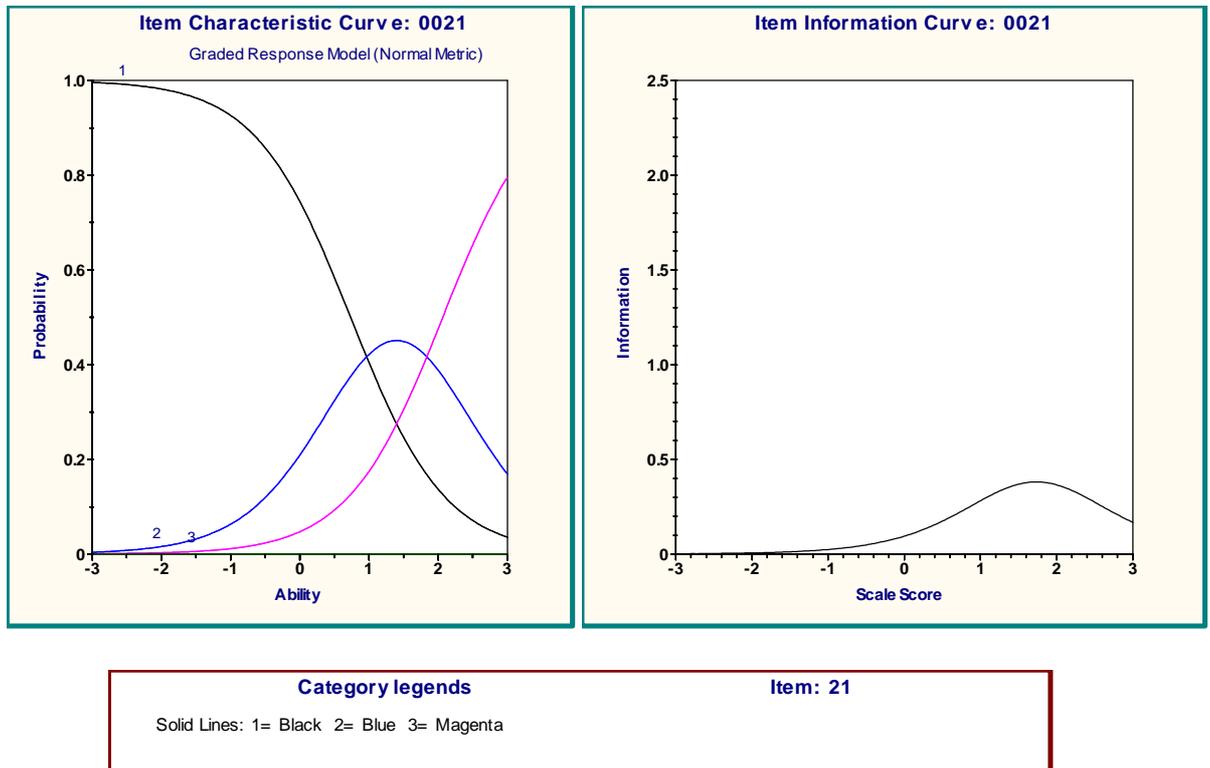
Categorias de resposta:

1 (curva preta) Não estou mais preocupado com minha saúde do que o habitual.

2 (curva azul) Estou preocupado com problemas físicos, tais como dores, indisposição do estômago ou constipação.

3 (curva rosa) {
Estou muito preocupado com problemas físicos e é difícil pensar em outra coisa.
Estou tão preocupado com meus problemas físicos que não consigo pensar em qualquer outra coisa.

**Figura 21 – Curvas de categoria de resposta e informação do item para o item 21
- PERDA DA LIBIDO –**

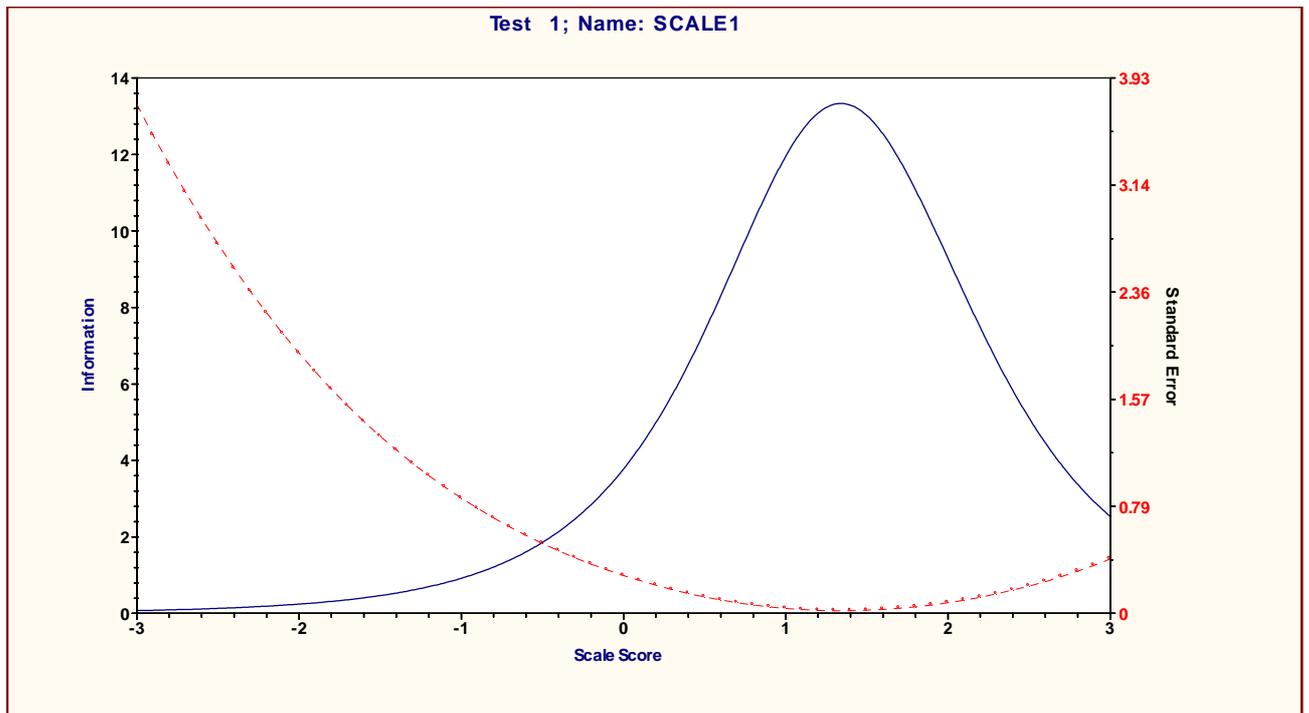


Categorias de resposta:

1 (curva preta) Não notei qualquer mudança recente no meu interesse por sexo.

2 (curva azul) { Estou menos interessado por sexo agora.
Estou muito menos interessado por sexo agora.

3 (curva rosa) Perdi completamente o interesse por sexo.

Figura 22 – Curva de informação do teste para o BDI.

Linha contínua – curva de informação do teste

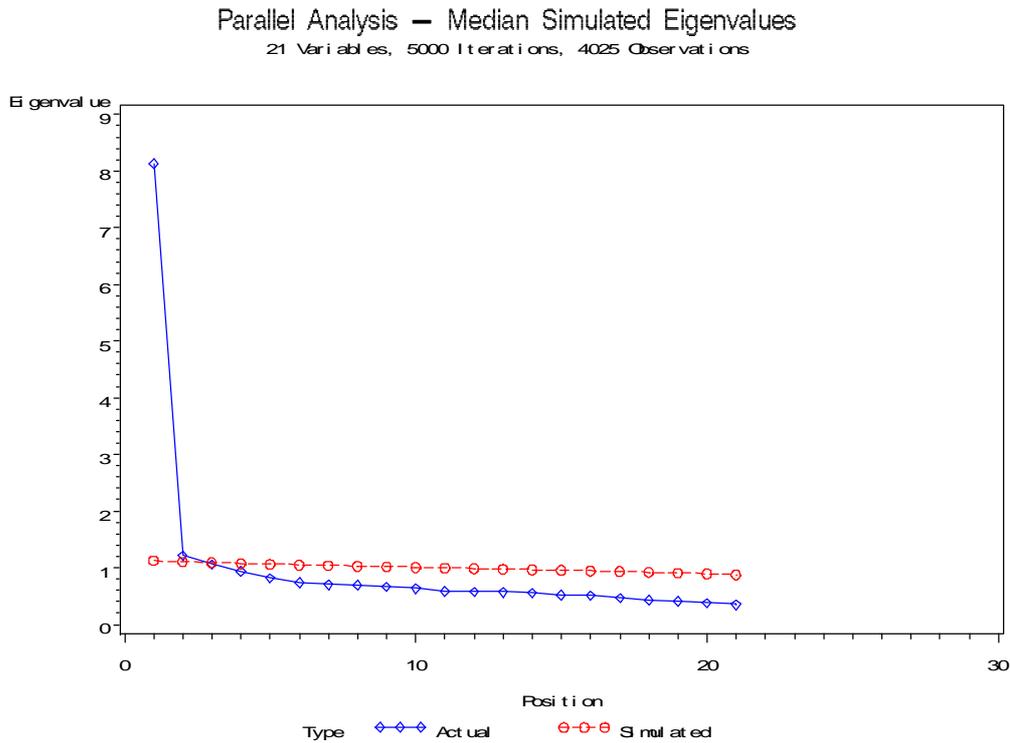
Linha pontilhada – erro padrão

ANEXO D - GRÁFICOS RESULTANTES DO PROCEDIMENTO DE ANÁLISE PARALELA UTILIZADO NOS DOIS ARTIGOS

Uma das suposições para a adequabilidade dos modelos TRI unidimensionais e, justamente, a unidimensionalidade do instrumento de medida (todos os itens do instrumento medem o mesmo traço latente). Existe evidência de que a suposição de unidimensionalidade pode ser relaxada, bastando que a mesma seja suficiente para que os modelos TRI possam ser utilizados. Em vista disso, foi avaliada a unidimensionalidade do BDI através do procedimento conhecido como Análise Paralela¹. 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, 4025 observações e 21 variáveis). As amostras simuladas são não correlacionadas e são geradas através do método de Monte Carlo (foram geradas 5000 amostras e a estatística resumo utilizada foi a mediana). O procedimento de Análise Paralela foi realizado utilizando uma macro do SAS criada para viabilizar o uso do mesmo, e seus resultados podem ser melhor visualizados através de um gráfico onde aparecem os autovalores resultantes dos dados reais e as medianas dos autovalores dos dados simulados.

1. Glorfeld LW. An improvement on Horn's Parallel Analysis Methodology for selecting the number of factors retain. *Educational and Psychological measurement*. 1995; 55: 377-393.

**Figura 1 - Resultado da Análise Paralela para a amostra total,
com 4025 observações e 21 itens.**



**Figura 2 - Resultado da Análise Paralela para a amostra total dicotomizada,
com 4025 observações e 21 itens.**

